Test-Retest Reliability of Resting-State Magnetoencephalography Power in Sensor and Source Space

María Carmen Martín-Buro,[1,2,3]* Pilar Garcés,[1,3,4] and Fernando Maestú[1,3,5]

[1]*Laboratory of Cognitive and Computational Neuroscience, Centre for Biomedical Technology, Madrid, Spain*
[2]*Psychology Division, Cardenal Cisneros University College, Complutense University of Madrid, Spain*
[3]*Biomedical Research Networking Center in Bioengineering Biomaterials and Nanomedicine (CIBER-BBN), Madrid, Spain*
[4]*Department of Applied Physics III, Faculty of Physics, Complutense University of Madrid, Spain*
[5]*Department of Basic Psychology II, Faculty of Psychology, Complutense University of Madrid, Spain*

Abstract: Several studies have reported changes in spontaneous brain rhythms that could be used as clinical biomarkers or in the evaluation of neuropsychological and drug treatments in longitudinal studies using magnetoencephalography (MEG). There is an increasing necessity to use these measures in early diagnosis and pathology progression; however, there is a lack of studies addressing how reliable they are. Here, we provide the first test-retest reliability estimate of MEG power in resting-state at sensor and source space. In this study, we recorded 3 sessions of resting-state MEG activity from 24 healthy subjects with an interval of a week between each session. Power values were estimated at sensor and source space with beamforming for classical frequency bands: delta (2–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), low beta (13–20 Hz), high beta (20–30 Hz), and gamma (30–45 Hz). Then, test-retest reliability was evaluated using the intraclass correlation coefficient (ICC). We also evaluated the relation between source power and the within-subject variability. In general, ICC of theta, alpha, and low beta power was fairly high (ICC > 0.6) while in delta and gamma power was lower. In source space, fronto-posterior alpha, frontal beta, and medial temporal theta showed the most reliable profiles. Signal-to-noise ratio could be partially responsible for reliability as low signal intensity resulted in high within-subject variability, but also the inherent nature of some brain rhythms in resting-state might be driving these reliability patterns. In conclusion, our results described the reliability of MEG

Contract grant sponsor: PICATA predoctoral contract, Moncloa Campus of International Excellence (UCM–UPM)

*Correspondence to: M. Carmen Martín-Buro, Laboratory of Cognitive and Computational Neuroscience, Centre for Biomedical Technology (CTB), Campus de Montegancedo s/n, Pozuelo de Alarcón 28223, Spain. E-mail: carmen.martinburo@ctb.upm.es

Received for publication 7 July 2015; Revised 19 September 2015; Accepted 29 September 2015.
DOI: 10.1002/hbm.23027
Published online 14 October 2015 in Wiley Online Library (wileyonlinelibrary.com).

power estimates in each frequency band, which could be considered in disease characterization or clinical trials. *Hum Brain Mapp 37:179–190, 2016.*

Key words: MEG; reliability; resting state; brain rhythms; test-retest; spectral power; signal to noise ratio; intraclass correlation coefficient

INTRODUCTION

Extending our understanding of the functional architecture of the human brain necessarily requires the study of brain rhythms. Brain rhythms are periodic fluctuations in the excitability of a group of neurons, which have the intrinsic ability to resonate and oscillate at multiple frequencies [Buzsáki, 2012]. The synchronous activity of neuronal ensembles works as a coordination and communication mechanism creating precise temporal windows for transmitting and representing information [Schnitzler and Gross, 2005]. Here, timing information is critical to the understanding of sensory and cognitive processes, both in response to a stimulus and in resting-state. Such time series data can be obtained by means of high temporal resolution imaging techniques such as magnetoencephalography (MEG) and electroencephalography (EEG).

Brain rhythms take place even in the absence of a task and provide valuable information about predicting successful performance in a cognitive task. Furthermore, they can help describe the organization of the brain in several neurological and psychiatric conditions. For these reasons, resting-state provides an interesting line of study for human brain function. Supporting this idea, several spectral measures derived from MEG and EEG recordings in resting, such as absolute or relative power of each frequency band, have been applied to brain rhythm analysis and have been shown to reflect changes related to normal and pathological states. These findings reported disturbed oscillatory activity in developmental disorders, such as autism [Cornew et al., 2013], in psychiatric disorders, such as schizophrenia [Fehr et al., 2001], and neurodegenerative diseases, such as Multiple Sclerosis [Van Der Meer et al., 2013] or Alzheimer's disease [Fernández et al., 2006] and Mild Cognitive Impairment [Garcés et al., 2013]. Additionally, drug-related changes versus placebo have been assessed in MEG pharmacological studies (e.g., the case of methylphenidate treatment in children with Attention Deficit Hyperactivity Disorder [Wienbruch et al., 2005]). Apart from providing important insights into differences between groups, resting-state recordings have recently monitored longitudinal changes like disease progression. For example, a slowing in oscillatory brain activity was found when cognitive decline advanced in patients with Parkinson's Disease [Olde Dubbelink et al., 2013].

This potential clinical application has led to the question of how reliable the power spectrum is at resting-state with MEG. Before implementing the use of brain oscillations as clinical biomarkers, follow up drug testing or health status monitoring, the demonstration that power spectrum measures are reliable is needed. If spectral measures vary significantly between sessions, the statistical power of these measures is decreased, which limits the attribution of the evaluated effect to the drug, treatment or disease [Deuker et al., 2009; Telesford et al., 2013].

To date, MEG reliability has been addressed for functional connectivity measures [Deuker et al., 2009; Jin et al., 2011; Leighton et al., 2011] whereas the reliability of the power in the classical frequency bands has only been reported with EEG. Overall, power estimates with EEG showed high reliability not only in the different frequency bands [McEvoy et al., 2000], but also across the whole composition of the spectrum as a set [Fingelkurts et al., 2006]. Furthermore, reliability values in these studies were lower in resting-state than during a cognitive task.

Prior literature about the reliability of the resting-state power is only available for EEG. In these studies, power in the eyes-closed condition was shown to be more reliable than eyes-open resting-state condition [Fingelkurts et al., 2006; Pollock et al., 1991]. In addition, reliability varied depending on the frequency bands. For EEG power, alpha and beta bands, followed by theta, were the most reliable frequency bands in different test-retest intervals [Burgess and Gruzelier, 1993; Cannon et al., 2012; Kondacs and Szabó, 1999; McEvoy et al., 2000]. However, other reports indicated similar, but higher, reliability values in beta [Pollock et al., 1991] and theta [Gudmundsson et al., 2007] than in alpha. Consistently among the studies however, power in gamma and delta was the least reliable in resting-state [Gasser et al., 1985; Gudmundsson et al., 2007; Kondacs and Szabó, 1999; Pollock et al., 1991]. Reliability has been assessed with both absolute and relative power. With absolute power, reliability tended to be greater than with relative power [Kondacs and Szabó, 1999; Pollock et al., 1991], however some studies did not find considerable differences [Gudmundsson et al., 2007]. In this regard, it was recommended [Pollock et al., 1991] to use absolute rather than relative measures because of their straightforward interpretability and high test-retest reliability.

