Koskinen, M.; Seppä, M.

Uncovering cortical MEG responses to listened audiobook stories

Published in:
NEUROIMAGE

DOI:
10.1016/j.neuroimage.2014.06.018

Published: 01/01/2014

Please cite the original version:
Koskinen, M., & Seppä, M. (2014). Uncovering cortical MEG responses to listened audiobook stories. NEUROIMAGE, 100, 263-270. DOI: 10.1016/j.neuroimage.2014.06.018
Uncovering cortical MEG responses to listened audiobook stories

M. Koskinen *, M. Seppä

Brain Research Unit at AMI Centre, and MEG Core, O.Y. Lounasmaa Laboratory, Aalto University School of Science, P.O. Box 13000, FI-00076 Aalto, Finland

ARTICLE INFO

Article history:
Accepted 6 June 2014
Available online 17 June 2014

Keywords:
Canonical correlation analysis (CCA)
Forward modeling
MEG
Single-trial analysis
Spatial filtering
Wavelet transform

ABSTRACT

Naturalistic stimuli, such as normal speech and narratives, are opening up intriguing prospects in neuroscience, especially when merging neuroimaging with machine learning methodology. Here we propose a task-optimized spatial filtering strategy for uncovering individual magnetoencephalographic (MEG) responses to audiobook stories. Ten subjects listened to 1-h-long recording once, as well as to 48 repetitions of a 1-min-long speech passage. Employing response replicability as statistical validity and utilizing unsupervised learning methods, we trained spatial filters that were able to generalize over datasets of an individual. For this blind-signal-separation (BSS) task, we derived a version of multi-set similarity-constrained canonical correlation analysis (SimCCA) that theoretically provides maximal signal-to-noise ratio (SNR) in this setting. Irrespective of significant noise in unaveraged MEG traces, the method successfully uncovered feasible time courses up to ~120 Hz, with the most prominent signals below 20 Hz. Individual trial-to-trial correlations of such time courses reached the level of 0.55 (median 0.33 in the group) at ~0.5 Hz, with considerable variation between subjects. By this filtering, the SNR increased up to 20 times. In comparison, independent component analysis (ICA) or principal component analysis (PCA) did not improve SNR notably. The validity of the extracted brain signals was further assessed by inspecting their associations with the stimulus, as well as by mapping the contributing cortical signal sources. The results indicate that the proposed methodology effectively reduces noise in MEG recordings to that extent that brain responses can be seen to nonrecurring audiobook stories. The study paves the way for applications aiming at accurately modeling the stimulus–response-relationship by tackling the response variability, as well as for real-time monitoring of brain signals of individuals in naturalistic experimental conditions.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Introduction

Good storytelling can be very captivating for the listener. Stories engage brain functions from perception, attention shifting, and comprehension to emotional immersion in the story, providing a versatile and modern approach for neuroscience research. However, natural speech and narratives have yet gained only modest attention in neuroscience experimentation. As speech is a relatively unpredictable one-time-only phenomenon, it offers a great challenge for reliably detecting related activation patterns from noisy functional brain-imaging data. Nonrecurring audio streams are in stark contrast to typical strictly controlled event-related analysis designs with simplified brief and repeated stimuli. Therefore, novel methodological approaches are required for the analysis.

Recent work has indicated that listened narratives may engage brain regions beyond those activated by non-speech sounds, single words, or short isolated sentences. These areas include, for example, medial and inferior frontal cortices, precuneus, cingulate cortex, and amygdala (Ben-Yakov et al., 2012; Boldt et al., 2013; Lerner et al., 2011; Regev et al., 2013; Wilson et al., 2008; Zion Columbic et al., 2013). Moreover, in functional magnetic resonance imaging (fMRI), listened stories have been used for mapping of brain areas linked to syntactic processing (Brennan et al., 2012) or responding to transitions in narration (Whitney et al., 2009). Recently, growing interest has emerged for electrophysiological signals that can provide superior temporal resolution sufficient for detecting activity patterns at syllable rates (~8 Hz) and above. This approach has, for example, enabled identifying segments of listened news with magnetoencephalography (MEG; Koskinen et al., 2013), finding brain correlates for sentence comprehension (Peelle et al., 2013), and resolving attention mechanisms in a “cocktail party” setup with electroencephalography (EEG; O’Sullivan et al., 2014), and with electrocorticography (ECoG; Zion Columbic et al., 2013).

