Published in final edited form as:

J Neurosci Methods. 2019 January 15; 312: 93–104. doi:10.1016/j.jneumeth.2018.11.006.

Assessing Recurrent Interactions in Cortical Networks: Modeling EEG Response to Transcranial Magnetic Stimulation

Jui-Yang Chang^{#a}, Matteo Fecchio^{#b}, Andrea Pigorini^b, Marcello Massimini^{b,c}, Giulio Tononi^d, Barry D. Van Veen^{a,*}

^aDepartment of Electrical and Computer Engineering, University of Wisconsin–Madison, Wisconsin, USA

^bDipartimento di Scienze Biomediche e Cliniche “L. Sacco”, Università degli Studi di Milano, Milan, Italy

^cIstituto Di Ricovero e Cura a Carattere Scientifico, Fondazione Don Carlo Gnocchi, Milan, Italy

^dDepartment of Psychiatry, University of Wisconsin–Madison, Wisconsin, USA

[#] These authors contributed equally to this work.

Abstract

Background: The basic mechanisms underlying the electroencephalography (EEG) response to transcranial magnetic stimulation (TMS) of the human cortex are not well understood.

New Method: A state-space modeling methodology is developed to gain insight into the network nature of the TMS/EEG response. Cortical activity is modeled using a multivariate autoregressive model with exogenous stimulation parameters representing the effect of TMS. An observation equation models EEG measurement of cortical activity. An expectation-maximization algorithm is developed to estimate the model parameters.

Results: The methodology is used to assess two different hypotheses for the mechanisms underlying TMS/EEG in wakefulness and sleep. The integrated model hypothesizes that recurrent interactions between cortical regions are the source of TMS/EEG, while the segregated model hypothesizes that the TMS/EEG results from excitation of independent cortical oscillators. The results show that the relatively simple EEG response to TMS recorded during non-rapid-eye-movement sleep is described equally well by either the integrated or segregated model. However, the integrated model fits the more complex TMS/EEG of wakefulness much better than the segregated model.

*Corresponding author at: 3546a Engineering Hall 1415 Engineering Drive, Madison, WI 53706, USA. jyryanchang@gmail.com (Jui-Yang Chang), fecchio.matteo@gmail.com (Matteo Fecchio), andreapigorini@gmail.com (Andrea Pigorini), marcello.massimini@gmail.com (Marcello Massimini), gtononi@wisc.edu (Giulio Tononi), bvanveen@wisc.edu (Barry D. Van Veen).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declaration of Interest

The authors declare no conflict of interests.

Comparison with Existing Method(s): Existing methods are limited to small numbers of cortical regions of interest or do not represent the effect of TMS. Our results are consistent with previous studies contrasting the complexity of TMS/EEG in wakefulness and sleep.

Conclusion: The new method strongly suggests that effective feedback connections between cortical regions are required to produce the TMS/EEG in wakefulness.

Keywords

Electroencephalography; Transcranial magnetic stimulation; Multivariate autoregressive model with exogenous stimulation; State-space model; Expectation-maximization algorithm

1. Introduction

The development of multichannel transcranial magnetic stimulation (TMS)-compatible electroencephalography (EEG) amplifiers (Virtanen et al., 1999; Iramina et al., 2003; Thut et al., 2005) has recently opened the possibility of recording the electrical response of the human brain to a direct cortical stimulation. TMS/high-density electroencephalography (hd-EEG) stimulates and records directly from the cerebral cortex, while by-passing sensory pathways and motor pathways. Unlike traditional sensory-evoked potentials and TMS-evoked muscle potentials, this approach does not depend on the integrity/status of sensory and motor systems and can be applied to directly assess changes in cortical reactivity and cortico-cortical connectivity in physiological and pathological conditions. For example, perturbing different cortical targets with TMS in healthy, awake humans always triggers a complex, compound response which involves a distributed set of cortical areas and lasts for about 300 ms. In contrast, during non-rapid eye movement (NREM) sleep the same stimulus elicits a much simpler response of comparable duration (Massimini et al., 2005). This simpler response to TMS has also been observed in other conditions in which consciousness is lost, such as general anesthesia (Ferrarelli et al., 2010) and the vegetative state (Rosanova et al., 2012). Based on these observations, the perturbational complexity index (PCI) was developed to quantify the complexity of the overall EEG response to TMS (Casali et al., 2013). PCI has proven to be a reliable index of consciousness (Casali et al., 2013; Sarasso et al., 2015; Bai et al., 2016). The slow-wave-like response typical of NREM sleep has always been found associated with low values of PCI.

Although these empirical findings have practical implications, the basic mechanisms underlying the physiological EEG response to TMS are still largely unknown. Previous *in computo* works suggest that the specificity of the complex EEG response to TMS elicited during wakefulness may be associated with long-range connections (Cona et al., 2011) and a combination of intrinsic neuronal properties and cortico-cortical circuits interactions (Esser et al., 2009). Intracranial EEG recordings in humans also point to the importance of long-range re-current connections (Pigorini et al., 2015). However, the extent to which these factors contribute to the physiological EEG response to TMS during wakefulness and to the breakdown of effective connectivity and complexity during loss of consciousness still remains to be clarified.

Here we address the role of recurrent connections by interpreting TMS-evoked potentials (TEPs) using a modeling perspective. Specifically, we estimate models from single-trial EEG responses to TMS while exploring two different hypotheses for the complex, long-lasting compound response observed in wakefulness and sleep: (1) a segregated model (Figure 1 (a)) in which TMS results in a feedforward sweep that engages a number of cortical regions with different properties whose independent responses result in the TEP; (2) an integrated model (Figure 1 (b)) in which the initial feedforward sweep due to TMS is followed by recurrent interactions among cortical regions to produce the TEP. The identical set of cortical regions are used in both models. A semi-data-driven procedure that is independent of the models is used to select the regions for each subject across both wakefulness and sleep conditions. Hence, the only difference between the integrated and segregated models is the presence/absence of recurrent interactions between the selected cortical regions.

We examined the integrated and segregated hypotheses by extending the linear state-space model (SSM) framework developed in Cheung et al. (2010) for spontaneous EEG to incorporate the feedforward pathways engaged by exogenous stimulation such as TMS. Our method estimates the coefficients associated with the feedforward pathways to the cortical regions and the (possible) recurrent interactions between regions. We use this new method to compare the two different models for TMS/EEG recordings during both wakefulness and NREM sleep. The integrated model (Figure 1 (b)) assumes the cortical regions involved in the response are fully connected and interacting. In contrast, the segregated model (Figure 1 (a)) assumes the cortical regions involved in the response do not interact with one another. The segregated and integrated models mimic the absence or presence of long-range connections, respectively.

The estimated models are compared using the TMS evoked response and cross-validation of one-step prediction errors on single trials. Our results show that the brain dynamics evoked by TMS during NREM sleep can be described equally well by both segregated and integrated models over the entire response duration. In contrast, integrated and segregated models provide comparable fit to the actual TEP in wakefulness only for the assumed feedforward path duration. After feedforward effects subside the integrated model provides a much better fit to the actual TEP. Similarly, the integrated model has lower one-step prediction error than the segregated model on single trials not used to train the models. The results strongly suggest that the high levels of complexity typical of TMS/EEG responses during wakefulness, as also assessed by PCI, requires the presence of effective recurrent interconnections.

2. Methods

2.1. Data

Subjects.—Seven healthy volunteers participated in the study. All subjects gave written informed consent, and the experiment was approved by the Comitato Etico Interaziendale Milano Area A, Milan, Italy. A clinical examination was performed before the experiment to exclude potential adverse effects of TMS. TMS was performed in accordance with current safety guidelines. TMS/EEG data was initially collected during wakefulness when subjects

were alert and relaxed, with eyes open, and then the same stimulation was performed after subjects entered a consolidated period of NREM sleep.

TMS Targeting.—Stimulation was performed by a focal figure-of-eight coil (mean/outer winding diameter 50/70 mm, biphasic pulse shape, pulse length 280 μ s, and focal area of the stimulation 0.68 cm²) driven by a Mobile Stimulator Unit (eXimia TMS Stimulator, Nexstim Ltd.). Cortical TMS targets were identified on MRI (magnetic resonance imaging) scans acquired on a 3T magnetic resonance scanner (Trio Tim, Siemens, Germany) using a T1-weighted MP- RAGE (magnetization-prepared rapid acquisition gradient echo) sequence. We controlled TMS parameters by means of a Navigated Brain Stimulation (NBS) system (Nexstim, Helsinki, Finland). A 3D infrared tracking position sensor unit was employed to locate the relative positions of the coil and subject's head referenced to the individual MRI scan with an error tolerance of 3 mm. The NBS system also calculated the distribution and the intensity of the intracranial electric field induced by TMS on the cortical surface in real time. The output of the stimulating unit was adjusted to induce an electric field of 90 V/m on the cortical surface, which is above the threshold (50 V/m) for a significant EEG response (Casali et al., 2010; Komssi et al., 2007). The stimulation coordinates were passed to a software aiming tool that ensured the reproducibility of position, direction, and angle of the stimulator throughout the session. At least 200 trials were collected. TMS was delivered with an interstimulus interval jittering randomly between 2000 and 2300 ms (0.4–0.5 Hz).

