Neuro-Current Response Functions: A Unified Approach to MEG Source Analysis under the Continuous Stimuli Paradigm

Proloy Das[a,b,*], Christian Brodbeck[a,b], Jonathan Z. Simon[a,b,c], Behtash Babadi[a,b,*]

[a] *Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA*
[b] *Institute for Systems Research, University of Maryland, College Park, MD 20742, USA*
[c] *Department of Biology, University of Maryland, College Park, MD 20742, USA*

Abstract

Characterizing the neural dynamics underlying sensory processing is one of the central areas of investigation in systems and cognitive neuroscience. Neuroimaging techniques such as magnetoencephalography (MEG) and Electroencephalography (EEG) have provided significant insights into the neural processing of continuous stimuli, such as speech, thanks to their high temporal resolution. Existing work in the context of auditory processing suggests that certain features of speech, such as the acoustic envelope, can be used as reliable linear predictors of the neural response manifested in MEG/EEG. The corresponding linear filters are referred to as temporal response functions (TRFs). While the functional roles of specific components of the TRF are well-studied and linked to behavioral attributes such as attention, the cortical origins of the underlying neural processes are not as well understood. Existing methods for obtaining the cortical representation of such linear speech processing work in a two-stage fashion: either the TRFs are first estimated at the sensor level, and then mapped to the cortex via source localization, or the neuroimaging data are first mapped to the cortex followed by estimating TRFs for each of the resulting cortical sources. Given that each stage is biased towards specific requirements, such as sparsity and smoothness, the end result typically suffers from destructive propagation of biases across stages, which in turn hinders a valid statistical interpretation of the results and requires significant post-hoc processing to summarize the results in a meaningful fashion. In this work we address this issue by estimating a linear filter representation of cortical sources directly from neuroimaging data. To this end, we introduce Neuro-Current Response Functions (NCRFs), which are three-dimensionally oriented linear filters spatially distributed throughout the cortex, that predict the cortical current dipoles giving rise to the observed MEG (or EEG) data in response to speech. NCRF estimation is cast within a Bayesian framework, which allows unification of the TRF and source estimation problems, and also facilitates the incorporation of prior information on the structural properties of the NCRFs. We present a fast estimation algorithm, which we refer to as the Champ-Lasso algorithm, by leveraging recent advances in optimization. We demonstrate the utility of the Champ-Lasso algorithm through application to simulated and experimentally recorded MEG data under auditory experiments. Our simulation studies reveal significant improvements over existing work, in terms of both spatial resolution and

reliance on fine-tuned coordinate co-registration. Application to experimentally-recorded data corroborates existing results, while also delineating the distinct cortical distribution of the underlying neural processes at high spatiotemporal resolution and obviating the need for post-processing steps such as clustering and denoising. In summary, we provide a principled modeling and estimation paradigm for MEG source analysis tailored to extracting the cortical origin of electrophysiological responses to continuous stimuli.
Keywords: MEG, Temporal Response Functions, Speech Processing, Source Localization, Bayesian Estimation

1. Introduction

The human brain routinely processes complex information as it unfolds over time, for example, when processing natural speech, information from lower levels has to be continuously processed to build higher level representations, from the acoustic signal to phonemes to words to sentence meaning. Quantitative characterization of the neural dynamics underlying such sensory processing is not only important in understanding brain function, but it is also crucial in the design of neural prostheses and brain-machine interface technologies.

Existing models for capturing these neural processes generally fall into the two categories of encoding and decoding models. Encoding models aim at predicting the neural response from the sensory stimuli, whereas decoding models are used to reconstruct the sensory stimuli from the neural response (Marmarelis, 2004). Within the framework of encoding, several classes of models have been proposed, which fall into the general category of system identification methodology (Wu et al., 2006), where the goal is to fit parametrized models of the neural response by taking the sensory stimuli as input, using training examples.

At the micro-scale, when predicting the response of a single neuron, non-linear models such as generalized linear models have been shown to possess higher predictive power than their linear counterparts, and are well matched to the underlying neurophysiological principles (Paninski, 2004). In modeling neural activity at the meso-scale using neuroimaging modalities such as electroencephalography (EEG) and magnetoencephalography (MEG), however, experimental evidence suggests that linear encoding models can be beneficial in predicting the key features of sensory processing; examples include encoding models of visual and auditory stimuli (Boynton et al., 1996; Lalor et al., 2006, 2009; Ding and Simon, 2012b).

*This work was supported by the National Science Foundation (Award No: 1552946 and 1734892), the Defense Advanced Research Projects Agency (Award No: N6600118240224) and the National Institutes of Health (Award No: R01-DC014085). A preliminary version of this work was presented at the 2018 Asilomar Conference on Signals, Systems, and Computers (Das et al., 2018).
*Corresponding authors
Email addresses: proloy@umd.edu (Proloy Das), christianbrodbeck@me.com (Christian Brodbeck), jzsimon@umd.edu (Jonathan Z. Simon), behtash@umd.edu (Behtash Babadi)

Arguably the earliest and most widely used technique to construct neural encoding models is the 'reverse correlation' technique, in which neural responses time-locked to multiple repetitions of simple stimuli (such as acoustic tones and visual gratings) are averaged, weighted by the instantaneous value of the preceding stimulus, to form the so-called evoked response function. Originally devised to study the tuning properties of sensory neurons (Aertsen and Johannesma, 1983; Aertsen et al., 1981; Ringach and Shapley, 2004), it was later incorporated into MEG/EEG analysis. In probing the neural response to more sophisticated stimuli such as continuous speech and video, the goal is to understand the encoding of the continuous stimuli as a whole, which is composed of both low level (e.g., acoustics) and high level (e.g., semantics) features which are bound together and distributed across time.

To address this issue, techniques from linear systems theory have been successfully utilized to capture neural encoding using MEG/EEG under the continuous stimuli paradigm. In this setting, the encoding model takes the form of a linear filter which predicts the MEG/EEG response from the features of the stimulus. For example, it has been shown that the acoustic envelope of speech is a suitable predictor of the EEG response (Lalor and Foxe, 2010). These filters, or impulse response functions, play a crucial role in characterizing the temporal structure of auditory information processing in the brain, and are often referred to as Temporal Response Function (TRF) (Ding and Simon, 2012b, 2013a,b). For instance, in a competing-speaker environment in the presence of two speech streams, it has been observed that the TRF extracted from MEG response to the acoustic power consists of an early component at around 50 ms representing the acoustic power of the speech mixture, while a later peak at around 100 ms preferentially encodes the acoustic power of the attended speech stream (Ding and Simon, 2012a; Akram et al., 2016). Building up on evidence from ferret electrophysiology (Mesgarani et al., 2008) and human electrocorticography (ECoG) (Mesgarani et al., 2014), recent studies have expanded the TRF framework beyond the acoustic level to account for phoneme-level processing (Di Liberto et al., 2015), lexical processing (Brodbeck et al., 2018a) and semantic processing (Broderick et al., 2018).

Thanks to the grounding of the TRF model in linear system theory, several techniques from the system identification literature have been utilized for TRF estimation, such as the *normalized reverse correlation* (Theunissen et al., 2001), *ridge regression* (Machens et al., 2004), *boosting* (David et al., 2007), and *SPARLS* (Akram et al., 2017), some of which are available as software packages (Theunissen, 2007, 2010; Crosse et al., 2016; Brodbeck, 2017). While these methods have facilitated the characterization of the functional roles of various TRF components in sensory and cognitive processing of auditory stimuli, they predominantly aim at estimating TRFs over the MEG/EEG sensor space. While recent studies, using electrophysiology in animal models and ECoG in humans, have provided new insights into the cortical origins of auditory processing (See, for example, Mesgarani et al., 2008; Mesgarani and Chang, 2012; Pasley et al., 2012; Mesgarani et al., 2014), they do not account for the whole-brain distribution of the underlying sources due to their limited spatial range. As such, the whole-brain cortical origins of the TRF components are not well studied.

To address this issue using neuroimaging, current dipole fitting methods have been utilized to map the sensor space distribution of the estimated TRF components onto cortical sources (Lalor et al., 2009; Ding and Simon, 2012a). Given that the processing of sophisticated stimuli such as speech is known to be facilitated by a widely distributed cortical network, single dipole sources are unlikely to capture the underlying cortical dynamics. More recent results have used the minimum norm estimate (MNE) source localization technique to first map the MEG activity onto the cortical mantle, followed by estimating a TRF for each of the resulting cortical sources (Brodbeck et al., 2018b). While these methods have shed new light on the cortical origins of the TRF, they have several limitations that need to be addressed. First, the ill-posed nature of the MEG/EEG source localization problem under distributed source models results in cortical estimates with low spatial resolution (Schoffelen and Gross, 2009; Baillet et al., 2001). Given the recent and ongoing advances in MEG/EEG source localization towards improving the spatial resolution of the inverse solutions (See, for example, Gorodnitsky et al., 1995; Friston et al., 2008; Haufe et al., 2008; Wipf et al., 2010; Wipf and Nagarajan, 2009; Wu et al., 2016; Gramfort et al., 2013; Babadi et al., 2014; Krishnaswamy et al., 2017; Pirondini et al., 2018), it is tempting to simply use more advanced source localization techniques followed by fitting TRFs to the resulting cortical sources. However, these techniques leverage specific prior knowledge on the source configurations and are typically developed for the event-related potential (ERP) paradigm (Handy, 2005; Gazzaniga and Ivry, 2009; Luck, 2014). As such, they are biased towards providing *source* estimates with high resolution under specific repetition-based experimental settings, and do not account for the key structural properties of the underlying *neural processes* that extract information from continuous sensory stimuli. These key properties include the smoothness and/or sparsity of the response functions in the lag domain, and their spatial correlation over the cortex.

Second, high resolution inverse solutions require precise forward models that are constructed based on high resolution magnetic resonance imaging (MRI) scans with accurate sensor registration, which may not be readily available. In addition, the single-trial nature of experiments involving continuous auditory stimuli, does not allow to leverage the time-averaging across multiple trials common in source localization of evoked responses from MEG/EEG. Finally, the two-stage procedures of first fitting TRFs over the sensor space followed by localizing the peaks using dipole fitting, or first finding source estimates over the cortex followed by fitting TRFs to cortical sources, results in so-called bias propagation: the inherent biases arising from the estimation procedure in the first stage propagate to the second stage, often destructively so, and limit, sometimes severely, the statistical interpretability of the resulting cortical TRFs.

In order to address these limitations, in this paper we provide a methodology for *direct* estimation of cortical TRFs from MEG observations, taking into account their sparsity and spatiotemporal structure. We refer to these cortical TRFs as neuro-current response functions (NCRFs). We construct a unified estimation framework by integrating the linear encoding and distributed forward source models, in which the NCRFs linearly process different features of the continuous stimulus and result in the observed neural responses

at the sensor level. We cast the inverse problem of estimating the NCRFs as a Bayesian optimization problem where the likelihood of the recorded MEG response is directly maximized over the NCRFs, thus eliminating the need for the aforementioned two-stage procedures. In addition, to guard against over-fitting and ensure robust recovery of the NCRFs, we design a regularizer that captures the spatial sparsity and temporal smoothness of the NCRFs and mitigates the need for accurate cortical surface patch statistics in the head model. While the resulting optimization problem turns out to be non-convex, we provide an efficient coordinate-descent algorithm that leverages recent advances in evidence maximization to obtain the solution in a fast and efficient manner.

We evaluate the performance of the proposed NCRF estimation framework using both simulation studies and application to experimentally recorded MEG data. Our simulation studies reveal that the proposed method is capable of producing cortical TRF estimates with higher spatial resolution and less variability compared to existing methods. Application to experimentally recorded MEG data from young adult subject listening to speech provides new insights into the cortical dynamics of speech processing.

2. Theory and Methods

We develop our theory and methods for a canonical MEG auditory experiment in which the subject is listening to a single speech stream. Our goal is to determine how the different features of the speech stream are processed at different cortical stages and evoke specific neural responses that give rise to the recorded MEG data. For clarity of description and algorithm development, we first consider a single-trial experiment, and take the momentary acoustic power of the speech stream, i.e., the speech envelope, as the feature of interest. We will discuss below the more general scenarios including multiple trials, multiple speech stimuli, and multiple, and possibly competing, features reflecting different levels of cognitive processing.

2.1. Preliminaries and Notation

Let $e_t,\ 1 \leq t \leq T$ denote the speech envelope at discrete time index t for a duration of T samples taken at a sampling frequency of f_s. We consider a distributed cortical source model composed of M dipole sources $\mathbf{d}_m = (\mathbf{r}_m, \mathbf{j}_{m,t}),\ 1 \leq m \leq M$, where $\mathbf{r}_m \in \mathbb{R}^3$ denotes the Euclidean coordinates of the mth dipole and $\mathbf{j}_{m,t} := [j_{m,t,R}, j_{m,t,A}, j_{m,t,S}]^\top \in \mathbb{R}^3$ denotes the dipole current vector at time t in the right-anterior-superior (RAS) coordinate system. The dipole locations can be obtained by standard tessellation of the 3D structural MR images of the cortex and assigning dipoles to the corresponding vertices (Baillet et al., 2001; Gramfort et al., 2014).

