

# Population Level Inference for Multivariate MEG Analysis

Anna Jafarpour<sup>1,2\*</sup>, Gareth Barnes<sup>3</sup>, Lluís Fuentemilla<sup>4,5</sup>, Emrah Duzel<sup>1,2</sup>, Will D. Penny<sup>3</sup>

**1** Institute of Cognitive Neuroscience, University College London, London, United Kingdom, **2** Institute of Cognitive Neurology and Dementia Research, Otto-von-Guericke University, Magdeburg, Germany, **3** Wellcome Trust Centre for Neuroimaging, University College London, London, United Kingdom, **4** Cognition and Brain Plasticity Unit, Institute of Biomedicine Research of Bellvitge, Barcelona, Spain, **5** Department of Basic Psychology, University of Barcelona, Barcelona, Spain

## Abstract

Multivariate analysis is a very general and powerful technique for analysing Magnetoencephalography (MEG) data. An outstanding problem however is how to make inferences that are consistent over a group of subjects as to whether there are condition-specific differences in data features, and what are those features that maximise these differences. Here we propose a solution based on Canonical Variates Analysis (CVA) model scoring at the subject level and random effects Bayesian model selection at the group level. We apply this approach to beamformer reconstructed MEG data in source space. CVA estimates those multivariate patterns of activation that correlate most highly with the experimental design; the order of a CVA model is then determined by the number of significant canonical vectors. Random effects Bayesian model comparison then provides machinery for inferring the optimal order over the group of subjects. Absence of a multivariate dependence is indicated by the null model being the most likely. This approach can also be applied to CVA models with a fixed number of canonical vectors but supplied with different feature sets. We illustrate the method by identifying feature sets based on variable-dimension MEG power spectra in the primary visual cortex and fusiform gyrus that are maximally discriminative of data epochs before versus after visual stimulation.

**Citation:** Jafarpour A, Barnes G, Fuentemilla L, Duzel E, Penny WD (2013) Population Level Inference for Multivariate MEG Analysis. *PLoS ONE* 8(8): e71305. doi:10.1371/journal.pone.0071305

**Editor:** Stefano L. Sensi, University G. D'Annunzio, Italy

**Received:** March 27, 2013; **Accepted:** June 25, 2013; **Published:** August 5, 2013

**Copyright:** © 2013 Jafarpour et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by the Wellcome Trust (grant number 091593/Z/10/Z) and L.F. is supported by the Spanish Government (PSI2010-15024) and the Ramón y Cajal program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: a.jafarpour@ucl.ac.uk

## Introduction

Multivariate analysis of Magnetoencephalography (MEG) data is a powerful technique which considers the relationship between multiple data features and multiple experimental conditions. Previously this approach has been used to study oscillatory representations of stimuli; such as visual stimuli at the time of perception [1,2] or the replay of oscillatory patterns during memory tasks [3,4]. A major problem with such multivariate analyses is the identification of discriminative data features from a high dimensional measurement space. In the above studies, the set of discriminatory features were allowed to vary from subject to subject. This between-subject variability, however, makes it difficult to interpret experimental findings in terms of a consistent set of underlying cognitive processes.

In this paper we present a principled approach to systematically select the most discriminatory and minimally complex feature set that is consistent over subjects i.e. at the 'group level'. This analysis also enables systematic inference on the dimensions of these feature-spaces. For example, if there is no dependence between data features and experimental condition the inferred dimension will be zero.

Our framework is based on Canonical Variates Analysis (CVA). CVA models multivariate dependencies between a set of class labels and data features. The order of the CVA model is then based on the number of significant canonical vectors, as determined by the Bayesian Information Criterion (BIC) [5,6]. Absence of a multivariate dependence (or no significant decodability) is indicated by the zeroth order model (null model or model 0) being the most likely. Here we apply CVA to

beamformer reconstructed MEG data in source space [7]. Nevertheless, in principle, it could be applied to data in sensor space or to data after various transformations, including the use of principal or independent component analysis [8]. (We return to this latter issue in the discussion.)

The model ranking approach allows us to test, at the group level, both whether there is a multivariate dependence between data features and experimental condition, and if there is, to find which feature sets maximise the strength of this dependence.