EEG Recordings During TMS.—We recorded TMS-evoked potentials by means of a TMS-compatible 60-channel amplifier (Nexstim), a device that prevents amplifier saturation by means of a proprietary sample-and-hold circuit (Virtanen et al., 1999). The analog output of the amplifier is held constant from 100 μ s before stimulus to 2 ms after stimulus. The impedance at all electrodes was kept below 5 k Ω . The EEG signals were referenced to an additional electrode on the forehead, bandpass filtered (0.1–350 Hz) and sampled at 1450 Hz with 16-bit resolution. Two extra sensors were used to record the electro-oculogram (EOG). Well established procedures were followed for collecting TMS/EEG. Subjects wore earplugs and a sound was played continuously to avoid contamination of TMS-evoked potential by auditory potentials evoked by the click associated with the TMS discharge (Massimini et al., 2005, 2007; Ferrarelli et al., 2008). Bone conduction was attenuated by placing a thin layer of foam between coil and scalp (ter Braack et al., 2015). These precautions ensure the measured EEG is due to direct cortical stimulation (Gosseries et al., 2015).

General Experimental Procedures.—During the experiment each subject was lying on a reclining chair with a head-rest that allowed a comfortable and stable head position. The navigation system was calibrated with a muscle artifact free target location in the left/right premotor (Brodmann area 6 or BA6) or the left/right posterior parietal (BA7) cortex identified prior to TMS/EEG data collection in wakefulness. These areas were stimulated along the midline, thus reducing the possibility of inducing muscular activation (Mutanen et al., 2013) and/or any possible secondary cortical response due to somatosensory perception (Fecchio et al., 2017). A second TMS/EEG data collection session using identical stimulation parameters was initiated after subjects entered a consolidated period (>5 min) of NREM sleep stage 3.

TMS evoked potential.—In order to measure the duration of TMS induced response, we calculated the evoked potential and the duration of significant response following procedures in Massimini et al. (2005). Figure 2 (a) and (b) show the TMS evoked potentials and the temporal extent of significant response (in red) during wakefulness and sleep for a single subject. In panel (c), we show maximum extent of the TMS-induced response among the seven subjects in wakefulness and sleep.

Data Preprocessing.—Data analysis was performed using MATLAB (MathWorks). First, TMS/EEG trials were visually inspected to detect and reject trials containing excessive noise, muscle activity, or eye movements. Next, trials were segmented into windows of ± 800 ms around the TMS stimulus. Channels with large residual artifacts or bad signal quality were excluded from further analysis. All sessions analyzed used a minimum of 52 channels. The EEG data were average referenced, baseline corrected and independent component analysis (ICA) was applied in order to remove residual artifacts.

The data was further downsampled by a factor of 15 in two stages using the MATLAB function `resample` to obtain an effective sampling rate of 96.67 Hz. The downsampled data was then zero-phase high-pass filtered with a Butter-worth filter with passband edge frequency of 2 Hz. The downsampled, filtered data was baseline-corrected once more to make each channel of each trial zero- mean. We then followed a procedure similar to that described in Chang et al. (2012) to identify outlying TEPs. The Mahalanobis distance (Anderson, 2003) between the data in the trial being tested as an outlier and the remaining trials was computed. Trials with Mahalanobis distances having probability less than 0.1 were excluded from the subsequent analysis. A minimum of 115 trials were available for each session. Each trial of the data used for analysis included 310.3 ms prior to stimulation onset, 403.4 ms post stimulation and contains 70 samples. The TMS onset was at the 31st sample.

Perturbational Complexity Index (PCI).—PCI (Casali et al., 2013) is a nonparametric measure of the data's spatio-temporal complexity. It is hypothesized to reflect the ability of many functionally specialized thalamo-cortical modules to interact producing a complex response. PCI was calculated by first band-pass filtering TEPs with a 0.1–45 Hz passband and then down-sampling to 362.5 Hz. Second, the cortical current density was estimated using a three-sphere head model (Berg and Scherg, 1994; Zhang, 1995) and an empirical Bayes solution with weighted minimum norm constraint (Friston et al., 2006; Mattout et al., 2006; Phillips et al., 2005). Next, significant cortical activations representing the deterministic pattern of TMS-evoked responses at the source level were obtained by applying a non-parametric bootstrap-based statistic (Lv et al., 2007; Pantazis et al., 2005), leading to a 2D binary space-time matrix. The normalized Lempel-Ziv complexity (Lempel and Ziv, 1976) of this matrix was then accumulated over space to obtain the temporal evolution of PCI, $PCI(t)$.

Figure 3 depicts the temporal evolution of PCI (cumulative PCI) and the rate of complexity divergence in the difference between cumulative PCI in wakefulness and sleep.

2.2. Model

The linear SSM framework developed in Cheung et al. (2010) is extended to TMS/EEG recordings by explicitly modeling the feedforward effects of TMS stimulation as illustrated in Figure 1. The SSM model for TMS/EEG consists of two linear equations. A state equation describes the evolution of cortical activity as a multivariate autoregressive process with exogenous stimulation (MVARX), where the stimulation is TMS. An observation equation characterizes the measured single-trial EEG recordings as a weighted sum of cortical activity and noise. The parameters of the SSM are unknown, but assumed to be constant during the measurement time. An expectation-maximization (EM) algorithm is employed to find the maximum likelihood estimates (MLE) of the unknown parameters.

State Equation.—The K cortical signals representing activity in the K cortical ROIs at time n and trial j are denoted by the K by 1 vector $\mathbf{x}_{n,j} = [x_{n,j}^1, x_{n,j}^2, \dots, x_{n,j}^K]^T$ where $x_{n,j}^k$ is the cortical signal in ROI k at time n and trial j . The cortical activity is modeled as an MVARX-(p, ℓ) process (Lütkepohl, 2005; Chang et al., 2012):

$$\mathbf{x}_{n,j} = \sum_{i=1}^p \mathbf{A}_i \mathbf{x}_{n-i,j} + \sum_{i=0}^{\ell-1} \mathbf{b}_i u_{n-i,j} + \mathbf{w}_{n,j} \quad (1)$$

for $n = 1, 2, \dots, N$ and $j = 1, 2, \dots, J$. TMS is represented by the input $u_{n,j}$. We have $u_{n_o,j} = 1$ if TMS is applied at time n_o in trial j , and $u_{n,j} = 0$ for $n \neq n_o$. The residual error $\mathbf{w}_{n,j}$ is modeled as a zero-mean normally distributed random variable with K by K covariance matrix $\mathbf{Q}$. The K by K matrix $\mathbf{A}_i$ characterizes how cortical signals from i time samples in the past influence present cortical signals. The $(m, k)^{\text{th}}$ element of $\mathbf{A}_i$, $a_i^{m,k}$, is a weight that models the contribution of the signal from ROI k at i time samples in the past to the prediction of the signal from ROI m at the current time. The segregated model (Figure 1 (a)) constrains the $\mathbf{A}_i$ to be diagonal while the integrated model has no constraints on the $\mathbf{A}_i$. The K by 1 vector $\mathbf{b}_i$ captures the influence of the TMS on each of the K ROIs i samples following stimulation onset. Thus, the $\mathbf{b}_i$ coefficients model the feedforward volley of activity induced by TMS in the i^{th} ROI. This feedforward volley is assumed to include indirect effects of TMS through brain regions not included in the cortical regions being modeled. The number of $\mathbf{b}_i$, ℓ models the maximum latency in significant feedforward connections from the stimulation site directly or indirectly to the K ROIs.

We collect the $\mathbf{b}_i$ into a K by ℓ matrix $\mathbf{B} = [\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_{\ell-1}]$ for notational convenience. Similarly, we collect the $\mathbf{A}_i$ into a K by Kp matrix $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p]$. We assume the initial state for each trial j , $[x_{0,j}^T, x_{-1,j}^T, \dots, x_{-p+1,j}^T]^T$, is normally distributed with identical, but unknown mean μ_0 and covariance matrix $\sigma_0^2 \mathbf{I}$. $\mathbf{A}$ and $\mathbf{Q}$ may be used to compute functional and effective connectivity and measures of integrated information (Cheung et al., 2010; Oizumi et al., 2016).