We assume that these current dipoles are in part stimulus-driven, i.e.,

$$j_{m,t,i} = f_i(e_t, e_{t-1}, \cdots, e_1) + v_{m,t,i} \tag{1}$$

where the placeholder i takes the values of one of the coordinate axes, $\{R, A, S\}$, f_i is a generic function, and $\mathbf{v}_{m,t} := [v_{m,t,R}, v_{m,t,A}, v_{m,t,S}]^\top$ accounts for the stimulus-independent background activity. Following the common modeling approaches in this context (Ringach and Shapley, 2004; Lalor et al., 2006; Lalor and Foxe, 2010; Lalor et al., 2009), we take f_i to represent a linear finite impulse response (FIR) filter of length L:

$$f_i(e_t, e_{t-1}, \cdots, e_1) = \sum_{l=0}^{L-1} \tau_{m,i,l} e_{t-l} = (\boldsymbol{\tau}_{m,i})^\top \mathbf{e}_t, \quad i \in \{R, A, S\}, \tag{2}$$

where $\boldsymbol{\tau}_{m,i} := [\tau_{m,i,0}, \tau_{m,i,1} \cdots, \tau_{m,i,L-1}]^\top$ and $\mathbf{e}_t := [e_t, e_{t-1}, \cdots, e_{t-L+1}]^\top$. Note that $\boldsymbol{\tau}_{m,i}$ can be thought of as a TRF corresponding to the activity of dipole source m along the coordinate axis determined by i. The length of the filter L is typically determined by a priori assumptions on the effective integration window of the underlying neural process. The 3D linear filters $\boldsymbol{\tau}_m := [\boldsymbol{\tau}_{m,R}, \boldsymbol{\tau}_{m,A}, \boldsymbol{\tau}_{m,S}] \in \mathbb{R}^{L\times 3}$ are vector-valued TRFs at each source m, capturing the linear processing of the stimuli at the cortical level. As such, we refer to these vector-valued filters as Neuro-Current Response Functions (NCRFs) henceforth.

Let $\mathbf{j}_t := [\mathbf{j}_{1,t}^\top, \mathbf{j}_{2,t}^\top, \cdots, \mathbf{j}_{M,t}^\top]^\top \in \mathbb{R}^{3M}$ be a vector containing all the current dipoles at time t, and $\mathbf{J} := [\mathbf{j}_1, \cdots, \mathbf{j}_T] \in \mathbb{R}^{3M\times T}$ be the matrix of current dipoles obtained by concatenating the instantaneous current dipoles across time $t = 1, 2, \cdots, T$. Similarly, let $\boldsymbol{V} \in \mathbb{R}^{3M\times T}$ denote the matrix of stimulus-independent background activity, $\boldsymbol{\Phi} := [\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \cdots, \boldsymbol{\tau}_M]^\top \in \mathbb{R}^{3M\times L}$ denote the matrix of NCRFs, and $\mathbf{S} := [\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_T] \in \mathbb{R}^{L\times T}$ denote the matrix of features. Eqs. (1) and (2) can then be compactly expressed as:

$$\mathbf{J} = \boldsymbol{\Phi}\mathbf{S} + \mathbf{V}. \tag{3}$$

As for the sensor space, we assume a conventional MEG setting with N sensors placed at different positions over the scalp, recording magnetic fields/gradients as a multidimensional time series. The MEG observation at the ith sensor at time t is denoted by $y_{i,t}$, $1 \le i \le N$ and $t \in [1, \cdots, T]$. Let $\mathbf{Y} \in \mathbb{R}^{N\times T}$ be the MEG measurement matrix with the $(i,t)^{\text{th}}$ element given by $y_{i,t}$. The MEG measurement matrix is related to the matrix of current dipoles $\mathbf{J}$ according to the following forward model (Sarvas, 1987; Mosher et al., 1999; Baillet et al., 2001):

$$\mathbf{Y} = \mathbf{L}\mathbf{J} + \mathbf{W}, \tag{4}$$

where $\mathbf{L} \in \mathbb{R}^{N\times dM}$ maps the *source space* activity to the *sensor space* and is referred to as the *lead-field matrix*, and $\mathbf{W} \in \mathbb{R}^{N\times T}$ is the matrix of additive measurement noise. The lead-field matrix can be estimated based on structural MRI scans by solving Maxwell's equations under the quasi-static approximation (Hämäläinen et al., 1993).

2.2. Problem Formulation

Given the stimulus-driven and current-driven forward models of Eqs. (3) and (4), our main goal is to estimate the matrix $\mathbf{\Phi}$, i.e., the NCRFs. To this end, we take a Bayesian approach, which demands distributional assumptions on the various uncertainties involved, i.e., the stimulus-independent background activity and the measurement noise. For the measurement noise, we adopt the common temporally uncorrelated multivariate Gaussian assumption, i.e.,

$$p\left(\mathbf{Y}|\mathbf{J}\right) = \left|(2\pi)\mathbf{\Sigma}_w\right|^{-T/2} \exp\left(-\frac{1}{2}\|\mathbf{Y}-\mathbf{L}\mathbf{J}\|^2_{\mathbf{\Sigma}_w^{-1}}\right), \tag{5}$$

where $\|\mathbf{A}\|^2_{\mathbf{B}} := \mathsf{tr}\left\{\mathbf{A}^\top\mathbf{B}\mathbf{A}\right\}$ and $\mathbf{\Sigma}_w \in \mathcal{S}_+^N$ denotes the unknown noise covariance matrix. The covariance matrix $\mathbf{\Sigma}_w$ can be estimated from either empty-room or pre-stimulus recordings (Engemann and Gramfort, 2015). Next, let $\mathbf{V}_m \in \mathbb{R}^{3\times T}$ denote the matrix of background activity at source m, for $m = 1, 2, \cdots, M$. We adopt the following distribution for the background activity $\mathbf{V}$:

$$p(\mathbf{V}|\mathbf{\Gamma}) = \left(\prod_{m=1}^{M} \left|(2\pi)\mathbf{\Gamma}_m\right|^{-T/2}\right) \exp\left(-\frac{1}{2}\sum_{m=1}^{M}\|\mathbf{V}_m\|^2_{\mathbf{\Gamma}_m^{-1}}\right), \tag{6}$$

i.e., the portion of the current dipoles reflecting the background activity are modeled as zero-mean independent Gaussian random vectors with unknown 3D covariance matrix $\mathbf{\Gamma}_m \in \mathcal{S}_+^3$. Under this assumption, Eq. (3) can be expressed as:

$$p\left(\mathbf{J}|\mathbf{\Phi},\mathbf{\Gamma}\right) = \left|(2\pi)\mathbf{\Gamma}\right|^{-T/2} \exp\left(-\frac{1}{2}\|\mathbf{J}-\mathbf{\Phi}\mathbf{S}\|^2_{\mathbf{\Gamma}^{-1}}\right), \tag{7}$$

where $\mathbf{\Gamma} \in \mathcal{S}_+^{3M}$ is a block-diagonal covariance matrix with its mth diagonal block given by $\mathbf{\Gamma}_m$, for $m = 1, 2, \cdots, M$.

Under these assumptions, the joint distribution of the MEG measurement and current dipole matrices is given by:

$$p\left(\mathbf{Y},\mathbf{J}|\mathbf{\Phi},\mathbf{\Gamma}\right) = \left|(2\pi)\mathbf{\Sigma}_w\right|^{-T/2}\left|(2\pi)\mathbf{\Gamma}\right|^{-T/2} \exp\left(-\frac{1}{2}\|\mathbf{Y}-\mathbf{L}\mathbf{J}\|^2_{\mathbf{\Sigma}_w^{-1}} - \frac{1}{2}\|\mathbf{J}-\mathbf{\Phi}\mathbf{S}\|^2_{\mathbf{\Gamma}^{-1}}\right), \tag{8}$$

By marginalizing over $\mathbf{J}$, we obtain the distribution of the MEG measurement matrix parametrized by the NCRF matrix $\mathbf{\Phi}$ and the source covariance matrix $\mathbf{\Gamma}$:

$$p\left(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Gamma}\right) = \left|(2\pi)\left(\mathbf{\Sigma}_w + \mathbf{L}\mathbf{\Gamma}\mathbf{L}^\top\right)\right|^{-T/2} \exp\left(-\frac{1}{2}\|\mathbf{Y}-\mathbf{L}\mathbf{\Phi}\mathbf{S}\|^2_{(\mathbf{\Sigma}_w+\mathbf{L}\mathbf{\Gamma}\mathbf{L}^\top)^{-1}}\right). \tag{9}$$

It is now possible to cast the problem of finding $\mathbf{\Phi}$ as a Bayesian estimation problem, in which a loss function fully determined by the posterior distribution of NCRF matrix $\mathbf{\Phi}$ given the MEG measurement matrix $\mathbf{Y}$ is minimized. In other words, if $\mathbf{\Gamma}$ were known, the NCRF matrix estimation would amount to the following *maximum likelihood* problem:

$$\min_{\mathbf{\Phi}} \qquad \frac{1}{2}\|\mathbf{Y}-\mathbf{L}\mathbf{\Phi}\mathbf{S}\|^2_{(\mathbf{\Sigma}_w+\mathbf{L}\mathbf{\Gamma}\mathbf{L}^\top)^{-1}} \ . \tag{10}$$

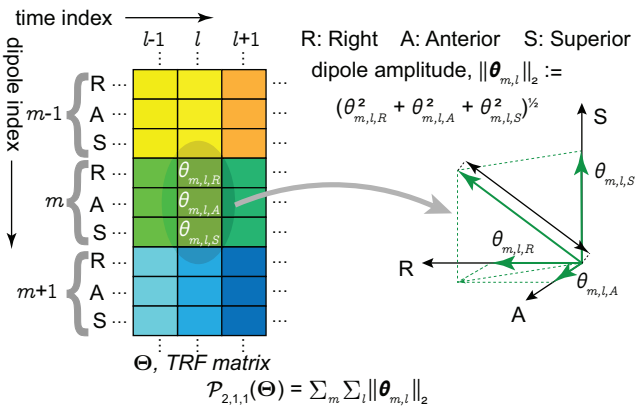

Figure 1: Mixed-norm penalty term $\mathcal{P}_{2,1,1}(\boldsymbol{\Theta})$ for regularizing the loss function. The penalty term is constructed by first isolating all 3D Gabor coefficient vectors across the dictionary elements and space, and then aggregating their ℓ_2 norm. As a result, it promotes sparsity in space and Gabor coefficients, while being invariant to the orientation of the dipole currents.

Another advantage of this Bayesian framework is the possibility of introducing regularization schemes that can mitigate the ill-posed nature of this problem, and instead work with regularized maximum likelihood problems. Note that this optimization problem makes a direct connection between the MEG measurement matrix,$\mathbf{Y}$ and the NCRF matrix $\boldsymbol{\Phi}$ and allows us to avoid the aforementioned two-stage procedures in finding TRFs at the cortical level (Lalor et al., 2009; Brodbeck et al., 2018b).

2.2.1. Regularization

As is the case in other source imaging methods, there are many fewer constraints than the free parameters determining the NCRFs. This makes the problem severely ill-posed. As such, proceeding with the maximum likelihood problem in Eq. (10) is likely to result in overfitting. In order to ensure robust recovery of a meaningful solution to this ill-posed problem, we need to include prior knowledge on the structure of the NCRFs in the form of regularization.

To this end, we construct regularizers based on a convex norm of the NCRF matrix $\boldsymbol{\Phi}$, to both capture the structural properties of the NCRFs and facilitate algorithm development. The structural properties of interest in this case are spatial sparsity over the cortical source space, sparsity of the peaks/troughs, smoothness in the lag domain, and rotational invariance (Ding and Simon, 2012b; Akram et al., 2017).

In order to promote smoothness in the lag domain and sparsity of the peaks/troughs, we adopt a concept from Chen et al. (2001), in which a temporally smooth time series is approximated by a small number of Gabor atoms over an over-complete dictionary $\mathbf{G} \in \mathbb{R}^{L \times \tilde{L}}$, for some $\tilde{L} \geq L$ (Feichtinger and Strohmer, 2012; Akram et al., 2017). To this end, we first perform a change of variables $\boldsymbol{\tau}_m := \mathbf{G}\boldsymbol{\theta}_m$, $\boldsymbol{\Phi} = \boldsymbol{\Theta}\mathbf{G}^\top$, and $\widetilde{\mathbf{S}} := \mathbf{G}^\top \mathbf{S}$, where $\boldsymbol{\theta}_m \in \mathbb{R}^{\tilde{L} \times 3}$ are the coefficients of the mth NCRF over the dictionary $\mathbf{G}$ and $\boldsymbol{\Theta} \in \mathbb{R}^{3M \times \tilde{L}}$ is a matrix containing $\boldsymbol{\theta}_m$s across its rows. Then, to enforce sparsity of the peaks/troughs, spatial sparsity,