In the current study, we aim at uncovering cortical MEG responses to listened 1-h-long nonrecurring audiobook recording. The work has three key focuses. First, our main emphasis is to uncover physiologically feasible time series from the raw MEG data. Topographic mapping is also considered, but only for validation of the main analysis. Second, as the listened stories have different effects on different brain regions, we aim for data-driven analysis and search for statistical regularities in brain responses, without constraining the search space by stimulus features that might or might not be reflected in the raw data. Third, our purpose is to pave the way for applications aiming at real-time signal analysis of...
individual subjects’ MEG data in naturalistic experimental conditions. Thus, we focus on subject-wise task-optimized analysis models.

The signal-to-noise ratio (SNR) of single-trial MEG recordings is very low ($\ll 1$). Thus to obtain reliable signals without averaging, we concentrate on signal replicability and require that a successful analysis model, reliably uncovering brain responses to nonrecurring speech, produces replicable MEG signatures for a repeated stimulus irrespective of the signal variability and prominent noise in the recorded data. Therefore, in addition to the nonrecurring speech stimulus, a supporting dataset with repeated trials (1-min-long independent stimulus) was recorded for model training. Notably, the analysis model should generalize over trials and over datasets.

We utilize a linear mapping model, representing spatial filtering of MEG signals in sensor space, combined with parameter estimation based on canonical correlation analysis (CCA). In its standard version, CCA finds linear projections for two distinct multivariate datasets so that the resulting random variables become maximally correlated between the sets. For the replicability requirement, however, our aim is to find a single weighting matrix (canonical basis vectors, i.e. spatial filters) that provides maximal correlations between projections of $N$ trials. This requirement relates to similarity-constrained CCA problem (SimCCA; Lahti et al., 2009). A similar idea has been utilized with EEG in computation of intra- and intersubject correlations (Dmochowski et al., 2012). Here, we derive a robust singular-value-decomposition (SVD) based solution that is practical for large datasets, with the aim of uncovering reliable brain responses from the recorded data, i.e. blind signal separation (BSS). Importantly, maximizing trial-to-trial correlations by SimCCA corresponds to maximizing the SNR (e.g. Bershad and Rockmore, 1974). Such maximization is expected to reduce noise sensitivity of the estimates, to prevent overfitting, and to improve the model performance with novel test datasets. Thus, the proposed approach differs from the well-known spatial filtering schemes, such as beamforming (van Veen et al., 1997) that is based on anatomical models, or independent-component (ICA) and principal-component analyses (PCA) that do not intrinsically utilize sample ordering (e.g. Karhunen and Hao, 2011). We show that our method succeeds in revealing feasible brain responses to nonrecurring audiobook stimuli in individual subjects.

Materials and methods

Subjects and recordings

Ten native Finnish-speaking and normally hearing participants gave their written informed consent for the study. The experimental setup was approved by the ethics committee of the Helsinki and Uusimaa Hospital district.

MEG was recorded with a 306-channel Elekta Neuromag™ system (Elekta Oy, Helsinki, Finland). The passband was 0.03–330 Hz and sampling frequency 1000 Hz. The auditory speech stimulus was presented through a non-magnetic open-field audio speaker (Panphonics Ltd., Tampere, Finland) ~2.5 m in front of the subject (~40° upwards). The recordings were arranged without visual stimulation or gaze wandering. The map had some relevance with respect to the story heard.

The MEG recordings were arranged in two 1-h sessions on separate days. The first half of the sessions was spent listening to a 1-min-long passage repeatedly 24 times, with at least 5-s spacing between the trials. After every eight trials, the subjects’ alertness was checked by short conversation and sitting position was adjusted if needed. The second half was reserved for the ongoing story, listened in two pieces with a break in between. Altogether, the recordings comprised 48 repeated trials (recorded in 2–6 pieces) and a 59.5-min nonrecurring single-trial (recorded in four pieces). To exclude possible trivial explanation for results by spurious magnetic fields or by other artifacts, we also collected 48 trials of empty-room data in a similar manner.