Observation Equation.—Let the M by 1 vector $\mathbf{y}_{n,j}$ denote the measured data in *MEEG* channels at time n and trial j . The observation equation models $\mathbf{y}_{n,j}$ as the sum of activity due to each ROI, $\mathbf{H}_k \lambda_k x_{n,j}^k$, plus observation noise $\mathbf{v}_{n,j}$:

$$\mathbf{y}_{n,j} = \sum_{k=1}^K \mathbf{H}_k \lambda_k x_{n,j}^k + \mathbf{v}_{n,j}. \quad (2)$$

The M by 1 vector of observation noise $\mathbf{v}_{n,j}$ is assumed to be zero-mean normally distributed with covariance matrix $\mathbf{R} = \sigma_R^2 \mathbf{I}$. Here $\mathbf{H}_k$ is an M by 3 matrix describing the forward model for the k^{th} cortical ROI. $\mathbf{H}_k$ is a cortical patch basis (Limpiti et al., 2006). The 3 by 1 vector λ_k species the orientation of the source with respect to the basis formed by the columns of $\mathbf{H}_k$ and is assumed to be unit norm but unknown.

We assume the leadfield matrix is based on known dipole orientations and thus $\mathbf{H}_k$ is a rank-3 approximation to the space spanned by the columns of the leadfield vectors associated with all dipole sources within ROI k . More detailed discussion of the observation model is available in Cheung et al. (2010).

2.3. EM Algorithm

The parameters to be estimated in the SSM are $\theta = \{A, B, Q, \lambda_1, \lambda_2, \dots, \lambda_K, R, \mu_0, \sigma_0^2\}$. Our goal is to find the MLE of θ . MLEs have the least variance of all unbiased estimates for sufficiently large data sets (Kay, 1993). We write the log likelihood function as

$$\begin{aligned} L(\theta) &= \log p(\mathbf{Y}|\mathbf{U}, \theta) = \log \int p(\mathbf{Y}, \mathbf{X}|\mathbf{U}, \theta) d\mathbf{X} \\ &= \sum_{j=1}^J \log p(\mathbf{y}_{1,j}, \mathbf{y}_{2,j}, \dots, \mathbf{y}_{N,j} | \mathbf{U}, \theta) \end{aligned} \quad (3)$$

where $\mathbf{Y}$ denotes the collection of measured data from all trials, $\mathbf{U}$ denotes the TMS input and $\mathbf{X}$ denotes the collection of cortical signals. The MLE is the solution to the optimization problem: $\max_{\theta} L(\theta)$. This optimization problem does not have a closed-form solution and in general is not convex as it involves latent variables $\mathbf{X}$. The EM algorithm is an iterative coordinate ascent algorithm for finding MLEs (Shumway and Stoffer, 1982; Cheung et al., 2010). The algorithm starts with an initial guess $\theta^{(0)}$ and iterates the E- and M-steps:

- E-step: Evaluate the probability distribution $p(\mathbf{X}|\mathbf{Y}, \mathbf{U}, \theta^{(k)})$ and the conditional expectation $Q(\theta, \theta^{(k)}) = \mathbb{E}_{\mathbf{X}|\mathbf{Y}, \mathbf{U}, \theta^{(k)}} [\log p(\mathbf{Y}, \mathbf{X}|\mathbf{U}, \theta)]$
- M-step: Find $\theta^{(k+1)} = \arg\max_{\theta} Q(\theta, \theta^{(k)})$.

A convergence criterion is employed to decide whether or not the algorithm should terminate at iteration k . The convergence criterion involves comparing $L(\theta^{(k)})$ and $L(\theta^{(k-1)})$ (Dias and Wedel, 2004) in the E-step. The EM algorithm is guaranteed to monotonically

increase the objective function $L(\theta)$. Thus, at worst the EM algorithm finds a local maximum of the log likelihood function. We start our EM algorithm at multiple initial guesses and choose the solution with largest log likelihood to improve the chances of finding the global maximum. The EM algorithm has been shown to give more accurate model estimates for spontaneous data than two-step methods that employ source reconstruction followed by estimation of the cortical multivariate autoregressive model (Cheung et al., 2010). Additional details of the modeling procedure and the EM algorithm are presented in Supplementary Text 1.

2.4. Region Selection

We choose the subject-specific ROIs using a semi-data-driven sequential method. First, a minimum norm estimate of source activity at each dipole is reconstructed. Second, patches with the largest power for both wakefulness and sleep are sequentially added into the subject's ROI set.

The cerebral cortex of each subject's brain was modeled as a three-dimensional grid of 3004 fixed dipoles oriented normally to cortical surface. This model was adapted to the anatomy of each subject using the Statistical Parametric Mapping software package (SPM5, <http://fil.ion.ucl.ac.uk/spm/>) using the same parameters used in Casali et al. (2010). Finally, the inverse transformation was applied to the Montreal Neurological Institute (MNI) canonical mesh of the cortex for approximating to real anatomy.

Cortical patches are defined as collections of dipoles. We exclude deep dipoles from the analysis to avoid source activity with very low signal-to-noise ratio. Deep dipoles are defined based on the distances from the dipole to all electrodes. Our criteria results in 2295 dipoles being included in the construction of patches. We generated 617 patches of geodesic radius 2 cm that are overlapping and approximately uniformly distributed.

The power in the measured data associated with each candidate patch is computed using the minimum norm method for each subject's wake and sleep data set. We select the ROI set according to patch signal power using a sequential approach. The first ROI is chosen as the patch with the largest normalized average power over wake and sleep from the Brodmann area targeted by TMS. In each subsequent iteration, the patch with the largest normalized average power over wake and sleep is added to the ROI set, under the constraint that this new patch does not physically overlap with any of the already selected patches. This procedure is repeated until $\lfloor M/3 \rfloor$ patches are selected where M is the smaller of the numbers of artifact free channels in wakefulness and sleep. $\lfloor \cdot \rfloor$ denotes the greatest integer less than operation. Note that $\lfloor M/3 \rfloor$ is the maximum number of regions that can be modeled without introducing linearly dependent components into the observation equation. A detailed description of the region selection procedure is described in Appendix A.

2.5. Model Selection

The number of regions included in the model, K , and the memory, p , are selected using the Akaike information criterion (AIC) (Lütkepohl, 2005). We considered $K = 3, 6, 9, \dots, \lfloor M/3 \rfloor$ and $p = 5, 10, 15, \dots, 30$. The model parameter ℓ represents the approximate duration the

feedforward volley of TMS- induced activity to the ROIs in samples. The actual duration may vary from one region to another; however, we chose a single value for all regions as a modeling compromise. We choose $\ell = 5$, $\ell = 10$, and $\ell = 15$ corresponding to maximum feedforward volley durations of 50 ms, 100 ms, and 150 ms, for the results presented in the paper. These choices are motivated by previous re-search that suggests feedforward effects are approximately limited to within 100 ms post-stimulus (Lamme and Roelfsema, 2000; Garrido et al., 2007; Gaillard, 1988).

2.6. Model Evoked Response

The evoked response of the model is obtained by setting $\mathbf{w}_{n,j} = \mathbf{0}$ in Eq. (1), $\mathbf{v}_{n,j} = \mathbf{0}$ in Eq. (2), and

$$u_{n,j} = \begin{cases} 1, & n = n_o \\ 0, & n \neq n_o \end{cases}$$

where n_o corresponds to the time that the stimulus is applied. Note that setting $\mathbf{w}_{n,j} = \mathbf{0}$ and $\mathbf{v}_{n,j} = \mathbf{0}$ eliminates the dependence of the model on j , so no averaging is utilized in obtaining the model evoked response.

Insight into the nature of the two models is obtained by expressing the evoked response in terms of the feedforward and feedback model parameters. For simplicity of presentation we assume $p=1$ and drop the trial subscript j . Denote the evoked response of the model in the cortical regions as $\hat{\mathbf{x}}_n$. Using $\hat{\mathbf{x}}_n = \mathbf{0}$ for $n < n_o$ we have

$$\begin{aligned} \hat{\mathbf{x}}_{n_o} &= \mathbf{b}_0; \\ \hat{\mathbf{x}}_{n_o+1} &= \mathbf{A}_1 \hat{\mathbf{x}}_{n_o} + \mathbf{b}_1; \\ \hat{\mathbf{x}}_{n_o+2} &= \mathbf{A}_1 \hat{\mathbf{x}}_{n_o+1} + \mathbf{b}_2; \\ &\vdots \\ \hat{\mathbf{x}}_{n_o+\ell-1} &= \mathbf{A}_1 \hat{\mathbf{x}}_{n_o+\ell-2} + \mathbf{b}_{\ell-1}; \\ \hat{\mathbf{x}}_{n_o+\ell} &= \mathbf{A}_1 \hat{\mathbf{x}}_{n_o+\ell-1}; \\ &\vdots \\ \hat{\mathbf{x}}_k &= \mathbf{A}_1 \hat{\mathbf{x}}_{k-1}, \text{ for } k \geq n_o + \ell; \end{aligned}$$

These expressions indicate that it is possible to exactly fit the first ℓ values of a measured evoked response $\mathbf{x}_n$ for any $\mathbf{A}_1$ by choosing the $\mathbf{b}_i$ appropriately. If we choose $\mathbf{b}_0 = \mathbf{x}_{n_o}$, and

set $\mathbf{b}_i = \mathbf{x}_{n_o+i} - \mathbf{A}_1 \mathbf{x}_{n_o+i-1}$ for $i=1,2,\dots,\ell-1$, then $\hat{\mathbf{x}}_n = \mathbf{x}_n$ for $n_o \leq n \leq n_o + \ell - 1$. That is,

structural constraints on $\mathbf{A}_1$ do not necessarily manifest in the first ℓ values of the model evoked response. The model evoked response after the first ℓ time steps evolves according to only $\mathbf{A}_1$, and this is when we expect differences between integrated and segregated model evoked responses to be most evident.