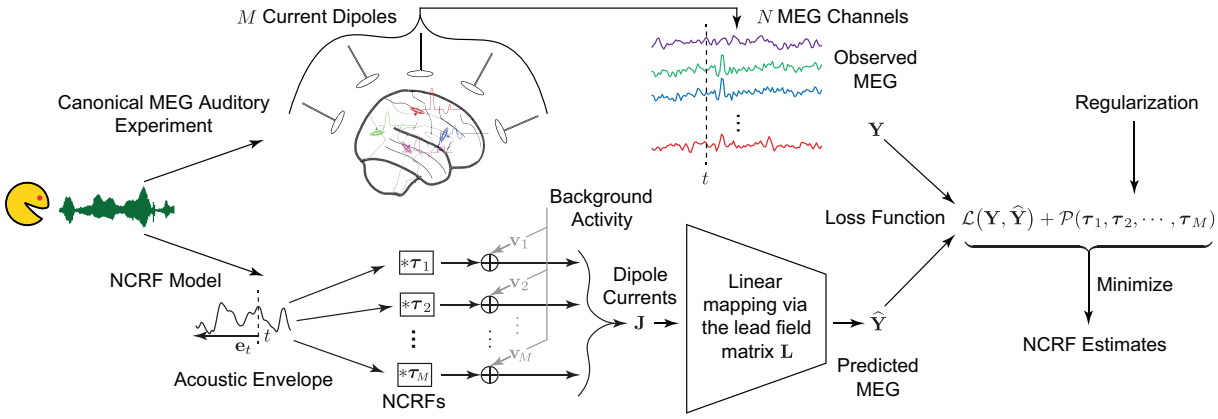

Figure 2: Schematic depiction of the proposed modeling and estimation framework. Upper branch: the experimental setting in which the subject is listening to speech while MEG neural responses are being recorded. Lower branch: the modeling framework in which the speech waveform is transformed into a feature variable representation, and is thereby processed via M linear filters (i.e., NCRFs) to generate time-varying current dipoles at each of the corresponding M source locations. These dipoles give rise to the predicted MEG response via a source-to-sensor mapping (i.e., the lead-field matrix). The two branches converge on the right hand side, where the NCRFs are estimated by minimizing a regularized loss function.

and rotational invariance, we use the following mixed-norm penalty over $\boldsymbol{\theta}_m$s, i.e., the Gabor coefficients:

$$\mathcal{P}_{2,1,1}(\boldsymbol{\Theta}) := \sum_{m=1}^{M} \sum_{l=1}^{\tilde{L}} \sqrt{\theta_{m,l,R}^2 + \theta_{m,l,A}^2 + \theta_{m,l,S}^2}. \tag{11}$$

Let $\boldsymbol{\theta}_{m,l} \in \mathbb{R}^3$ be the lth Gabor coefficient vector for the mth NCRF. Note that the summand is $\|\boldsymbol{\theta}_{m,l}\|_2$, which is a rotational invariant norm with respect to the choice of dipole RAS coordinate system. This structural feature allows the estimates to be robust to coordinate rotations (See Appendix A). The inner summation of $\|\boldsymbol{\theta}_{m,l}\|_2$ (as opposed to $\|\boldsymbol{\theta}_{m,l}\|_2^2$) over $l = 1, 2, \cdots, \tilde{L}$ enforces group sparsity of the Gabor coefficients (i.e., the number of peaks/troughs), akin to the effect of ℓ_1-norm. Finally, the outer summation over $m = 1, 2, \cdots, M$ promotes spatial sparsity of the NCRFs (See Fig. 1, and also Appendix A).

Using this change of variables and regularization scheme, we can reformulate (10) as the following regularized maximum likelihood problem:

$$\min_{\boldsymbol{\Theta}} \ \frac{1}{2} \|\mathbf{Y} - \mathbf{L}\boldsymbol{\Theta}\widetilde{\mathbf{S}}\|^2_{(\boldsymbol{\Sigma}_w + \mathbf{L}\boldsymbol{\Gamma}\mathbf{L}^\top)^{-1}} + \eta \mathcal{P}_{2,1,1}(\boldsymbol{\Theta}). \tag{12}$$

The parameter $\eta > 0$ controls the trade-off between data fidelity and regularization, i.e., the complexity of the resulting model grows inversely with the magnitude of η. This parameter can be chosen in a data-driven fashion using cross-validation (See Section 3.2). Fig. 2 provides a visual illustration of the proposed modeling and estimation paradigm.

2.2.2. Source Covariance Matrix Adaptation

Note that the objective function in Eq. (12) is convex in $\boldsymbol{\Theta}$ and thus one can proceed to solve for $\boldsymbol{\Theta}$ by standard convex optimization techniques. However, this requires the knowledge of the source covariance

matrix $\mathbf{\Gamma}$, which is unknown in general. From Eq. (7), it is evident that $\mathbf{\Gamma}$ implicitly offers adaptive penalization over the source space through spatial filtering. As such, the source covariance matrix serves as a surrogate for depth compensation (Lin et al., 2006), by reducing the penalization level at locations with low SNR. One data-independent approach for estimating $\mathbf{\Gamma}$ is based on the lead-field matrix (Haufe et al., 2008). Here, thanks to the Bayesian formulation of our problem, we take a data-driven approach to adapt the source covariance matrix to the observations. One principled way to do so is to estimate both $\mathbf{\Theta}$ and $\mathbf{\Gamma}$ from the observed MEG data by solving the following optimization problem:

$$\min_{\mathbf{\Theta},\mathbf{\Gamma}} \quad \frac{T}{2}\log\left|\mathbf{\Sigma}_w + \mathbf{L}\mathbf{\Gamma}\mathbf{L}^\top\right| + \frac{1}{2}\|\mathbf{Y} - \mathbf{L}\mathbf{\Theta}\widetilde{\mathbf{S}}\|^2_{(\mathbf{\Sigma}_w + \mathbf{L}\mathbf{\Gamma}\mathbf{L}^\top)^{-1}} + \eta\mathcal{P}_{2,1,1}(\mathbf{\Theta}) \tag{13}$$

Unfortunately, the loss function in Eq. (13) is not convex in $\mathbf{\Gamma}$. However, given an estimate of $\mathbf{\Theta}$, solving for the minimizer of (13) in $\mathbf{\Gamma}$ is a well-known problem in Bayesian estimation and is referred to as evidence maximization or empirical Bayes (Berger, 1985). Although a general solution to this problem is not straightforward to obtain, there exist several Expectation-Maximization (EM)-type algorithms, such as ReML (Friston et al., 2008), sMAP-EM (Lamus et al., 2012), and the conjugate function-based algorithm called Champagne (Wipf et al., 2010), which might be employed to estimate $\mathbf{\Gamma}$ given an estimate of $\mathbf{\Theta}$. In the next section, we present an efficient recursive coordinate descent-type algorithm that leverages recent advances in evidence maximization and proximal gradient methods to solve the problem of Eq. (13).

2.3. Inverse Solution: The Champ-Lasso *Algorithm*

Since simultaneous minimization of (13) with respect to both $\mathbf{\Theta}$ and $\mathbf{\Gamma}$ is not straightforward, we instead aim to optimize the objective function by alternatingly updating $\mathbf{\Theta}$ and $\mathbf{\Gamma}$, keeping one fixed at a time. Suppose after the rth iteration, the updated variable pair is given by $\left(\mathbf{\Theta}^{(r)}, \mathbf{\Gamma}^{(r)}\right)$, then the update rules for $(r+1)^{\mathsf{th}}$ iteration are as given as follows:

Updating $\mathbf{\Gamma}$

With $\mathbf{\Theta} = \mathbf{\Theta}^{(r)}$, Eq. (13) reduces to the following optimization problem:

$$\min_{\mathbf{\Gamma}} \quad \mathsf{tr}\left(\mathbf{\Sigma}_v^{-1}\mathbf{C}_v\right) + \log|\mathbf{\Sigma}_v|, \tag{14}$$

with $\mathbf{C}_v = (\mathbf{Y} - \mathbf{L}\mathbf{\Theta}^{(r)}\widetilde{\mathbf{S}})(\mathbf{Y} - \mathbf{L}\mathbf{\Theta}^{(r)}\widetilde{\mathbf{S}})^\top/T$ and $\mathbf{\Sigma}_v = \mathbf{\Sigma}_w + \mathbf{L}\mathbf{\Gamma}\mathbf{L}^\top$. Although the problem is non-convex in $\mathbf{\Gamma}$, it can be solved via the Champagne algorithm (Wipf et al., 2010), which solves for $\mathbf{\Gamma}$ by updating a set of auxiliary variables iteratively. Though the solution $\mathbf{\Gamma}^{(r+1)}$ is not guaranteed to be a global minimum, the convergence rate is fast (with computation cost per iteration being linear in N), and more importantly each iteration is guaranteed not to increase the loss function in Eq. (14).

Updating $\mathbf{\Theta}$

Fixing $\mathbf{\Gamma} = \mathbf{\Gamma}^{(r+1)}$, results in the following convex optimization problem over $\mathbf{\Theta}$:

$$\min_{\mathbf{\Theta}} \quad \frac{1}{2}\|\mathbf{L}\mathbf{\Theta}\widehat{\mathbf{S}} - \mathbf{Y}\|^2_{\mathbf{\Sigma}_v^{(r+1)^{-1}}} + \eta\mathcal{P}_{2,1,1}(\mathbf{\Theta}), \tag{15}$$

where $\mathbf{\Sigma}_v^{(r+1)} = \mathbf{\Sigma}_w + \mathbf{L}\mathbf{\Gamma}^{(r+1)}\mathbf{L}^\top$. The first term in Eq. (15) is a smooth differentiable function whose gradient is straightforward to compute, and the proximal operator for the penalty term $\mathcal{P}_{2,1,1}(\mathbf{\Theta})$ has a closed-form expression and can be computed in an efficient manner (Gramfort et al., 2012). Regularized optimization problems of this nature can be efficiently solved using an instance of the forward-backward splitting (FBS) method (Beck and Teboulle, 2009; Nesterov, 2005). We use an efficient implementation of FBS similar to FASTA (Fast Adaptive Shrinkage/Thresholding Algorithm) software package (Goldstein et al., 2014) to obtain $\mathbf{\Theta}^{(r+1)}$ from Eq. (15).

Although the loss function is not jointly-convex in $(\mathbf{\Theta}, \mathbf{\Gamma})$, the foregoing update steps ensure that the loss in (13) is not increased at any iteration and stops changing when a fixed-point or limit-cycle is reached (Wright, 2015). Finally, note that due to the efficiency of the overall solver, it is possible to restart the optimization with a randomized initialization, and choose the best solution among several potential alternatives.

2.4. Extension to Multiple Feature Variables

The preceding sections focused on the case of a single stimulus feature variable, i.e., the speech envelope. However, complex auditory stimuli such as natural speech, are processed at various levels of hierarchy. Upon entering the ears, the auditory signal is decomposed into an approximate spectrogram representation at the cochlear level prior to moving further into the auditory pathway (Yang et al., 1992). Beyond these low-level acoustic features, higher-level phonemic, lexical, and semantic features of the natural speech are also processed in the brain. Thus, to obtain a complete picture of complex auditory cortical processing, it is desirable to consider response functions corresponding to more than one feature variable.

One can proceed to estimate response functions for each feature variable separately. But, since many of these features have significant temporal correlations, the resulting response functions do not readily provide unique information regarding the different levels of the processing hierarchy. To investigate simultaneous processing of these various feature variables and allow them to compete in providing independently informative encoding models, we consider a multivariate extension of the response functions (Ding and Simon, 2012b; Di Liberto et al., 2015).

Suppose that there are $F \geq 1$ feature variables of interest. We modify Eq. (3) by replacing each column of the NCRF matrix $\mathbf{\Phi}$ by F columns (one for each temporal response function) and each row of the stimulus matrix by F rows (one for each feature variable). As we will demonstrate below in Section 3, this will enable

Algorithm 1 The Champ-Lasso Algorithm over Multiple Trials

Input: MEG observations $\mathbf{Y}^k$, modified stimuli matrix $\widehat{\mathbf{S}}^k$, for $k = 1, 2, \cdots, K$; Lead-field matrix $\mathbf{L}$; Regularization parameter η; initial values of $\mathbf{\Theta}^0$; Tolerance parameter $\mathsf{tol} \in (0, 10^{-3})$, Maximum number of outer iterations $R_{\max} \in \mathbb{N}^+$.

$r = 0$.

repeat

  for $k = 1, \cdots, K$ **do**

    $\mathbf{C}_v^{k(r)} = \frac{1}{T}(\mathbf{Y}^k - \mathbf{L}\mathbf{\Theta}^{(r)}\widehat{\mathbf{S}}^k)(\mathbf{Y}^k - \mathbf{L}\mathbf{\Theta}^{(r)}\widehat{\mathbf{S}}^k)^\top$

    $\mathbf{\Gamma}^{k(r+1)} = \arg\min_{\mathbf{\Gamma}} \ \mathsf{tr}\left(\mathbf{\Sigma}_v^{-1}\mathbf{C}_v^{k(r)}\right) + \log|\mathbf{\Sigma}_v| \quad \text{s.t. } \mathbf{\Sigma}_v = \mathbf{\Sigma}_w + \mathbf{L}^\top\mathbf{\Gamma}\mathbf{L}$ (Champagne iterations)

    $\mathbf{\Sigma}_v^{k(r+1)} = \left(\mathbf{\Sigma}_w + \mathbf{L}\mathbf{\Gamma}^{k(r+1)}\mathbf{L}^\top\right)^{-1}$

  end for

  $\mathbf{\Theta}^{(r+1)} = \arg\min_{\mathbf{\Theta}} \ \sum_{k=1}^{K} \frac{1}{2}\|\mathbf{L}\mathbf{\Theta}\widehat{\mathbf{S}}^k - \mathbf{Y}^k\|^2_{\mathbf{\Sigma}_v^{k(r+1)-1}} + \eta\mathcal{P}_{2,1,1}(\mathbf{\Theta})$ (FASTA iterations)

until $\frac{\|\mathbf{\Theta}^{(r+1)} - \mathbf{\Theta}^{(r)}\|_2}{\|\mathbf{\Theta}^{(r)}\|_2} < \mathsf{tol}$ or $r = R_{\max}$.

Set $r \leftarrow r + 1$.