Finally, with regard to the reliability distribution, Fingelkurts et al. (2006) described a decrease from frontal to occipital sensors, whereas recordings with more than twenty sensors [Gudmundsson et al., 2007; McEvoy et al., 2000] found the opposite pattern, that is, power in sensors covering the frontal area of the scalp was less stable than in the occipital ones. In consequence, the spatial distribution of the reliability of resting-state power spectrum remains unclear. Moreover, there is little data evaluating the reliability of these parameters in the source space. Initial attempts focused on the reliability of somatosensory [Schaefer et al., 2002] and auditory [Atcherson et al., 2006] evoked responses—identifiable active stable sources. Although, the only resting-state source space study was restricted to the reliability of eight selected regions [Cannon et al., 2012]. Following this approach, resting-state absolute power was highly reliable in all of the selected regions with a decrease in the anterior cingulate and left prefrontal in eyes-closed condition. In summary, despite having worthwhile results with sensor space EEG, there is no available evidence about the reliability of the power in resting-state in MEG and its spatial distribution. Moreover, although EEG and MEG share the same signal origin, the signal recorded in each technique is quite different (i.e., electrical vs. magnetic component and necessity of reference in EEG vs. reference-free in MEG). This could lead to different reliability findings between techniques. Here, we provide the first reliability assessment of resting-state power with MEG at both sensor and source space. To achieve this aim, beamforming was used to estimate MEG power distribution in source space. Following published procedures [Telesford et al., 2013], we calculated intraclass correlation coefficient (ICC) as the index of reliability [Shrout and Fleiss, 1979]. In addition, to explore how reliability parameters are modulated by the intensity of the MEG signal, we evaluated the relationship between signal-to-noise ratio (SNR) and within-subject variability. The initial hypothesis is that sensor space reliability could be similar to those previously obtained with EEG and may even be higher in the frequency bands with better SNR for instance in alpha band. Similar findings could be found at the source space with higher spatial resolution.

MATERIALS AND METHODS

Subjects

Twenty-four healthy volunteers (14 female, 10 male; mean age 28.86 years; range 20–41; 2 left-handed) with normal or corrected-to-normal vision participated in this experiment. None of the participants had histories of psychiatric, neurologic, or chronic medical conditions. All of the participants were informed about the aim of the study and signed a written informed consent before participating. The local Ethics Committee approved the investigation.

MEG Data Acquisition and Preprocessing

Each subject participated in three MEG recording sessions at the Centre for Biomedical Technology (Madrid, Spain) with a test-retest interval of seven days between each session. To minimize the impact of their circadian rhythm, the time of the day was constant across the recordings. For each subject and session, three blocks of MEG data were recorded successively: (1) 4 min of eyes-open resting-state, (2) 4 min of eyes-closed resting-state, and (3) 2 min of empty room signal.

MEG data was acquired inside a magnetically shielded room (Vacuumschmelze GmbH, Hanau, Germany) with a 306-channel Vectorview system (Elekta Neuromag, Helsinki) which includes 204 planar gradiometers and 102 magnetometers. Data was sampled at 1000 Hz with an online band-pass filtering at 0.1–300 Hz.

Four HPI coils attached to the subject (two on the forehead and one on each mastoid) continuously monitored head position inside the MEG helmet. Their position with respect to three fiducial points (nasion, left- and right preauricular) was determined prior to the MEG recording with a 3D digitizer (Fastrack Polhemus), along with the subject's headshape. To track eye movements and blinks, vertical electrooculograms were acquired from two electrodes placed above and below the left eye with a reference placed on the left earlobe.

Maxfilter 2.2 software (Elekta Neuromag) was used to remove external noise from MEG data using the temporal extension of signal space separation with movement compensation (t-SSS) [Taulu and Simola, 2006], correlation limit of 0.9 and a time window of 10 seconds. Then, blinks, muscle and jumps artifacts were automatically detected with FieldTrip toolbox [Oostenveld et al., 2011] for Matlab software. The artifacts were located and continuous time series were segmented into artifact-free epochs of 4 s. Each subject obtained, per condition, a minimum of 13 epochs (mean 26.8 ± 6.3) in eyes-open condition, 20 epochs (mean 29.9 ± 3.6) in eyes-closed, and 7 (mean 21.3 ± 8.5) in the empty room. We checked the possible influence of the number of clean epochs on reliability and no strong correlation was found (see Supporting Information Table SI).

Sensor Space Power

Power spectra were obtained for all artifact-free epochs with a multitaper method using discrete prolate spheroidal sequences tapers and 1 Hz smoothing, as implemented in FieldTrip. Frequencies of interest were defined in 0.5 Hz steps from 1 to 20 Hz and 2 Hz steps from 22 to 100 Hz. Then, the power in delta (2–4Hz), theta (4–8Hz), alpha (8–13Hz), beta1 (13–20), beta2 (20–30), and gamma (30–45Hz) were obtained by averaging power estimates over trials. The mean alpha frequency was calculated as the center of gravity of the power spectrum within the (8–13Hz), following [Klimesch, 1999]. This was performed for every MEG sensor, subject, and session separately.

Source Space Power

To ensure an accurate source and forward model, source reconstruction was only performed for 16/24 subjects, for which a T1-weighted MRI was available. For these subjects, source locations were placed regularly over their individual cortical surface with 6mm spacing by using Freesurfer (version 5.1.0, Fischl et al., 2002; Ségonne et al., 2007) and MNE softwares [Gramfort et al., 2014]. The forward model was solved with a 3-shell boundary element method: inner skull, outer skull and skin surfaces were extracted from the subject's MRI with NFT software [Acar and Makeig, 2010], and leadfields were computed with MNE.

Then, the absolute power for each source location and frequency band was computed with a frequency-domain beamformer (DICS) [Gross et al., 2001]. To avoid mixing sensor information with different noise profiles or resort to an arbitrary scaling, we performed source reconstruction with magnetometer and gradiometer data separately. As both sensor types produced similar results, the main manuscript presents the source space reliability obtained with magnetometers, although gradiometer results can be found in the supplementary material. We note that magnetometer and gradiometer data are not independent measures after preprocessing, since they both were employed in the t-SSS filtering and result from the back-projection of the same inside components.

As required for the DICS computation, sensor space cross-spectral density matrices were first computed for each frequency band, using FieldTrip. Then, beamformer filters $\boldsymbol{w}(\boldsymbol{r_i},f_b)$ ($N_{\text{sensors}}\times 3$ matrices) were computed for each source location $\boldsymbol{r_i}$ and frequency band f_{b}, following [Gross et al., 2001] and using 5% regularization and an unconstrained source orientation. The power for each source location $\boldsymbol{r_i}$ and frequency band f_{b} can then be written as [Sekihara and Nagarajan, 2008]:

$$P(\boldsymbol{r_i},\ f_{\text{b}})=\vartheta_{\max}\left(\boldsymbol{w}^H(\boldsymbol{r_i},f_{\text{b}})\boldsymbol{C}(f_{\text{b}})\boldsymbol{w}(\boldsymbol{r_i},f_{\text{b}})\right) \tag{1}$$

where, $\boldsymbol{C}(f_{\text{b}})$, superscript H and $\vartheta_{\max}(\ldots)$ refer to the cross-spectral density matrix for frequency band f_{b}, the Hermitian transpose and the maximum eigenvalue of a matrix $(\ldots)$, respectively.