3. Results

3.1. Model Performance in Capturing TEPs

A detailed description of results for one subject and $\ell = 10$ is given first and is followed by a summary of results for all seven subjects and values of ℓ studied.

Figure 4 depicts the ROIs selected for the analysis of the subject (see 2.4). The numerical label on each ROI indicates the order in which the patch was chosen during the region selection process. The TMS target was in the left parietal cortex (BA7), which is shown in red. The TMS-evoked and model-evoked responses in a representative set of six channels are shown in Figure 5 during both wakefulness and sleep. The number of ROIs K and order p of the segregated model were set equal to the values selected by AIC for the integrated model (see 2.5). Figures 6 and 7 show per-channel normalized squared response error (NSRE) measured over the interval of 100 to 400 ms post stimulus for each channel. Let $\bar{y}_n$ be the TMS-evoked response and $\hat{y}_n(\hat{\theta})$ be the model evoked response where $(\hat{\theta})$ denotes the functional dependence of the model evoked response on the estimated parameters $\hat{\theta}$. The per-channel NSRE in the m^{th} channel is defined as

$$\text{per-channel NSRE}_m = \frac{\sum_{n=41}^{70} (\bar{y}_{n,m} - \hat{y}_{n,m}(\hat{\theta}))^2}{\sum_{n=41}^{70} \bar{y}_{n,m}^2}$$

where $\bar{y}_{n,m}$ is the TMS-evoked response from the m^{th} channel at time n and $\hat{y}_{n,m}(\hat{\theta})$ is the model evoked response of the m^{th} channel at time n . The interval 100 to 400 ms post stimulus was chosen to isolate the effect of feedback interactions from the feedforward TMS distribution modeled by the $\mathbf{B}_f$ (Figure 1). The channels depicted in Figure 5 are marked with colored bars. Channel indices that are not shown were identified as artifactual during data preprocessing.

Scatter plots of global NSRE for models with different ℓ 's are depicted in Figure 8 as a function of the global mean field power (GMFP) (Esser et al., 2006) ratio. The global NSRE is measured over all channels and the time interval of [100, 400] ms poststimulus. More specifically, it is defined as

$$\text{global NSRE}(\ell) = \frac{\sum_{n=\ell+31}^{70} \|\bar{y}_n - \hat{y}_n(\hat{\theta})\|_2^2}{\sum_{n=\ell+31}^{70} \|\bar{y}_n\|_2^2}$$

where $\|\mathbf{y}\|_2$ denotes the Euclidean norm of the vector $\mathbf{y}$. The GMFP ratio is formed as the ratio of the average GMFP over the interval of 0 ms to 400 ms post stimulation to a constant representing a bootstrapped estimate (McCubbin et al., 2008) of the maximum GMFP over the pre-stimulation interval -300 ms to -50 ms. Panels (a) and (b) depict global NSRE measured from 50 to 400 ms for models with $\ell = 5$. Panels (c) and (d) show global NSRE measured from 100 to 400 ms for models with $\ell = 10$. Panels (e) and (f) depict global NSRE

measured from 150 to 400 ms for models with $\ell = 15$. Robust linear fits of global NSRE as a function of GMFP ratio are also shown in Figure 8. The robust linear fits are estimated with the iteratively reweighted least squares algorithm.

The cumulative NSRE was computed as a function of time by summing the SRE over all channels from stimulus onset to the current time

$$\text{cumulative NSRE}(n) = \frac{\sum_{n'=31}^n \|\bar{\mathbf{y}}_{n'} - \hat{\mathbf{y}}_{n'}(\hat{\theta})\|_2^2}{\sum_{n'=31}^n \|\bar{\mathbf{y}}_{n'}\|_2^2}.$$

Figure 9 depicts cumulative NSRE for all subjects with models of $\ell = 5$, $\ell = 10$, and $\ell = 15$ in wakefulness and sleep. There is variability in the cumulative NSRE as a function of time across subjects in both wakefulness and sleep. A clear difference is apparent between integrated and segregated models in wakefulness when feedforward effects subside.

We also compared the integrated and segregated models using cross-validated mean squared prediction error (MSPE). The MSPE evaluates the models' ability to predict single trials of the scalp measurements one step into the future. The model fitting procedure (see 2.3) minimizes the error between model predictions and measured single trials in a probabilistic sense. This motivates the MSPE as an intuitive measure for the model to single trials.

Cross validation is used to control for the possibility of over fitting by the integrated model and involves splitting the single trials into test and training sets. The training set is used to estimate the model parameters while the test set is used to evaluate the models' ability to generalize to new data. We used ten-fold cross validation, so the data is partitioned into ten groups, with each group containing roughly the same number of trials. For the i^{th} fold of validation, the trials in the i^{th} group are used as the test set, and the other nine groups of trials are used as the training set. The model parameters are first estimated from the training set. Then the test trials withheld from training are used to evaluate one-step MSPE of the model estimated from the training set. The one-step MSPE of fold i is defined as

$$\text{MSPE}_i = \frac{1/|\text{Group}_i| \cdot \sum_{j \in \text{Group}_i} \sum_{n=1}^{70} \|\mathbf{y}_{n,j} - \tilde{\mathbf{y}}_{n,j}(n-1, \hat{\theta}_{-i})\|_2^2}{1/J \cdot \sum_{j \in J} \sum_{n=1}^{70} \|\mathbf{y}_{n,j} - \tilde{\mathbf{y}}_{n,j}(n-1, \hat{\theta}_{-i})\|_2^2}$$

where Group i denotes the set of trials assigned to group i , $\tilde{\mathbf{y}}_{n,j}(n-1, \hat{\theta}_{-i})$ denotes the one-step prediction of $\mathbf{y}_{n,j}$ using the measurements up to time point $n-1$ in trial j and the estimated parameters $\hat{\theta}_{-i}$. Here subscript $-i$ denotes that the model parameters are estimated from all trials except for those in the i^{th} group. The one-step prediction error is computed with Kalman-filtering procedures (see Supplementary Text 1 and Cheung et al. (2012)).

The difference between segregated and integrated models in MSPE for each group of trials and subject are depicted in Figure 10. The differences are greater in wake than sleep for all

subjects and choices of ℓ . A nonparametric Wilcoxon signed-rank test was used to test the hypothesis that the mean of the difference is zero. This hypothesis is rejected at an $\alpha = 0.01$ significance level in all subjects in wake and four of the seven subjects in sleep, for all three choices of ℓ . We also performed a two-way ANOVA and obtained a significant ($p < 0.01$) difference between integrated and segregated models in wakefulness but not in sleep.

4. Discussion

We have presented a method for assessing recurrent interactions in the cortex based on an MVARX model for the cortical activity induced by TMS and patch-based forward models for mapping the cortical activity to the scalp EEG. The MVARX model describes the cortical activity in each ROI as a weighted combination of past activity in all ROIs plus feedforward activation by TMS. The MVARX model approximates the complex interactions in the cortex with a linear model—the simplest causal model that can account for the rich temporal dynamics associated with EEG.

We previously reported on MVARX models for intracranial EEG with direct electrical current stimulation (Chang et al., 2012). The extension of the MVARX approach to modeling cortical activity due to TMS from scalp EEG is nontrivial because the cortical activity is observed indirectly through EEG physics. Furthermore, the active cortical regions are unknown *a priori*. Our approach is novel in use of a ROI identification method that is tuned specifically to the observation equation in the MVARX model and in use of a one-step EM algorithm for estimating the model parameters from single trial scalp TMS-EEG data.