Output: $\mathbf{\Theta}^{(R)}$ where R is the index of the last outer iteration of the algorithm.

us to distinguish between different cortical regions in terms of their response latency across a hierarchy of features.

2.5. Extension to Multiple Trials with Different Stimuli

Next, we consider extension to K different trials corresponding to possibly different auditory stimuli. Let the stimuli, MEG observation, and background activity covariance matrices for the kth trial be denoted by $\widetilde{\mathbf{S}}^k$, $\mathbf{Y}^k$, and $\mathbf{\Gamma}^k$, respectively, for $k = 1, \cdots, K$. We can extend the optimization problem of Eq. (13) as follows:

$$\min_{\mathbf{\Theta}, \mathbf{\Gamma}} \quad \sum_{k=1}^{K} \left[\frac{T}{2} \log\left|\mathbf{\Sigma}_w + \mathbf{L}\mathbf{\Gamma}^k\mathbf{L}^\top\right| + \frac{1}{2}\|\mathbf{Y}^k - \mathbf{L}\mathbf{\Theta}\widehat{\mathbf{S}}^k\|^2_{(\mathbf{\Sigma}_w + \mathbf{L}\mathbf{\Gamma}^k\mathbf{L}^\top)^{-1}} \right] + \eta\mathcal{P}_{2,1,1}(\mathbf{\Theta}). \tag{16}$$

In doing so, we have assumed that the NCRFs remain unchanged across trials, which promotes integration of complementary information from different trials (without direct averaging). Note that this assumption intentionally suppresses the trial-to-trial variability of the NCRFs by adaptively weighting the contribution of each trial according to its noise level (i.e., $\mathbf{\Gamma}^k$), in favor of recovering NCRFs that can explain common cortical patterns of auditory processing. In contrast, if all the trials were to be concatenated or directly averaged to form a unified trial (with a single covariance matrix $\mathbf{\Gamma}$), the trial-to-trial variability would not necessarily be suppressed, especially when there are few trials available. Furthermore, this formulation allows to incorporate trials with different lengths into the same framework.

The optimization problem of Eq. (16) can be solved via a slightly modified version of that presented in Section 2.3. The resulting algorithm is summarized in Algorithm 1, which we refer to as the Champ-Lasso algorithm. A python implementation of the Champ-Lasso algorithm is archived on the open source repository Github (Das, 2019) to ease reproducibility and facilitate usage by the broader systems neuroscience community.

2.6. Subjects, Stimuli, and Procedures

The data used in this work are a subset of recordings presented in (Presacco et al., 2016), and is publicly available in the Digital Repository at the University of Maryland (Presacco et al., 2018). The auditory experiments were conducted under the participation of 17 young adult subjects (aged 18-27 years), recruited from the Maryland, Washington D.C. and Virginia area. The participants listened to narrated segments from the audio-book, *The Legend of Sleepy Hollow by Washington Irving* (https://librivox.org/the-legend-of-sleepy-hollow-by-washington-irving/), while undergoing MEG recording. Although the dataset contains recordings under different background noise levels, for the current work we considered recordings of two 1 min long segments of the audio-book with no background noise presented as single-speaker audio. Each of these segments was repeated three times to every individual, yielding a total 6 min of data per subject. To ensure that the participants actively engage in the listening task, they were tasked to also silently count the number of specific words that they would hear in the story.

2.7. Recording and Preprocessing

The data were acquired using a whole head MEG system (KIT, Nonoichi, Ishikawa, Japan) consisting of 157 axial gradiometers, at the University of Maryland Neuroimaging Center, with online low-pass filtering (200 Hz) and notch filtering (60 Hz) at a sampling rate of 1 kHz. Data were pre-processed with MNE-python 0.18.1 (Gramfort et al., 2013, 2014). After excluding flat and noisy channels, temporal signal space separation was applied to remove extraneous artifacts (Taulu and Simola, 2006). Data were then filtered between 1 Hz to 80 Hz using a zero-phase FIR filter. Independent component analysis (extended infomax, Bell and Sejnowski, 1995) was applied to remove ocular and cardiac artifacts. Finally, 60 s long data epochs corresponding to the stimuli were extracted and downsampled to 200 Hz.

2.8. Source Space Construction

At the beginning of each recording session, each participant's head shape was digitized with a 3-Space Fastrak system (Polhemus), including 3 fiducial points and 5 marker positions. Five marker coils attached to these five marker positions were used to localize the head position of the participant relative to the MEG sensors. The head position measurement was recorded twice: at the beginning and end of the recording session and the average measured head positions were used. Since MR scans of the participants were not

performed, the 'fsaverage' brain model (Fischl, 2012) was co-registered (via uniform scaling, translation and rotation) to each participant's head, using the digitized head shapes.

A volumetric source space for the 'fsaverage' brain was defined on a regular grid with spacing of 7 mm between two neighboring points, and then morphed to individual participants. These morphed source spaces were then used to compute lead-field matrices by placing 3 orthogonal virtual current dipoles on each of the grid points. The computed lead-field matrices contained contribution from 3222 virtual current dipoles, after removing those within subcortical structures along the midline. No cortical patch statistics were available due to the lack of MR scans, so the current dipoles were allowed to have arbitrary orientations in 3D.

2.9. Stimulus Feature Variables

We included predictor variables reflecting three different hierarchical levels of speech processing, including acoustic, lexical, and semantic features. These feature variables are described in detail in (Brodbeck et al., 2018b):

- *Envelope:* The speech envelope was found by averaging the auditory spectrogram representation generated using a model of the auditory periphery (Yang et al., 1992) across frequency bands. This continuous univariate feature variable reflects the momentary acoustic power of the speech signal.
- *Word Frequency:* First, logarithmic word frequency measures, $\log_{10} \mathsf{wf}$, were extracted from the SUBTLEX database (Brysbaert and New, 2009) for each word. Then, a piecewise-continuous feature variable was constructed by representing each word in the speech segment by a rectangular pulse with height given by $6.33 - \log_{10} \mathsf{wf}$. Note that in this coding scheme, infrequent words are assigned higher values, while common words get lower values. Windows of silence were assigned 0.
- *Semantic Composition:* Lastly, to probe semantic processing, the semantic composition patterns identified by (Westerlund et al., 2015), including adjective-noun, adverb-verb, adverb-adjective, verb-noun, preposition-noun and determiner-noun pairs, were used. To generate the feature variable, the second word in each pair was represented by a rectangular window of height 1, and 0 elsewhere. This binary-valued feature variable identifies the semantic binding of word pairs within the speech stream.

All three variables were constructed from the speech segments at the same sampling frequency as the preprocessed MEG data (i.e. 200 Hz). All feature variables were centered and scaled by their mean absolute value, to facilitate comparison of NCRF components pertaining to different feature variables.

3. Results

3.1. Simulation Studies

Before applying the Champ-Lasso algorithm to localize NCRFs from experimentally recorded data, we assessed its performance using realistic simulation studies with known ground truth. In accordance with our

experimental settings, we synthesized six 1 min long MEG data segments according to the forward model of Eqs. (3) and (4), mimicking neural processing of the speech envelope. To this end, we simulated temporal response functions of length 500 ms (with significant M50 and M100 components) associated with dipole current sources within the auditory and motor cortices. Fig. 3–A shows the simulated activity over the cortical surface at specific time lags (top and bottom panels) as well as the temporal profile of the NCRFs

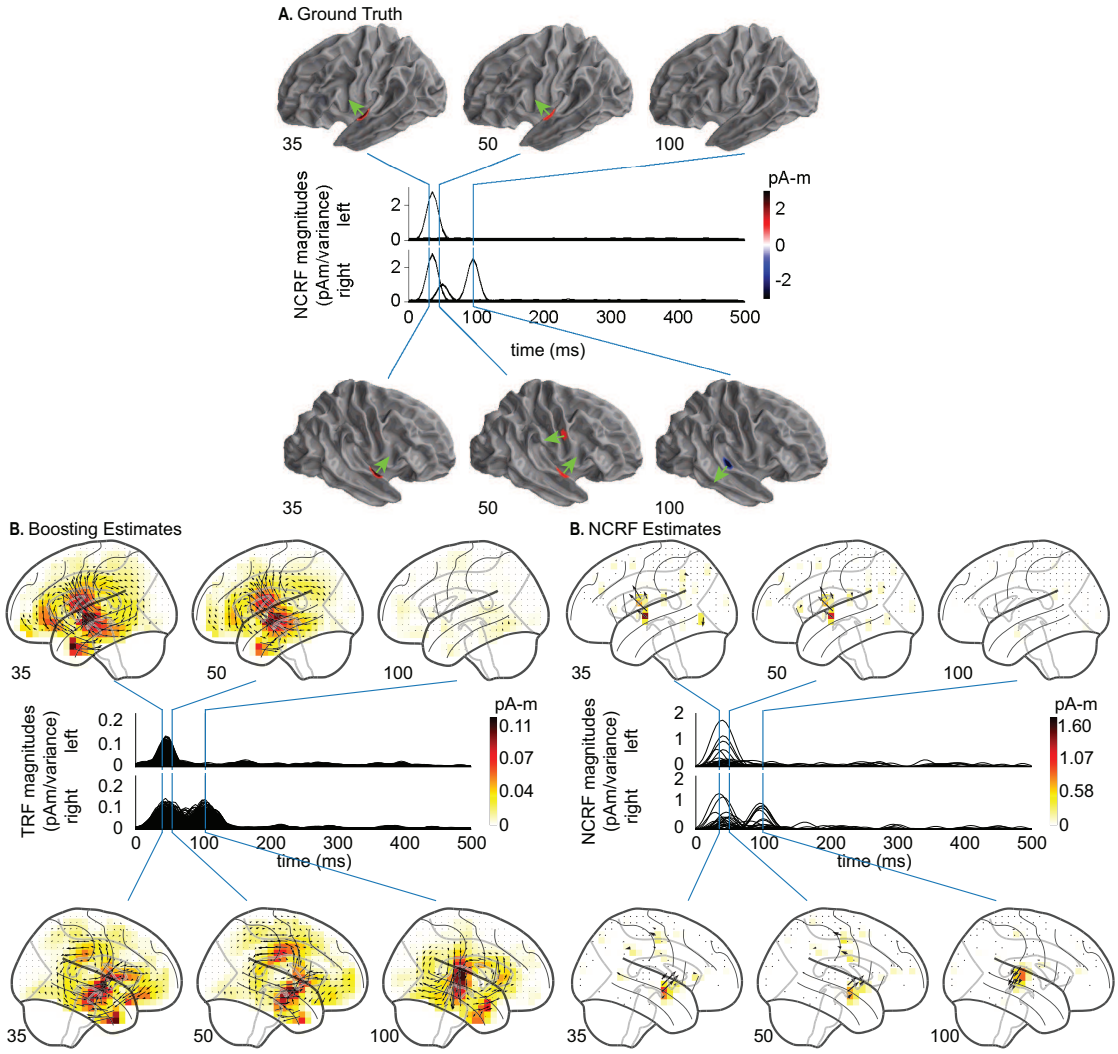

Figure 3: Results for a simulated auditory experiment. The top and bottom portions of each subplot pertain to the left and right hemispheres, respectively. A. The anatomical plots show the simulated neural sources normal to cortical surface, and the traces show the overlaid temporal profiles. The colorbar encodes directional intensity normal to the cortical surface. B. The two-stage localized TRFs (free-orientation) are shown on the anatomical plots, where the 3D dipoles are projected onto the lateral plane. The colorbar encodes dipole magnitudes. C. NCRF estimates from the Champ-Lasso algorithm. The spatial extent, dipole moment scale, temporal profile, and orientations of the neural sources are faithfully recovered by the Champ-Lasso algorithm, whereas the two-stage localized TRFs are spatially dispersed and exhibit spurious peaks in the anterior temporal and inferior frontal lobes.

(middle panel). The cortical activity was simulated using patches defined over a finely-discretized source space (namely, ico-5, with average dipole spacing of 3.1 mm) and constrained the dipole directions normal to the 'fsaverage' surface patches. To make the simulation as realistic as possible, we used real MEG recordings corresponding to a different speech stream as stimulus independent background activity.

In order to avoid any favorable bias in the inverse solution, we used a different source space for NCRF estimation, i.e., the aforementioned volumetric source space with unconstrained dipole orientations (Subsection 2.8), than the one used for simulating the data, i.e., ico-5. As a comparison benchmark, we also applied the two-stage method of Brodbeck et al. (2018b) to first localize the cortical sources via MNE, followed by independently estimating TRFs for all sources.

The two-stage localized TRFs and estimated NCRFs are shown in Fig. 3–B and 3–C, respectively. Since boosting tends to result in temporally sparse response functions (David et al., 2007), response functions were smoothed with a 50 ms Hamming window. The anatomical plots show the spatial response function profile at the same temporal peaks selected in Fig. 3–A, with direction of the vectors projected onto the lateral plane. The Champ-Lasso algorithm successfully recovers both the smooth temporal profile of the NCRFs and the spatial extent and location of the active sources, and provides estimates that closely resemble the ground truth. The two-stage localized TRFs, however, fail to recover the true extent of the sources and exhibit spurious peaks in the anterior temporal and inferior frontal lobes. It is worth noting that despite the fact that the Champ-Lasso algorithm is unaware of the true dipole orientations, the resulting NCRF orientations closely match the normal directions of the patches in Fig. 3–A (shown using the green arrows). The Champ-Lasso also successfully suppresses spurious peaks in the anterior temporal and inferior frontal lobes, demonstrating its robustness to over-fitting.

