Contrasting functional imaging parametric maps: The mislocation problem and alternative solutions

Mathieu Bourguignon [a,b,*], Nicola Molinaro [a,c], Vincent Wens [b]

[a] Basque Center on Cognition, Brain and Language (BCBL), Donostia/San Sebastian, Spain
[b] Laboratoire de Cartographie fonctionnelle du Cerveau, UNI – ULB Neuroscience Institute, Université libre de Bruxelles (ULB), Brussels, Belgium
[c] Ikerbasque, Basque Foundation for Science, Bilbao, Spain

ARTICLE INFO

ABSTRACT

In the field of neuroimaging, researchers often resort to contrasting parametric maps to identify differences between conditions or populations. Unfortunately, contrast patterns mix effects related to amplitude and location differences and tend to peak away from sources of genuine brain activity to an extent that scales with the smoothness of the maps. Here, we illustrate this mislocation problem on source maps reconstructed from magnetoencephalographic recordings and propose a novel, dedicated location-comparison method. In realistic simulations, contrast mislocation was on average ~10 mm when genuine sources were placed at the same location, and was still above 5 mm when sources were 20 mm apart. The dedicated location-comparison method achieved a sensitivity of ~90% when inter-source distance was 12 mm. Its benefit is also illustrated on real brain-speech entrainment data. In conclusion, contrasts of parametric maps provide precarious information for source location. To specifically address the question of location difference, one should turn to dedicated methods as the one proposed here.

1. Introduction

A recurrent question in neuroscience pertains to how much brain activity differs between conditions or groups. In the field of neuroimaging, this question is often addressed by contrasting parametric maps (Friston et al., 1994). However contrast patterns mix effects related to amplitude and location differences, and tend to peak away from sources of genuine brain activity to an extent that scales with the smoothness of the maps; this fact will be demonstrated in the **Theory** section (Section 2) with a one-dimensional toy model of contrast. In this report, we focus on maps derived from magnetoencephalographic (MEG) recordings, simply because the problem of contrast mislocation is expected to be particularly salient for those maps due to their intrinsically high smoothness.

MEG is a widely used non-invasive neuroimaging technique. It records the extracranial magnetic fields generated by synchronized electrical currents flowing though the apical dendrites of pyramidal neurons (Hämäläinen et al., 1993). The uniqueness of MEG lays in its high temporal resolution combined with a good spatial resolution for focal cortical sources. These characteristics allow researchers to track brain activity millisecond by millisecond during tasks and at rest, and to assess functional and effective brain connectivity.

One critical step in MEG analysis is to solve the inverse problem and estimate the location of cortical sources generating observed scalp distributions. Nowadays, source estimation is often performed with imaging methods such as linearly constrained minimum variance beamformer (LCMVB; Hillebrand et al., 2005; Van Veen et al., 1997) or minimum norm estimation (MNE; Dale and Sereno, 1993; Hämäläinen and Ilmoniemi, 1994). These inversion schemes make it possible to reconstruct source time-series, based on which one can build parametric maps of, e.g., task-induced power changes or coherence with an external reference signal. These source-space MEG parametric maps are analogous to activation maps obtained by functional magnetic resonance imaging (fMRI) or positron emission tomography (PET) but their spatial properties are different. They are affected by anisotropic long-range correlation (i.e., high spatial smoothness) that depends on the source reconstruction method used (Gross et al., 2003). Contrasts of such smooth maps are therefore expected to be particularly impacted by the contrast mislocation problem.

Here, we illustrate with simulated data the hazard of interpreting the location of peak contrast as a location wherein a source is more active in one condition compared to the other. We also present a novel method specific to the question of location difference, which may provide useful

standalone information or additional information to help interpret contrast results. In that method, we use a bootstrap procedure (Efron, 1979) to estimate a sample distribution of coordinates of maxima in the to-be-compared conditions, and we test the null hypothesis that the distance is zero. The sensitivity and specificity of this test are evaluated on simulated data and on real brain-speech entrainment data wherein brain activity is coherent with heard speech in a natural speech listening task (Ahissar et al., 2001; Bourguignon et al., 2013a; Luo and Poeppel, 2007; Molinaro et al., 2016; Peelle et al., 2013; Vander Ghinst et al., 2016).

2. Theory

We start this section by illustrating with a simple 1-D toy model why peak contrasts tend to mislocate (Subsection 2.1). We then review a classical method to assess the statistical significance of contrasts, but which is by construction hampered by the mislocation problem (Subsection 2.2). Finally, we introduce a novel method designed to complement contrast analyses by specifically testing for location differences between conditions (Subsection 2.3). In the present report, we shall focus on the comparison of group-level maps, using individual maps that have been morphed to match a reference anatomy (e.g., the MNI brain) to generate the group-level statistics. The same approach would obviously apply to within-subject maps, with trials taking the role of subjects.

2.1. Comparing parametric maps in two conditions

The issue of comparing maps via contrasts can be understood using the following one-dimensional toy model. Consider two genuine brain sources located at $x_1 = 0$, $x_2 = -D$ and thus separated by a distance D. Assume that their activation appears in maps as Gaussian profiles $G_1(x) = \alpha\, e^{-\frac{x^2}{2\sigma^2}}$ and $G_2(x) = e^{-\frac{(x+D)^2}{2\sigma^2}}$, where α denotes their relative amplitude and σ their spatial spread (linked to the full-width at half-maximum by FWHM $= \sqrt{8\log(2)}\,\sigma$). Contrasting the two maps yields a contrast profile $C(x) = G_1(x) - G_2(x)$ and we denote by x_C its peak coordinate. Fig. 1 provides a comprehensive view of the impact of the amplitude α and separation D parameters on the peak location x_C. For sources with similar amplitude ($\alpha \sim 1$) and not well separated ($D < 0.5$ FWHM), the contrast peaks $x_C = 0.3$–0.5 FWHM away from the first source. This clearly illustrates why peak contrasts tend to localize away from genuinely active brain areas and can thus be rather uninformative.

2.2. Direct contrast method

We now review a practical implementation of a commonly used contrast method to compare MEG maps that includes proper statistical assessment of the maximum contrast value (Maris and Oostenveld, 2007; Nichols and Holmes, 2002).

In the direct contrast method, the contrast of group-averaged maps in two conditions is compared to a permutation distribution of maximum contrast value, which is obtained as follows (Maris and Oostenveld, 2007; Nichols and Holmes, 2002).

1) For each subject, randomly assign maps of both conditions to two dummy condition sets. An exemplary dummy set 1 for $n = 5$ subjects could consist of maps in condition 2, 1, 1, 2, 1 from subjects 1, 2, 3, 4, 5 respectively, leaving maps from conditions 1, 2, 2, 1, 2 to dummy set 2.
2) Generate the group-averaged map for both dummy condition sets.
3) Contrast these two group maps and extract the maximum value for one-sided tests or the maximum absolute value for two-sided tests.
4) Repeat this procedure N_{rep} times (practically $N_{\text{rep}} = \sim 1000$) to build a permutation distribution of maximum contrast value.

Brain areas corresponding to a genuine contrast above the $(1 - p) \times 100$ th percentile of the permutation distribution are considered more activated in one condition than in the other at statistical significance level p.

This approach is perfectly rigorous for the question it answers, i.e., whether the two group maps are significantly different at some location, but as explained in Subsection 2.1 it is subject to the problem of contrast mislocation. We now present a novel, complementary approach that focuses on the separation parameter D rather than on the contrast values.

2.3. Location-comparison method

In the location-comparison method, we generate bootstrap group-averaged maps to build a distribution of location difference between local maxima in the two conditions, and test the null hypothesis that this distance is zero. The procedure works as follows.

1) Randomly draw with replacement a sample of n pairs of individual maps, n being the number of subjects. An exemplary sample for $n = 5$ could consist of maps from subjects 2, 2, 2, 4, 5.
2) Generate the group-averaged map in both conditions for that drawn sample.
3) Identify the local maximum of each group-averaged map that is closest to the mean coordinates of genuine parametric maps maximum and save the coordinate difference. Practically, the search for the local maximum is performed within a predefined search
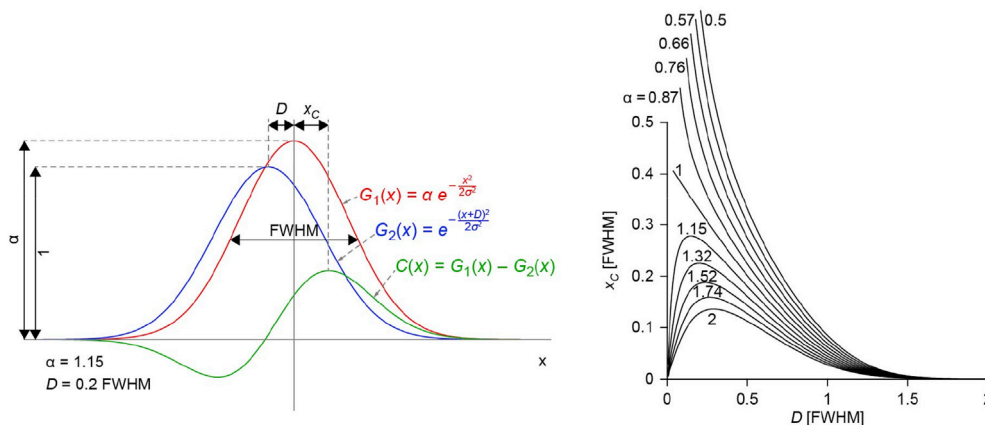

Fig. 1. One-dimensional toy model of contrast. *Left* — Two Gaussian profiles (red and blue traces) characterized by their relative amplitude (α), their full-width at half-maximum (FWHM), and the distance separating their maximum (D). Their contrast (green trace) peaks at a distance x_C from the maximum of the first profile. *Right* — Impact of α and D on x_C. Values of x_C are displayed only when the corresponding contrast takes a value higher than $\alpha/20$.

volume centered on the mean coordinate of genuine parametric maps maximum.