To test whether these multivariate dependencies are consistent over a group of subjects, we use random effects Bayesian model selection [9], based on the BIC values. We illustrate this method to determine the spectral resolution (number of frequency bands) that maximizes decodability of data features into the experimental conditions. We used MEG power spectra at each voxel in source space, within the regions of interest (ROIs), the primary visual cortex (V1) and fusiform gyrus (FFG), and the experimental conditions indicating whether the data were from a pre- or post-stimulus epoch of a simple visual processing paradigm.

## Materials and Methods

### Methods

This section describes the data processing pipeline we propose. This comprises five steps

- MEG Source Reconstruction.** This step is optional, as MEG data can also be analysed in sensor space or, for example, projected onto principal or independent component

spaces. In this paper the features of the MEG signal we use are power spectra. More generally, these can be any function of the MEG data, such as phase and/or amplitude or more exotic nonlinear measures.

2. **Canonical Variates Analysis.** Here we apply a CVA model at each point in source space as our goal is brain mapping. The maps indicate which areas show consistent relationships between multivariate data features and experimental condition.
3. **Bayes factors.** The order of a CVA model is determined by the number of canonical vectors. This step computes the evidence of a model with  $m$  canonical vectors in relation to the evidence of a model with zero canonical vectors. The ratio of these evidences is known as a Bayes factor.
4. **Feature Set Selection.** The optimal model will depend not only on the number of canonical vectors but also on the features to which these vectors map. In this paper we compare models with single canonical vectors but with a different fractionation of the MEG power spectrum.
5. **Random Effects Bayesian Model Selection (RFX-BMS).** The previous steps are applied to data from multiple subjects to produce Bayes factor maps for each subject and model comparison. This allows for single alternative models to be compared with a null model, or for any number of models to be simultaneously compared with each other. This final step computes the frequency with which models are used in the population from which the subjects are drawn.

The following subsections describe each of the above steps in more detail.

**MEG Source Reconstruction.** We source reconstructed data for each subject using the SPM8 implementation of the Linearly Constrained Minimum Variance (LCMV) beamformer [10–12]. The software for source reconstruction, and computation of Bayes factors for CVA models is available in the SPM Beamforming toolbox (<http://code.google.com/p/spm-beamforming-toolbox/>). This produces a log Bayes factor image for each subject and model. The software for implementing Random effects Bayesian model selection is available in the latest release of SPM [13] (<http://www.fil.ion.ucl.ac.uk/spm/software/>). This takes the log Bayes factor images for all subjects and produces expected frequency maps (\*\_xpm.img) and, optionally, exceedance probability maps (\*\_epm.img).

The forward model used in source reconstruction was defined using an inverse normalized canonical head-shape brain for all subjects [14]. At each source location we selected the orientation that maximises projected power [12] which gives a single weight vector for each source location. Briefly, the weights for location  $s$  were given by

$$w_s = (L_s C^{-1} L_s^T)^{-1} L_s C^{-1} \tag{1}$$

where  $L_s \in \mathbb{R}^{1 \times m}$  is the lead field matrix for  $m$  channels at source location  $s$  and  $C \in \mathbb{R}^{m \times m}$  is the sensor covariance matrix. This corresponds to an LCMV beamformer with zero for the regularisation parameter [11]. Accordingly, the source level estimate of activity for trial  $n$  at location  $s$  is given by

$$y = w_s B_n \tag{2}$$

where  $B_n \in \mathbb{R}^{m \times p}$  comprises  $p$  complex valued Fourier coefficients describing the signal at  $m$  MEG sensors on trial  $n \in 1, \dots, N$ . In the next section we go on to look at multivariate dependence between the experimental design and the spectral features and this source level estimate across the brain.

**Canonical Variates Analysis.** CVA is a method for detecting dependencies between a set of variables  $X \in \mathbb{R}^{N \times q}$ , and a set of variables  $Y \in \mathbb{R}^{N \times v}$ . The aim of CVA is to find the linear projections of  $X$  and  $Y$  with maximal correlation. Given  $z_x = XU_x$  and  $z_y = YU_y$ , we can compute the canonical correlation

$$r = \frac{\|z_x^T z_y\|^2}{\|z_x^T z_x\| \|z_y^T z_y\|} \tag{3}$$

The projections  $U_x$  and  $U_y$  which maximise this correlation are known as the canonical vectors and the resulting  $z_x$  and  $z_y$  are the canonical variates. If  $\Sigma = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}$  is the sample covariance matrix and  $V_Y$  and  $V_X$  are the left and right singular vectors of  $\Sigma_{XX}^{-1/2} \Sigma_{XY} \Sigma_{YY}^{-1/2}$  in decreasing order, then the canonical vectors can be computed as  $U_Y = \Sigma_{YY}^{-1/2} V_Y$  and  $U_X = \Sigma_{XX}^{-1/2} V_X$  [15]. There are  $i = 1..h$  pairs of canonical vectors where  $h = \min(q, v)$ . The canonical correlations  $r_i$  for  $i = 1..h$  are used to compute Bayes factors, as described in the following section.