MEG preprocessing

The recordings were preprocessed with the signal-space-separation technique (SSS; Taulu et al., 2004; Taulu and Kajola, 2005) as implemented in the MaxFilter program (Elekta Oy, Helsinki, Finland) with the default parameter settings to reduce noise and to convert each file into the standard head position. All 204 gradiometer channels of the device were selected for further processing. Principal component analysis (PCA) was applied subject-wise to reduce data dimensionality from 204 down to the maximum degrees of freedom (dof) remaining after the SSS procedure, the value provided by the MaxFilter program. All subject-wise recordings were used for PCA computation and the minimum dof of these (~68) determined the number of selected PCA basis vectors ($N_r \times \text{dof}$ matrix $\mathbf{A}$, $N_r = 204$ i.e. the number of channels).

Of the 48 trials with the recurrent stimuli, 33 were randomly assigned to training and 15 to test sets. All data were wavelet-transformed with the Mexican-hat wavelet into 17 scales with exponential spacing corresponding to center frequencies approximately between 0.5 and 120 Hz (Torrence and Compo, 1998).

Spatial filtering with CCA

Generally, the spatial filtering corresponds to a linear mapping

$$\mathbf{Z} = \mathbf{W}^T \mathbf{X},$$

where the $N_r \times N_c$ matrix $\mathbf{X}$ is the multi-channel MEG time series, $N_r$ is the number of channels, and $N_c$ is the number of time points. $N_r \times n$ matrix $\mathbf{W}$ contains a set of $n$ spatial filters $w_1, \ldots, w_n$ that set weights to the different MEG sensors (or to PCA mappings in this case with $N_r$ replaced by dof, see the MEG preprocessing section) resulting in a filtered $n \times N_c$ signal matrix

$$\mathbf{Z} = (z_1 \ldots z_n)^T$$

where $z_i = (\mathbf{w}^T \mathbf{X})^T$. Using CCA nomenclature, we call the spatial filters $\mathbf{w}_i$ the CCA basis vectors and the respective results $z_i$ the canonical variables. Our aim is to estimate weights $\mathbf{W}$ that provide variates $z_i$ correlating maximally between the single-trial responses. Here, the CCA (Hardoon et al., 2004; Hotelling, 1936) is utilized with the constraint that the matrix $\mathbf{W}$ should be same for all trial datasets (i.e. similarity-constrained CCA, see e.g. Dmochowski et al., 2012; Lahti et al., 2009). In the following, we briefly summarize our implementation based on data whitening and singular value decomposition (SVD) (Karhunen and Hao, 2011; Hyvärinen et al., 2001).

As usual, $\mathbf{X}$ is a zero-mean multi-dimensional dataset, $\mathbf{E} = (e_1 \ldots e_n)$ the matrix containing the unit-norm eigenvectors of the covariance matrix $\mathbf{C}_{xx} = \mathbf{E} \mathbf{E}^T$, and $\mathbf{D} = \text{diag}(d_1, \ldots, d_n)$ is the diagonal matrix of the eigenvalues of $\mathbf{C}_{xx}$. The whitened data $\mathbf{X}$ can be expressed as

$$\mathbf{\tilde{X}} = \mathbf{D}^{-1/2} \mathbf{E}^T \mathbf{X}.$$ 

In a standard (i.e. not similarity-constrained) CCA, data whitening is performed separately for a second dataset (in a similar fashion) and the between-sets covariance matrix $\mathbf{C}_{zw}$ of the whitened data is factorized by SVD as

$$\mathbf{C}_{zw} = \mathbf{U} \mathbf{Z} \mathbf{V}^T = \sum_{i=1}^{L} \lambda_i \mathbf{u}_i \mathbf{v}_i^T.$$ 