Our EM approach directly estimates the cortical model from the scalp EEG using the maximum likelihood criterion. This avoids the suboptimal nature of two-step methods Hui et al. (2010) that first attempt to solve the inverse problem to obtain cortical activity, and then solve a second problem to fit a model to the estimated cortical activity. Our one-step approach has the potential for significantly better performance at modest and low SNR than two-step approaches. This is because the inverse problem is ill-posed and its solution amplifies noise. Noise in the estimated cortical activity contaminates and biases the model parameter estimates in the second step. While the cortical signals and models are unknown in human data, a comparison between EM and two-step approaches in a related problem (Cheung et al., 2010) provides clear support for this reasoning. The results in this paper show that our approach leads to models with good fidelity to TEP.

Linear models benefit from reduced computational complexity, simpler parameter estimation approaches, and increased robustness to noise compared to nonlinear models, although nonlinear models such as DCM (Friston et al., 2003) offer the potential of parameters with physiological meaning and higher degrees of freedom in modeling complex responses. DCM is rarely applied to more than six or seven ROIs due to the computational complexity of estimating model parameters and the difficulty of ensuring convergence of estimated parameters to good solutions. We did not consider DCM or other nonlinear models in this study because of this effective limitation to a smaller number of ROIs. The TMS-evoked response is relatively widely distributed throughout the cortex, especially in wakefulness (Massimini et al., 2005), and a small number of ROIs is unlikely to capture the full extent of

cortical activity. Consequently in this paper we routinely use twelve or more ROIs to capture as many sources of activity as possible. The number of ROIs considered is constrained to be smaller than the number of measurement channels divided by three, otherwise the observation equation would become degenerate. Using the maximum possible number of ROIs also helps reduce sensitivity to the ever present hidden node problem with network models.

The role of recurrent interactions in cortical networks is assessed by comparing and contrasting the performance of two MVARX models—an integrated model with all possible interactions between ROIs (Figure 1 (b)) and a segregated model with no interactions between ROIs (Figure 1 (a))—in both wakefulness and sleep. We evaluated the effect of reentrant connectivity on a given set of ROIs using a model independent, semi-data-driven procedure to choose a common set of ROIs across both conditions for each subject. Using different numbers or choices of ROIs across models or conditions would confound the respective choices with the effect of network structure.

Our results show that the integrated model provides significantly better visual agreement with the measured TEP than the segregated model for the more complex responses associated with wakefulness (Figures 5 (a), 6, 8), especially later than 100ms post stimulus for all three values of ℓ studied. Consistent with previous intracranial recordings in monkeys (Lamme and Roelfsema, 2000) and humans (Gaillard et al., 2009), a previous non-invasive TMS/EEG experiment showed that a maximum spread of activation, possibly reflecting the first feed-forward sweep, can occur 80–100 ms after the stimulation (Casali et al., 2010). In line with these studies, we considered durations for the exogenous parts of both integrated and segregated models of 50 ms ($\ell = 5$), 100 ms ($\ell = 10$), and 150 ms ($\ell = 15$).

Modeling of the feedforward effects accounts for the fidelity of the model responses during the first 100ms post stimulus (Figure 9). The EM algorithm chooses the model parameters using the single trials and does not explicitly fit the TEP. However, the TEP is a large component of the response right after the stimulus, and thus it is not surprising that both integrated and segregated models choose their feedforward parameters to closely fit the TEP in the first 100ms as discussed in Section 2.6. However, after the first 100ms the interactions between ROIs provided by the integrated model make a significant difference in modeling the more complex responses of wakefulness. The TEP in sleep is of comparable duration (Figure 2), but is not as complex and thus is nearly equally well described by the integrated and segregated models.

The NSRE is computed over the interval 100ms - 400 ms post stimulus to quantitatively assess the impact of the recurrent interactions in the model on the TEP. The per-channel NSRE for Subject 1 (Figure 6) shows that the integrated model performs better in every single channel during wakefulness. Normalization enables us to see how the error compares to the amplitude of the signal. This avoids the deceptive implications of very small errors that are associated with very low amplitude channels. In contrast, the largest normalized errors in Figure 6 correspond to channels with very weak amplitude signals. The global NSRE across all channels shows that the integrated model performs significantly better for all seven subjects in wakefulness (Figure 8 (a), (c), (e)). These results highlight the

important role of recurrent interactions in producing the complex TMS response patterns associated with wakefulness. In contrast, the difference between integrated and segregated models is much less in sleep (Figures 5 (b), 7, 8 (b), (d), (f)) than in wakefulness.

The cumulative NSRE (Figure 9) shows that the performance of integrated and segregated models is nearly identical in the first 100ms post TMS in both wakefulness and sleep. This is due to the identical duration assumed for the feedforward volley in both models. In wakefulness the integrated and segregated model NSREs diverge after 100ms post TMS and continue to diverge as time increases (Figure 9 (a), (c), (e)). Interestingly, these results are paralleled by the time course of the PCI metric shown in Figure 3. The population cumulative PCI for wakefulness and sleep is similar up to 75–100 ms, maximally diverge between 100 and 125 ms and then the cumulative PCI for wakefulness increases significantly, while that for sleep does not. The cumulative PCI thus suggests that a more complex model is required to describe the TEP in wakefulness than sleep. Overall, our findings seem to suggest that the build-up of the complex responses observed during wakefulness after feedforward effects subside requires the engagement of recurrent interactions among cortical nodes and that these interactions are impaired during NREM sleep. This mechanism and its timeframe are generally consistent with the results of a recent intracranial study employing single-pulse electrical stimulations and stereo EEG recordings (Pigorini et al., 2015). This study showed that during wakefulness electrical stimulation triggers a sequence of deterministic phase-locked activations in its cortical targets. In contrast, during NREM sleep cortical neurons have the tendency to fall into a silent down-state upon receiving an input due to underlying bistability. The downstate occurs as early as 100 ms after the pulse and obliterates the deterministic effects of the initial input, as indicated by a sharp drop of phase-locked activity. Thus, one may hypothesize that while during wakefulness the initial feedforward activation triggered by cortical stimulation evolves after 100 ms into a chain of deterministic interactions among cortical ROIs leading to a complex response, during NREM sleep the same feedforward sweep simply triggers down-state in target neurons which blocks the build up of complex interactions.

The integrated model has more parameters than the segregated model. We used cross validation on the MSPE (Cheung et al., 2012) to assess whether the improved performance of the integrated model in wakefulness could be due to over fitting. The reduced MSPE for the integrated model is statistically significant in all of the seven subjects in wakefulness. Figure 10 shows the ten differences between segregated and integrated model cross-validated MSPE are positive for every subject in wakefulness. This strongly suggests that the improved performance of the integrated model in wakefulness is not due to overfitting the data; it shows that the integrated model generalizes better to data not used to train the model. Note that cross validation provides a very robust approach to model selection that naturally controls for model complexity. If the more complex model is fitting noise, then it will have poor performance describing data not used to train the model. In contrast, if the cross-validated MSPE is lower for the more complex model, then additional parameters are genuinely contributing to better modeling the data. Cross validation in principle also provides a more robust approach to selecting other model parameters, such as the number of regions and the memory p . However, cross validation has a very high computational cost and

thus we chose to limit its use here to a binary comparison of integrated and segregated models.

It is possible that an MVARX model with a subset of the feedback connections in the integrated model—that is, some connections between ROIs constrained to zero—could improve upon the performance of the fully integrated model. Reducing model degrees of freedom potentially reduces the variance associated with model parameter estimation. We did not explore limiting the number of feedback connections due to the very large number of such models possible for even twelve ROIs. Furthermore, it would not change the conclusion that effective feedback connections between ROIs appear to be required to produce the TMS/EEG in wakefulness. The contrast between fully integrated and segregated models provides evidence for the importance of at least some feedback connections. It does not imply that a fully connected model is necessary.

Modeling involves a compromise between computation and parameter estimation considerations and the faithfulness of the model to the underlying, typically unknown, phenomenon. This tradeoff is manifest in our decisions to employ a linear model versus a nonlinear model, only consider a fully connected model versus partially connected models, and use the same values of p in all connections.

A challenging aspect of applying the modeling methodology described here is identification of the ROIs to include in the cortical network model. Any prior information of brain regions involved in the paradigm being studied should be used to select ROIs as illustrated in the spontaneous EEG studies of Dentico et al. (2014); Kundu et al. (2015). Semi-data-driven ROI selection, such as used in the present study, is closely related to the source localization problem and thus, any source localization method may be used. However, we strongly recommend that the ROI selection process be based on mapping source localization results onto the bases used to represent the ROIs in the observation equation. This ensures ROI selection maps directly to ROI representation in the model.

Source localization methods are unlikely to identify ROIs that are relatively silent in the scalp EEG due to depth or weak electrical activity. Similarly, our method for MVARX model estimation will have difficulty identifying network interactions involving ROIs that do not contribute measurable activity to the scalp EEG.