**Bayes Factors.** We first introduce some terminology. The dimension of a CVA model is given by the number of significantly non-zero canonical vectors. If there exists a linear multivariate dependence between  $X$  and  $Y$  then the dimension of the corresponding CVA model is non-zero. Thus one can test for linear multivariate dependence by estimating CVA model dimension.

A standard approach from classical inference here is Bartlett’s test for dimensionality [15]. However, to our knowledge, there is no simple way to carry over these results to the group level. We therefore prefer a Bayesian method, as this integrates seamlessly with established methods for group level inference (see final subsection).

These Bayesian methods first compute the evidence for a model,  $p(Y|m)$ , with  $m$  canonical vectors. Various methods exist for computing the Bayesian model evidence for a CVA model. These include the Bayesian Information Criterion (BIC) [5,6] and variational approximations [16]. This paper uses a BIC approximation which we now derive.

If there is no relation between dependent variable (or ‘data’)  $Y$  and independent variable  $X$ , then the log-likelihood of the data is

$$\log p(Y) = -\frac{N}{2} \log |\Sigma_{YY}| \tag{4}$$

where  $\Sigma_{YY}$  is the data covariance. If there is a relation between  $X$  and  $Y$  then the log-likelihood can be calculated as follows. The maximum likelihood coefficients are  $\beta_{ML} = (X^T X)^{-1} X^T Y$  and the log-likelihood is

$$\log p(Y|\beta_{ML}) = -\frac{N}{2} \log |\Sigma_{Y|X}| \tag{5}$$

where  $\Sigma_{Y|X} = \Sigma_{YY} - \Sigma_{XY}^T \Sigma_{XX}^{-1} \Sigma_{XY}$ ,  $\Sigma_{XY}$  is the covariance between  $X$  and  $Y$ , and  $\Sigma_{XX}$  is the covariance of  $X$ . The log-likelihood ratio,  $\Lambda$ , is therefore

$$\begin{aligned} \Lambda &= \log \frac{p(Y|\beta_{ML})}{p(Y)} \\ &= \frac{N}{2} \log |\Sigma_{Y|X}^{-1} \Sigma_{YY}|. \end{aligned} \tag{6}$$

If  $\sigma_i$  is the  $i$ th eigenvalue of  $\Sigma_{Y|X}^{-1}\Sigma_{YY}$  we can write

$$\Lambda = \frac{N}{2} \sum_{i=1}^h \log \sigma_i \tag{7}$$

where  $h = \min(q, v)$ . This is also known as Wilk's Lambda [15]. We also define the quantity

$$\Lambda_{j,t} = \frac{N}{2} \sum_{i=j}^t \log \sigma_i \tag{8}$$

so  $\Lambda_{1,t}$  is the log-likelihood ratio for a CVA model with  $t$  canonical variates. The quantity  $\Lambda_{j,t}$  is used to compute the BIC and can also be expressed in other forms. We next show how it is computed in our implementation, and finally show how it is related to canonical correlations.

A second expression for  $\Lambda_{j,t}$  can be derived as follows. Let  $\Sigma_{YY} = \Sigma_{\hat{Y}\hat{Y}} + \Sigma_{Y|X}$ , where  $\Sigma_{\hat{Y}\hat{Y}}$  is the covariance explained by the model and  $\Sigma_{Y|X}$  is the covariance not explained by the model. Then if  $\lambda_i$  are eigenvalues of  $\Sigma_{Y|X}^{-1}\Sigma_{\hat{Y}\hat{Y}}$  then the above relationship can be used to show that  $\sigma_i = \lambda_i + 1$  (see Appendix B of [17]). Hence an alternative expression for Wilk's Lambda is

$$\Lambda = \frac{N}{2} \sum_{i=1}^h \log(1 + \lambda_i) \tag{9}$$

This expression is used in the current paper and has been implemented in the SPM software [17]. Accordingly,  $\Sigma_{\hat{Y}\hat{Y}}$  can be formed directly from model predictions

$$\begin{aligned} \hat{Y} &= X\beta_{ML} \\ \Sigma_{\hat{Y}\hat{Y}} &= \hat{Y}^T \hat{Y} \end{aligned} \tag{10}$$

and  $\Sigma_{Y|X}$  from the residuals

$$\begin{aligned} R &= Y - \hat{Y} \\ \Sigma_{Y|X} &= R^T R. \end{aligned} \tag{11}$$