The singular column vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ in matrices $\mathbf{U}$ and $\mathbf{V}$ are the basis vectors for CCA in respective whitened signal spaces. Corresponding
singular values \( \rho \) in the diagonal of \( S \) represent the canonical correlations. For the constrained solution, we first (i) used the training trials for estimating a common whitening matrix from the average covariance matrix \( C_{\text{ave}} = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} X_i X_j^T \) and thereby provide common linear transform for all trials, we formed the between-sets covariance matrix \( C_s \) by averaging over all the combinations of two different training data sets. Denoting the whitened version of data-set \( X \) by \( \hat{X} \), we have

\[
C_s = \frac{1}{N(N-1)(N_t-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{X}_i \hat{X}_j^T = \frac{NXX^T - I}{(N(N-1)(N_t-1))},
\]

where \( X \) is the whitened time series averaged over trials and \( I \) is the identity matrix. It is useful to note that the condition \( j \neq i \) in the summations above (the cause of matrix \( I \)) makes our method seemingly different from the principal component analysis (PCA) method applied to \( \hat{X} \), i.e. our method considers only between-trials covariance of the whitened data. However, the identity matrix only shifts the singular values by a constant, and does not change the singular vectors or their ordering. Thus, our method is essentially equivalent to performing PCA on the averaged whitened data. Finally, the matrix \( W \) in the original signal space is

\[
W = ED^{-1/2}U.
\]

In our implementation, the model performance was assessed subject-wise by applying the trained model to the data of the 15 test trials and by calculating their pair-wise correlations separately for \( n \) CCA projections in each wavelet scale. The \( t \)-test (two-tailed) was employed to find statistically significant deviations from zero correlation (\( p < 0.01/n \)).

**Signal-to-noise ratio (SNR)**

Notably, maximizing trial-to-trial correlations \( \rho \) corresponds to maximizing the SNR, given as (Bershad and Rockmore, 1974)

\[
\text{SNR} = J \frac{\rho}{1-\rho} + K.
\]

Here, the coefficients for the unbiased estimate are \( J = \frac{1}{2} \) and \( K = \frac{1}{2} (1-J) \approx 0 \) with large \( N_t \). SNR was quantified for test trials in consecutive analysis stages after SSS, after PCA, after wavelet transform, and after SimCCA. Prior work has shown that CCA as a preprocessing stage before ICA may improve SNR in some cases (Karhunen and Hao, 2011; Karhunen et al., 2012). Thus, for comparison, SNR computations with ICA were included.

**Stimulus–response relationship**

We assessed the applicability of trained spatial filters in revealing plausible brain responses to 1-h long nonrecurring audiobook stories by computing stimulus–response correlations. The analysis was limited to the first canonical variates in each scale, corresponding to the highest trial-to-trial correlations with training data. The stimulus signal was rectified, downsampled to 1000 Hz, wavelet-transformed and split into 1-min epochs with 5-s spacing to reduce the temporal dependency between the consecutive epochs. Similar splitting was done with MEG time courses. The Pearson’s correlation coefficient was computed in each epoch. The delay in the brain signals with respect to stimulus was considered by shifting MEG data 0–249 ms in 1-ms steps to find the lag for maximal correlation. Two-tailed \( t \)-test with Bonferroni correction was used for evaluating the statistical significance of the correlations deviating from zero (with 250 time shifts \( p < 0.01/250 \)). The procedure was repeated for each of 17 wavelet scales separately. To verify the validity of \( t \)-test results, we formed control-data null distribution by computing correlations between the same signals but taken from different epochs in random.

**Current sources contributing to canonical variates**

We constructed neuroanatomical sensitivity maps to visualize the influence of the source currents on the amplitude of each canonical variate. Let \( A \) represent \( N_s \times dof \) matrix of PCA basis vectors earlier used for preprocessing, \( W \) is \( dof \times N_s \) matrix holding the CCA basis vectors, and \( G \) is \( N_s \times 3N_{\text{source}} \) gain matrix relating the source currents to the sensor signals, i.e. the forward solution. Here, \( N_s \) is the number of statistically significant canonical basis vectors, \( N_{\text{source}} \) is the number of source locations, and the multiplier 3 comes from three geometric dipole orientations \( (x, y, z) \). We compute

\[
S = W^T G.
\]

where \( N_s \times 3N_{\text{source}} \) matrix \( S \) represents the sensitivity of the different basis vectors (rows) to unit current dipoles in three orthogonal orientations \( x, y, \) and \( z \), for each location. For visualization, these orientations are combined into single sensitivity values \( v_j = \sqrt{s_{1,jx}^2 + s_{1,jy}^2 + s_{1,jz}^2} \) for each basis vector \( j \) and are further normalized into range 0 ... 1 as \( v_j = \left( v_{1,j} \ldots v_{N_{\text{source}},j} \right)^T / \max(v_{i,j}) \). In other words, \( G \) introduces anatomical mapping extending the idea of plotting the spatial filter weights alone (e.g. O’Sullivan et al., 2014). The approach is a simplified alternative, e.g. to visualizing correlations between the canonical variates and the time series at source points (Lankinen et al., 2014).