3.2. Application to Experimentally Recorded MEG Data

As mentioned in Section 2.9, we consider the three feature variables of acoustic envelope, word frequency, and semantic composition and aim at estimating 1000 ms-long NCRFs. The regularization parameter is tuned on the basis of generalization error via a 3-fold cross-validation procedure. To check whether inclusion of each feature variable improves the overall NCRF model significantly, the original model fit (i.e., the cost function evaluated at the estimated NCRF parameters) is compared against the average of three other model fits constructed by deliberately misaligning one feature variable via 4-fold cyclic permutations, using one-tailed t-tests.

To evaluate the group-level significance of the estimated NCRF components, the NCRF estimates are first smoothed with a Gaussian kernel (with standard deviation of 10 mm) over the source locations to compensate for possible head misalignments and anatomical differences across subjects. Then, at each dipole location and time index, the magnitudes of the vector-valued NCRFs are tested for significance using a permutation test via the threshold-free cluster-enhancement (TFCE) algorithm (Smith and Nichols, 2009)

(See Appendix B).

3.2.1. Analysis of the Acoustic Envelope NCRFs

Fig. 4–A depicts the group average of estimated NCRFs, corresponding to the acoustic envelope, masked by a significance level of $p = 0.05$. The time traces show the magnitude of the average NCRFs (gray segments are statistically insignificant) and the anatomical plots show the spatial NCRF profile at selected temporal peaks, with direction of the vectors projected onto the lateral plane. Fig. 4–B shows the temporal profiles (masked at a significance level of $p = 0.05$) of 6 selected NCRFs exhibiting peak spatial activity (collapsed across time). The colored dots on the anatomical plots show the locations of these NCRFs, with matching colors to those of the traces. The traces are grouped by hemisphere and dorsoventrally ordered. The left and right NCRFs in the motor areas are referred to as $l\mathsf{M}_{\mathsf{env}}$ and $r\mathsf{M}_{\mathsf{env}}$, respectively. The left and right auditory NCRF pairs are labeled as $l\mathsf{A1}_{\mathsf{env}}$, $l\mathsf{A2}_{\mathsf{env}}$ and $r\mathsf{A1}_{\mathsf{env}}$, $r\mathsf{A2}_{\mathsf{env}}$, respectively.

The NCRFs corresponding to the acoustic envelope in Fig. 4–A exhibit two prominent temporal peaks: an early peak at around $30 - 35$ ms, bilaterally centered over the auditory cortex (AC), and a later peak

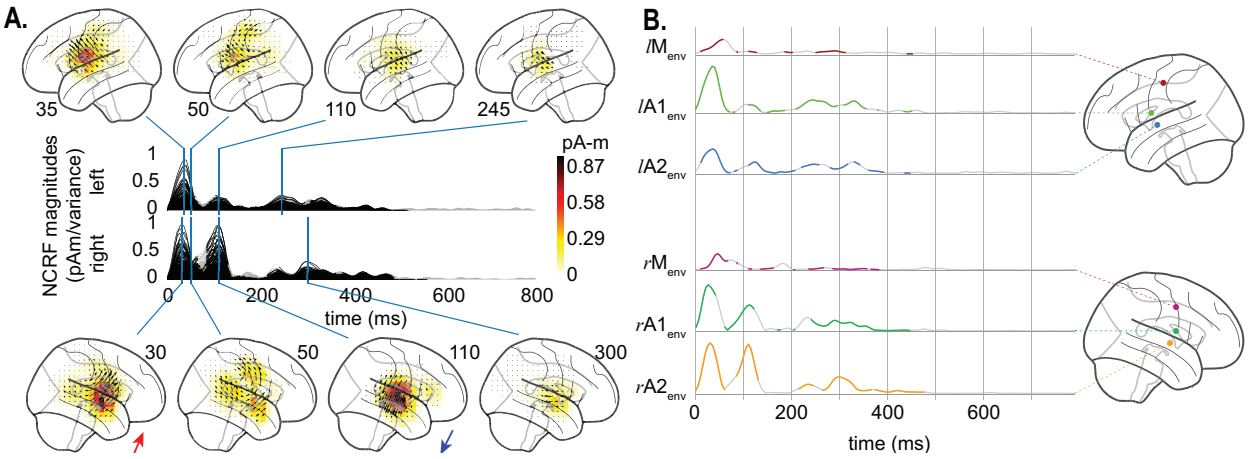

Figure 4: Estimated NCRFs for the acoustic envelope feature variable. A. The anatomical plots show the group-level average NCRFs projected onto the lateral plane (top and bottom panels) corresponding to selected visually salient peaks in the temporal profiles (middle panels). The top and bottom portions of the subplot pertain to left and right hemisphere, respectively. Numerical labels of each anatomical subplot indicates the corresponding time lag in ms. B. The time traces show the temporal profile of 6 selected NCRFs exhibiting peak spatial activity (collapsed across time), grouped by hemisphere and dorsoventrally ordered. The locations of the selected NCRFs are shown on the anatomical plots, with colors matching the time traces and linked by dashed lines. The gray portions of the traces in both subplots indicate statistically insignificant NCRFs at the group level (significance level of 5%). Note that the last 200 ms segments of the temporal profiles are cropped, as they did not correspond to any significant components at the group level. The prominent NCRFs consist of a bilateral auditory component at $\sim 30-35$ ms, a bilateral motor component at ~ 50 ms, and an auditory component at ~ 110 ms (stronger in right hemisphere and with nearly opposing polarity with respect to the earlier auditory component, indicated by the two colored arrows pointing to the average direction of the NCRFs in subplot A). See Supplementary Movie 01 for a detailed animation showing how the acoustic envelope NCRF components change as function of time lags.

at around 100 ms, dorsal to the first peak and stronger in the right hemisphere. The latter is evident from comparing the left temporal profiles $l\mathsf{A1}_{\mathsf{env}}$ and $l\mathsf{A2}_{\mathsf{env}}$, with their right hemisphere counterparts $r\mathsf{A1}_{\mathsf{env}}$ and $r\mathsf{A2}_{\mathsf{env}}$. Note that the orientations of the NCRFs at the second peak (blue arrow, bottom panel of Fig. 4–A) are nearly the opposite of those at the first peak (red arrow, bottom panel of Fig. 4–A), which accounts for the negative polarity of the M100 peak with respect to M50 in standard TRF analysis. Furthermore, after the appearance of the first peak ($\sim$ 35 ms, auditory traces $l\mathsf{A1}_{\mathsf{env}}$, $l\mathsf{A2}_{\mathsf{env}}$, $r\mathsf{A1}_{\mathsf{env}}$, and $r\mathsf{A2}_{\mathsf{env}}$) at the AC, the activity appears to gradually shift towards the primary motor cortex (PMC) in both hemispheres ($\sim$ 50 ms, motor traces $l\mathsf{M}_{\mathsf{env}}$ and $r\mathsf{M}_{\mathsf{env}}$). Additionally, the NCRFs show small bilateral late auditory components at around $\sim$ 250–350 ms.

3.2.2. Analysis of the Word Frequency NCRFs

Fig. 5–A shows the NCRFs for the word frequency feature variable, in the same format as in Fig. 4. Fig. 5–B shows the temporal profiles (masked at a significance level of $p = 0.05$) of 4 selected NCRFs exhibiting peak spatial activity (collapsed across time) also in the same format as in Fig. 4. These include a left auditory ($l\mathsf{A}_{\mathsf{wf}}$), a left frontal ($l\mathsf{F}_{\mathsf{wf}}$), a left inferior temporal (IT) ($l\mathsf{IT}_{\mathsf{wf}}$), and a right auditory ($r\mathsf{A}_{\mathsf{wf}}$)

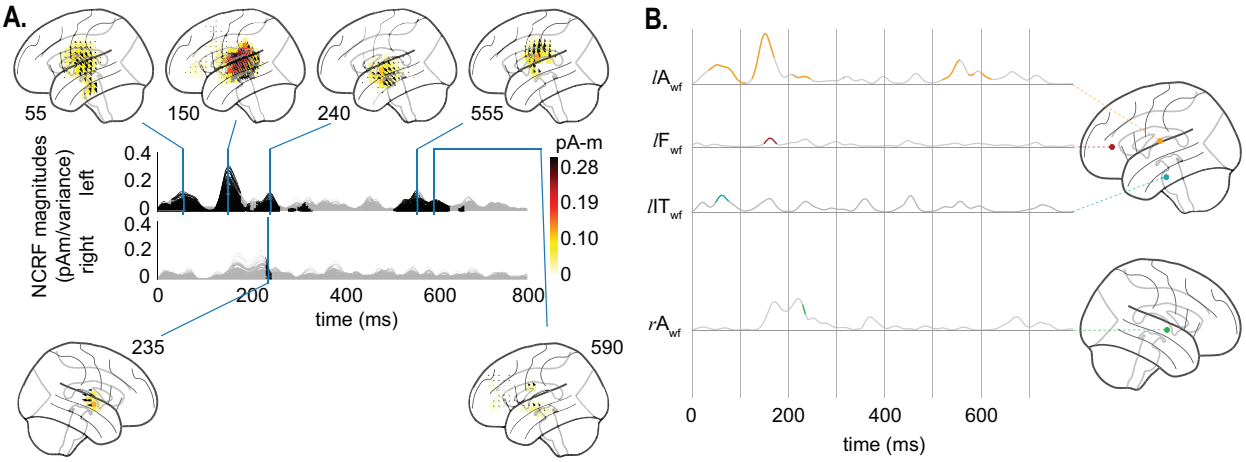

Figure 5: Estimated NCRFs for the word frequency feature variable. A. The anatomical plots show the group-level average NCRFs projected onto the lateral plane (top and bottom panels) corresponding to selected visually salient peaks in the temporal profiles (middle panels). The top and bottom portions of the subplot pertain to left and right hemisphere, respectively. Numerical labels of each anatomical subplot indicates the corresponding time lag in ms. B. The time traces show the temporal profile of 4 selected NCRFs exhibiting peak spatial activity (collapsed across time), grouped by hemisphere and and dorsoventrally ordered. The locations of the selected NCRFs are shown on the anatomical plots, with colors matching the time traces and linked by dashed lines. The anatomical plots show the locations of the selected NCRFs with colors matching the time traces. The gray portions of the traces in both subplots indicate statistically insignificant NCRFs at the group level (significance level of 5%). The prominent NCRFs manifest in the left hemisphere, dominated by an auditory component at $\sim$ 150 ms. See Supplementary Movie 02 for a detailed animation showing how the word frequency NCRF components change as function of time lags.

NCRF. The significant NCRF components manifest predominantly in the left hemisphere (Fig. 5–A). The earliest peak in the left AC occurs at around 50 ms, followed by a much stronger peak at around 150 ms, slightly posterior to the former (See $l\mathsf{A}_{\mathsf{wf}}$ in Fig. 5–B). The earlier peak also has contributions from the inferior temporal gyrus, as indicated by $l\mathsf{IT}_{\mathsf{wf}}$. In addition, the left frontal cortex exhibits weak activity at around 150 ms (See $l\mathsf{F}_{\mathsf{wf}}$ in Fig. 5–B). A weak but localized peak centered over the left superior temporal sulcus (STS) is visible at around 240 ms. The only significant component in the right hemisphere occurs around the same time. Finally, the late NCRF components (at around 500–600 ms) mostly originate from the left AC and STS, with weak contributions from the right frontal cortex.