4) Repeat this procedure N_{rep} times (practically $N_{rep} \square = \sim 1000$) to obtain a sample distribution of the coordinates of the local maximum of the group-averaged parametric maps.

Based on this three-dimensional distribution, we can test the null hypothesis that local maxima in the two conditions co-localize by performing a multivariate test on the sampling distribution of coordinates difference. The test is described in details in the Appendix. A matlab implementation of if is available at https://osf.io/ue6ws, and integration with other open toolboxes is planned.

At this point, it should be noted that more commonly used permutation-based statistics would not apply here. Indeed, peak co-ordinates in group maps obtained after randomly permuting conditions (i.e., maps at step 2 of the procedure described in Subsection 2.2) would naturally tend to locate at the peak of the group map of the condition of highest values. Hence, the ensuing location-comparison test would not be specific to differences in location.

3. Simulations

To evaluate the performance of the two comparison methods described in the previous paragraph, we applied them to simulated data sets generated on the basis of 5-min resting-state data recorded from 17 subjects (group 1, *rest* data; see Section 4). We ran simulations where the MEG maps represent the coherence with a reference signal (Subsection 3.1) and also the relative power increase resulting from an event-related synchronization (Subsection 3.2). We considered these two measures because both are commonly used in current MEG studies but they differ in their properties (e.g., the noise on individual coherence maps is always above 0, while that on individual maps of power increase fluctuates around 1). These properties may affect differently the performance of the methods, which we quantified in terms of specificity and sensitivity (Subsection 3.3) as function of the inter-source distance (Subsection 3.4), sample size (Subsection 3.5), and difference in signal-to-noise ratio (SNR) between conditions (Subsection 3.6).

3.1. Procedure to generate simulated maps of coherence with a reference signal

Our simulations of coherence with a reference signal were designed to tightly match real brain-speech entrainment data in which MEG signals from auditory regions of a listener are coherent with the temporal envelope of heard speech at ~0.5 Hz and 4–8 Hz (Ahissar et al., 2001; Bourguignon et al., 2013a; Luo and Poeppel, 2007; Molinaro et al., 2016; Peelle et al., 2013; Vander Ghinst et al., 2016). Accordingly, we simulated coherence at 0.5 Hz between MEG data and a fixed reference signal taken as the speech envelope of a 5-min audio recording of a person reading a continuous text.

Every individual simulated dataset was obtained by superimposing a coherent source signal to a noise background. We started by generating the *noise* Fourier coefficients for 150 epochs. Specifically, we estimated first the cross-spectral density matrix at 0.5 Hz of the 5-min of rest data, and performed a Cholesky factorization of it. Then we applied the obtained triangular matrix to 150 vectors of non-correlated random numbers drawn from a centered complex Gaussian distribution. This procedure generated the *noise* Fourier coefficients for 150 epochs, and their covariance matrix is on average the subject's original cross-spectral density matrix at rest. Obviously, these *noise* coefficients are unrelated to the reference signal.

To introduce some degree of coherence with the reference signal, we added to the *noise* coefficients the signal of a perfectly coherent source whose coordinates in the MNI space will be denoted by R_S. To that aim, we computed the leadfield of a source placed at R_S with a direction along the superior-inferior (Z) axis. This leadfield multiplied by the Fourier

coefficients of the reference signal (obtained from adjacent 2-s epochs) yielded the *signal* Fourier coefficients for 150 epochs. By construction, the coherence of these *signal* coefficients with those of the reference signal is exactly one. They were rescaled so that the mean SNR across all sensors is equal to a given SNR value and then added to the *noise* coefficients.

Source-space coherence maps with speech temporal envelope where then obtained with LCMVB. Importantly, LCMVB weights were derived from the mean covariance matrix across the two conditions, so that the inverse solution was identical for both conditions (Gross et al., 2013). We used a 2-mm grid source space within a 40-mm sphere centered on mean R_S. The final maps were interpolated to a 1-mm grid and smoothed with a Gaussian kernel of 5-mm FWHM. Individual maps were finally morphed to match the MNI template. More details about source reconstruction and normalization to MNI templates are provided in Section 4.

Importantly, this whole process was repeated twice in order to obtain coherence maps in two conditions for which source coordinates R_S or the SNR may or may not differ (as explained in Subsections 3.4–3.6).

3.2. Procedure to generate simulated maps of power increase

We used a procedure similar to that described above to generate maps of power increase. In these simulations, we used the Cholesky factorization to generate two independent sets of Fourier coefficients for 300 epochs, assimilated to *pre-* and *post-*stimulus *noise*. To simulate power increase, we added to the *post-*stimulus *noise* coefficients the *signal* coefficients arising from a source positioned at R_S and with amplitude rescaled so as to fix a given SNR value. Individual source-space parametric maps were then computed as the ratio between *post-* and *pre-*stimulus power maps reconstructed using LVMCB as briefly discussed above. This process was again repeated twice to obtain power-ratio maps in two conditions.

3.3. Methods performance

For all the simulations considered, we compared the group maps in the two conditions with the two methods described in Section 2. To evaluate their performance, each simulation was repeated 100 times (for each set of parameter values, see Subsections 3.4–3.6) so that sensitivity and specificity could be estimated by counting the number of simulations in which a significant effect was detected at $p < .05$. More precisely, specificity was evaluated as the proportion of simulations for which a significant location difference was found in the absence of genuine source location difference ($D = 0$). Sensitivity was assessed by that proportion for $D \neq 0$.

For the results obtained with the location-comparison method, we also examined the proportion of simulations in which Cond 1 and 2 were found to localize in significantly different locations as function of the theoretical false positive rate (α level). The ensuing curves are widely known as receiver operating characteristic (ROC) curves, which can be seen as distributions of *p*-values. In these ROC plots, the main diagonal corresponds to a uniform distribution of *p*-values, which is the distribution expected if the test performs at chance level. ROC curves significantly above the main diagonal indicate that the test detects a significant difference more often than expected by chance. This should occur when genuine inter-source distance (D) is non-zero if the test is sensitive to some extent. Curves that do not deviate from the main diagonal indicate that the test detects a significant difference at a rate expected by chance. This should occur when $D = 0$. Finally, curves below the main diagonal indicate that the test detects a significant difference less often than expected by chance, meaning that the test is too conservative. To estimate the statistical significance of those deviations from the main diagonal, we compared the distribution of the *p*-values to the uniform distribution using the Kolmogorov-Smirnov test.

3.4. Effect of genuine inter-source distance (simulation 1)

Simulation 1 was designed to quantify the effect of genuine inter-source distance D on methods' specificity and sensitivity in realistic situations in which one condition yields slightly stronger activation than the other. To that aim, we varied the distance D between sources of condition 1 and 2 and set the average SNR to higher values in condition 1 than in condition 2. Individual coordinates of genuine source locations $R_S = [X\ Y\ Z]$ were drawn from a random Gaussian distribution with mean [60 −20 10] mm and standard deviation [1 8 4] mm. To set the inter-source distance between conditions to $D = 0, 4, 8, 12, 16$, and 20 mm, R_S was shifted $D/2$ forward ($Y \leftarrow Y + D/2$) in condition 1 and backward ($Y \leftarrow Y − D/2$) in condition 2. The SNR in condition 1 was randomly chosen between 0.0005 and 0.003 with constant probability on a logarithmic scale for the coherence maps, and likewise between 0.002 and 0.01 for the power-ratio maps (hence simulating a higher noise level on coherence maps). The SNR in condition 2 was set to that in condition 1 multiplied by a random number in the range 0.3–1.2, hence ensuring a lower SNR on average.

With the direct contrast method, we expected the maximum contrast to mislocate as hinted by our 1-D toy model. We also expected the magnitude of the mislocation to decrease as D increases.

At $D = 0$, we expected the location-comparison method to reveal a significant difference in location at a rate compatible with chance level (i.e., to be unbiased). We also expected the location-comparison method to reveal a significant difference in location in an increasing proportion of simulations (i.e., to become more sensitive) as D increased.

3.5. Effect of sample size (simulation 2)

Simulation 2 was designed to quantify the impact of the number of subjects on the specificity and sensitivity of the location-comparison method. To that aim, we repeated simulation 1 for $D = 0$ and $D = 4$ mm with a varying number $n = 10, 15, 20, 25$, and 30 of subjects. Since simulations were based on data from 17 subjects only, to generate $n = 20$, 25, or 30 independent maps, we used the data of the first 3, 8, or 13 subjects (respectively) to generate two simulated maps with independent drawings of R_S and SNR.

At $D = 0$, we expected the location-comparison method to reveal a significant difference in location at a rate compatible with chance level (i.e., to be unbiased) regardless of n. At $D = 4$ mm, we expected the location-comparison method to reveal a significant difference in location in an increasing proportion of simulations (i.e., to become more sensitive) as n increases.

3.6. Effect of SNR difference between conditions (simulation 3)

Simulation 3 was designed to evaluate whether the location-comparison method is robust with respect to difference in SNR between conditions. To that aim, sources in both conditions were placed at the same position ($D = 0$ mm) but we varied the SNR in condition 1 while maintaining the SNR in condition 2. Individual R_S were thus identical in both conditions. It was drawn from a random Gaussian distribution with mean [60 −20 10] mm and standard deviation [1 8 4] mm. We tested 4 different sets of SNR bounds. In the first set, SNR bounds were identical in condition 1 and 2 (0.001–0.002 for coherence, 0.002–0.004 for power ratio). In the subsequent SNR sets, bounds were not changed in condition 2, and were multiplied by 2 (or 2.5 for power ratio maps), 5, and 10 in condition 1.