T1-weighted anatomical MR images of seven subjects out of ten were available from prior studies. With the permission of subjects, the images were reanalyzed to estimate matrix \( G \) by using FreeSurfer software (http://surfer.nmr.mgh.harvard.edu) and MNE Suite software package (http://www.martinos.org/mne/). For settings, we used 7 mm grid spacing and smoothing constant 5 for graphics. The registration of MEG coordinates with MRI coordinate system was done separately for \( n \) CCA projections in each wavelet scale. The \( t \)-test (two-tailed) was employed to find statistically significant deviations from zero correlation (\( p < 0.01/n \)).

**Results**

All MEG data were used in the analysis, except with one subject –11-min piece of the ongoing story was discarded due to recording artifacts.

**SimCCA**

The model training and primary performance testing was carried out with 33 and 15 1-min-long repeated trials, respectively. Fig. 1A represents the maximal trial-to-trial correlations (i.e. between the 1st canonical variates) for each subject and the wavelet scale. Notably, statistically significant correlations (\( p < 0.01 \), \( t \)-test with Bonferroni correction) could be found throughout the 0.5–120 Hz frequency range, most prominent ones residing below ~20 Hz. The correlations varied considerably between subjects, e.g. from 0.11 up to 0.55 (median 0.33) at 0.5 Hz. For comparison, the same procedure did not reveal any statistically significant correlations with empty-room data, discounting the possibility of spuriously induced magnetic fields. Fig. 2 demonstrates canonical variate time series.
Closers inspections of SNR in different processing stages revealed that spatial filtering with SimCCA was the most influential step in the pipeline (Fig. 3). With the test trials at 0.5 Hz, SNR increased 9.1–20.0 fold (median 12.5 in the group), from 0.021–0.074 (median 0.047) to 0.29–1.21 (median 0.49). These numbers refer to the first canonical variates at 0.5 Hz providing maximal effects. Notably, PCA or ICA themselves did not improve SNR considerably in this context and the ICA after SimCCA decreased the SNR of that with SimCCA alone.

**Stimulus–response correlations**

Spatially filtered MEG data (the 1st canonical variates), computed for the nonrecurring 1-h-long audiobook stimuli, correlated significantly with the speech envelope ($p < 0.01$, $t$-test with Bonferroni correction for 250 tests, i.e. lags; Fig. 4A). More specifically, out of 40,250 $t$-tests (number of statistically significant canonical basis vectors in the groups $\times$ 250 lags), 24,049 (60%, i.e. over relatively wide range of lags) were statistically significant. The corresponding number with the randomized control data was 2 with an expected value of 1.6. The peak correlations were typically distributed around ~100 ms lag for the MEG signal, limiting the analysis to lags between 0 and 249 ms (Fig. 4B). The results were significant with wavelet scales up to 31 Hz with all subjects. However, the most pronounced correlations resided below 3 Hz with considerable inter-individual differences (up to 0.61, median 0.45 at 0.5 Hz). Practically equal correlations were gained with 1-min training trials data analyzed with the same lags for reference. For interpretation, it is useful to note that the individual signals are not necessarily consistent in the group or may not represent activity in primary auditory regions. Instead, the results merely indicate that the tested time series were stimulus-related at the scales up to 31 Hz.

**Anatomical mapping**

We computed topographic maps for qualitative feasibility evaluation with five subjects’ data. These maps represent the sensitivity of individually optimized spatial filters to source amplitude at different locations on the cortex. Fig. 5 demonstrates selected results; complete data are presented in the Supplementary material. Characteristically, the spatial filters showed selectivity with different cortical regions and hemispheres, and the influence of auditory regions is obvious. Sources in superior temporal sulcus/gyrus, inferior frontal regions, motor/somatosensory, and parietal cortices consistently contribute to filters across subjects. Heterogeneity of subject-wise optimized models is considerable.