3.2.3. Analysis of the Semantic Composition NCRFs

The estimated NCRFs corresponding to the semantic composition feature variable are shown in Fig. 6–A, along with 5 representative NCRFs in Fig. 6–B. These include two left auditory ($l\mathsf{A1}_{\mathsf{sc}}$ and $l\mathsf{A2}_{\mathsf{sc}}$), two right frontal ($r\mathsf{F1}_{\mathsf{sc}}$ and $r\mathsf{F2}_{\mathsf{sc}}$), and one right middle temporal ($r\mathsf{MT}_{\mathsf{sc}}$) NCRF. The main NCRF components in the left AC peak at around 155 ms and 475 ms, with the earlier peak being ventral to the later one (See $l\mathsf{A1}_{\mathsf{sc}}$ and $l\mathsf{A2}_{\mathsf{sc}}$ in Fig. 6–B). The significant right hemispheric NCRFs are temporally concentrated between

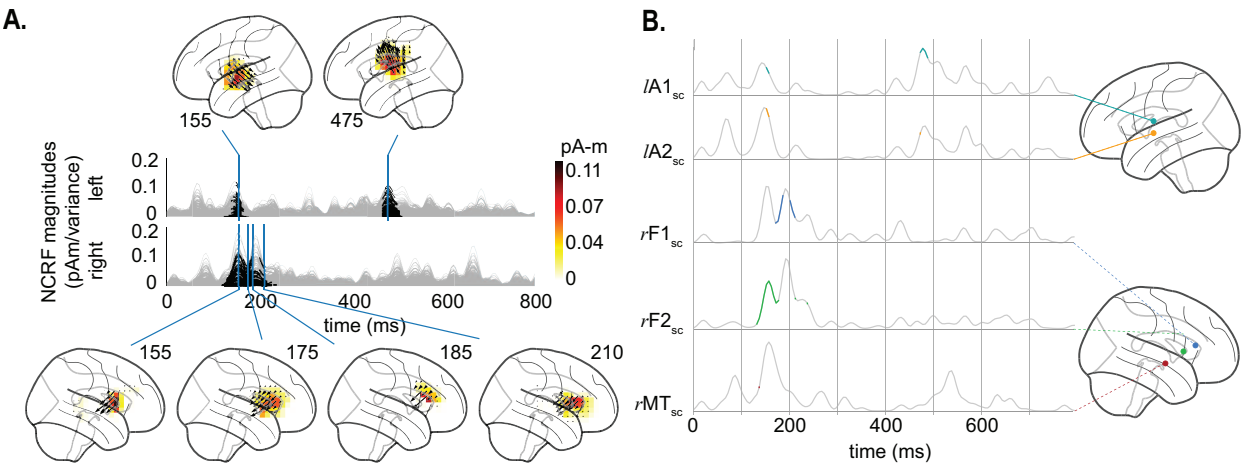

Figure 6: Estimated NCRFs for the semantic composition feature variable. A. The anatomical plots show the group-level average NCRFs projected onto the lateral plane (top and bottom panels) corresponding to selected visually salient peaks in the temporal profiles (middle panels). The top and bottom portions of the subplot pertain to left and right hemisphere, respectively. Numerical labels of each anatomical subplot indicates the corresponding time lag in ms. B. The time traces show the temporal profile of 5 selected NCRFs exhibiting peak spatial activity (collapsed across time), grouped by hemisphere and dorsoventrally ordered. The locations of the selected NCRFs are shown on the anatomical plots, with colors matching the time traces and linked by dashed lines. The gray portions of the traces in both subplots indicate statistically insignificant NCRFs at the group level (significance level of 5%). The prominent NCRF components consist of a bilateral auditory component at $\sim$ 155 ms, a right auditory-frontal component from $\sim$ 180 ms to $\sim$ 210 ms, and a left auditory late component at $\sim$ 475 ms. The sequence of peaks given by $r\mathsf{MT}_{\mathsf{sc}} \rightarrow r\mathsf{F2}_{\mathsf{sc}} \rightarrow r\mathsf{F1}_{\mathsf{sc}} \rightarrow r\mathsf{F2}_{\mathsf{sc}}$ show the back and forth movement of the NCRFs from the auditory to frontal cortices. See Supplementary Movie 03 for a detailed animation showing how the semantic composition NCRF components change as function of time lags.

155 to 210 ms, and appear superior to those in the left hemisphere, involving inferior frontal gyrus (IFG). Strikingly, these NCRFs in the right hemisphere seem to move in the anterosuperior direction until around 185 ms, at which point the right hemisphere exhibits strong frontal activity (Fig. 6–A). The NCRFs return to their initial location afterwards at around 210 ms. This sequence of spatiotemporal changes is also evident in the sequence of temporal peaks in Fig. 6–B, given by $r\mathsf{MT}_{\mathsf{sc}} \rightarrow r\mathsf{F2}_{\mathsf{sc}} \rightarrow r\mathsf{F1}_{\mathsf{sc}} \rightarrow r\mathsf{F2}_{\mathsf{sc}}$.

4. Discussion and Concluding Remarks

Characterizing the dynamics of cortical activity from noninvasive neuroimaging data allows us to probe the underlying mechanisms of sensory processing at high spatiotemporal resolutions. In this work, we demonstrated a framework for direct estimation of such cortical dynamics in response to various features of continuous auditory stimuli from the MEG response. To this end, we developed a fast inverse solution under a Bayesian estimation setting, the Champ-Lasso algorithm, for inferring the Neuro-Current Response Functions (as spatiotemporal models of cortical processing) in a robust and scalable fashion.

One of the key features of the Champ-Lasso algorithm is the ability to simultaneously estimate cortical source covariances in a data-driven fashion (as opposed to relying on data-agnostic depth-weighting procedures) and finding the NCRF model parameters. The interplay between the two as well as incorporating the structural properties of the NCRFs into the model, taking advantage of the Bayesian nature of the estimation framework, ultimately leads to spatially focal NCRFs, with smooth temporal profiles. In other words, the NCRF and source covariance estimation procedures work in tandem to best explain the observed MEG data while minimizing the spatial leakage and capturing the smoothness of the temporal responses. In contrast, previously existing methodologies result in estimates that are spatially broad, which then require post-hoc clustering procedures to meaningfully summarize the underlying spatiotemporal cortical dynamics. These serialized procedures in turn introduce biases to the estimates, and hinder meaningful statistical interpretation of the results.

To demonstrate the utility of our proposed framework, we estimated NCRFs corresponding to several feature variables of speech, reflecting different levels of cognitive processing and comprehension from MEG. The data analyzed in this paper were analyzed by an earlier method in (Brodbeck et al., 2018b), where a two-stage procedure was utilized to probe the cortical processing of speech: the MEG data was first cortically localized using an MNE inverse solver, followed by estimating individual temporal response functions for each source. In order to summarize the resulting estimates in a meaningful fashion, yet another processing step was necessary to disentangle the different spatially dispersed and highly overlapping cortical sources. Our results corroborate those obtained in (Brodbeck et al., 2018b), while obviating the need for any such post-processing, by providing a one-step estimation procedure with the substantial benefit of greatly improved spatial resolution. In addition, the three-dimensional nature of the NCRFs in our framework allows the

segregation of different spatial activation patterns that are temporally overlapping. For example, the bilateral activity components in the primary motor cortex in response to the acoustic envelope are automatically clearly distinguishable from the early activation in the auditory cortex, without the need for any post-hoc processing.

Our results also support other neuroimaging evidence for the hierarchical model of speech processing, involving not only the temporal lobe, but also the motor and frontal cortices (Scott and Johnsrude, 2003; Hickok and Poeppel, 2004; Davis and Johnsrude, 2007; Okada et al., 2010; Peelle et al., 2010; de Heer et al., 2017). To probe the functional organization of this hierarchy, we estimated NCRFs corresponding to features extracted from speech at the acoustic, lexical and semantic levels and found distinct patterns of cortical processing at high spatiotemporal resolutions. Our results indeed imply that while the acoustic and lexical features are processed primarily within the temporal and motor cortical regions (Fadiga et al., 2002; Wilson et al., 2004; Pulvermüller et al., 2006; Crinion et al., 2003; Dewitt and Rauschecker, 2012; Mesgarani et al., 2014; Hullett et al., 2016), phrase-level processing, assessed here using the semantic composition variable, is carried out through the involvement of the frontal cortex (Kaas and Hackett, 1999; Hickok and Poeppel, 2007; Rauschecker and Scott, 2009).

Another advantage of our proposed methodology is mitigating the dependence of the solution on the precise geometry of the underlying cortical source models. In conventional neuromagnetic source imaging, individual structural MR images are utilized in the construction of source space models, particularly for retrieving the cortical surface segmentation. The normal direction to the so-called cortical patches in these models is key in determining the lead-field matrix, which are often referred to as orientation-constrained source models. However, in many available neuroimaging datasets (including the one analyzed in this work), MR images are not available, relying only on an average head model, instead of one informed by the subject-specific cortical geometry. In order to mitigate the need for such information, we utilized a free-orientation volumetric source space in our estimation framework. While this makes the underlying optimization problem more involved and computationally intensive, it adds more than a compensatory amount of flexibility to the underlying models and allows them to recover missing information regarding the cortical source space geometry. To this end, we used rotationally invariant sparsity-inducing priors to regularize the spatiotemporal distribution of the NCRFs. Together with the aforementioned data-driven source covariance adaptation, this regularization scheme results in consistent source orientation estimates and provides a degree of immunity to unwanted side-effects of error-prone coordinate-frame rotations. To confirm these theoretical expectations, we validated this feature of our framework using simulation studies with known ground truth. In light of the above, posing NCRF estimation over an orientation-free volumetric source space can also be thought of as unifying the virtues of distributed source imaging and single dipole fitting: we aim at estimating both the orientations and magnitudes of spatially sparse dipole currents within the head volume that can best linearly predict the MEG responses to continuous stimuli.

A python implementation of the Champ-Lasso algorithm is archived on the open source repository Github (Das, 2019) to ease reproducibility and facilitate usage by the broader systems neuroscience community. The current implementation of our algorithm uses the aforementioned regularization scheme to recover temporally smooth and spatially sparse NCRFs. Due to the plug-and-play nature of the proposed Bayesian estimation framework, one can easily utilize other relevant regularization schemes to promote spatial smoothness or incorporate spectro-temporal prior information, by just modifying the penalty term.

Appendix A. Details of the Regularization Scheme

In this appendix, we provide more details on the regularization scheme used for NCRF estimation. Recall that the NCRF matrix estimation amounts to the following maximum likelihood problem:

$$\min_{\boldsymbol{\Theta}} \ \frac{1}{2}\|\mathbf{Y} - \mathbf{L}\boldsymbol{\Theta}\widehat{\mathbf{S}}\|^2_{(\boldsymbol{\Sigma}_w + \mathbf{L}\boldsymbol{\Gamma}\mathbf{L}^\top)^{-1}} \ . \tag{A.1}$$

given a particular choice of $\boldsymbol{\Gamma}$. With this choice, one can find the gradient of the objective as:

$$\mathbf{L}^\top(\boldsymbol{\Sigma}_w + \mathbf{L}\boldsymbol{\Gamma}\mathbf{L}^\top)^{-1}\left(\mathbf{Y} - \mathbf{L}\boldsymbol{\Theta}\widehat{\mathbf{S}}\right)\widehat{\mathbf{S}}^\top \tag{A.2}$$

and thus can attempt to solve the maximum likelihood problem using gradient descent techniques. The following observations on the gradient, however, show that the problem is ill-conditioned:

1. The left multiplier of $\boldsymbol{\Theta}$, i.e., $\mathbf{L}^\top(\boldsymbol{\Sigma}_w + \mathbf{L}\boldsymbol{\Gamma}\mathbf{L}^\top)^{-1}\mathbf{L}$ is singular.
2. The right multiplier of $\boldsymbol{\Theta}$, i.e, $\widehat{\mathbf{S}}\widehat{\mathbf{S}}^\top$, which is the empirical stimulus correlation matrix is likely to be rank-deficient for naturalistic stimuli (Crosse et al., 2016).

Therefore, a direct attempt at solving the problem via the gradient descent results in estimates of $\boldsymbol{\Theta}$ with high variability. In estimation theory, such ill-conditioning is handled by introducing a bias to the estimator, which contains a priori information about the problem, in order to reduce the estimation variance. In addition, the NCRF model typically has many more free parameters than the observed data points, and without introducing prior information, the estimation problem is prone to over-fitting.

The prior information is often incorporated in the form of regularization. A commonly used regularization scheme in this context is the Tikhonov regularization and its variants for promoting smoothness (Lalor et al., 2006). Other estimation schemes such as boosting and ℓ_1-regularization promote sparse solutions (David et al., 2007; Akram et al., 2017). Here, we introduce a structured regularization by penalizing a specific mixed-norm of the NCRF matrix to recover spatio-temporally sparse solutions over the Gabor coefficients:

$$\mathcal{P}_{2,1,1}(\boldsymbol{\Theta}) = \sum_{m=1}^{M}\sum_{l=1}^{L}\|\boldsymbol{\theta}_{m,l}\|_2 = \sum_{m=1}^{M}\sum_{l=1}^{L}\sqrt{\theta^2_{m,l,R} + \theta^2_{m,l,A} + \theta^2_{m,l,S}}. \tag{A.3}$$

In words, for each current dipole location, we penalize the vector-valued response function by sum of the magnitude of its corresponding Gabor coefficients.

Note that the ℓ_1-regularization in this case, i.e., $\sum_{m=1}^{M}\sum_{l=1}^{L}\|\boldsymbol{\theta}_{m,l}\|_1$, is not compatible with the expected cortical distribution of the NCRFs. Since the ℓ_1-norm is separable with respect to the three 3 coordinates of $\boldsymbol{\theta}_{m,l}$, it tends to select a sparse subset of the 3D coordinates, rendering the recovered NCRF components parallel to the coordinate axes. In contrast, the proposed penalty aims to select the NCRF components as a single entity by penalizing the vector magnitudes at each lag. Indeed, if the current dipoles are constrained to be normal to the cortical patches in the NCRF formulation, the proposed penalty coincides with ℓ_1-regularization.

Another advantage of this mixed-norm penalty is its rotational invariance when working with 3D vector-valued response functions. Suppose the coordinate system is rotated by an orthogonal matrix $\boldsymbol{U} \in \mathbb{R}^{3\times 3}$. Then, the lead-field and NCRF matrices are transformed by: $\mathbf{L} \to \mathbf{L}\widetilde{\boldsymbol{U}}^\top =: \mathbf{L}'$, $\boldsymbol{\Theta} \to \widetilde{\boldsymbol{U}}\boldsymbol{\Theta} =: \boldsymbol{\Theta}'$ where $\widetilde{\boldsymbol{U}} = \boldsymbol{I}_M \otimes \boldsymbol{U}$. Then, $\mathbf{L}'\boldsymbol{\Theta}' = \mathbf{L}\widetilde{\boldsymbol{U}}^\top\widetilde{\boldsymbol{U}}\boldsymbol{\Theta} = \mathbf{L}\boldsymbol{\Theta}$ and

$$\mathcal{P}_{2,1,1}(\boldsymbol{\Theta}') = \eta\sum_{m=1}^{M}\sum_{l=1}^{L}\|\boldsymbol{\theta}'_{m,l}\|_2 = \eta\sum_{m=1}^{M}\sum_{l=1}^{L}\|\boldsymbol{U}\boldsymbol{\theta}_{m,l}\|_2 = \eta\sum_{m=1}^{M}\sum_{l=1}^{L}\|\boldsymbol{\theta}_{m,l}\|_2 = \mathcal{P}_{2,1,1}(\boldsymbol{\Theta}),$$

which implies the aforementioned rotational invariance. As a result, the solutions are not dependent on any particular choice of coordinate system. Also, since the penalty does not prefer specific source orientations, it makes the solution more resilient to co-registration error than other approaches that do not consider the vector-valued nature of the current dipoles or constrain the solutions to be normal to the cortical surface.