However, $P(\boldsymbol{r_i},\ f_{\text{b}})$ was not employed directly in the reliability analysis. In fact, beamforming estimates are biased, particularly towards the center of the brain, where the SNR of MEG signals is lowest. Therefore, the following normalized power estimates were employed:

$$Z(\boldsymbol{r_i},\ f_{\text{b}})=\frac{P(\boldsymbol{r_i},\ f_{\text{b}})}{N(\boldsymbol{r_i},\ f_{\text{b}})} \tag{2}$$

where $N(\boldsymbol{r_i},\ f_{\text{b}})$ is a noise estimate in source space, and is obtained by employing (1) and substituting the original cross-spectral density matrix $\boldsymbol{C}(f_{\text{b}})$ by $\boldsymbol{C_N}(f_{\text{b}})$ (the cross-spectral density matrix of a noise estimate). Although the noise is sometimes assumed to be independent and uncorrelated across sensors [yielding a diagonal $\boldsymbol{C_N}(f_{\text{b}})$], we considered that this assumption is simplistic and poorly reflects the specific noise characteristics of our data. For instance, we preprocessed the raw MEG recordings with a t-SSS filtering, which reduces the dimensionality of the data. Therefore, we employed empty room recordings for the noise estimation: $\boldsymbol{C_N}(f_{\text{b}})$ was computed from the empty-room recordings following the same analysis pipeline as for the resting-state data: t-SSS, artifact detection, segmentation into clean epochs and spectral estimation.

Finally, the noise-normalized power estimates $Z(\boldsymbol{r_i},\ f_{\text{b}})$ were transformed into MNI space. First, a template mesh of source locations was created from the subjects with Freesurfer. Then, $Z(\boldsymbol{r_i},\ f_{\text{b}})$ were transformed from the subject's surface to the standard and smoothing with a 15 mm moving average filter was applied. Overall, this yielded $N_{\text{subjects}}\times N_{\text{sessions}}=16\times 3$ power estimates for each template source location, frequency band, and condition. We note that these power values were computed for each subject and session separately.

Additionally, relative powers were estimated for each frequency band. For that, beamformer filters $\boldsymbol{w}(\boldsymbol{r_i})$ were computed from the average cross-spectral density matrices over the (2–45 Hz) range, and then applied to individual-band cross-spectral density matrices $\boldsymbol{C}(f_{\text{b}})$ as in (1). Then, the relative power of each frequency band was obtained by normalizing the power estimate $P(\boldsymbol{r_i},\ f_{\text{b}})$ with the overall sum power over all frequency bands.

Intraclass-Correlation

Test-retest reliability of any score can be defined as [Weir, 2005]:

$$R=\frac{\sigma_{\text{t}}{}^2}{\sigma_{\text{t}}{}^2+\sigma_{\text{e}}{}^2} \tag{3}$$

where $\sigma_{\text{t}}{}^2$ and $\sigma_{\text{e}}{}^2$ are the true score variance and the error variance, respectively. It ranges from 0 to 1, where 0 indicates no reliability and 1 indicates perfect reliability. R was assessed from the between- and within-subject variability [McGraw and Wong, 1996; Shrout and Fleiss, 1979; Weir, 2005] with a type 1-1 ICC:

$$\text{ICC}=\frac{\text{MSR}-\text{MSW}}{\text{MSR}+(N_{\text{sessions}}-1)\text{MSW}} \tag{4}$$

where MSR and MSW are the between-subjects and within-subject mean square values and derive from the $N_{\text{subjects}}\times N_{\text{sessions}}$ score matrix. This matrix contains the power estimations of a given sensor (or source), frequency band, and condition (eyes-closed or eyes-open resting-state). Additionally, the null-hypothesis $H_0:\ \rho=0$ was tested by computing the F-value $F_0=\frac{\text{MSR}}{\text{MSW}}$, which follows an F-distribution with N_{subjects}-1 and $N_{\text{subjects}}(N_{\text{sessions}}\text{-}1)$ degrees of freedom, and then the corresponding P-value was computed as in Shrout and Fleiss [1979].

As computed in (4), the ICC, depends both on the between- and within- subject variability. It increases with decreasing within-subject variability and with increasing

TABLE I. ICC of the average power over five MEG sensor regions, for each frequency band and condition

		Delta	Theta	Alpha	Low beta	High beta	Gamma	Mean alpha frequency
Eyes-open	Occipital	0.52	0.79	0.86	0.86	0.75	0.64	0.91
	Left temporal	0.72	0.75	0.92	0.85	0.85	0.59	0.85
	Right temporal	0.79	0.76	0.83	0.82	0.79	0.59	0.89
	Parietal	0.76	0.85	0.86	0.86	0.85	0.77	0.93
	Frontal	0.78	0.82	0.85	0.76	0.74	0.48	0.70
Eyes-closed	Occipital	0.78	0.86	0.94	0.91	0.89	0.59	0.88
	Left temporal	0.69	0.74	0.92	0.87	0.75	0.50	0.87
	Right temporal	0.55	0.83	0.95	0.83	0.75	0.49	0.89
	Parietal	0.66	0.82	0.93	0.85	0.79	0.63	0.90
	Frontal	0.90	0.54	0.84	0.74	0.70	0.53	0.82
Empty room	Occipital	0.02	0.19	0.38	0.42	0.17	0.34	−0.09
	Left temporal	0.06	0.01	0.03	0.07	−0.01	0.08	−0.08
	Right temporal	−0.01	−0.01	0.05	0.14	0.02	0.17	−0.03
	Parietal	−0.07	−0.07	−0.02	−0.01	0.01	−0.00	−0.15
	Frontal	−0.09	−0.11	−0.05	−0.01	−0.04	−0.03	0.03

The right column contains the ICC of the mean alpha frequency extracted.

between-subject variability. It characterizes therefore the within-subject variations within a given sample. It was computed for each sensor, source, frequency band, and condition separately.

Within-Subject Variability vs. Source Power

To test whether within-subject variability is dependent on the source intensity, we evaluated the joint distribution of both magnitudes. For a given subject, source position and frequency band, the representative power was simply defined as the average source power over the three MEG sessions: $\bar{Z}=\text{mean}(Z_s)|_{s=1,2,3}$, where Z_s represents the source power for a session s. The corresponding within-subject variability was defined as the relative inter-sessions variations: $\Delta z=\frac{\text{std}(Z_s)|_{s=1,2,3}}{\bar{Z}}$.

Furthermore, for a given frequency band, we computed the bivariate histogram of $\bar{Z}$ and Δz across all subjects and source positions. Then, in order to estimate the conditional probability distribution of Δz given $\bar{Z}$, we normalized the histogram by the sum of its counts for each $\bar{Z}$ bin separately.

RESULTS

Power Reliability in Sensor Space

We first estimated sensor space power reliability in eyes-open, eyes-closed, and empty room conditions. Table I shows ICC values for average power over the five helmet areas for all frequency bands (see Supporting Information Table SII for average relative power). Overall, ICC values ranged from 0.48 to 0.95 in resting-state and, as expected, the empty room ICC values were appreciably lower and nearly zero. Figure 1 displays the ICC distribution in the sensor space. In general, the power in sensors covering the parieto-occipital area of the scalp remained reliable among all the frequency bands. Moreover, sensor space power was also found to be highest for sensors in parietal areas (see Supporting Information Figure S1 for the topographies of the average sensor power for all frequency bands).