**Discussion**

We utilized a task-optimized spatial filtering strategy for uncovering single-trial brain responses to naturalistic stimuli. The method succeeded in revealing relevant individual MEG time-courses from audiobook stimuli. The feasibility of the analysis was demonstrated in three ways. First, the spatial filters were capable of uncovering replicable brain signals with
Fig. 2. An example of spatial filtering (i.e. canonical variates) at wavelet scales corresponding to central frequencies of 0.54 Hz, 1.1 Hz, 5.7 Hz, 11 Hz, and 120 Hz in one subject. In the first three bands, the first and the second canonical variates are shown in the same time window. All 48 trials are superimposed. Note different time scales for the canonical variates.

Fig. 3. Median SNR over the group in selected stages of the analysis chain (SSS-PCA-wavelet-SimCCA-ICA). Letters in the left corner represent group medians of individually maximal SNR (A) over 204 channels, (B) over ICA components after PCA, and (C) over PCA components, SSS signals as a starting point.
repeated 1-min-long test trials. With this test data, consistent brain responses were detected up to ~120 Hz in 6/10 subjects, with the most prominent signals residing below ~20 Hz. Statistically significant trial-to-trial correlations, especially at the upper end of the spectrum, demonstrate the efficiency of the method in poor SNR conditions. The detected frequency range far exceeded the frequencies typically observed in MEG/EEG studies with narrative speech stimuli (~<10 Hz; e.g. O’Sullivan et al., 2014; Gross et al., 2013). Second, with 1-h-long nonrecurring recordings, signal components with the largest trial-to-trial correlations (i.e. the 1st canonical variates) were shown to be stimulus-related up to ~30 Hz. Thus our trained filters were able to generalize over independent datasets and to reveal feasible brain responses to relatively long-lasting stream of nonrecurring speech stimuli. Third, the spatial filters were found to be sensitive to source currents in cortical regions previously known to participate in speech processing.

The proposed approach has several distinctions compared with the well-known spatial filtering schemes. First, compared with beamforming (van Veen et al., 1997), the analysis can be done solely in sensor-space signals avoiding errors in anatomical source models. This procedure may broaden the applicability of the method, for example, to EEG recordings. However, the mapping of the signal sources in post-hoc analysis can reveal clusters of mutually correlated signal sources that are important in e.g. functional connectivity analysis. Second, SimCCA has some similarities, for example, with ICA or with PCA. Although replicability requirement favors the proposed methodology, thereby preventing fair comparison with the other methods, we were somewhat surprised to find out that standard ICA or PCA did not noticeably improve SNR, and thus could not generalize over training and testing sets, or to reveal consistent trial-to-trial responses. This result, however, does not exclude the possibility that e.g. independent components estimated for a particular dataset could be stimulus-related. Possibly, much of the difficulty stems from the low SNR of the unaveraged single-trial MEG data. These methods, however, do not intrinsically utilize sample ordering (Kerhunen and Hao, 2011), and they are thus principally suboptimal in revealing consistent temporal structures between trials. As a solution, previous studies have pointed out advantages of combining ICA and CCA (Kerhunen et al., 2012; Kerhunen and Hao, 2011). In-depth analysis with ICA is out of the scope of this paper.

Our approach aimed at uncovering both instantaneous changes, and longer temporal patterns of brain signals in detail, in contrast to typical topographic mapping per se. In this line of research, previous studies indicate that sound onsets (Lakatos et al., 2005) or other characteristics in natural sound stimuli, for example in rock music (Szymanski et al., 2011), in environmental sounds, or in animal vocalizations (Kayser et al., 2009) often cause replicable transient brain responses or enforce abrupt phase resetting of oscillations (Lakatos et al., 2008). Moreover, the temporal structure of the stimulus, such as syllabic variations in the heard speech (4–8 Hz; e.g. Luo and Poeppel, 2007; Cogan and Poeppel, 2011), is reflected in the brain oscillation patterns. On this basis, we selected the Mexican-hat wavelet transform for its capability to react both to the oscillatory and to the aperiodic waveforms with minimal temporal blurring. The downside is that the effective passbands broaden, even to tens of Hz in the gamma range. Here, the redundancy between adjacent scales was reduced by exponential wavelet scaling.