Appendix B. Statistical Testing Procedures

To asses the statistical significance of the estimated NCRF components at the group level and across the source space, they need to be compared against suitable null hypotheses. The fact that the NCRF components are 3D vectors requires technical care in choosing the null hypotheses. Here, we provide two possible null hypotheses and testing methodologies: the Length Test, that only considers the length or magnitude of the NCRFs, and the Vector Test that takes into account both the magnitude and direction of the NCRFs. The corresponding source codes that implement these tests can be found at (Brodbeck, 2017).

The Length Test

This test aims to assess the statistical significance of the NCRF components by comparing their magnitudes against a baseline 'null' NCRF model at the group level. To control for false positives arising from over-fitting, instead of using an all-zero null model of the NCRFs, we aim to learn the null model from the dataset itself. The time-series of the feature variables are split into four equal segments, and these segments are permuted cyclically to yield three 'misaligned' feature time-series. Then, for each feature variable, three 'misaligned' time-series are constructed by swapping its original time-series with the 'misaligned' ones, while keeping the other two feature variables intact. Then, the average NCRF magnitudes estimated from these

three 'misaligned' time-series are considered as the null model for that feature variable. The NCRF magnitude pairs from the original data and the null model are tested for significance using mass-univariate tests based on related measures t-tests.

To control for multiple comparisons, nonparametric permutation tests (Nichols and Holmes, 2002; Maris and Oostenveld, 2007) based on the threshold-free cluster-enhancement (TFCE) algorithm (Smith and Nichols, 2009) are used. First, at each dipole location and time point, the t-statistic is computed from the difference between the NCRF magnitude pairs. The resulting statistic-map is then processed by the TFCE algorithm, which boosts contiguous regions with high test statistic as compared to isolated ones, based on the assumption that spatial extent of the true sources is typically broader than those generated by noise. To find the distribution of these TFCE values under the null hypothesis, TFCE values are calculated following the same procedure, on 10000 different random permutations of the data. In each permutation, the sign of the NCRF magnitude differences is flipped for a randomly selected set of subjects, without resampling the same set of subjects. Then, at every permutation, the maximum value of the obtained TFCE values is recorded, thereby constructing a non-parametric distribution of the maximum TFCE values under the null hypothesis. The original TFCE values that exceed the $(1-\alpha)$ percentile of the null distribution are considered significant at a level of α corrected for multiple comparisons across the sources.

The Vector Test

This test aims at quantifying the significance of the estimated 3D NCRF components at the group level, based on the one-sample Hotelling's T^2 test. In the one-sample Hotelling's T^2 test, the population mean of the sample vectors is tested against the null hypothesis of mean zero, i.e. $\boldsymbol{\mu}_0 = 0$. To control for multiple comparisons, a similar strategy based on nonparametric permutations as in the case of the Length Test is used. At every time lag, the T^2 statistic for each dipole is computed as:

$$T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top \bar{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \tag{B.1}$$

where $\bar{\mathbf{x}}$, $\bar{\Sigma}$ are the population mean and covariance matrix of the vector-valued NCRF components, respectively. The T^2 statistic quantifies the variability of vector-valued samples, akin to the role of the t-statistic for 1D samples (Mardia, 1975). The resulting T^2-maps are then processed by the TFCE algorithm. As before, to construct a non-parametric distribution of maximum TFCE values under the null hypothesis, maximum values of the TFCE-processed T^2 maps on 10000 different random permutations of the data are recorded. In each permutation, the vector-valued NCRF components of each subject undergo uniform random rotations in 3D (Miles, 1965). The original TFCE values that exceed the $(1-\alpha)$ percentile of the null distribution are considered significant at a level of α, corrected for multiple comparisons across the sources.

Traditionally, response functions are estimated as scalar functions of the data, either over the sensor space or over the source space by orientation-constrained inverse solvers. Considering the directional vari-

ability of the NCRF estimates at the group level, however, takes into account the group level anatomical variability that may effect the current dipole orientations. In addition, the Vector Test is less computationally demanding than the Length Test, because it does not require refitting NCRFs for permuted models. In the Results section of the manuscript, we presented the NCRFs masked at a significance level of 5%, based on the Length Test. To demonstrate the difference between these two tests, here we also present the sames results using the Vector Test (Fig B.7).

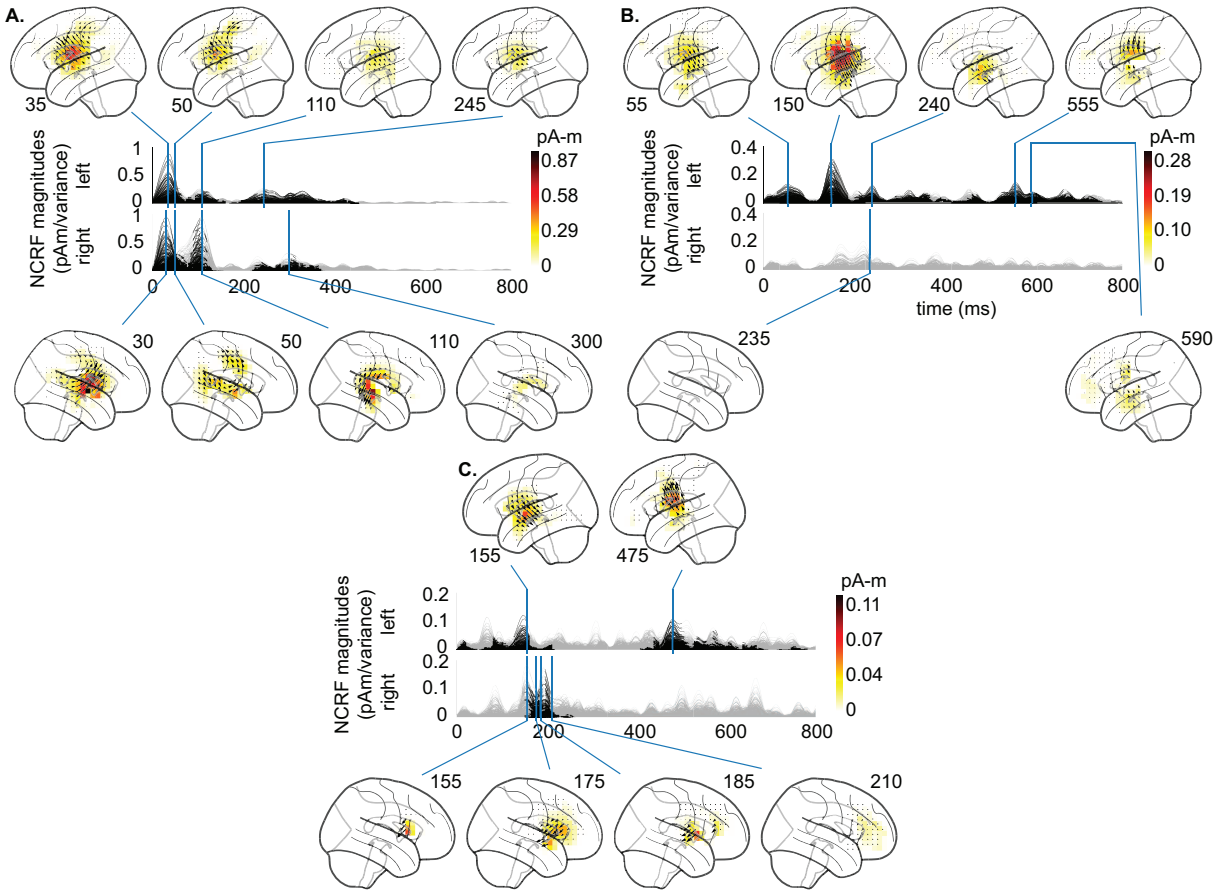

Figure B.7: Estimated NCRFs for acoustic envelope (A), word frequency (B), semantic composition (C): The anatomical plots show the group-level average NCRFs projected onto the lateral plane (top and bottom panels) corresponding to selected visually salient peaks in the temporal profiles (middle panels). The top and bottom portions of the subplot pertain to left and right hemisphere, respectively. Numerical labels of each anatomical subplot indicates the corresponding time lag in ms. The gray portions of the traces indicate statistically insignificant NCRFs at the group level (significance level of 5%). The significance levels are computed using the Vector Test, as opposed to the main manuscript where the significance levels are based on the Length Test. The main features of the NCRFs discussed in the Results section are similarly recovered by the Vector Test.

References

References

Aertsen, A.M.H.J., Johannesma, P.I.M., 1983. Reverse-correlation methods in auditory research. Quarterly Reviews of Biophysics 16, 341–414. https://doi.org/10.1017/S0033583500005126.

Aertsen, A.M.H.J., Olders, J.H.J., Johannesma, P.I.M., 1981. Biological cybernetics spectro-temporal receptive fields of auditory neurons in the grassfrog III. analysis of the stimulus-event relation for natural stimuli. Biological Cybernetics 39, 195–209. https://doi.org/10.1007/bf00342772.

Akram, S., Presacco, A., Simon, J.Z., Shamma, S.A., Babadi, B., 2016. Robust decoding of selective auditory attention from MEG in a competing-speaker environment via state-space modeling. NeuroImage 124. https://doi.org/10.1016/j.neuroimage.2015.09.048.

Akram, S., Simon, J.Z., Babadi, B., 2017. Dynamic estimation of the auditory temporal response function from MEG in competing-speaker environments. IEEE Transactions on Biomedical Engineering 64, 1896–1905. https://doi.org/10.1109/TBME.2016.2628884.

Babadi, B., Obregon-Henao, G., Lamus, C., Hämäläinen, M.S., Brown, E.N., Purdon, P.L., 2014. A subspace pursuit-based iterative greedy hierarchical solution to the neuromagnetic inverse problem. NeuroImage 87, 427–443. https://doi.org/10.1016/j.neuroimage.2013.09.008.

Baillet, S., Mosher, J.C., Leahy, R.M., 2001. Electromagnetic brain mapping. IEEE Signal processing magazine 18, 14–30. https://doi.org/10.1109/79.962275.

Beck, A., Teboulle, M., 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM journal on imaging sciences 2, 183–202. https://doi.org/10.1137/080716542.

Bell, A.J., Sejnowski, T.J., 1995. An information-maximization approach to blind separation and blind deconvolution. Neural Computation 7, 1129–1159. https://doi.org/10.1162/neco.1995.7.6.1129, arXiv:https://doi.org/10.1162/neco.1995.7.6.1129.

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis. Springer. https://doi.org/10.1007/978-1-4757-4286-2.

Boynton, G.M., Engel, S.A., Glover, G.H., Heeger, D.J., 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. Journal of Neuroscience 16, 4207–4221. https://doi.org/10.1523/JNEUROSCI.16-13-04207.1996.

Brodbeck, C., 2017. Eelbrain: 0.27, zenodo. https://doi.org/10.5281/zenodo.1444075.

Brodbeck, C., Hong, L.E., Simon, J.Z., 2018a. Rapid transformation from auditory to linguistic representations of continuous speech. Current Biology 28, 3976–3983. https://doi.org/10.1016/j.cub.2018.10.042.

Brodbeck, C., Presacco, A., Simon, J.Z., 2018b. Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension. NeuroImage 172, 162 – 174. https://doi.org/10.1016/j.neuroimage.2018.01.042.

Broderick, M.P., Anderson, A.J., Di Liberto, G.M., Crosse, M.J., Lalor, E.C., 2018. Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. Current Biology 28, 803–809. https://doi.org/10.1016/j.cub.2018.01.080.

Brysbaert, M., New, B., 2009. Moving beyond kučera and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. Behavior research methods 41, 977–990. https://doi.org/10.3758/BRM.41.4.977.

Chen, S.S., Donoho, D.L., Saunders, M.A., 2001. Atomic decomposition by basis pursuit. SIAM review 43, 129–159. https://doi.org/10.1137/S1064827596304010.

Crinion, J.T., Lambon-Ralph, M.A., Warburton, E.A., Howard, D., Wise, R.J., 2003. Temporal lobe regions engaged during normal speech comprehension. Brain 126, 1193–1201. https://doi.org/10.1093/brain/awg104.

Crosse, M.J., Di Liberto, G.M., Bednar, A., Lalor, E.C., 2016. The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. Frontiers in human neuroscience 10, 604. https://doi.org/10.3389/fnhum.2016.00604.

Das, P., 2019. neuro-currentRF: A Unified Approach to MEG Source Analysis under the Continuous Stimuli Paradigm. GitHub. URL: https://github.com/proloyd/neuro-currentRF.

Das, P., Brodbeck, C., Simon, J.Z., Babadi, B., 2018. Cortical localization of the auditory temporal response function from MEG via non-convex optimization. Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA https://doi.org/10.1109/ACSSC.2018.8645204.

David, S.V., Mesgarani, N., Shamma, S.A., 2007. Estimating sparse spectro-temporal receptive fields with natural stimuli. Network: Computation in Neural Systems 18, 191–212. https://doi.org/10.1080/09548980701609235.