We expected the location-comparison method to reveal a significant difference in location at a rate compatible with chance level (i.e., to be unbiased) regardless of SNR differences.

4. Materials and methods

We also illustrated the performance of the location-comparison methods on real brain-speech entrainment data.

4.1. Subjects

Two groups of subjects without any history of neuropsychiatric disease or language disorders were studied. Group 1 consisted of seventeen healthy native Spanish speakers (mean 24 years, range 20–32 years; 9 females and 8 males). Group 2 consisted of twenty healthy native French speakers (mean 30 years, range 23–40 years; 10 females and 10 males) previously reported in Vander Ghinst et al. (2016). All subjects were right-handed according to self-report. The study was approved by the BCBL Ethics Committee and the ethics committee of ULB-Erasme hospital. Subjects participated after informed consent. Participants from group 1 were measured at the BCBL, San Sebastian, Spain. Participants from group 2 were measured at the MEG unit of the ULB-Erasme hospital, Brussels, Belgium.

4.2. Experimental paradigm

Among other tasks, subjects underwent a speech listening session and a rest session while they were sitting in a chair, head inside a MEG helmet. In the speech-listening session, participants heard an audio recording of a text read in their native language (group 1: Spanish; group 2: French) by a native speaker for 5 min. In the rest session, subjects were asked to fixate the gaze at a point on the wall of the magnetically shielded room and try to reduce blinks and saccades to the minimum for 5 min. The order of sessions was randomized among all recordings.

4.3. Data acquisition

Neuromagnetic signals were acquired with a whole-scalp-covering neuromagnetometer (Vectorview; Elekta Oy, Helsinki, Finland) in a magnetically shielded room. The recording pass-band was 0.1–330 Hz and the signals were sampled at 1 kHz. The head position inside the MEG helmet was continuously monitored by feeding current to 4–5 head-tracking coils located on the scalp. Head position indicator coils, three anatomical fiducials, and at least 150 head-surface points (covering the whole scalp and the nose surface) were localized in a common coordinate system using an electromagnetic tracker (Fastrak, Polhemus, Colchester, VT, USA).

Electrooculograms (EOG) monitored vertical and horizontal eye movements, and electrocardiogram (ECG) recorded heartbeat signals, time-locked to MEG signals.

High-resolution 3D-T1 cerebral magnetic resonance images (MRI) were acquired on a 3 T MRI scan for group 1 (Siemens Medical System, Erlangen, Germany) and on a 1.5 T MRI scan for group 2 (Intera, Philips, the Netherlands).

4.4. Data preprocessing

Continuous MEG data were first preprocessed off-line using the temporal signal space separation method (correlation coefficient, 0.9; segment length set to recording duration) to suppress external interferences and to correct for head movements (Taulu et al., 2005; Taulu and Simola, 2006). To further suppress heartbeat, eye-blink, and eye-movement artifacts, thirty independent components were then evaluated from the MEG data low-pass filtered at 25 Hz, and independent components displaying a correlation exceeding ±0.15 with any EOG or ECG signal were subtracted from the MEG data.

Voice temporal envelope was obtained as the sound signals high–pass filtered at 50 Hz, rectified, low-pass filtered at 50 Hz, and resampled at 1000 Hz time lock to the MEG signals.

To perform frequency and coherence analyses, continuous data obtained during listening were split into 2-s epochs with 1.6-s epoch overlap. We used overlapping epochs as that leads to decreased noise on coherence estimates (Bortel and Sovka, 2014). MEG epochs exceeding

5 pT (magnetometers) or 1 pT/cm (gradiometers) were excluded from further analyses to avoid contamination of our data by any other source of artifact that would not have been dealt with by the temporal signal space separation or independent component analysis based artifact suppression. These steps led to 703 ± 45 (mean ± SD, group 1) and 736 ± 20 (group 2) artifact-free epochs of MEG and voice envelope signals for each subject.

4.5. Coherent source analysis

Coherence was previously used to assess the coupling between voice signal and brain signals at the frequencies corresponding to sentence-level prosody and syllable production rate (Bourguignon et al., 2013a; Luo and Poeppel, 2007; Poeppel, 2003). We evaluated the coherence at the sensor level using the formulation of Halliday et al. (1995). Data from gradiometer pairs were combined in the direction of maximum coherence as done in Bourguignon et al. (2015).

We also estimated the coherence at the source level. To do so, individual MRIs were first segmented using the Freesurfer software (Reuter et al., 2012). Then, the MEG forward model was computed for two orthogonal tangential current dipoles placed on a homogeneous 2-mm grid source space covering the whole brain (MNE suite; Gramfort et al., 2014). Coherence maps were produced within the computed source space at delta (0.5 Hz; frequency corresponding to sentence level prosody) and theta (4–8 Hz; frequency corresponding to syllable production rate) using a LCMVB (Hillebrand et al., 2005; Van Veen et al., 1997). The theta coherence was obtained by simply averaging the coherence values across all frequency bins falling into that band. Source maps were then interpolated to a 1-mm homogenous grid and smoothed with a Gaussian kernel of 5-mm FWHM. Both planar gradiometers and magnetometers were used for inverse modeling after dividing each sensor signal (and the corresponding forward-model coefficients) by its noise standard deviation. The noise variance was estimated from the continuous *rest* MEG data band-passed through 1–195 Hz, for each sensor separately.

4.6. Group analyses

A non-linear transformation from individual MRIs to the MNI brain was first computed using the spatial normalization algorithm implemented in Statistical Parametric Mapping (SPM8; Ashburner et al., 1997; Ashburner and Friston, 1999) and then applied to individual MRIs and coherence maps. This procedure generated a normalized coherence map in the MNI space for each subject and frequency. Coherence maps were then averaged across subjects.

5. Results

The results are organized as follows. In Subsection 5.1 we illustrate the limitations of the direct-contrast approach on simulation 1 data. In Subsections 5.2–5.4 we evaluate how the performance of the location-comparison method is affected by the inter-source distance (Subsection 5.2), the sample size (Subsection 5.3) and the SNR difference between conditions (Subsection 5.4). Finally, in Subsection 5.5, we illustrate the performance of the location-comparison method on real brain-speech entrainment data.

5.1. Mislocation of peak contrast

In this section, we illustrate and quantify the extent to which peak contrasts locate away from genuine brain sources. For that purpose, we used simulation 1 data in which we varied the distance D between sources in two conditions and imposed that the SNR was on average higher in condition 1 than in condition 2 (Subsection 4.4). For each simulation, we contrasted source-reconstructed MEG maps and assessed the significance of the contrast map with the non-parametric permutation test (Maris and Oostenveld, 2007; Nichols and Holmes, 2002) reviewed

Table 1
Summary of direct contrast results for simulation 1 data when inter-source distance is zero ($D = 0$). The three last rows provide mean ± SD values. n^*: Percentage of simulations in which significant contrast was uncovered ($p < .05$). R_1: Coordinates of the peak coherence or power ratio in condition 1. R_2: Coordinates of the peak coherence or power ratio in condition 2. R_C: Coordinates of the peak contrast. $|\cdot|$ and $\hat{}$ denote the vector length and angle, respectively.

	Coherence	Power ratio
n^*	47 %	42 %
$\|R_C – R_1\|$ [mm]	13.6 ± 7.5	6.7 ± 5.3
$\|R_1 – R_2\|$ [mm]	4.5 ± 2.5	3.5 ± 2.4
$\cos(R_2\widehat{R_1}R_C)$	−0.74 ± 0.27	−0.77 ± 0.25

in Subsection 2.2. We then characterized the relative locations of the peaks in condition and contrast maps.

Table 1 summarizes the results obtained when simulated sources were placed at the same location in the two conditions ($D = 0$). In that setting, a significant contrast was detected in ~45% of the simulations, all of them positive (i.e., condition 1 > condition 2). This was expected because the SNR, and hence coherence and power-ratio values, were higher in condition 1 than in condition 2. However, in these simulations that led to a positive contrast, the peak contrast (coordinates R_C) and the peak value in condition 1 (coordinates R_1) were distant of $|R_C - R_1|$ = ~14 mm for the coherence maps and of ~7 mm for the power-ratio maps, whilst the distance between R_1 and the peak value in conditions 2 (coordinates R_2) was 2–3 times smaller. This clearly illustrates the mislocation problem predicted by our 1D toy model (Subsection 2.1). Also predicted by this model, the contrast consistently peaked opposite from condition 2 with respect to condition 1 (see Fig. 1). Fig. 2 presents the outcome of a representative simulation where this configuration of maxima is clearly visible. As confirmation of this observation, $\cos(R_2\widehat{R_1}R_C)$ (the cosine of the angle formed by the vector pointing from R_1 to R_2 and the vector pointing from R_1 to R_C) was significantly negative ($ps < 0.0001$; Wilcoxon signed rank test).

We now turn to the simulations with varying inter-source distances ($D = 0, 4, 8, 12, 16, 20$ mm). Fig. 3 A and B present the results for both types of parametric maps. Globally, the sensitivity of contrast detection increased as D increased, both for the dominant positive (condition 1 > condition 2) and the negative (condition 2 > condition 1) effects. In other words, significant contrasts were uncovered in an increasingly high proportion of simulations. At $D = 8$ mm, significant positive contrasts were uncovered in over ~60% of the simulations. In positive contrasts, the mislocation $|R_C - R_1|$ was stable or even increased for the smallest values of D we tested (4 and 8 mm); at $D = 8$ mm, it was of the order of 10 mm. It then decreased with increasing D values, demonstrating better contrast localization for well-separated activations. At $D = 20$ mm, the mislocation was twice as high for coherence maps (~10 mm) as compared with power-ratio maps (~5 mm). Not surprisingly, fewer negative contrasts were uncovered for the smallest values of D (less than 5% at $D = 8$ mm) but the detection rate of negative peaks increased up to over 50% at $D = 20$ mm; the associated mislocation $|R_C - R_2|$ were of ~10 mm for coherence maps and ~5 mm for power-ratio maps.