Reliability varied somewhat across the frequency bands and the scalp areas. Delta power showed the highest ICC values in the frontal and parietal areas. Theta power remained highly reliable (range 0.74–0.86) except in the frontal area in the eyes-open condition (ICC = 0.54). Alpha power showed the highest ICC (range 0.83–0.95) among all the frequency bands and scalp areas. Moreover, ICC values were slightly higher in the eyes-closed condition. Low beta power revealed high ICC values (range 0.74–0.91) in almost all the sensors, especially in the occipital and parietal areas. ICC values in high beta were slightly smaller (range 0.70–0.89) than in low beta, especially in the frontal and temporal sensors, although the ICC distribution was quite similar. Finally, gamma power showed the lowest ICC across all the frequency bands and only the sensors covering the parietal area of the scalp showed fairly high ICC values (range 0.63–0.77).

Power Reliability in Source Space

ICC was calculated for the power estimates of each source location and frequency band, and is represented in Figures 2 and 3 for eyes-open and eyes-closed condition, respectively (see Supporting Information Figures S4 and S5 for gradiometer data). In general, these source space results were similar to the previously described sensor space results. However, highest ICC values were obtained in widespread regions for alpha, low beta and theta bands. However, for delta, high beta and gamma bands, reliability was medium to low

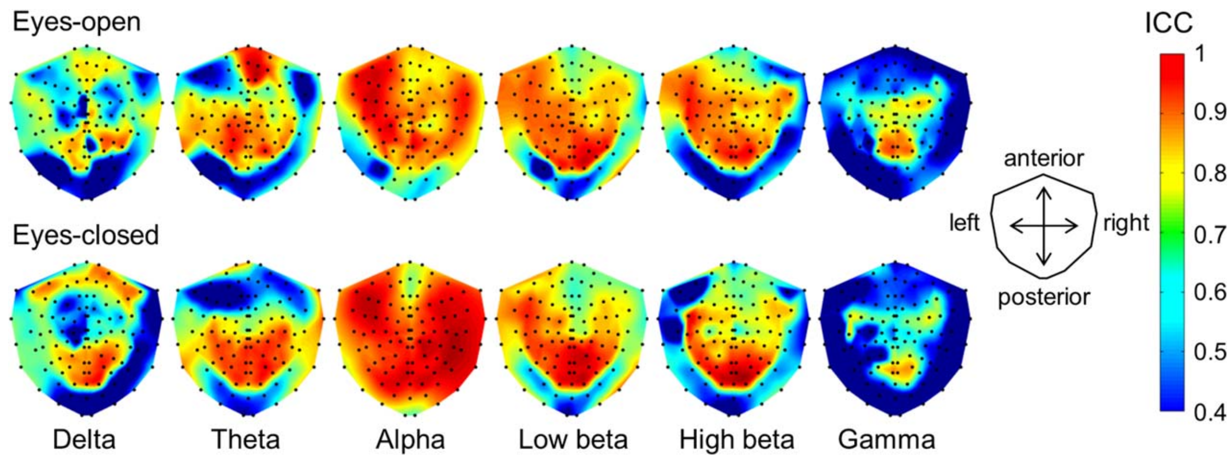

Figure 1.
Topography map of the ICC of sensor-space power for each resting state condition, frequency band, sensor. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

(ICC < 0.6) for most brain areas, although high ICC values (ICC > 0.6) were found in restricted brain regions.

Although power in delta band showed medium to low reliability, some frontal regions such as superior and orbitofrontal cortex revealed fairly high ICC values. The power reliability maps in the theta band showed a peak of ICC in regions surrounding the central sulcus such as superior parietal and superior frontal gyrus, paracentral, and posterior cingulate. Note that parahippocampal gyrus showed high ICC values in the eyes-open condition, whereas it decreased along with other temporal regions in the eyes-closed one.

As in the sensor space analysis, alpha power showed high ICC for most brain areas, especially in frontal and parietal cortex. Although values remained high in the eyes-closed condition, fewer regions presented high ICC.

In the low beta band, the most reliable regions were the left parietal, precuneus, and the isthmus of the cingulate gyrus. The power reliability distribution seemed to be more anterior and bilateral in the eyes-closed condition especially in the medial orbitofrontal, superior frontal, and paracentral gyri. High beta and gamma were the frequency bands with fewer regions with high reliability. The former showed medium to low ICC values except in precuneus, paracentral, and parahippocampal gyrus (ICC > 0.6) in the eyes-open condition. Gamma band power showed widespread low ICC values, except in the left precentral gyrus (ICC > 0.7) in the eyes-open condition.

Absolute vs. Relative Power

In addition to the previous analysis of absolute power estimates, the ICC of the relative power was obtained for each source location and frequency band (see Supporting Information Figures S2 and S3), by normalizing power in each frequency band with the overall power in 2–45 Hz. Highest ICC values were found in alpha, low beta and high beta, especially over occipital and parietal regions. For high beta, ICC values were higher in more regions than for the absolute power. Conversely, relative power in delta and theta showed smaller ICC values in some regions such as the frontal cortex.

On the whole, ICC for relative power estimates seems to present widespread patterns, especially in low beta and high beta, whereas high ICC values for the absolute power were restricted to specific regions. Moreover, relative power in parietal and occipital cortex was fairly reliable even in cases with low ICC values for absolute power, such as the right cuneus in low beta. Conversely, the relative power in frontal cortex showed lower ICC values than the absolute power in almost all frequency bands.

Dependence between within-Subject Variability and Source Power

To determine whether the within-subject variability depends on the source power, the joint distribution of within-subject variability and average power for each frequency band is displayed in Figure 4 for eyes-open condition (see Supporting Information Figure S6 for eyes-closed). In general, low power levels result in high within-subject variability. This trend was present in all frequency bands, and was particularly evident in gamma, where power values were small (<2) throughout the brain. However, the relation between power and within-subject variability was not linear. In fact, although power and within-subject variability were inversely related for low power values, this tendency was not maintained for moderate to high power values (3–6), for which within-subject variability remained rather constant. In addition, the lowest within-subject variability was not invariably found for