Davis, M.H., Johnsrude, I.S., 2007. Hearing speech sounds: Top-down influences on the interface between audition and speech perception. Hearing Research 229. https://doi.org/10.1016/j.heares.2007.01.014.

de Heer, W.A., Huth, A.G., Griffiths, T.L., Gallant, J.L., Frédéric, X., Theunissen, E., 2017. The hierarchical cortical organization of human speech processing. Journal of Neuroscience 37, 6539–6557. https://doi.org/10.1523/JNEUROSCI.3267-16.2017.

Dewitt, I., Rauschecker, J.P., 2012. Phoneme and word recognition in the auditory ventral stream. Proceedings of the National Academy of Sciences 109, 2709. https://doi.org/10.1073/pnas.1113427109.

Di Liberto, G.M., O'Sullivan, J.A., Lalor, E.C., 2015. Low-frequency cortical entrainment to speech reflects phoneme-level processing. Current Biology 25, 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030.

Ding, N., Simon, J.Z., 2012a. Emergence of neural encoding of auditory objects while listening to competing speakers. Proceedings of the National Academy of Sciences 109, 11854–11859. https://doi.org/10.1073/pnas.1205381109.

Ding, N., Simon, J.Z., 2012b. Neural coding of continuous speech in auditory cortex during monaural and

dichotic listening. Journal of Neurophysiology 107. https://doi.org/10.1152/jn.00297.2011.

Ding, N., Simon, J.Z., 2013a. Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. Journal of Neuroscience 33, 5728–5735. https://doi.org/10.1523/JNEUROSCI.5297-12.2013.

Ding, N., Simon, J.Z., 2013b. Robust cortical encoding of slow temporal modulations of speech, in: Basic Aspects of Hearing. Springer, pp. 373–381. https://doi.org/10.1007/978-1-4614-1590-9_41.

Engemann, D.A., Gramfort, A., 2015. Automated model selection in covariance estimation and spatial whitening of MEG and EEG signals. NeuroImage 108, 328–342. https://doi.org/10.1016/j.neuroimage.2014.12.040.

Fadiga, L., Craighero, L., Buccino, G., Rizzolatti, G., 2002. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. European journal of Neuroscience 15, 399–402. https://doi.org/10.1046/j.0953-816x.2001.01874.x.

Feichtinger, H.G., Strohmer, T., 2012. Gabor analysis and algorithms: Theory and applications. Springer Science & Business Media.

Fischl, B., 2012. Freesurfer. NeuroImage 62, 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021.

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. https://doi.org/10.1016/j.neuroimage.2007.09.048.

Gazzaniga, I., Ivry, R., 2009. Cognitive Neuroscience: The Biology of the Mind. New York: WW Norton and Company Press.

Goldstein, T., Studer, C., Baraniuk, R., 2014. A field guide to forward-backward splitting with a FASTA implementation. arXiv preprint arXiv:1411.3406 .

Gorodnitsky, I.F., George, J.S., Rao, B.D., 1995. Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm. Electroencephalography and clinical Neurophysiology 95, 231–251. https://doi.org/10.1016/0013-4694(95)00107-A.

Gramfort, A., Kowalski, M., Hämäläinen, M., 2012. Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. Physics in Medicine & Biology 57, 1937. https://doi.org/10.1088/0031-9155/57/7/1937.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., et al., 2013. MEG and EEG data analysis with MNE-python. Frontiers in neuroscience 7, 267. https://doi.org/10.3389/fnins.2013.00267.

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. https://doi.org/10.1016/j.neuroimage.2013.10.027.

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. Reviews of Modern Physics 65, 413–497. https://doi.org/10.1103/RevModPhys.65.413.

Handy, T.C., 2005. Event-related potentials: A methods handbook. MIT press.

Haufe, S., Nikulin, V.V., Ziehe, A., Müller, K.R., Nolte, G., 2008. Combining sparsity and rotational invariance in EEG/MEG source reconstruction. NeuroImage 42, 726–738. https://doi.org/10.1016/j.neuroimage.2008.04.246.

Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. Cognition 92, 67–99. https://doi.org/10.1016/j.cognition.2003.10.011.

Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nature Reviews Neuroscience 8, 393–402. https://doi.org/10.1038/nrn2113.

Hullett, P.W., Hamilton, L.S., Mesgarani, N., Schreiner, C.E., Chang, E.F., 2016. Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. Journal of Neuroscience 36, 2014–2026. https://doi.org/10.1523/JNEUROSCI.1779-15.2016.

Kaas, J.H., Hackett, T.A., 1999. 'What' and 'where' processing in auditory cortex. Nature Neuroscience 2, 1045–1047. https://doi.org/10.1038/15967.

Krishnaswamy, P., Obregon-Henao, G., Ahveninen, J., Khan, S., Babadi, B., Iglesias, J.E., Hämäläinen, M.S., Purdon, P.L., 2017. Sparsity enables estimation of both subcortical and cortical activity from MEG and EEG. Proceedings of the National Academy of Sciences 114, E10465–E10474. https://doi.org/10.1073/pnas.1705414114.

Lalor, E.C., Foxe, J.J., 2010. Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. European journal of neuroscience 31, 189–193. https://doi.org/10.1111/j.1460-9568.2009.07055.x.

Lalor, E.C., Pearlmutter, B.A., Reilly, R.B., McDarby, G., Foxe, J.J., 2006. The VESPA: a method for the rapid estimation of a visual evoked potential. NeuroImage 32, 1549–1561. https://doi.org/10.1016/j.neuroimage.2006.05.054.

Lalor, E.C., Power, A.J., Reilly, R.B., Foxe, J.J., 2009. Resolving precise temporal processing properties of the auditory system using continuous stimuli. Journal of Neurophysiology 102, 349–359. https://doi.org/10.1152/jn.90896.2008.

Lamus, C., Hämäläinen, M.S., Temereanca, S., Brown, E.N., Purdon, P.L., 2012. A spatiotemporal dynamic distributed solution to the MEG inverse problem. NeuroImage 63, 894–909. https://doi.org/10.1016/j.neuroimage.2011.11.020.

Lin, F.H., Witzel, T., Ahlfors, S.P., Stufflebeam, S.M., Belliveau, J.W., Hämäläinen, M.S., 2006. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm

estimates. NeuroImage 31, 160–171. https://doi.org/10.1016/j.neuroimage.2005.11.054.

Luck, S.J., 2014. An introduction to the event-related potential technique. MIT press.

Machens, C.K., Wehr, M.S., Zador, A.M., 2004. Linearity of cortical receptive fields measured with natural sounds. Journal of Neuroscience 24, 1089–1100. https://doi.org/10.1523/JNEUROSCI.4445-03.2004.

Mardia, K., 1975. Assessment of multinormality and the robustness of hotelling's T^2. test. Journal of the Royal Statistical Society: Series C (Applied Statistics) 24, 163–171.

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG-and MEG-data. Journal of Neuroscience Methods 164, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024.

Marmarelis, V.Z., 2004. Nonlinear dynamic modeling of physiological systems. volume 10. John Wiley & Sons. https://doi.org/10.1002/9780471679370.

Mesgarani, N., Chang, E.F., 2012. Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485, 233. https://doi.org/10.1038/nature11020.

Mesgarani, N., Cheung, C., Johnson, K., Chang, E.F., 2014. Phonetic feature encoding in human superior temporal gyrus. Science 343. https://doi.org/10.1126/science.1245994.

Mesgarani, N., David, S.V., Fritz, J.B., Shamma, S.A., 2008. Phoneme representation and classification in primary auditory cortex. The Journal of the Acoustical Society of America 123. https://doi.org/10.1121/1.2816572.

Miles, R.E., 1965. On random rotations in $\mathbb{R}^3$. Biometrika 52, 636–639. https://doi.org/10.1093/biomet/52.3-4.636.

Mosher, J.C., Leahy, R.M., Lewis, P.S., 1999. EEG and MEG: forward solutions for inverse methods. IEEE Transactions on Biomedical Engineering 46, 245–259. https://doi.org/10.1109/10.748978.

Nesterov, Y., 2005. Smooth minimization of non-smooth functions. Mathematical programming 103, 127–152. https://doi.org/10.1007/s10107-004-0552-5.

Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Human brain mapping 15, 1–25. https://doi.org/10.1002/hbm.1058.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.H., Saberi, K., Serences, J.T., Hickok, G., 2010. Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. Cerebral Cortex 20, 2486–2495. https://doi.org/10.1093/cercor/bhp318.

Paninski, L., 2004. Maximum likelihood estimation of cascade point-process neural encoding models. Network: Computation in Neural Systems 15, 243–262. https://doi.org/10.1088/0954-898X_15_4_002.

Pasley, B.N., David, S.V., Mesgarani, N., Flinker, A., Shamma, S.A., Crone, N.E., Knight, R.T., Chang, E.F., 2012. Reconstructing speech from human auditory cortex. PLoS Biology 10. https://doi.org/10.1371/journal.pbio.1001251.

Peelle, J.E., Johnsrude, I.S., Davis, M.H., 2010. Hierarchical processing for speech in human auditory cortex and beyond. Frontiers in Human Neuroscience 4. https://doi.org/10.3389/fnhum.2010.00051.

Pirondini, E., Babadi, B., Obregon-Henao, G., Lamus, C., Malik, W.Q., Hämäläinen, M.S., Purdon, P.L., 2018. Computationally efficient algorithms for sparse, dynamic solutions to the EEG source localization problem. IEEE Transactions on Biomedical Engineering 65, 1359–1372. https://doi.org/10.1109/TBME.2017.2739824.

Presacco, A., Simon, J., Anderson, S., 2018. EEG-MEG. available in the Digital Repository at the University of Maryland (DRUM) URL: http://hdl.handle.net/1903/21184.

Presacco, A., Simon, J.Z., Anderson, S., 2016. Effect of informational content of noise on speech representation in the aging midbrain and cortex. Journal of Neurophysiology 116, 2356–2367. https://doi.org/10.1152/jn.00373.2016.

Pulvermüller, F., Huss, M., Kherif, F., Del Prado Martin, F.M., Hauk, O., Shtyrov, Y., 2006. Motor cortex maps articulatory features of speech sounds. Proceedings of the National Academy of Sciences 103, 7865–7870. https://doi.org/10.1073/pnas.0509989103.

Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. Nature Neuroscience 12, 718–724. https://doi.org/10.1038/nn.2331.

Ringach, D., Shapley, R., 2004. Reverse correlation in neurophysiology. Cognitive Science 28, 147–166. https://doi.org/10.1207/s15516709cog2802_2.

Sarvas, J., 1987. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. Physics in Medicine & Biology 32, 11. https://doi.org/10.1088/0031-9155/32/1/004.

Schoffelen, J.M., Gross, J., 2009. Source connectivity analysis with MEG and EEG. Human brain mapping 30, 1857–1865. https://doi.org/10.1002/hbm.20745.

Scott, S.K., Johnsrude, I.S., 2003. The neuroanatomical and functional organization of speech perception. Trends in Neurosciences 26, 100–107. https://doi.org/10.1016/S0166-2236(02)00037-1.

Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage 44, 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061.

Taulu, S., Simola, J., 2006. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Physics in Medicine and Biology 51, 1759–1768. https://doi.org/10.1088/0031-9155/51/7/008.

Theunissen, F.E., 2007. STRFPak: 5.3. URL: http://strfpak.berkeley.edu/.

Theunissen, F.E., 2010. STRFlab: 1.45. URL: http://strflab.berkeley.edu/.

Theunissen, F.E., David, S.V., Singh, N.C., Hsu, A., Vinje, W.E., Gallant, J.L., 2001. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Network: Computation in Neural Systems 12, 289–316. https://doi.org/10.1080/net.12.3.289.316.

Westerlund, M., Kastner, I., Al Kaabi, M., Pylkkänen, L., 2015. The LATL as locus of composition: MEG evidence from english and arabic. Brain and Language 141, 124–134. https://doi.org/10.1016/j.bandl.

2014.12.003.

Wilson, S.M., Saygin, A.P., Sereno, M.I., Iacoboni, M., 2004. Listening to speech activates motor areas involved in speech production. Nature Neuroscience 7, 701–702. https://doi.org/10.1038/nn1263.

Wipf, D., Nagarajan, S., 2009. A unified bayesian framework for MEG/EEG source imaging. NeuroImage 44, 947–966. https://doi.org/10.1016/j.neuroimage.2008.02.059.

Wipf, D.P., Owen, J.P., Attias, H.T., Sekihara, K., Nagarajan, S.S., 2010. Robust bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG. NeuroImage 49, 641–655. https://doi.org/10.1016/j.neuroimage.2009.06.083.

Wright, S.J., 2015. Coordinate descent algorithms. Mathematical Programming 151, 3–34. https://doi.org/10.1007/s10107-015-0892-3.

Wu, M.C.K., David, S.V., Gallant, J.L., 2006. Complete functional characterization of sensory neurons by system identification. Annual Review of Neuroscience 29, 477–505. https://doi.org/10.1146/annurev.neuro.29.051605.113024.

Wu, W., Nagarajan, S., Chen, Z., 2016. Bayesian machine learning: EEG/MEG signal processing measurements. IEEE Signal Processing Magazine 33, 14–36. https://doi.org/10.1109/MSP.2015.2481559.

Yang, X., Wang, K., Shamma, S.A., 1992. Auditory representations of acoustic signals. IEEE Transactions on Information Theory 38, 824–839. https://doi.org/10.1109/18.119739.