5.2. Specificity and sensitivity of the location-comparison method (simulation 1)

We now turn to an examination of the specificity and sensitivity of the location-comparison method, first in the context of simulation 1. Table 2 and Fig. 4 A summarize the results.

In the absence of genuine source location difference ($D = 0$), the location-comparison method revealed spurious significant ($p < .05$) difference in location at a rate expected by chance in simulated coherence maps (8%), and in none of the simulations of power-ratio maps (see Table 2). Importantly, the ROC curve did not deviate significantly from the main diagonal (meaning that the test was not biased) for coherence
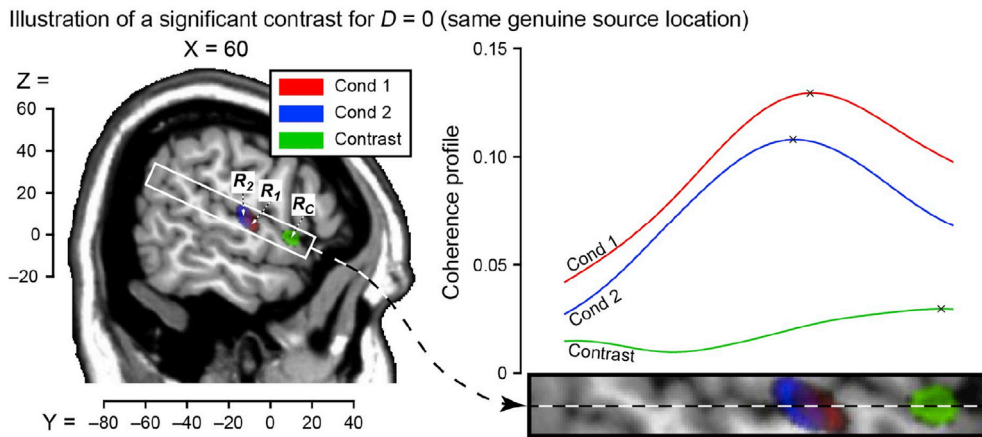
Fig. 2. Illustration of the outcome of the direct contrast analysis when genuine inter-source distance is zero ($D = 0$; simulation 1). In simulation 1, sources in two conditions (Cond 1 & 2) were centered on [–60 20 –10] mm and the SNR was on average higher in Cond 1 than in Cond 2. The figure presents a representative instance of simulation 1 in which a significant positive contrast (Cond 1 > Cond 2) was detected. *Left* — Location of peak coherence (Cond 1: red, Cond 2: blue) and contrast (green) superimposed on a parasagittal slice ($X = 60$ mm). *Right* — Coherence and contrast profiles along the least-square line fitted to the coordinates of the peak coherence in Cond 1 (R_1) and Cond 2 (R_2), and of the peak contrast (R_C). It is noteworthy that the contrast localized far off from Cond 1, where no genuine sources were placed.
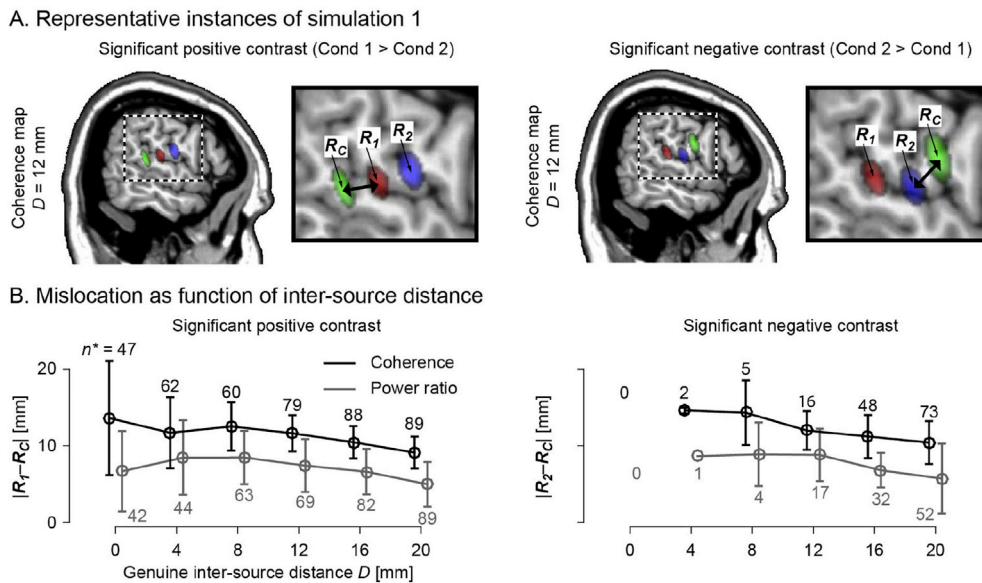

Fig. 3. Performance of the direct contrast method as function of the genuine inter-source distance ($D = 0, 4, 8, 12, 16,$ and 20 mm; simulation 1). In simulation 1 the SNR was on average higher in Condition (Cond) 1 than in Cond 2. A — Representative instances of simulation 1 at $D = 12$ mm. The peak of coherence and contrast maps are superimposed on a parasagittal slice ($X = 60$ mm). The enlargement clarifies the color code. The left column shows a significant positive contrast of coherence maps (Cond 1 > Cond 2; left) and the right column a significant negative contrast (Cond 2 > Cond 1; right). B — Mean ± SD contrast mislocation across simulations that led to significant contrast (n^*; indicated next to SD bars) as function of D for both coherence maps (black traces) and power ratio maps (gray traces). The contrast mislocation is here defined as the distances from peak contrast (coordinates R_C) to peak in Cond 1 (coordinates R_1; for positive contrasts) or Cond 2 (coordinates R_2; for negative contrasts).

simulations ($p = .38$; Kolmogorov-Smirnov test, see Fig. 4 A) but it was significantly below the main diagonal (meaning that the test was too conservative) for power-ratio simulations ($p < .001$).

The sensitivity of the test increased as D increased from 4 to 20 mm (see Table 2) and the ROC curves were largely above the main diagonal ($ps < 0.001$, see Fig. 4 A) as is expected when the null hypothesis (here, that $D = 0$) is not true. At $D = 12$ mm, the test uncovered a significant difference in ~90% of the simulations for $\alpha = 0.05$.

5.3. Effect of the sample size (simulation 2)

We next applied the location-comparison method to simulation 2 data to quantify its specificity and sensitivity for different sample sizes (number of subjects $n = 10, 15, 20, 25,$ and 30).

Fig. 4 B presents the results obtained with the location-comparison method in the absence of genuine location difference ($D = 0$). In that

setting, the ROC curves did not deviate significantly from the main diagonal for coherence maps based on 10, 15 and 30 subjects ($ps > 0.05$;

Table 2
Summary of location-comparison results for simulation 1. Ideally, for a given value of the genuine inter-source distance (D), the distance between estimated peak coordinates in condition 1 and 2 (R_1 and R_2) should be $R_1 - R_2 = [0 -D \ 0]$. n^*: Percentage of simulations in which the test detected a statistically significant difference between R_1 and R_2.