Founding the model training on replicable brain responses has noteworthy physiological implications. The analysis is inherently steered towards revealing activity in brain areas where the responses remain relatively similar from trial to trial. Evidently, the extrinsic networks (Golland et al., 2007; Boldt et al., 2013) most directly influenced by the stimuli, are involved. In this respect, association between the spatially filtered signals and the speech stimulus was expected. Notably, stimulus–response relation was not explicitly included in models (in contrast, for example, to Ding and Simon, 2013; Koskinen et al., 2013; O’Sullivan et al., 2014; Zion Golumbic et al., 2013). Thus, the parameter search space was not constrained by an incomplete and partly unknown set of stimulus features that could lead to the exclusion of unknown contributing sources.

In prior studies, the most common speech–signal feature correlating with MEG or EEG has been signal power or amplitude envelope (Abrams et al., 2008; Ahissar et al., 2001; Nourski et al., 2009; Ding and Simon, 2013; Gross et al., 2013), as used here for validation. Interestingly, envelope was found to provide only partial explanation for the signal components, as suggested by statistically significant but weak stimulus–response correlations above 3 Hz. Specifying a more complete set of contributing speech features is out of the scope of this paper as the aim was merely to demonstrate the dependency. Considering the stimulus–response-relationship, it is useful to note that the proposed methodology may help tackling the response variability that inherently reduces the mutual dependency between the stimulus and the response, and thus, the performance of stimulus–response models (see e.g. Belitski et al., 2010; Cogan and Poeppel, 2011; Kayser et al., 2009; Schyns et al., 2011; Szymanski et al., 2011).

We make two remarks on the used experimental setup. First, in experiments where the same stimuli are repeated, both brain responses and the subject’s attentiveness tend to decrease with stimulus repetition. At the same time, changes occur in response time courses and in the brain areas that react to the stimulus. Second, with narratives, a
A relation exists between the extent of consistently responding brain regions and the lengths of temporal receptive windows (TRWs), i.e. time-spans that accumulate contextual information (Hari et al., 2010; Lerner et al., 2011). Generally, early auditory regions tend to process momentary features while high-level perceptual and cognitive areas more likely favor longer coherent structures in the story (e.g. 38 ± 17 s in Lerner et al., 2011). In our work, the SimCCA models were trained with repeated stimuli that was relatively long, 1-min. Our data provide

---

**Fig. 5.** Selected sensitivity maps of Subject #2. Sensitivity values starting from 0.5 are colored red, turning to orange and yellow at ~0.9. (A) This specific filter is most sensitive to sources at auditory and superior temporal regions. (B and C) Clearly lateralized, nonspecific activity across regions typically associated with speech processing. (D and E) These two spatial filters at the same wavelet scale pick mutually uncorrelated signals from motor/sensory regions.
some indications that the detected brain signals may reflect activity not only in early auditory regions, but also in more extensively temporal and posterior cortices, and in inferior frontal regions, in addition to motor/somatosensory areas. These results are concordant with previous electrophysiological findings with narrative speech (Gross et al., 2013; Zion Columbic et al., 2013).

In conclusion, the proposed spatial filtering strategy offers apparent methodological advantages. Spatial filter parameter estimation based on replicability constraint effectively reduces noise, thereby uncovering consistent time courses hidden in multivariate MEG datasets. In this framework, the method provides the maximum SNR given the modeling constraints. Importantly, the spatial filters were robust enough to generalize over independent datasets, revealing plausible single-trial time courses up to ~120 Hz range. The very topics of replicability of analysis results, noise resiliency of models, and the ability to reveal stimulus-related brain activity all play crucial roles in attempts to accurately and unambiguously link brain responses with stimulus features.

Acknowledgments

The authors thank Profs. Riitta Hari and Aapo Hyvärinen for valuable suggestions and comments, MSc. Mia Illman for assistance in MEG recordings. Prof. Matti Hämäläinen, Prof. Lauri Parkkonen, MSc. Kaisu Lankinen, and MSc. Jukka Saari for discussions. The study was supported by the Academy of Finland (post-doctoral project #134655) and aivoAALTO project of the Aalto University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2014.06.018.

References