The problem of hidden nodes is endemic to any network model. If an important ROI is not included in the model, then it is possible to draw false inferences about causal influences and connectivity within the true underlying network. Our procedure for selecting ROIs includes those with the most significant contributions to the measured scalp EEG. It is possible that an ROI with very weak activity, or a deep ROI, plays a significant role in network interactions in this study. However, the difference we found between integrated and segregated models is likely insensitive to hidden nodes since a hidden node would only change the connectivity of the integrated model. In contrast, the potential impact of hidden nodes must be considered in studies that rely on comparing connectivity within a single model.

Theoretical considerations have been used to argue complex network inter-actions as a basis for consciousness (Oizumi et al., 2014). This study concurs with others (Chang et al., 2012) that highlight the role of network interactions for supporting consciousness. The TEP in sleep, when the subjects are un-conscious, is explained almost as well by the segregated model as it is by the integrated model. Including network interactions provides marginal benefit in describing the measured TEP or to the model's one-step prediction performance of single trials. In contrast, during wakefulness the network interactions of the integrated model result in substantially better fit to the TEP and improved one-step prediction of single trials.

5. Conclusions

This paper describes a modeling methodology for single-trial TMS-EEG response. We show that our approach generates model evoked responses that characterize measured TMS evoked potentials. We use our modeling methodology to compare the performance of integrated (fully connected) and segregated (completely disconnected) models for the TMS-EEG obtained during both wakefulness and NREM sleep. Significant differences in modeling performance strongly suggests that the complex TMS evoked response is a consequence of recurrent network connections among different brain regions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Acknowledgments

This work was supported in part by the National Institutes of Health under awards R21EB015542 and R01GM116916, the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 720270 (HBP SGA1) and the grant "Sinergia" CRSII3_160803/1 by the Swiss National Science Foundation.

Appendix A.

Region Selection Procedure

We select ROIs for inclusion in the model by first using a minimum norm method to reconstruct source activity at 3004 dipoles tessellating the cortical surface. Next the estimated source activity is used to estimate the combined normalized power in each cortical patch across both wakefulness and sleep data sets for each subject. The ROIs in the model are selected based on the combined normalized power across the entire patch set.

Consider the M by P lead field matrix $\mathbf{L}$ where $P=3004$ is the number of dipoles. Let the dipole activity originating from voxel β be represented by the signal $r_{n,j}^\beta$, and the concatenation of the dipole signals be the P by 1 vector $\mathbf{r}_{n,j} = [r_{n,j}^1, r_{n,j}^2, \dots, r_{n,j}^P]^\top$. The measured data can be expressed as

$$\mathbf{y}_{n,j} = \mathbf{L}\mathbf{r}_{n,j} + \mathbf{n}_{n,j}$$

where $\mathbf{n}_{n,j}$ is noise. The minimum norm solution for $\mathbf{r}_{n,j}$ satisfies

$$\min_{\mathbf{r}_{n,j}} \sum_{j=1}^J \sum_{n=1}^N \|\mathbf{y}_{n,j} - \mathbf{L}\mathbf{r}_{n,j}\|_2^2 + \eta M \|\mathbf{r}_{n,j}\|_2^2.$$

We choose η using the generalized cross-validation (GCV) method proposed in Golub et al. (1979). The GCV objective is given as

$$\hat{\eta} = \underset{\eta}{\operatorname{argmin}} \frac{\frac{1}{M} \sum_{j=1}^J \sum_{n=1}^N \|(\mathbf{I} - \mathcal{A}(\eta))\mathbf{y}_{n,j}\|_2^2}{\left[\frac{1}{M} \operatorname{tr}(\mathbf{I} - \mathcal{A}(\eta)) \right]^2}$$

where $\mathcal{A}(\eta) = \mathbf{L}(\mathbf{L}^\top \mathbf{L} + \eta M \cdot \mathbf{I})^{-1} \mathbf{L}^\top$. Once $\hat{\eta}$ is selected, the signals from all dipoles are reconstructed as (Sekihara and Nagarajan, 2008)

$$\mathbf{r}_{n,j} = \mathbf{L}^\top (\mathbf{L}\mathbf{L}^\top + (\hat{\eta}M)\mathbf{I})^{-1} \mathbf{y}_{n,j}.$$

Define $\mathbf{L}^\ell$ as the M by G^ℓ matrix containing the G^ℓ columns of $\mathbf{L}$ corresponding to the G^ℓ dipoles within patch ℓ and let $\mathbf{L}^\ell = \mathbf{U} \Sigma \mathbf{V}^\top$ be the singular value decomposition. Define the M by 3 patch basis (Limpiti et al., 2006) $\mathbf{D}^\ell = \tilde{\mathbf{U}}^\ell \tilde{\Sigma}^\ell$ where the M by 3 matrix $\tilde{\mathbf{U}}^\ell$ contains the three dominant left singular vectors and $\tilde{\Sigma}^\ell$ is a 3 by 3 diagonal matrix of the three largest singular values. Assuming $\mathbf{L}^\ell$ is well approximated by a rank 3 matrix (Limpiti et al., 2006), then the Gram matrix $\mathbf{L}^\ell (\mathbf{L}^\ell)^\top$ can be approximated as $\mathbf{D}^\ell (\mathbf{D}^\ell)^\top$. This result will be used below.

Next we use the reconstructed dipole activity to estimate the cortical signal power associated with the ℓ^{th} patch. The signals from dipoles within patch ℓ are

$\mathbf{r}_{n,j}^\ell = (\mathbf{L}^\ell)^\top (\mathbf{L}\mathbf{L}^\top + (\hat{\eta}M)\mathbf{I})^{-1} \mathbf{y}_{n,j}$. We model the cortical signal from each patch as a scalar $s_{n,j}^\ell$ with the orientation of the signal with respect to the columns of $\mathbf{D}^\ell$ given by λ^ℓ . That is, the measured signal components originating from patch ℓ are modeled as $\mathbf{y}_{n,j}^\ell \approx \mathbf{D}^\ell \lambda^\ell s_{n,j}^\ell$.

Ignoring noise, we have $\mathbf{y}_{n,j}^\ell = \mathbf{L}^\ell \mathbf{r}_{n,j}^\ell$ and thus

$$\begin{aligned} s_{n,j}^\ell &= (\lambda^\ell)^\top (\tilde{\Sigma}^\ell)^{-1} (\tilde{\mathbf{U}}^\ell)^\top \mathbf{y}_{n,j}^\ell \\ &= (\lambda^\ell)^\top (\tilde{\Sigma}^\ell)^{-1} (\tilde{\mathbf{U}}^\ell)^\top \mathbf{L}^\ell (\mathbf{L}^\ell)^\top (\mathbf{L}\mathbf{L}^\top + (\hat{\eta}M)\mathbf{I})^{-1} \mathbf{y}_{n,j} \\ &= (\lambda^\ell)^\top (\mathbf{D}^\ell)^\top (\mathbf{L}\mathbf{L}^\top + (\hat{\eta}M)\mathbf{I})^{-1} \mathbf{y}_{n,j} \end{aligned}$$

Where we used $\mathbf{L}^\ell (\mathbf{L}^\ell)^\top \approx \mathbf{D}^\ell (\mathbf{D}^\ell)^\top = \mathbf{U}^\ell (\mathbf{\Sigma}^\ell)^2 (\mathbf{U}^\ell)^\top$ and $(\mathbf{\Sigma}^\ell)^{-1} (\mathbf{U}^\ell)^\top \mathbf{U}^\ell (\mathbf{\Sigma}^\ell)^2 (\mathbf{U}^\ell)^\top = \mathbf{\Sigma}^\ell (\mathbf{U}^\ell)^\top = (\mathbf{D}^\ell)^\top$. We choose the orientation λ^ℓ to maximize the power associated with $s_{n,j}^\ell$, that is

$$\lambda^\ell = \underset{\lambda^\ell}{\operatorname{argmax}} \left(\lambda^\ell \right)^\top (\mathbf{D}^\ell)^\top (\mathbf{L}\mathbf{L}^\top + (\hat{\eta}M)\mathbf{I})^{-1} \mathbf{R}_y (\mathbf{L}\mathbf{L}^\top + (\hat{\eta}M)\mathbf{I})^{-1} \mathbf{D}^\ell \lambda^\ell$$

Subject to $(\lambda^\ell)^\top \lambda^\ell = 1$ as in Sekihara and Nagarajan (2008). Here the spatial covariance matrix $\mathbf{R}_y = 1/JN \cdot \sum_{j=1}^J \sum_{n=1}^N \mathbf{y}_{n,j} (\mathbf{y}_{n,j})^\top - \mu_y \mu_y^\top$ and $\mu_y = 1/JN \cdot \sum_{j=1}^J \sum_{n=1}^N \mathbf{y}_{n,j}$. The solution is to choose λ^ℓ as the eigenvector corresponding to the maximum eigenvalue of $(\mathbf{D}^\ell)^\top (\mathbf{L}\mathbf{L}^\top + (\hat{\eta}M)\mathbf{I})^{-1} \mathbf{R}_y (\mathbf{L}\mathbf{L}^\top + (\hat{\eta}M)\mathbf{I})^{-1} \mathbf{D}^\ell$. The power $\sigma(\ell)$ associated with is equal to the maximum eigenvalue.