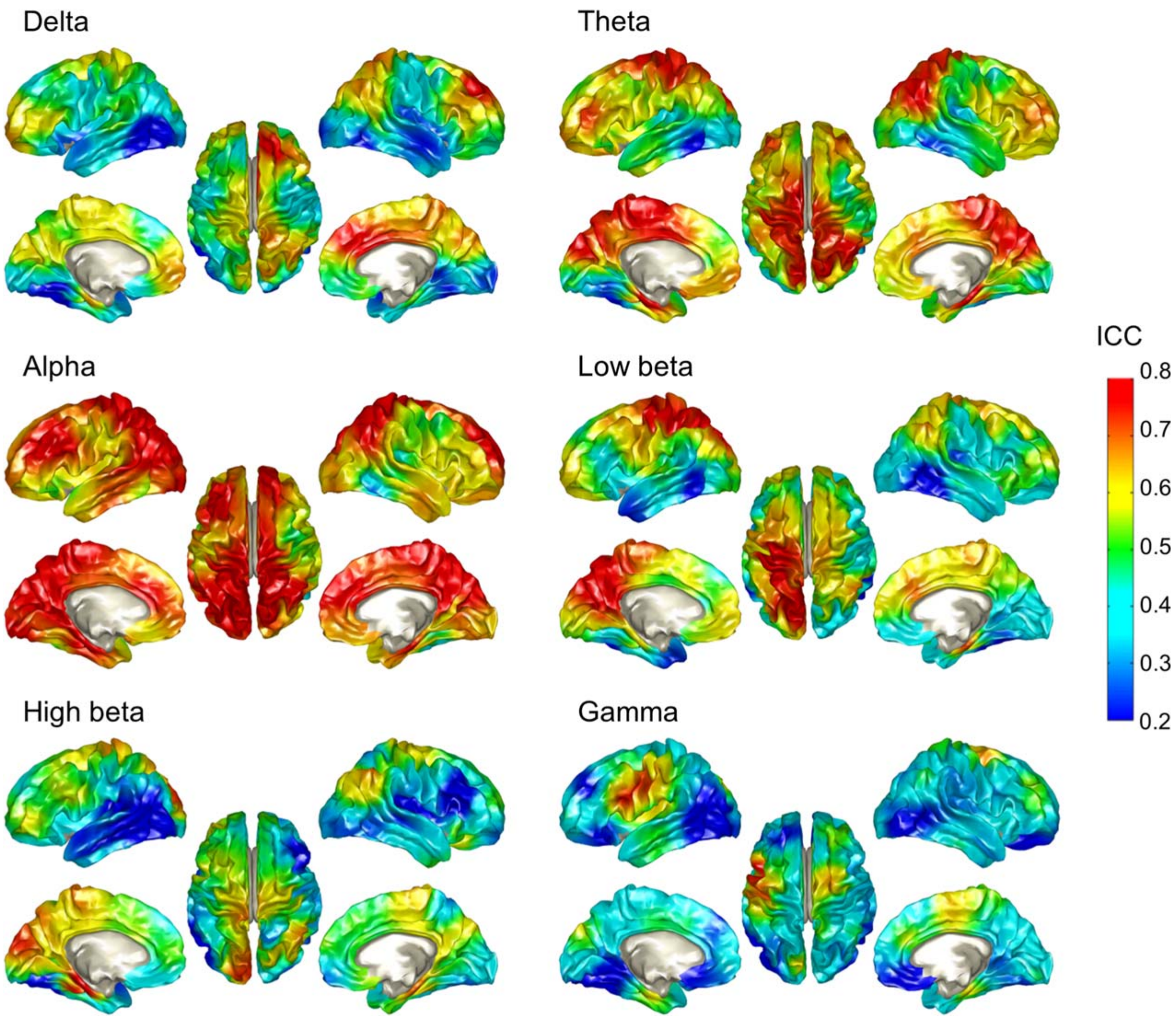

Figure 2.

ICC of source space power for the resting state eyes-open condition. ICC values were computed for each source location and frequency band separately. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

highest power values. For instance, in alpha band eyes-closed condition, high power values (8–11) resulted in higher within-subject variability than moderate power values (3–8). Overall, this indicates that, although a general inverse relation was found between within-subject variability and power values, within-subject variability does not exclusively result from power intensity.

DISCUSSION

In this article, we examined the test-retest reliability of the resting-state power in classical frequency bands at sensor and source space with MEG. To achieve this aim, three weekly MEG recordings were performed and the ICC of power values at each sensor and source location was calculated. Moreover, to evaluate how power magnitudes modulated reliability values, we explored the relation between source intensity and within-subject variability for each frequency band. To our knowledge, this is the first time that reliability of source space MEG resting-state power has been evaluated. There are two main findings in this study. Firstly, theta, alpha, and low beta were the most reliable brain rhythms at sensor and source space, in contrast to gamma power, which showed poor reliability in resting-state. Secondly, within subject variability was partially dependent on power intensities, as shown by the inverse relation found between within-subject variability and power intensity.

Our results are in line with previous sensor space EEG test-retest literature, which also found reliable theta, alpha,

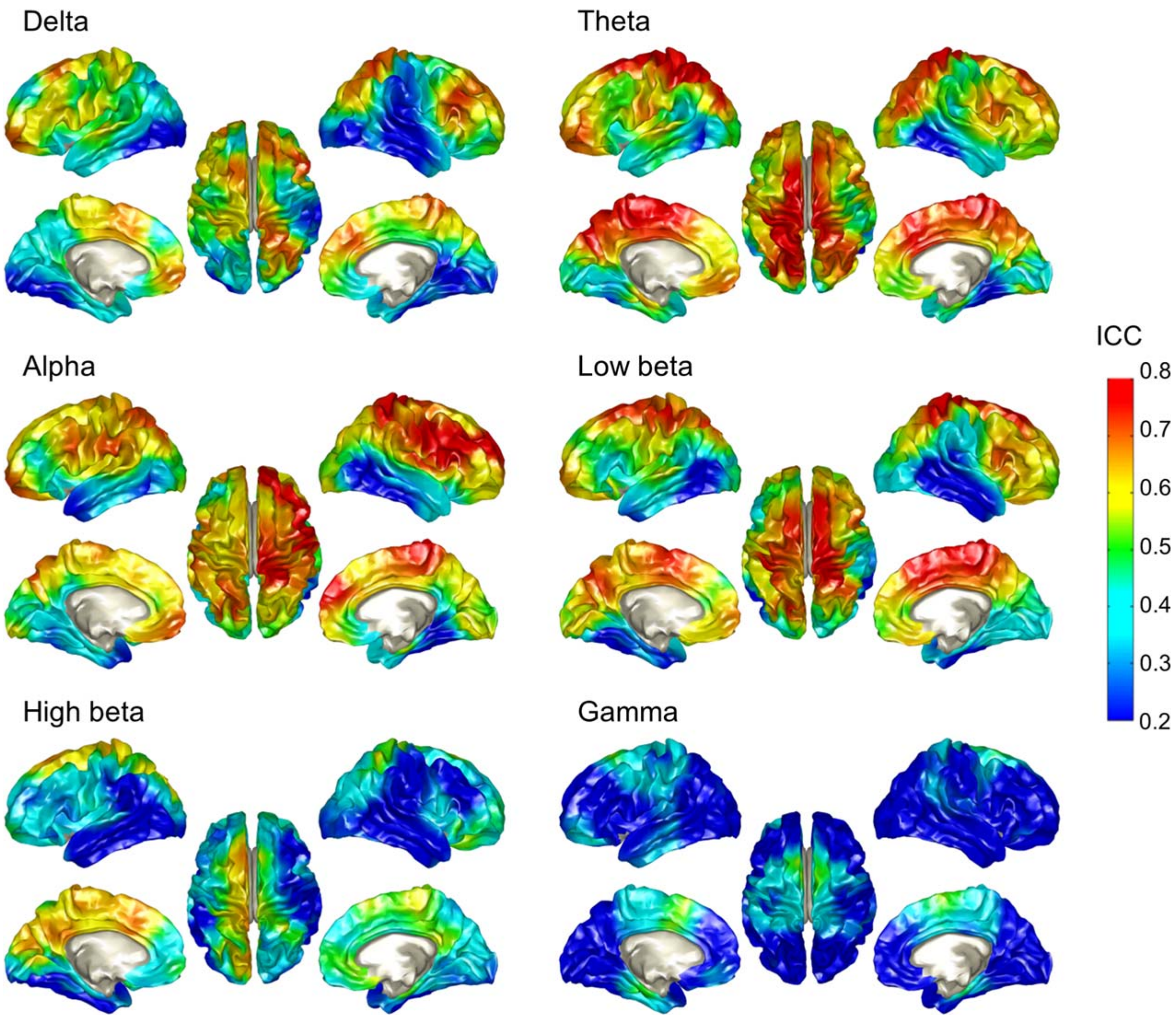

Figure 3.
ICC of source space power for the resting state eyes-closed condition. ICC values were computed for each source location and frequency band separately. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