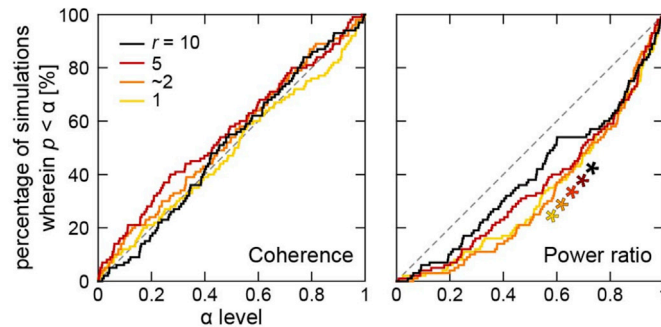| D [mm] =	Coherence		Power ratio	
	n^*	R_1–R_2[mm]	n^*	R_1–R_2[mm]
0	8	[–0.2 0.2–0.1]	0	[0.1 0.1 0.1]
4	46	[–0.3–4.0 0.1]	17	[–0.2–4.0 –0.6]
8	85	[–0.3–7.0 –0.1]	60	[–0.2–8.5 –0.7]
12	96	[–1.2–10.3 0.3]	84	[–0.3–12.9–1.2]
16	00	[–1.0–14.4 0.6]	94	[–0.9–16.6–1.3]
20	98	[–1.4–18.8 1.2]	99	[–0.8–20.9–1.2]
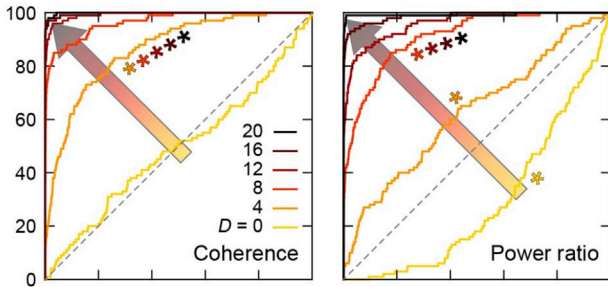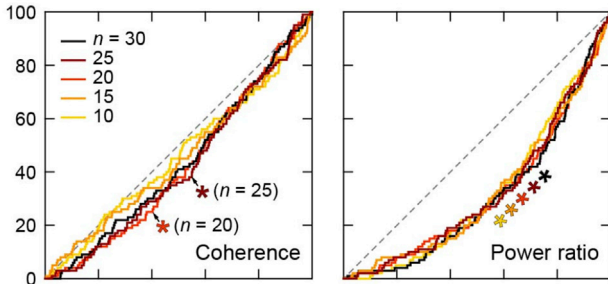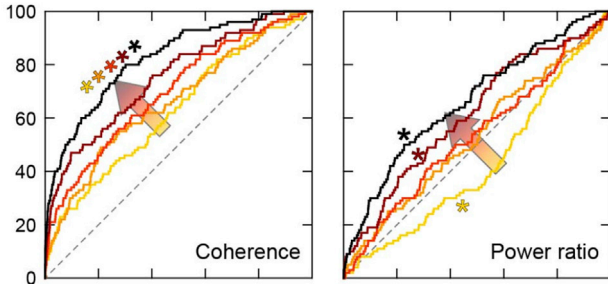
Fig. 4. Performance of the location-comparison method. In each plot, solid traces (ROC curves) present the proportion of simulations in which Condition 1 and Condition 2 localized in significantly different locations—according to the location-comparison test—as function of the theoretical false positive rate (α level). The dashed gray diagonal line is the proportion expected when there is no location difference. A star placed next to a curve indicates that it deviates significantly from the diagonal. Subsection 3.3 explains how to interpret the shape of these curves. Plots on the left are relative to simulations of coherence and plots on the right to simulations of power-ratio. A — Effect of the genuine inter-source distance (D) on the performance of the location-comparison method. In simulation 1, we varied D (0, 4, 8, 12, 16, and 20 mm), kept the sample size to $n = 17$ subjects, and set the ratio between SNR in condition 2 (SNR_2) and 1 (SNR_1) of each subject to a random value in 0.3–1.2. Large arrows indicate that significance detection rate increased as D increased. B and C — Effect of the sample size on the performance of the location-comparison method (simulation 2). In simulation 2, we varied n (10, 15, 20, 25, and 30 subjects), set SNR_2/SNR_1 to a random value in 0.3–1.2, and set D to 0 (B) or 4 mm (C). Large arrows indicate that significance detection rate increased as n increased for $D = 4$ mm. D — Effect of the SNR difference on the performance of the location-comparison

Kolmogorov-Smirnov test). The ROC curves were significantly below the main diagonal (meaning that the test was too conservative) for coherence maps based on 20 and 25 subjects ($p = .013$ and $p = .0055$ respectively) and for all power-ratio maps ($ps < 0.05$). This demonstrates that the location-comparison test had a good specificity regardless of the sample size, but that it is also too conservative.

Fig. 4 C presents the results obtained with the location-comparison method for $D = 4$ mm. In that setting, the ROC curves were significantly above the main diagonal ($ps < 0.05$; meaning that the test was sensitive to some extent) for all values of n for coherence maps, and only for large enough sample sizes ($n = 25$ and 30) for power ratio maps. These results speak to the good sensitivity of the dedicated location-comparison test. Importantly, the test became more sensitive as sample size increased.

5.4. Effect of SNR difference between conditions (simulation 3)

We now turn to the question of whether the location-comparison method is robust with respect to differences in SNR between the compared conditions. To that aim, we applied the location-comparison method to simulation 3 data in which $D = 0$ and the SNR in condition 1 was set to values 1, ~2, 5, and 10 times higher than that in condition 2.

Fig. 4 D summarizes the results. The ROC curves did not deviate significantly from the main diagonal for coherence maps ($ps > 0.05$, meaning that the test was not biased) and were significantly below the main diagonal for power-ratio maps ($ps < 0.05$, meaning that the test was too conservative). Hence, this demonstrates that the location-comparison test has a good specificity regardless of the SNR differences between conditions.

5.5. Location comparison in genuine brain-speech entrainment data

In this last result section, we illustrate the applicability of the location-comparison method by applying it to real brain-speech entrainment data acquired from 37 participants. As previously reported (Ahissar et al., 2001; Bourguignon et al., 2013a; Luo and Poeppel, 2007; Molinaro et al., 2016; Peelle et al., 2013; Vander Ghinst et al., 2016), MEG signals were coherent with speech temporal envelope at delta and theta frequencies (see Fig. 5). Source reconstruction revealed maxima of coherence in bilateral auditory cortices (see Table 3).

Obviously, given that coherence values were about 4 times higher at delta than theta frequencies, contrasting these maps would yield rather non-informative information as was illustrated in Subsection 5.1. The location-comparison method does not suffer from this limitation and thus can be used to compare the location of the sources corresponding to delta vs. theta entrainment in the left and right hemispheres despite the strong difference in SNR. This analysis revealed that the focus of delta entrainment in the right hemisphere was ~15 mm more posterior than that of the theta entrainment ($p < .0001$; see Table 3). No location difference was detected for the left hemisphere.

6. Discussion

The present study demonstrates that peak contrasts of MEG group-averaged maps tend to localize in inactive remote brain areas. More alarming, when not interpreted carefully, this inaccurate localization comes with a false positive rate inflated in an intractable manner. It is therefore crucial to remember that a significant contrast is not relevant for source location but simply implies that conditions differ on at least one parameter of the global source activation pattern. Here, we have introduced a new method to explicitly compare source location between two conditions. The associated statistical test tended to be too
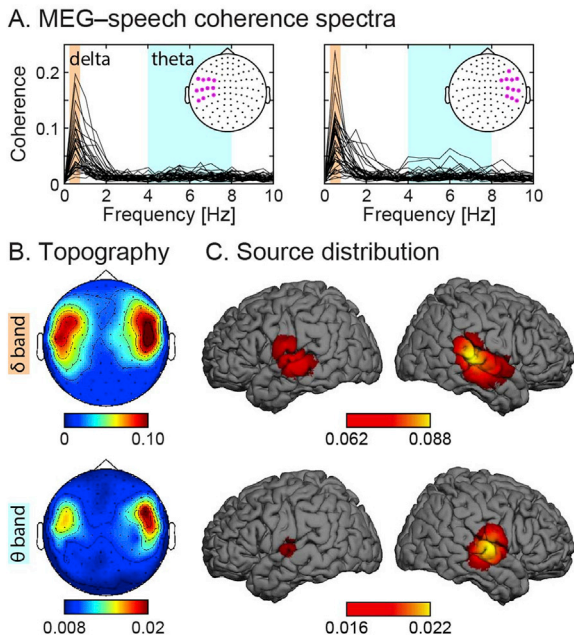
method (simulation 3). In simulation 3, we varied the SNR ratio ($r = SNR_1/SNR_2 = 1$, ~2, 5, and 10) and set n to 17 subjects and D to 0.

A. MEG–speech coherence spectra

B. Topography   C. Source distribution

Fig. 5. Illustration of the results of genuine brain-speech entrainment obtained in 37 subjects. A — Individual spectra of coherence estimated between MEG and speech temporal envelope (one trace per subject). A *left* — For each subjects the graph presents the maximum coherence across a left sensor selection. The sensor selection was the union of the 9 sensors with maximum coherence at delta frequencies (0.5 Hz) and the 9 sensors with maximum coherence at theta frequencies (4–8 Hz); 11 sensors in total. A *right* — Same for the right sensors. B — Sensor topography of the coherence at delta and theta frequencies (viewed from the top). C — Surface rendering of the coherence at delta and theta frequencies projected in the source space with a linearly constrained minimum variance beamformer. Maps are arbitrarily threshold at 70% of their maximum value.

Table 3

Assessment of genuine brain-speech entrainment data with the location-comparison method. From left to right are the coordinates of the peak coherence at delta (0.5-Hz) and theta (4–8-Hz) frequencies, their difference, and the results of the location-comparison test (F statistic and associated p-value). Results and presented for both left (upper row) and right (lower row) hemispheres.

	delta [mm]	theta [mm]	distance [mm]	$F_{3,998}$	p
left	[−65.6 −13.2 10.1]	[−64.1 −15.2 6.5]	[0.5 −2.0 −3.6]	1.11	0.35
right	[66.6 −28.7 5.9]	[65.5 −12.8 3.9]	[−1.2 15.9 −2.0]	18.3	<0.0001

conservative but still reached a statistical power of ~90% when genuine sources were distant of 12 mm in plausible simulations. Based on this method, we could demonstrate that right-hemisphere brain-speech entrainment in the delta band is actually located ~15 mm more anterior compared with that in the theta band.

6.1. Interpretation of contrast maps

The present study illustrates in practical cases that it is not straightforward to interpret contrasts of source-reconstructed MEG data. In MEG studies nowadays, conditions are often contrasted and the significance of contrast values is assessed statistically, e.g., with non-parametric maximum-based permutation tests (Maris and Oostenveld, 2007; Nichols and Holmes, 2002). In this approach, contrast values are compared to a permutation distribution of the maximum contrast value across a given volume, often the whole brain and sometimes a region of interest (Nichols and Holmes, 2002). This approach yields significance thresholds intrinsically corrected for the family-wise error. It is therefore tempting to ascribe loci of supra-threshold contrast to a significant difference in the involvement of the corresponding brain areas, but this interpretation is

inaccurate as exemplified by our results. In fact, the presence of a significant contrast between two conditions simply indicates a change in the statistical distribution of source activation, which may be related to a local change or to other parameters such as the mean and variability in sources' location and amplitude, and interactions thereof (Maris and Oostenveld, 2007).