The power associated with the ℓ^{th} patch of the data collected from a subject during wakefulness and sleep is denoted $\sigma_w(\ell)$ and $\sigma_s(\ell)$, respectively. We form a weighted average of wakefulness and sleep patch powers to obtain a normalized Index

$$\bar{\sigma}(\ell) = \frac{\sigma_w(\ell)}{\operatorname{tr}(\mathbf{R}_{y,w})} + \frac{\sigma_s(\ell)}{\operatorname{tr}(\mathbf{R}_{y,s})}$$

where the normalizing weights $\operatorname{tr}(\mathbf{R}_{y,w})$ and $\operatorname{tr}(\mathbf{R}_{y,s})$ are overall signal power of the data in wakefulness and sleep, respectively.

The first ROI is the patch with the largest $\bar{\sigma}(\ell)$ from the Brodmann area targeted by TMS. In each subsequent iterations, a new ROI with the largest $\bar{\sigma}(\ell)$ is sequentially added, under the constraint that it is not physically overlapping with the previously selected ROIs. This procedure continues until $\lfloor M/3 \rfloor$ ROIs are selected.

References

- Anderson T, 2003 An Introduction to Multivariate Statistical Analysis. Wiley Series in Probability and Statistics, Wiley.
- Bai Y, Xia X, Kang J, Yin X, Yang Y, He J, Li X, 2016 Evaluating the effect of repetitive transcranial magnetic stimulation on disorders of consciousness by using tms-eeg. *Frontiers in neuroscience* 10, 473. [PubMed: 27812319]
- Berg P, Scherg M, 1994 A fast method for forward computation of multiple-shell spherical head models. *Electroencephalography and clinical neurophysiology* 90, 58–64.
- ter Braack EM, de Vos CC, van Putten MJ, 2015 Masking the auditory evoked potential in tms-eeg: a comparison of various methods. *Brain topography* 28, 520–528. [PubMed: 23996091]
- Casali AG, Casarotto S, Rosanova M, Mariotti M, Massimini M, 2010 General indices to characterize the electrical response of the cerebral cortex to tms. *Neuroimage* 49, 1459–1468. [PubMed: 19770048]

- Casali AG, Gosseries O, Rosanova M, Boly M, Sarasso S, Casali KR, Casarotto S, Bruno MA, Laureys S, Tononi G, et al., 2013 A theoretically based index of consciousness independent of sensory processing and behavior. *Science translational medicine* 5, 198ra105–198ra105.
- Chang JY, Pigorini A, Massimini M, Tononi G, Nobili L, Van Veen B, 2012 Multivariate autoregressive models with exogenous inputs for intracerebral responses to direct electrical stimulation of the human brain. *Frontiers in Human Neuroscience* 6, 317. [PubMed: 23226122]
- Cheung BLP, Nowak R, Lee HC, Drongelen W, Van Veen BD, 2012 Cross validation for selection of cortical interaction models from scalp eeg or meg. *IEEE Transactions on Biomedical Engineering* 59, 504–514. [PubMed: 22084038]
- Cheung BLP, Riedner BA, Tononi G, Van Veen BD, 2010 Estimation of cortical connectivity from eeg using state-space models. *Biomedical Engineering, IEEE Transactions on* 57, 2122–2134.
- Cona F, Zavaglia M, Massimini M, Rosanova M, Ursino M, 2011 A neural mass model of interconnected regions simulates rhythm propagation observed via tms-eeg. *NeuroImage* 57, 1045–1058. [PubMed: 21600291]
- Dentico D, Cheung BL, Chang JY, Guokas J, Boly M, Tononi G, Van Veen B, 2014 Reversal of cortical information flow during visual imagery as compared to visual perception. *Neuroimage* 100, 237–243. [PubMed: 24910071]
- Dias JG, Wedel M, 2004 An empirical comparison of em, sem and mcmc performance for problematic gaussian mixture likelihoods. *Statistics and Computing* 14, 323–332.
- Esser S, Huber R, Massimini M, Peterson M, Ferrarelli F, Tononi G, 2006 A direct demonstration of cortical ltp in humans: a combined tms/eeg study. *Brain research bulletin* 69, 86–94. [PubMed: 16464689]
- Esser SK, Hill S, Tononi G, 2009 Breakdown of effective connectivity during slow wave sleep: investigating the mechanism underlying a cortical gate using large-scale modeling. *Journal of Neurophysiology* 102, 2096–2111. [PubMed: 19657080]
- Fecchio M, Pigorini A, Comanducci A, Sarasso S, Casarotto S, Premoli I, Derchi CC, Mazza A, Russo S, Resta F, et al., 2017 The spectral features of eeg responses to transcranial magnetic stimulation of the primary motor cortex depend on the amplitude of the motor evoked potentials. *PloS one* 12, e0184910. [PubMed: 28910407]
- Ferrarelli F, Massimini M, Peterson MJ, Riedner BA, Lazar M, Murphy MJ, Huber R, Rosanova M, Alexander AL, Kalin N, et al., 2008 Reduced evoked gamma oscillations in the frontal cortex in schizophrenia patients: a tms/eeg study. *American Journal of Psychiatry* 165, 996–1005. [PubMed: 18483133]
- Ferrarelli F, Massimini M, Sarasso S, Casali A, Riedner BA, Angelini G, Tononi G, Pearce RA, 2010 Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness. *Proceedings of the National Academy of Sciences* 107, 2681–2686.
- Friston K, Henson R, Phillips C, Mattout J, 2006 Bayesian estimation of evoked and induced responses. *Human brain mapping* 27, 722–735. [PubMed: 16453291]
- Friston KJ, Harrison L, Penny W, 2003 Dynamic causal modelling. *Neuroimage* 19, 1273–1302.
- Gaillard A, 1988 Problems and paradigms in erp research. *Biological psychology* 26, 91–109. [PubMed: 3061482]
- Gaillard R, Dehaene S, Adam C, Clemenceau S, Hasboun D, Baulac M, Cohen L, Naccache L, 2009 Converging intracranial markers of conscious access. *PLoS Biol* 7, e1000061.
- Garrido MI, Kilner JM, Kiebel SJ, Friston KJ, 2007 Evoked brain responses are generated by feedback loops. *Proceedings of the National Academy of Sciences* 104, 20961–20966.
- Golub GH, Heath M, Wahba G, 1979 Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* 21, 215–223.
- Gosseries O, Sarasso S, Casarotto S, Boly M, Schnakers C, Napolitani M, Bruno MA, Ledoux D, Tshibanda JF, Massimini M, et al., 2015 On the cerebral origin of eeg responses to tms: insights from severe cortical lesions. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation* 8, 142–149.
- Hui HB, Pantazis D, Bressler SL, Leahy RM, 2010 Identifying true cortical interactions in meg using the nulling beamformer. *NeuroImage* 49, 3161–3174. [PubMed: 19896541]