and beta power estimates [Gasser et al., 1985; Kondacs and Szabó, 1999; McEvoy et al., 2000]. Amongst them, highest reliability was obtained in alpha for groups of children or young adults and in theta when including healthy elderly people [Gudmundsson et al., 2007]. Reliability has been mainly assessed with the ICC, and ICC values for theta, alpha and low beta power ranged from 0.54 to 0.95, depending on the brain rhythm and sensor location. Furthermore, gamma and delta power values presented lower reliability, in accordance with previous studies [Gasser et al., 1985; Gudmundsson et al., 2007; Kondacs and Szabó, 1999; Pollock et al., 1991].

Additionally, reliability differed between absolute and relative power estimates. In line with previous EEG studies [Gudmundsson et al., 2007; Kondacs and Szabó, 1999; Pollock et al., 1991], sensor space reliability was similar for absolute and for relative power in theta, alpha, and beta bands, although relative power yielded generally lower ICC in alpha and higher ICC in high beta and gamma. Since relative power values are normalized with the overall power, which is dominated by the high intensities in theta and alpha bands, it is possible that the relative power in gamma becomes reliable because of the high reliability of theta and alpha bands intensities. Nonetheless, source space results followed a different trend: higher ICC values were found for relative than for absolute power for all frequency bands. This could be attributed to a bias in the source reconstruction, beamformer solutions are biased in regions with low signal to noise ratio [Sekihara and Nagarajan, 2008]. Beamformer intensities are therefore

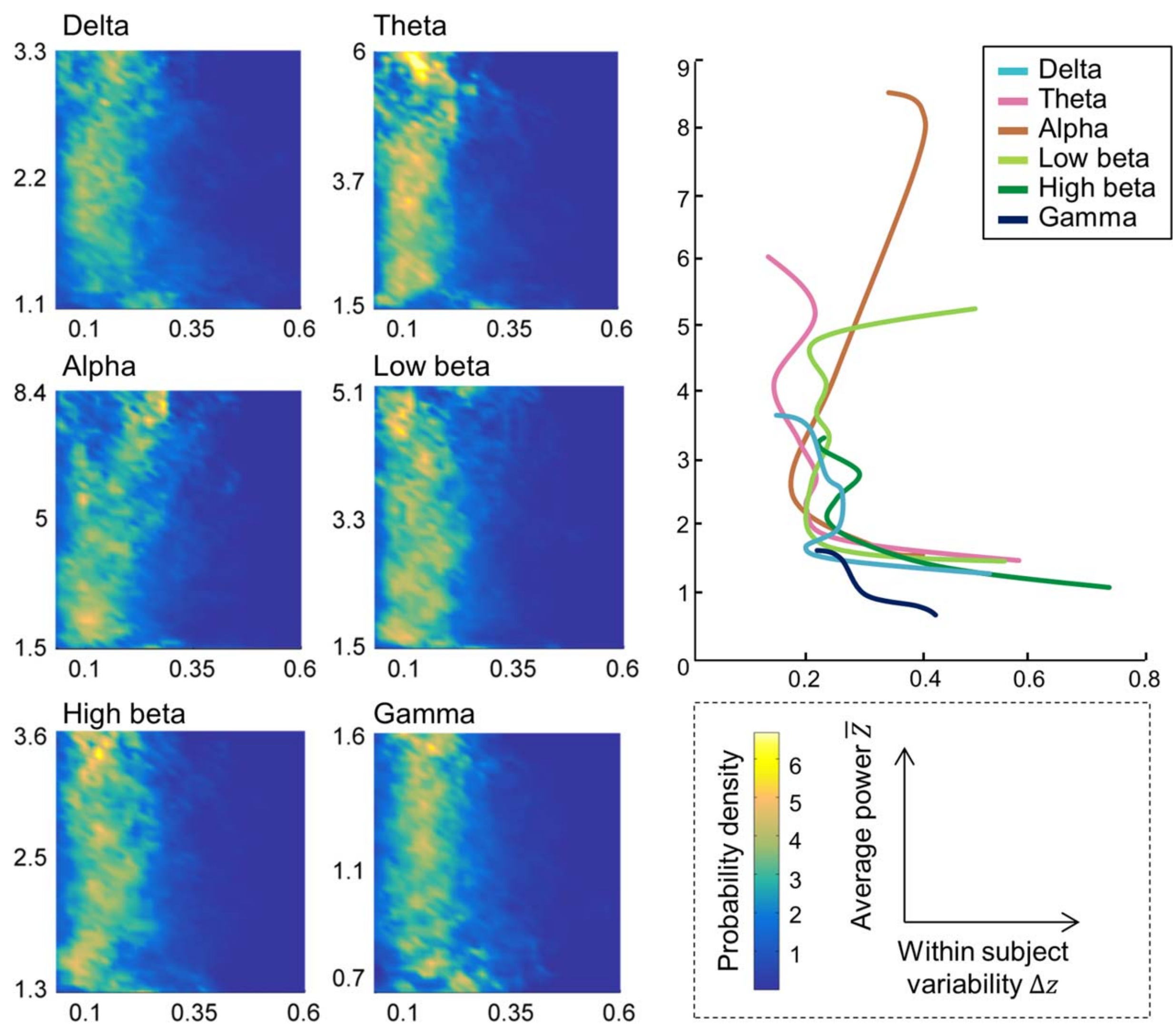

Figure 4.

Dependence of within-subject variability with the average power, for the resting state eyes-open condition. The surface plots estimate the conditional probability of obtaining a given within-subject variability Δz for a source power $\overline{\boldsymbol{Z}}$. The right plot represents the average within-subject variability as a function of the source power $\overline{\boldsymbol{Z}}$. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

usually not directly employed in any statistical analysis: they are rather normalized with another condition or with a noise estimate [Luckhoo et al., 2014]. Although this is often performed by assuming uncorrelated noise, we employed empty room recordings for normalization, since they are a more realistic estimate of the noise present the MEG data. Empty room data fail however to account for biological noise emerging from the subject. Relative power escapes this issue by normalizing with the overall source power, thereby avoiding any a priori assumptions on the noise characteristics. Nevertheless, relative powers are also less specific, as their mix intensities from all frequency bands, and they do not enable the separate inspection of brain rhythms. For instance, changes in the relative power of low intensity frequency bands (high beta or gamma), could be overshadowed by small variations in alpha or theta bands.