Here, we have quantified the hazard of the naïve association of significant contrasts to significant difference in the involvement of the underlying brain areas. To that aim we have simulated a group of 17 subjects in whom a single source was inserted at the exact same location in two conditions and with the SNR in one condition randomly set to 30–120% of that in the other condition (simulation 1). In that setting, significant contrasts were rightfully observed in ~45% of the simulations but the contrast peaked on average 7–14 mm away from the genuine source. In a speech experiment, such mislocation could lead to interpreting a difference in the activity of the primary auditory cortex as a preferential recruitment of inferior frontal gyrus (Broca's area). Importantly, the mislocation problem persists for non-zero inter-source distance.

6.2. Contrast maps in other imaging modalities

As illustrated in the 1D toy model described in Section 2 (see also Fig. 1), the extent of contrast mislocation scales with the smoothness of the maps (quantified here by their FWHM) being contrasted. Importantly, mislocation errors are especially high when peaks in the maps of compared conditions are slightly off and of similar amplitude. That is, when profiles have similar amplitude separated by less than 0.5 FWHM, their contrast peaks 0.3–0.5 FWHM away from the closest profiles' peak.

Our toy model speaks to the generality of the problem of contrast mislocation. Peak contrasts in group-averaged parametric maps obtained with any neuroimaging modality are thus expected to localize away from genuinely active areas under the circumstances discussed above. Within-subject activation maps obtained with fMRI or PET are characterized by point-spread functions of less than 5-mm FWHM. The smoothness of group-averaged maps is however higher due to unavoidable variability in brain structural and functional anatomy (Steinmetz and Seitz, 1991; Uylings et al., 2005). Moreover, for group analyses, maps are often smoothed with Gaussian kernels of 5–15 mm FWHM. This is in line with the matched filter theorem according to which optimal smoothing kernel should match the spatial extent of the signal to be detected (Rosenfeld and Kak, 1982). Summing up all the sources of variability, the effective smoothness of group-averaged fMRI or PET maps is expected to be characterized by FWHMs of ~10–20 mm. Consequently, contrasting such maps could lead to location errors of at least few millimeters, and of up to ~10 mm in the worst-case scenario. Of notice, the contrast mislocation problem we are here referring to is different from the "hair-cut" problem in which peak of significance is shifted towards low-variance regions (Reimold et al., 2006; Ridgway et al., 2012).

In addition to all factors that contribute to blurring fMRI and PET maps, MEG is also affected by other factors, so that the extent of the mislocation is expected to be greater for contrasts of MEG maps. Indeed, the MEG signal generated by a single source affects many sensors (field spread), and in turn, source maps reconstructed using the most typical inverse solvers (LCMVB or MNE) are fraught with long-range correlation (spatial leakage). The way this correlation vanishes with distance depends on the location and orientation of the source (Hari et al., 1988), on the functional parameter under consideration, on the source reconstruction method used (Sekihara et al., 2005; Wens, 2015), and on the SNR in adaptive approaches (Barnes and Hillebrand, 2003; Gross et al., 2001; Van Veen et al., 1997). Also, inconsistent errors across subjects in MEG-MRI coregistration further increase the smoothness of group-averaged MEG maps (Singh et al., 1997). Consequently, maps of source-projected MEG data are typically much smoother than PET and fMRI maps, and their smoothness is difficult to quantify. The MEG technique is therefore more susceptible to mislocation errors in contrasts

of conditions than other imaging modalities. Even larger mislocation errors would be foreseeably obtained for maps reconstructed from electroencephalographic (EEG) signals since they spread more than MEG signals and their propagation is harder to model (Hämäläinen et al., 1993).

One way to avoid the extreme smoothness of MEG maps is to use spatially sparse imaging methods such as, e.g., minimum current estimation (Uutela et al., 1999), minimum mixed-norm estimation (Gramfort et al., 2013, 2012), or the multiple sparse priors approach (Friston et al., 2008). In this case the MEG maps are very focal, but this opposite extreme poses other problems to contrast two different conditions at the group level (unless maps are smoothed afterwards). In that respect, these data-driven sparse imaging techniques provide results that compare more to equivalent current dipoles (ECDs), which we shall discuss later on.

6.3. Direct contrast of normalized maps

A work around to be more specific to the question of location comparison with the contrast-based approach consists in normalizing individual maps to a common maximum value prior to contrasting them between conditions. However, this approach still fails to control the false positive rate when the SNR differs between conditions (data not shown). This is likely because when the SNR is too low, individual maps peak less consistently at the location of the genuine source. Consequently, the mean normalized map across subjects displays lower peak value (and potentially higher surrounding values) in the condition with the lowest SNR compared to that in the condition with the highest SNR, leading to statistically significant contrast values. In addition, peak contrast tends to locate away from genuine sources, simply because normalizing the maps tends to equalize the amplitude of the peaks in the to-be-compared group-averaged maps; a configuration that is susceptible to generate high mislocation errors. It is therefore mandatory to accept that significant contrast maps cannot provide more information than there being some difference between conditions in the global source activation pattern. In general, the location and amplitude of the peak contrast do not provide interpretable information.

6.4. Applicability of the proposed location-comparison procedure

In the present paper, we have presented a new approach that could complement the direct contrast approach in that it focuses exclusively on sources' location. This also means that other methods should be used to test for differences in e.g. amplitude or source orientation. The location-comparison method estimates the probability that sources in two conditions be located at different locations. It is based on a bootstrap estimation of the distance between map peaks in the two conditions that is assessed with a multivariate location test. Cluster-based methods have been developed previously that also capitalize on map peaks but their aim was rather to estimate a confidence volume for source location in individual conditions (Alikhanian et al., 2013; Gilbert et al., 2012).

Our location-comparison method appears to be a sensible tool to test for differences in location between conditions or groups. Indeed, we have demonstrated that this method tends to be too conservative, for unclear reasons especially for power-ratio maps and less so for coherence maps, although a sensitivity of ~90% was still achieved in realistic situations in which the genuine inter-source distance is 12 mm in a population of 17 subject. Even though increasing the number of subjects rendered the test more specific, it also made it more sensitive at the smallest inter-source distance tested ($D = 4$ mm). Finally, the test appeared to be robust to, and conservative enough over a wide range of difference in power-ratio.

To function properly, the location-comparison method necessitates that the location of the peaks in the to-be-compared conditions are not affected by a different location bias. Source maps reconstructed with classical inverse solutions might be fraught with such location bias due to many factors (Sekihara et al., 2005). The magnitude of such bias may also

be modulated by the SNR, which could lead to false positive detection of a location difference. Therefore, for the sake of comparing the location between conditions, it is necessary to opt for a source reconstruction method whose location bias is minimally affected by potential differences between conditions. To that aim, the same inverse solution should be applied to both conditions, as is common practice in MEG studies (Gross et al., 2013), and LCMVB appears to be a method of choice since it has no location bias even in the presence of noise (Sekihara et al., 2005). This theoretical nicety is however not always verified in practice and a depth bias may persist and be modulated by the SNR. Rendering the location-comparison test insensitive to source depth provides avenue for further developments. One simple possibility would be to focus only on location differences in the plane parallel to the skull so as to obtain a test blind to source depth. In practice this plane may be identified by the two first principal directions of the forward model at the mean coordinate of maxima in compared conditions ($(\mathbf{R}_1 + \mathbf{R}_2)/2$). The location-comparison test could then be applied to coordinates projected on this plane, making it robust to depth bias.

The method to specifically compare source location we here propose relates to, and extends traditional coordinate comparison methods based on ECD modeling. In ECD modeling, an algorithm seeks the ECD (coordinates, orientation and amplitude) that best explains the field topography (or a restriction of it) at a given time (Hämäläinen et al., 1993). Coordinates of ECDs in two different conditions can then be compared directly with a multivariate location test such as Hotelling T^2 test (Hotelling, 1931) or with other previously proposed methods (Litvak et al., 2007; Singh and Harding, 1996). Such procedure, however, requires that an ECD can actually be fitted in every subject, which is not always the case in real data. Also, ECD modeling has a restricted scope of application. It works only on linear parameters that preserve the amplitude and polarity of the magnetic field to reconstruct, such as averaged responses or cross-correlogram with a given reference signal, but not with nonlinear parameters such as power or coherence. The method we devised is free of these limitations. First, data from subjects with poorer results can be integrated in the comparison procedure and their weight will depend on the parameter under consideration and possible transformation (e.g., a square transformation emphasizes subjects' data that are highly significant whereas normalization by the maximum value gives the same weight to all subjects' data). Second, the method is not restricted to the type of parameter being imaged since it is based on source-space maps. Last, as maps are obtained with data-driven methods (e.g., LCMVB or MNE), the result of the comparison depends minimally on researchers' input as opposed to any analysis performed based on ECDs.

6.5. Brain-speech entrainment results

We have illustrated the yield of the location-comparison method on real brain-speech entrainment data and discuss now the implications of our finding that delta and theta brain-speech entrainment in the right hemisphere locate in different cortical areas. Importantly, the direct-contrast approach would have been of little use in this context since delta and theta entrainments are characterized by very different levels of coherence.

When listening to speech, human auditory cortical activity in the delta and theta bands entrain to speech rhythmicity, that is the brain-speech entrainment phenomenon (Ahissar et al., 2001; Bourguignon et al., 2013a; Luo and Poeppel, 2007; Molinaro et al., 2016; Peelle et al., 2013; Vander Ghinst et al., 2016). The strength of brain-speech entrainment is enhanced when listening to intelligible speech compared to non-intelligible speech (Ahissar et al., 2001; Luo and Poeppel, 2007; Peelle et al., 2013). As delta and theta frequency windows match with word/phrase and syllable repetition rates, it has been hypothesized that corresponding brain oscillations subserve chunking of speech into relevant segments for further speech recognition (Ahissar et al., 2001). In line with this view, delta fluctuations were found to track

sentence boundaries, even in the absence of prosodic cues (Ding et al., 2016). This finding demonstrates that auditory delta oscillations can be driven by syntax, and hence that they relate to some extent to language processing (Ding et al., 2016). In other words, the delta tracking does not only relate to low-level auditory encoding.