- Iramina K, Maeno T, Nonaka Y, Ueno S, 2003 Measurement of evoked electroencephalography induced by transcranial magnetic stimulation. *Journal of applied physics* 93, 6718–6720.
- Kay SM, 1993 Fundamentals of statistical signal processing, volume i: estimation theory
- Komssi S, Savolainen P, Heiskala J, Kähkönen S, 2007 Excitation threshold of the motor cortex estimated with transcranial magnetic stimulation electroencephalography. *Neuroreport* 18, 13–16. [PubMed: 17259853]
- Kundu B, Chang JY, Postle BR, Van Veen BD, 2015 Context-specific differences in fronto-parieto-occipital effective connectivity during short-term memory maintenance. *NeuroImage* 114, 320–327. [PubMed: 25863155]
- Lamme VA, Roelfsema PR, 2000 The distinct modes of vision ordered by feedforward and recurrent processing. *Trends in neurosciences* 23, 571–579. [PubMed: 11074267]
- Lempel A, Ziv J, 1976 On the complexity of finite sequences. *IEEE Transactions on information theory* 22, 75–81.
- Limpiti T, Van Veen BD, Wakai RT, 2006 Cortical patch basis model for spatially extended neural activity. *Biomedical Engineering, IEEE Transactions on* 53, 1740–1754.
- Lütkepohl H, 2005 New introduction to multiple time series analysis. Springer Science & Business Media
- Lv J, Simpson DM, Bell SL, 2007 Objective detection of evoked potentials using a bootstrap technique. *Medical engineering & physics* 29, 191–198. [PubMed: 16621656]
- Massimini M, Ferrarelli F, Esser SK, Riedner BA, Huber R, Murphy M, Peterson MJ, Tononi G, 2007 Triggering sleep slow waves by transcranial magnetic stimulation. *Proceedings of the National Academy of Sciences* 104, 8496–8501.
- Massimini M, Ferrarelli F, Huber R, Esser SK, Singh H, Tononi G, 2005 Breakdown of cortical effective connectivity during sleep. *Science* 309, 2228–2232. [PubMed: 16195466]
- Mattout J, Phillips C, Penny WD, Rugg MD, Friston KJ, 2006 Meg source localization under multiple constraints: an extended bayesian framework. *NeuroImage* 30, 753–767. [PubMed: 16368248]
- McCubbin J, Yee T, Vrba J, Robinson S, Murphy P, Eswaran H, Preissl H, 2008 Bootstrap significance of low snr evoked response. *Journal of neuroscience methods* 168, 265–272.
- Mutanen T, Mäki H, Ilmoniemi RJ, 2013 The effect of stimulus parameters on tms-eeG muscle artifacts. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation* 6, 371–376.
- Oizumi M, Albantakis L, Tononi G, 2014 From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Comput Biol* 10, e1003588. [PubMed: 24811198]
- Oizumi M, Amari S.i., Yanagawa T, Fujii N, Tsuchiya N, 2016 Measuring integrated information from the decoding perspective. *PLoS Comput Biol* 12, e1004654. [PubMed: 26796119]
- Pantazis D, Nichols TE, Baillet S, Leahy RM, 2005 A comparison of random field theory and permutation methods for the statistical analysis of meg data. *NeuroImage* 25, 383–394. [PubMed: 15784416]
- Phillips C, Mattout J, Rugg MD, Maquet P, Friston KJ, 2005 An empirical bayesian solution to the source reconstruction problem in eeg. *NeuroImage* 24, 997–1011. [PubMed: 15670677]
- Pigorini A, Sarasso S, Proserpio P, Szymanski C, Arnulfo G, Casarotto S, Fecchio M, Rosanova M, Mariotti M, Russo GL, et al., 2015 Bistability breaks deterministic responses to intracortical stimulation during non-rem sleep. *Neuroimage* 112, 105–113. [PubMed: 25747918]
- Rosanova M, Gosseries O, Casarotto S, Boly M, Casali AG, Bruno MA, Mariotti M, Boveroux P, Tononi G, Laureys S, et al., 2012 Recovery of cortical effective connectivity and recovery of consciousness in vegetative patients. *Brain* 135, 1308–1320. [PubMed: 22226806]
- Sarasso S, Boly M, Napolitani M, Gosseries O, Charland-Verville V, Casarotto S, Rosanova M, Casali AG, Brichant JF, Boveroux P, et al., 2015 Consciousness and complexity during unresponsiveness induced by propofol, xenon, and ketamine. *Current Biology* 25, 3099–3105. [PubMed: 26752078]
- Sekihara K, Nagarajan SS, 2008 Adaptive spatial filters for electromagnetic brain imaging. Springer Science & Business Media

- Shumway RH, Stoer DS, 1982 An approach to time series smoothing and forecasting using the em algorithm. *Journal of time series analysis* 3, 253–264.
- Thut G, Ives JR, Kampmann F, Pastor MA, Pascual-Leone A, 2005 A new device and protocol for combining tms and online recordings of eeg and evoked potentials. *Journal of neuroscience methods* 141, 207–217. [PubMed: 15661302]
- Virtanen J, Ruohonen J, Nääätänen R, Ilmoniemi R, 1999 Instrumentation for the measurement of electric brain responses to transcranial magnetic stimulation. *Medical & biological engineering & computing* 37, 322–326. [PubMed: 10505382]
- Zhang Z, 1995 A fast method to compute surface potentials generated by dipoles within multilayer anisotropic spheres. *Physics in medicine and biology* 40, 335. [PubMed: 7732066]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Highlights

- A modeling methodology for single-trial electroencephalogram (EEG) response to transcranial magnetic stimulation (TMS) is proposed.
- The model generates model responses that closely characterize the TMS evoked potentials.
- Performance difference between fully connected and disconnected network specifications indicates that the complex TMS evoked potentials in wakefulness are supported by recurrent network connections among different nodes.

Figure 1:
Schematic diagrams of two multivariate autoregressive with exogenous stimulation (MVARX) network models. Each gray circle represents a cortical ROI, and directed edges de-note non-zero MVARX coefficients. The $\mathbf{B}_i$ represent the direct or indirect feedforward effect of TMS on each ROI. (a) Segregated model assumes each cortical region acts independently of all others. (b) Integrated model represents network interactions between cortical regions.

Figure 2:

Example TMS-evoked potentials and the duration of significant response across the subject population. (a) and (b) Butter y plots of TMS evoked potentials from 60 channels, recorded from the same subject during wakefulness and sleep, respectively. The red portions of the traces indicate the time intervals during which TMS induced a statistically significant response (calculated as in Massimini et al. (2005) - bootstrap statistics, $p < 0.01$). (c) Extent of TMS-induced response for the seven subjects in wakefulness (red) and sleep (blue). There is no statistically significant difference between wakefulness and sleep (Wilcoxon rank-sum test, $p < 0.81$).

Figure 3:

(a) Cumulative PCI for each subject (thin lines) and the group (thick lines) for both wakefulness (red) and sleep (blue). (b) The rate of divergence of difference in cumulative PCI between wakefulness and sleep, Δ PCI, calculated from single-subject differences between cumulative PCI during wakefulness and sleep, with 25-ms time bins. Statistical significance with respect to zero across bins is indicated in asterisks (significance level $\alpha = 0.01$, Mann-Whitney).

Figure 4:
The TMS stimulation site and ROIs chosen for the analysis of Subject 1. The number associated with each ROI (in yellow) represents the order in which the ROIs were identified by the ROI selection procedure (see Appendix). TMS stimulation was applied to the left parietal cortex, which is shown in red.

Figure 5:

Measured and modeled TMS-evoked response of Subject 1. (a) and (b) Butter y display of 60-channel TMS-evoked response of subject 1 in wakefulness and sleep, respectively. (c) and (d) Measured and modeled TMS-evoked response in selected channels for Subject 1 in wakefulness and sleep, respectively.

Figure 6:

Per-channel normalized squared response error (NSRE) between the measured and modeled TMS-evoked response for Subject 1 in wakefulness assuming a maximum feedforward volley duration of 100 ms ($\ell = 10$). The light red vertical bars indicate channels whose waveforms are displayed in Figure 5.

Figure 7:

Per-channel NSRE between the measured and modeled TMS-evoked response for Subject 1 in sleep assuming a maximum feedforward volley duration of 100 ms ($\ell = 10$). The light blue vertical bars indicate channels whose waveforms are displayed in Figure 5.

Figure 8:

Global NSRE between measured and modeled TMS-evoked responses summed over all channels as a function of global mean eld power (GMFP) ratio for all seven subjects. (a) Wakefulness, maximum feedforward volley duration 50 ms ($\ell=5$). (b) Sleep, maximum feedforward volley duration 50 ms ($\ell=5$). (c) Wakefulness, maximum feedforward volley duration 100 ms ($\ell=10$). (d) Sleep, maximum feedforward volley duration 100 ms ($\ell=10$). (e) Wakefulness, maximum feedforward volley duration 150 ms ($\ell=15$). (f) Sleep, maximum feedforward volley duration 150 ms ($\ell=15$).

Figure 9:

Cumulative normalized squared response error (NSRE). Thin lines indicate individual subject values while thick lines indicate averages across seven subjects. (a) Wakefulness, maximum feedforward volley duration 50 ms ($\ell=5$). (b) Sleep, maximum feedforward volley duration 50 ms ($\ell=5$). (c) Wakefulness, maximum feedforward volley duration 100 ms ($\ell=10$). (d) Sleep, maximum feedforward volley duration 100 ms ($\ell=10$). (e) Wakefulness, maximum feedforward volley duration 150 ms ($\ell=15$). (f) Sleep, maximum feedforward volley duration 150 ms ($\ell=15$).

Figure 10:

Difference between segregated and integrated models as measured by cross-validated mean squared prediction error (MSPE) for ten partitions of the trials. (a) Maximum feedforward volley duration 50 ms ($\ell=5$). (b) Maximum feedforward volley duration 100 ms ($\ell=10$). (c) Maximum feedforward volley duration 150 ms ($\ell=15$).