The SNR of power estimates may be partially responsible for the variability in its reliability across brain regions and rhythms. This was previously proposed to explain the low reliability of delta band power (Gasser et al., 1985; Pollock et al., 1991), as delta measurements are greatly affected by environmental and biological noise. Despite selection of artifact-free epochs, delta power may be more affected by this source of variance. In this work, we demonstrated that the reliability of power estimates was modulated by their intensities. Low power was in fact related to high within-subject variability. This was particularly evident in the gamma band, which presented low power throughout the brain. However, power intensity was not entirely responsible for its reliability, since source locations with the highest power did not consistently present the lowest within-subject variability. For instance, the alpha power in occipital regions presented both higher intensities and higher within-subject variability during the eyes-closed condition than during the eyes-open condition, therefore yielding smaller ICC.

Moreover, reliability might result from the inherent nature of brain oscillations. In general, the highest reliability values were found in those regions where the brain rhythms have been described as dominant in resting-state [Hillebrand et al., 2005] and during cognitive processing [Başar et al., 2001; Başar and Güntekin, 2012]. For instance, although the biophysical origin of alpha rhythm remains unclear, some studies have pointed out that it could be paced by the thalamus [Buzsáki, 2006; Hughes and Crunelli, 2005]. The spatial extent of corticothalamocortical circuit leads the dissemination of this rhythm in cortical areas and might support the widespread cortical distribution of high reliability in the alpha band. We found that theta power ICC was especially high in the central sulcus and the parahippocampal gyrus, while the theta rhythm has been classically identified in medial temporal structures such as enthorrinal and perirhinal cortex [Buzsáki, 2002; Wang, 2010] and midfrontal regions as the dorsal anterior cingulate cortex [Cavanagh and Frank, 2014; Congedo et al., 2010; Wang, 2010]. Likewise, low beta power showed high reliability in frontal regions such as the superior frontal or the paracentral gyrus in the eyes-closed condition. In agreement, beta rhythm has been typically identified in the primary motor cortex [Wang, 2010] and frontal regions associated with the inhibitory control of movement [Sauseng and Klimesch, 2008]. The nature of the delta oscillations could also explain its low reliability. It has been suggested that delta supports the continuous reorganization of the system due to its relation to cortical plasticity [Assenza et al., 2013] and maturation [Vlahou et al., 2014].

Delta, theta, and alpha are considered global processing rhythms because they recruit distributed neuronal populations and are identified in widespread brain regions. In contrast, low amplitude and high frequency oscillations such as beta and gamma are related to local processing and limited to restricted brain regions [Buzsáki and Wang, 2012; Knyazev, 2012]. This may explain the poor reliability of gamma band power. In fact, gamma oscillations are associated with high level processing such as perceptual binding, episodic retrieval and working memory [Jensen et al., 2007]. Fast oscillations adapt rapidly to the presence of incoming events or stimuli, such as when subjects have to maintain the representation of a stimulus in a working memory paradigm. Then, gamma power could be expected to be more reliable under an experimentally controlled task condition in which the external stimuli or the behavioral response leads to a specific brain configuration than in resting-state [Snyder and Raichle, 2012]. In a previous study with functional connectivity measures, greater reliability was found in the gamma band during an n-back task compared to resting-state [Deuker et al., 2009]. When the reliability between measures obtained in resting-state and those obtained under task performance were compared, greater reliability was found in the latter [Deuker et al., 2009; McEvoy et al., 2000]. In fact, attention or alertness variations in resting-state could yield a more variable brain activity pattern than when the subjects have to maintain a stimulus in a working memory paradigm.

Regarding the differences between resting-state conditions, it has been suggested that eyes-closed condition may be more reliable because of its higher SNR [Deuker et al., 2009]. Although eyes-closed condition showed higher SNR in the alpha and theta band, we found slightly high reliability in the eyes-open condition as it has been reported in a previous study with graphs metrics [Jin et al., 2011]. However, in our study highest signal intensities were related to higher within-subject variability, and eyes-closed power values were higher than eyes-open ones. However, eyes-open condition is not as experimentally controlled as a task condition, but in comparison to the eyes-closed one, the subjects were instructed to keep their gaze directed at a fixation cross and this situation provides a sort of control, which may reduce the variance.

From this study, we can derive three main conclusions that can influence future studies with MEG: (1) theta, alpha, and low beta power is more reliable than delta and gamma in resting-state; (2) frontal-beta, fronto-posterior alpha, medial-temporal, and midfrontal theta are the most reliable profiles in the source space; (3) absolute power in the source space seems to be more specific than relative power, but the latter may yield slightly higher reliability values.

A potential limitation of this study is that we evaluated reliability in a very specific population of healthy young subjects. It could be possible that reliability studies in subjects of different ages or patients with neurological or psychiatric diseases could present different reliability patterns not allowing for a direct generalization of our results. Thus, an important question for future studies is to determine the reliability of each brain rhythm across the lifespan. Describing the healthy trajectories in cognitive development and aging might allow detection of possible pathological processes and early diagnosis. Future studies may be performed with larger and/or clinical samples to confirm and extend the present results. Furthermore, it would be interesting to observe if those frequency bands with low reliability in resting-state, such as gamma, are reliable under task conditions. Finally, there is an increasing use of functional connectivity metrics with MEG and their reliability has not been addressed yet in the source space; such an investigation would help guide investigators in the analysis selection and the interpretation of the results derived from MEG data.

In conclusion, we studied for the first time the test-retest reliability of power measures with MEG at sensor and source space. We evaluated the effect of a number of factors on reliability –frequency band, power intensity, absolute and relative measures–, which can guide researchers and clinicians on obtaining reliable results in future MEG studies. Our study supports the use of the resting-state power of brain rhythms to assess the changes produced by drug treatments, neuropsychological rehabilitation, degenerative diseases, or developmental disorders.

ACKNOWLEDGMENTS

The authors are grateful to the research participants for their participation in this study and to Benjamin James Griffiths for his comments and recommendations that definitely help to improve the readability of the article.

REFERENCES

Acar ZA, Makeig S (2010): Neuroelectromagnetic Forward Head Modeling Toolbox. J Neurosci Methods 190:258–270.

Assenza G, Pellegrino G, Tombini M, Di Pino G, Tomasevic L, Tecchio F, Di Lazzaro V (2013): Delta waves increase after cortical plasticity induction during wakefulness. Clin Neurophysiol 124:e71–e72.

Atcherson SR, Gould HJ, Pousson Ma, Prout TM (2006): Long-term stability of N1 sources using low-resolution electromagnetic tomography. Brain Topogr 19:11–20.

Başar E, Güntekin B (2012): A short review of alpha activity in cognitive processes and in cognitive impairment. Int J Psychophysiol 86:25–38.

Başar E, Başar-Eroglu C, Karakaş S, Schürmann M (2001): Gamma, alpha, delta, and theta oscillations govern cognitive processes. Int J Psychophysiol 39:241–248.

Burgess A, Gruzelier J (1993): Individual reliability of amplitude distribution in topographical mapping of EEG. Electroencephalogr Clin Neurophysiol 86:219–223.

Buzsáki G (2002): Theta Oscillations in the Hippocampus Review. Neuron 33:325–340.

Buzsáki G (2006): Rhythms of the Brain. New York: Oxford University Press.

Buzsáki G (2012): Brain rhythms and neural syntax: Implications for efficient coding of cognitive content. Dialogues Clin Neurosci 14:345–367.