Here, we have demonstrated that right-hemisphere sources of delta and theta brain-speech entrainment seen with MEG localize in different cortical areas: in the supratemporal auditory cortex for theta entrainment, and ~15 mm more posterior, in the posterior part of the superior temporal gyrus (pSTG) for delta entrainment. The right pSTG has been long known to play a critical role in processing of speech prosody (Meyer et al., 2002). As we here demonstrate that right pSTG is the core structure underpinning delta entrainment, this cortical area could also play an important role in speech processing. Importantly, a previous study demonstrated that the information available at the right pSTG is forwarded to the left-hemisphere auditory and speech regions (Molinaro et al., 2016). Accordingly, the right pSTG appears to be one of the key nodes in the hierarchical structure necessary to process speech. It likely informs other brain areas about syntactic boundaries during continuous speech listening.

6.6. Perspectives

The present simulations of MEG coherence maps were designed to tightly match real brain-speech entrainment data (Ahissar et al., 2001; Bourguignon et al., 2013a; Luo and Poeppel, 2007; Molinaro et al., 2016; Peelle et al., 2013; Vander Ghinst et al., 2016). Our method could however apply to cortex–muscle coherence (Conway et al., 1995; Salenius et al., 1997, 1996), corticokinmeatic coherence with active, passive or observed movements (Bourguignon et al., 2013b, 2012, 2011; Jerbi et al., 2007; Marty et al., 2015; Piitulainen et al., 2015, 2013a, 2013b), and brain connectivity obtained with seed-based correlation or coherence (Brookes et al., 2012; de Pasquale et al., 2010; Hipp et al., 2012; Wens et al., 2014a, 2014b). Also, we have demonstrated that our method of location comparison works with power-ratio maps, although it is too conservative in that context. Of notice, in the case of seed-based connectivity mapping based on source envelope correlations, the spatial smoothness is typically much higher than for coherence or power maps (Wens et al., 2015), so the contrast mislocation problem is presumably even more stringent there.

Still, further developments are needed to optimize the location-comparison method. Here, statistical significance was assessed under the hypothesis that peak coordinates derived from the bootstrap procedure follow a multivariate normal distribution. Non-parametric tests could be developed that do not require this assumption. Also, statistics could be based on a generalization of the bias-corrected and accelerated (BC_a) bootstrap interval (Efron and Tibshirani, 1993). The BC_a bootstrap improves the percentile bootstrap interval by introducing some asymmetry in the lower and upper percentile bounds (Efron and Tibshirani, 1993). However, to the best of our knowledge, BC_a bootstrap—and percentile bootstrap more generally—applies only to univariate data. In addition, in our location-comparison method, resampling is performed on subjects' parametric maps rather than on subjects' individual coordinates as classically studied in bootstrap theory. Further work would be required to establish whether BC_a bootstrap applies in this context.

We did restrict the scope of the present article to situations in which only one source is active. This is clearly too simplistic with regard to brain complexity. Nevertheless, this simplification allowed us to demonstrate the potential flaws of the direct contrast approach, and to design and introduce our alternative location-comparison method. Future studies should extend the applicability of the location-comparison method to compare source location in multiple-source configurations.

7. Conclusion

Our study begs for a paradigm shift in the way we compare source-projected MEG maps across conditions or populations. The direct contrast approach is a crucial tool to reveal the existence of a difference between recruited networks but does not inform about the nature of this difference. In particular, the location of the maximum contrast should not be given any interpretation. Instead, when a contrast comes significant, additional analyses should be conducted on non-contrasted maps to identify in what aspect they differ. Our location-comparison method implements one such analysis; it specifically addresses the question of location difference between two conditions or populations.

Importantly, the problems related to the direct contrast approach are expected to occur with functional activation maps of other neuroimaging modalities (fMRI, PET), but admittedly, to a lower extent. Again, our location-comparison method also applies to maps of these other modalities and may thus provide a novel approach to the analysis of functional imaging data.

Conflicts of interest

None of the authors have potential conflicts of interest to be disclosed.

Acknowledgments

Appendix. Location-comparison test

Here, we derive a statistical test to estimate the probability that the distance between local maxima in two conditions is zero. This test is based on the three-dimensional bootstrap distribution of the coordinate difference Δr obtained as described in Subsection 2.3. There is an extensive literature on those tests and the simplest of them is Hotelling T^2 test, the direct generalization of Student t-test to multivariate data (Hotelling, 1931). However, location tests evaluate the null hypothesis that the cloud of points is centered on the origin 0, e.g., that the average coordinate difference $\overline{\Delta r} = 0$ for Hotelling T^2 test. What we rather need to examine here is whether 0 is a likely point of the cloud built from the sample distribution, or more precisely, whether 0 is included in a confidence volume containing $(1 - p) \times 100\%$ of the points, where p is the classical p-value associated to the test.

If Δr follows a multivariate Gaussian distribution, the most natural volumes to consider are ellipsoids of constant Mahalanobis distance (Mahalanobis, 1936)

$$d = \sqrt{\left(\Delta r - \overline{\Delta r}\right) C_{\Delta r}^{-1} \left(\Delta r - \overline{\Delta r}\right)}$$

to the center $\overline{\Delta r}$, where $C_{\Delta r}$ is the 3×3 sample covariance matrix of Δr. In the parametric setup, the Gaussian assumption implies that

$$d^2 \sim T_{3,\nu}^2 = \frac{3\nu}{\nu - 2} F_{3,\nu-2},$$

where $T_{3,\nu}^2$ is Hotelling T^2 distribution with three dimensions and $\nu = N_{\text{rep}}-1$ degrees of freedom, which relates to the better known F distribution as indicated in the above equation (Hotelling, 1931). Hence, to test the null hypothesis that 0 is included in the confidence volume, it suffices to compare the squared Mahalanobis distance $\overline{\Delta r} C_{\Delta r}^{-1} \overline{\Delta r}$ between 0 and the ellipsoid center $\overline{\Delta r}$ to the $(1 - p) \times 100$ th percentile of $T_{3,\nu}^2$. This is exactly the classical Hotelling T^2 test except that the sample covariance matrix $C_{\Delta r}$ of Δr is used instead of the sample covariance matrix $C_{\Delta r}/N_{\text{rep}}$ of $\overline{\Delta r}$. Of notice, in this framework, we can also estimate the volume of confidence of the maximum in one condition, as already proposed by Gross et al. (2003).

References

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, a, Mahncke, H., Merzenich, M.M., 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc. Natl. Acad. Sci. U. S. A 98, 13367–13372.

Alikhanian, H., Crawford, J.D., DeSouza, J.F.X., Cheyne, D.O., Blohm, G., 2013. Adaptive cluster analysis approach for functional localization using magnetoencephalography. Front. Neurosci. 7, 73.

Ashburner, J., Friston, K.J., 1999. Nonlinear spatial normalization using basis functions. Hum. Brain Mapp. 7, 254–266.

Ashburner, J., Neelin, P., Collins, D.L., Evans, A., Friston, K., 1997. Incorporating prior knowledge into image registration. Neuroimage 6, 344–352.

Barnes, G.R., Hillebrand, A., 2003. Statistical flattening of MEG beamformer images. Hum. Brain Mapp. 18, 1–12.

Bortel, R., Sovka, P., 2014. Approximation of the null distribution of the multiple coherence estimated with segment overlapping. Signal Process. 96, 310–314.

Bourguignon, M., De Tiège, X., De Beeck, M.O., Ligot, N., Paquier, P., Van Bogaert, P., Goldman, S., Hari, R., Jousmäki, V., 2013a. The pace of prosodic phrasing couples the listener's cortex to the reader's voice. Hum. Brain Mapp. 34, 314–326.

Bourguignon, M., De Tiège, X., Op de Beeck, M., Pirotte, B., Van Bogaert, P., Goldman, S., Hari, R., Jousmäki, V., 2011. Functional motor-cortex mapping using corticokinematic coherence. Neuroimage 55, 1475–1479.

Bourguignon, M., De Tiège, X., Op de Beeck, M., Van Bogaert, P., Goldman, S., Jousmäki, V., Hari, R., 2013b. Primary motor cortex and cerebellum are coupled with the kinematics of observed hand movements. Neuroimage 66, 500–507.

Bourguignon, M., Jousmäki, V., Op de Beeck, M., Van Bogaert, P., Goldman, S., De Tiège, X., 2012. Neuronal network coherent with hand kinematics during fast repetitive hand movements. Neuroimage 59, 1684–1691.

Bourguignon, M., Piitulainen, H., De Tiège, X., Jousmäki, V., Hari, R., 2015. Corticokinematic coherence mainly reflects movement-induced proprioceptive feedback. Neuroimage 106, 382–390.

Brookes, M.J., Woolrich, M.W., Barnes, G.R., 2012. Measuring functional connectivity in MEG: a multivariate approach insensitive to linear source leakage. Neuroimage 63, 910–920.

Conway, B.A., Halliday, D.M., Farmer, S.F., Shahani, U., Maas, P., Weir, A.I., Rosenberg, J.R., 1995. Synchronization between motor cortex and spinal motoneuronal pool during the performance of a maintained motor task in man. J. Physiol. 489, 917–924.

Dale, A.M., Sereno, M.I., 1993. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. J. Cognit. Neurosci. 5, 162–176.

de Pasquale, F., Della Penna, S., Snyder, A.Z., Lewis, C., Mantini, D., Marzetti, L., Belardinelli, P., Ciancetta, L., Pizzella, V., Romani, G.L., Corbetta, M., 2010. Temporal dynamics of spontaneous MEG activity in brain networks. Proc. Natl. Acad. Sci. U. S. A 107, 6040–6045.

Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2016. Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci. 19, 158–164.

Efron, B., 1979. Bootstrap methods: another look at the jacknife. Ann. Stat. 7, 1–26.

Efron, B., Tibshirani, R.J., 1993. An introduction to the bootstrap. Refrig. Air Cond. 57, 436.

Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., Henson, R., Flandin, G., Mattout, J., 2008. Multiple sparse priors for the M/EEG inverse problem. Neuroimage 39, 1104–1120.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P., Frith, C.D., Frackowiak, R.S.J., 1994. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.

Gilbert, J.R., Shapiro, L.R., Barnes, G.R., 2012. A peak-clustering method for MEG group analysis to minimise artefacts due to smoothness. PLoS One 7.

Gramfort, A., Kowalski, M., Hämäläinen, M., 2012. Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. Phys. Med. Biol. 57, 1937–1961.

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

Gramfort, A., Strohmeier, D., Haueisen, J., Hämäläinen, M.S., Kowalski, M., 2013. Time-frequency mixed-norm estimates: sparse M/EEG imaging with non-stationary source activations. Neuroimage 70, 410–422.

Gross, J., Baillet, S., Barnes, G.R., Henson, R.N., Hillebrand, A., Jensen, O., Jerbi, K., Litvak, V., Maess, B., Oostenveld, R., Parkkonen, L., Taylor, J.R., van Wassenhove, V., Wibral, M., Schoffelen, J.M., 2013. Good practice for conducting and reporting MEG research. Neuroimage 65, 349–463.

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. U. S. A 98, 694–699.

Gross, J., Timmermann, L., Kujala, J., Salmelin, R., Schnitzler, A., 2003. Properties of MEG tomographic maps obtained with spatial filtering. Neuroimage 19, 1329–1336.

Halliday, D., Rosenberg, J., Amjad, A., Breeze, P., Conway, B., Farmer, S., 1995. A framework for the analysis of mixed time series/point process data - theory and application to the study of physiological tremor, single motor unit discharges and electromyograms. Prog. Biophys. Mol. Biol. 64, 237–278.

Hämäläinen, M., Hari, R., Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V., 1993. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev. Mod. Phys. 65, 413–497.

Hämäläinen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Med. Biol. Eng. Comput. 32, 35–42.

Hari, R., Joutsiniemi, S.L., Sarvas, J., 1988. Spatial resolution of neuromagnetic records: theoretical calculations in a spherical model. Electroencephalogr. Clin. Neurophysiol. Evoked Potentials 71, 64–72.

Hillebrand, A., Singh, K.D., Holliday, I.E., Furlong, P.L., Barnes, G.R., 2005. A new approach to neuroimaging with magnetoencephalography. Hum. Brain Mapp. 25, 199–211.

Hipp, J.F., Hawellek, D.J., Corbetta, M., Siegel, M., Engel, A.K., 2012. Large-scale cortical correlation structure of spontaneous oscillatory activity. Nat. Neurosci. 15, 884–890.

Hotelling, H., 1931. The generalization of Student's ratio. Ann. Math. Stat. 2, 360–378.

Jerbi, K., Lachaux, J.-P., N'Diaye, K., Pantazis, D., Leahy, R.M., Garnero, L., Baillet, S., 2007. Coherent neural representation of hand speed in humans revealed by MEG imaging. Proc. Natl. Acad. Sci. U. S. A 104, 7676–7681.

Litvak, V., Zeller, D., Oostenveld, R., Maris, E., Cohen, A., Schramm, A., Gentner, R., Zaaroor, M., Pratt, H., Classen, J., 2007. LTP-like changes induced by paired associative stimulation of the primary somatosensory cortex in humans: source analysis and associated changes in behaviour. Eur. J. Neurosci. 25, 2862–2874.

Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54, 1001–1010.

Mahalanobis, P.C., 1936. On the generalised distance in statistics. Proc. Natl. Inst. Sci. India 2, 49–55.

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

Marty, B., Bourguignon, M., Jousmäki, V., Wens, V., Op de Beeck, M., Van Bogaert, P., Goldman, S., Hari, R., De Tiège, X., 2015. Cortical kinematic processing of executed and observed goal-directed hand actions. Neuroimage 119, 221–228.

Meyer, M., Alter, K., Friederici, A.D., Lohmann, G., Von Cramon, D.Y., 2002. FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum. Brain Mapp. 17, 73–88.

Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., Carreiras, M., 2016. Out-of-synchrony speech entrainment in developmental dyslexia. Hum. Brain Mapp. 37, 2767–2783.

Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15, 1–25.

Peelle, J.E., Gross, J., Davis, M.H., 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cerebr. Cortex 23, 1378–1387.

Piitulainen, H., Bourguignon, M., De Tiège, X., Hari, R., Jousmäki, V., 2013a. Corticokinematic coherence during active and passive finger movements. Neuroscience 238, 361–370.

Piitulainen, H., Bourguignon, M., De Tiège, X., Hari, R., Jousmäki, V., 2013b. Coherence between magnetoencephalography and hand-action-related acceleration, force, pressure, and electromyogram. Neuroimage 72, 83–90.

Piitulainen, H., Bourguignon, M., Hari, R., Jousmäki, V., 2015. MEG-compatible pneumatic stimulator to elicit passive finger and toe movements. Neuroimage 112, 310–317.

Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as "asymmetric sampling in time". In: Speech Communication, pp. 245–255.

Reimold, M., Slifstein, M., Heinz, A., Mueller-Schauenburg, W., Bares, R., 2006. Effect of spatial smoothing on t-maps: arguments for going back from t-maps to masked contrast images. J. Cerebr. Blood Flow Metabol. 26, 751–759.

Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61, 1402–1418.

Ridgway, G.R., Litvak, V., Flandin, G., Friston, K.J., Penny, W.D., 2012. The problem of low variance voxels in statistical parametric mapping; a new hat avoids a "haircut". Neuroimage 59, 2131–2141.

Rosenfeld, A., Kak, A.C., 1982. Digital picture processing. In: Digital Picture Processing, p. 435.

Salenius, S., Portin, K., Kajola, M., Salmelin, R., Hari, R., 1997. Cortical control of human motoneuron firing during isometric contraction. J. Neurophysiol. 77, 3401–3405.

Salenius, S., Salmelin, R., Neuper, C., Pfurtscheller, G., Hari, R., 1996. Human cortical 40 Hz rhythm is closely related to EMG rhythmicity. Neurosci. Lett. 213, 75–78.

Sekihara, K., Sahani, M., Nagarajan, S.S., 2005. Localization bias and spatial resolution of adaptive and non-adaptive spatial filters for MEG source reconstruction. Neuroimage 25, 1056–1067.

Singh, K.D., Harding, G.F.A., 1996. Monte-Carlo analysis and confidence region ellipsoids for equivalent current dipole solutions to MEG and EEG data. In: Biomag, 96, pp. 346–349.

Singh, K.D., Holliday, I.E., Furlong, P.L., Harding, G.F.A., 1997. Evaluation of MRI-MEG/EEG co-registration strategies using Monte Carlo simulation. Electroencephalogr. Clin. Neurophysiol. 102, 81–85.

Steinmetz, H., Seitz, R.J., 1991. Functional anatomy of language processing: neuroimaging and the problem of individual variability. Neuropsychologia 29, 1149–1161.

Taulu, S., Simola, J., 2006. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys. Med. Biol. 51, 1759–1768.

Taulu, S., Simola, J., Kajola, M., 2005. Applications of the signal space separation method. October 53, 3359–3372.

Uutela, K., Hämäläinen, M., Somersalo, E., 1999. Visualization of magnetoencephalographic data using minimum current estimates. Neuroimage 10, 173–180.

Uylings, H.B.M., Rajkowska, G., Sanz-Arigita, E., Amunts, K., Zilles, K., 2005. Consequences of large interindividual variability for human brain atlases: converging macroscopical imaging and microscopical neuroanatomy. In: Anatomy and Embryology, pp. 423–431.

Van Veen, B.D., van Drongelen, W., Yuchtman, M., Suzuki, A., 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans. Biomed. Eng. 44, 867–880.

Vander Ghinst, M., Bourguignon, M., Op de Beeck, M., Wens, V., Marty, B., Hassid, S., Choufani, G., Jousmäki, V., Hari, R., Van Bogaert, P., Goldman, S., De Tiège, X., 2016. Left superior temporal gyrus is coupled to attended speech in a cocktail-party auditory scene. J. Neurosci. 36, 1596–1606.

Wens, V., 2015. Investigating complex networks with inverse models: analytical aspects of spatial leakage and connectivity estimation. Phys. Rev. E - Stat. Nonlinear Soft Matter Phys. 91.

Wens, V., Bourguignon, M., Goldman, S., Marty, B., Op De Beeck, M., Clumeck, C., Mary, A., Peigneux, P., Van Bogaert, P., Brookes, M.J., De Tiège, X., 2014a. Inter- and intra-subject variability of neuromagnetic resting state networks. Brain Topogr. 27, 620–634.

Wens, V., Marty, B., Mary, A., Bourguignon, M., Op de Beeck, M., Goldman, S., Van Bogaert, P., Peigneux, P., De Tiège, X., 2015. A geometric correction scheme for spatial leakage effects in MEG/EEG seed-based functional connectivity mapping. Hum. Brain Mapp. 36, 4604–4621.

Wens, V., Mary, A., Bourguignon, M., Goldman, S., Marty, B., Op de Beeck, M., Van Bogaert, P., Peigneux, P., De Tiège, X., 2014b. About the electrophysiological basis of resting state networks. Clin. Neurophysiol. 125, 1711–1713.