Buzsáki G, Wang XJ (2012): Mechanisms of gamma oscillations. Annu Rev Neurosci 35:203–225.

Cannon RL, Baldwin DR, Shaw TL, Diloreto DJ, Phillips SM, Scruggs AM, Riehl TC (2012): Reliability of quantitative EEG (qEEG) measures and LORETA current source density at 30 days. Neurosci Lett 518:27–31.

Cavanagh JF, Frank MJ (2014): Frontal theta as a mechanism for cognitive control. Trends Cogn Sci 18:414–421.

Congedo M, John RE, De Ridder D, Prichep L (2010): Group independent component analysis of resting state EEG in large normative samples. Int J Psychophysiol 78:89–99.

Cornew L, Roberts TPL, Edgar JC (2013): Resting-State Oscillatory Activity in Autism Spectrum Disorders. J Autism Dev Disord 42:1884–1894.

Deuker L, Bullmore ET, Smith M, Christensen S, Nathan PJ, Rockstroh B, Bassett DS (2009): Reproducibility of graph metrics of human brain functional networks. Neuroimage 47:1460–1468.

Fehr T, Kissler J, Moratti S, Wienbruch C, Rockstroh B, Elbert T (2001): Source distribution of neuromagnetic slow waves and MEG-delta activity in schizophrenic patients. Biol Psychiatry 50:108–116.

Fernández A, Hornero R, Mayo A, Poza J, Gil-Gregorio P, Ortiz T (2006): MEG spectral profile in Alzheimer's disease and mild cognitive impairment. Clin Neurophysiol 117:306–314.

Fingelkurts A, Fingelkurts A, Ermolaev V, Kaplan A (2006): Stability, reliability and consistency of the compositions of brain oscillations. Int J Psychophysiol 59:116–126.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, Van Der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM (2002): Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. Neuron 33:341–355.

Garcés P, Vicente R, Wibral M, Pineda-Pardo JÁ, López ME, Aurtenetxe S, Marcos A, de Andrés ME, Yus M, Sancho M, Maestú F, Fernández A (2013): Brain-wide slowing of spontaneous alpha rhythms in mild cognitive impairment. Front Aging Neurosci 5:100.

Gasser T, Bächer P, Steinberg H, Bacher P (1985): Test-retest reliability of spectral parameters of the EEG. Electroencephalogr Clin Neurophysiol 60:312–319.

Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L, Hämäläinen MS (2014): MNE software for processing MEG and EEG data. Neuroimage 86:446–460.

Gross J, Kujala J, Hamalainen M, Timmermann L, Schnitzler A, Salmelin R (2001): Dynamic imaging of coherent sources: Studying neural interactions in the human brain. Proc Natl Acad Sci USA 98:694–699.

Gudmundsson S, Runarsson TP, Sigurdsson S, Eiriksdottir G, Johnsen K (2007): Reliability of quantitative EEG features. Clin Neurophysiol 118:2162–2171.

Hillebrand A, Singh KD, Holliday IE, Furlong PL, Barnes GR (2005): A new approach to neuroimaging with magnetoencephalography. Hum Brain Mapp 25:199–211.

Hughes SW, Crunelli V (2005): Thalamic mechanisms of EEG alpha rhythms and their pathological implications. Neuroscientist 11:357–372.

Jensen O, Kaiser J, Lachaux J-P (2007): Human gamma-frequency oscillations associated with attention and memory. Trends Neurosci 30:317–324.

Jin S-H, Seol J, Kim JS, Chung CK (2011): How reliable are the functional connectivity networks of MEG in resting states? J Neurophysiol 106:2888–2895.

Klimesch W (1999): EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res Brain Res Rev 29:169–195.

Knyazev GG (2012): EEG delta oscillations as a correlate of basic homeostatic and motivational processes. Neurosci Biobehav Rev 36:677–695.

Kondacs A, Szabó M (1999): Long-term intra-individual variability of the background EEG in normals. Clin Neurophysiol 110: 1708–1716.

Leighton BN, Vinogradov S, Guggisberg AG, Fisher M, Findlay AM, Nagarajan SS (2011): Clinical symptoms and alpha band

resting-state functional connectivity imaging in patients with schizophrenia: Implications for novel approaches to treatment. Biol Psychiatry 70:1134–1142.
Luckhoo HT, Brookes MJ, Woolrich MW (2014): Multi-session statistics on beamformed MEG data. Neuroimage 95:330–335.
McEvoy LK, Smith ME, Gevins A (2000): Test-retest reliability of cognitive EEG. Clin Neurophysiol 111:457–463.
McGraw K, Wong S (1996): Forming inferences about some intraclass correlation coefficients. Psychol Methods 1:30–42.
Van Der Meer ML, Tewarie P, Schoonheim MM, Douw L, Barkhof F, Polman CH, Stam CJ, Hillebrand a (2013): Cognition in MS correlates with resting-state oscillatory brain activity: An explorative MEG source-space study. Neuroimage Clin 2:727–734.
Olde Dubbelink KTE, Stoffers D, Deijen JB, Twisk JWR, Stam CJ, Berendse HW (2013): Cognitive decline in Parkinson's disease is associated with slowing of resting-state brain activity: A longitudinal study. Neurobiol Aging 34:408–418.
Oostenveld R, Fries P, Maris E, Schoffelen JMM (2011): FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011: 156869
Pollock VE, Schneider LS, Lyness SA (1991): Reliability of topographic quantitative EEG amplitude in healthy late-middle-aged and elderly subjects. Electroencephalogr Clin Neurophysiol 79:20–26.
Sauseng P, Klimesch W (2008): What does phase information of oscillatory brain activity tell us about cognitive processes? Neurosci Biobehav Rev 32:1001–1013.
Schaefer M, Mühlnickel W, Grüsser SM, Flor H (2002): Reproducibility and Stability of Neuroelectric Source Imaging in Primary Somatosensory Cortex. Brain Topography 14:179–189.
Schnitzler A, Gross J (2005): Normal and pathological oscillatory-communication in the brain. Nat Rev Neurosci 6:285–296.
Ségonne F, Pacheco J, Fischl B (2007): Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. IEEE Trans Med Imaging 26:518–529.
Sekihara K, Nagarajan SS (2008): Adaptive Spatial Filters for Electromagnetic Brain Imaging. Berlin, Germany: Springer.
Shrout P, Fleiss J (1979): Intraclass correlations: Uses in assessing rater reliability. Psychol Bull. 86:420–428.
Snyder AZ, Raichle ME (2012): A brief history of the resting state: The Washington University perspective. Neuroimage 62:902–910.
Taulu S, Simola J (2006): Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys Med Biol 51:1759–1768.
Telesford QK, Burdette JH, Laurienti PJ (2013): An exploration of graph metric reproducibility in complex brain networks. Front Neurosci 7:67.
Vlahou EL, Thurm F, Kolassa I-T, Schlee W (2014): Resting-state slow wave power, healthy aging and cognitive performance. Sci Rep 4:5101.
Wang X (2010): Neurophysiological and computational principles of cortical rhythms in cognition. Physiol Rev 1195–1268.
Weir JPJ (2005): Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 19:231–240.
Wienbruch C, Paul I, Bauer S, Kivelitz H (2005): The influence of methylphenidate on the power spectrum of ADHD children - an MEG study. BMC Psychiatry 5:29.