The  $i$ th canonical correlation can be expressed as  $r_i = \sqrt{\frac{\lambda_i}{\lambda_i + 1}}$ . Hence, a third equivalent form for the log likelihood ratio is

$$\Lambda = -\frac{N}{2} \sum_{i=1}^h \log(1 - r_i^2). \tag{12}$$

In summary, we can write

$$\begin{aligned} \Lambda_{j,t} &= \frac{N}{2} \sum_{i=j}^t \log \sigma_i \\ &= \frac{N}{2} \sum_{i=j}^t \log(1 + \lambda_i) \\ &= -\frac{N}{2} \sum_{i=j}^t \log(1 - r_i^2). \end{aligned} \tag{13}$$

This last expression appears in [5,6].

The log evidence for a model with no parameters (null model) is simply the log likelihood of the data,  $L_0 = \log p(Y)$ . The log evidence for model  $m$  with parameters  $\beta$  is given by  $L_m = \log \int p(Y|\beta)p(\beta)d\beta$ . This can be approximated by the Bayesian Information Criterion (BIC) as

$$BIC = \log p(Y|\beta_{ML}) - \frac{k}{2} \log N \tag{14}$$

where  $k$  is the number of parameters in the model and  $\beta_{ML}$  are the maximum likelihood parameters. A Bayes factor is the ratio of model evidences. Here we define

$$BF(m) = \frac{p(Y|m)}{p(Y)} \tag{15}$$

Log Bayes factors can therefore be approximated as differences in the BIC scores. Under BIC, the log Bayes factor for a CVA model of dimension  $m$  versus a model with dimension zero (null model) is given as

$$LogBF(m)_{BIC} = \Lambda_{1,m} - \frac{k}{2} \log N. \tag{16}$$

where

$$\Lambda_{1,m} = -\frac{N}{2} \sum_{i=1}^m \log(1 - r_i^2). \tag{17}$$

and  $N$  is the number of data points and  $r_i$  are the canonical correlations at each dimension  $i$ . This expression has been used in previous studies [5,6] and is derived from equation 13. The estimated model order is the one which has the largest LogBF. Negative values of  $LogBF(m)$  express evidence in favour of the null model. Intuitively, better CVA models will have stronger canonical correlations ( $r_i$ ) and fewer parameters ( $k$ ).

**Feature Set Selection.** It is also possible to compute Bayes factors for models with the same number of canonical vectors but supplied with different feature sets. Bayesian model comparison here allows the models to vary but the data must stay the same. Feature set selection therefore requires that we set up CVA models such that  $Y$  is a design matrix encoding experimental conditions and  $X$  are independent variables comprising the neuroimaging data features (in other words we switch the traditional roles of these variables ( $X$  and  $Y$ ) to make it clear that we are searching for optimal data features for a fixed experimental design). In this paper these features are  $d$ -dimensional power spectra. We then compute  $LogBF(d)$  images where each is the log Bayes Factor for a model with a single canonical vector and  $d$ -dimensional features  $X$ , versus a model with zero canonical vectors. We can then use the same Bayes factor images to compare different feature dimensions. For example, for pairwise comparisons

$$\begin{aligned} BF(d_1, d_2) &= \frac{p(Y|d_1)}{p(Y|d_2)} \\ &= \frac{p(Y|d_1)}{p(Y)} \frac{p(Y)}{p(Y|d_2)} \end{aligned} \tag{18}$$

Hence

$$LogBF(d_1, d_2) = LogBF(d_1) - LogBF(d_2) \tag{19}$$

We can also implement multi-way comparisons as described in the next section.

**Random effects Bayesian model selection.** Random effects Bayesian model selection (RFX-BMS) [9] views the assignment of models to subjects as a random process in which each subject is assigned to model  $i$  with probability  $f_i$ . Here  $f_i$  is the frequency with which model  $i$  is used in the population from which the subjects were drawn. The Bayesian algorithm for estimating model frequencies  $f_i$  from the table of log model evidence values [9] uses a Dirichlet prior

$$p(f) = \text{Dir}(f; \alpha^0) \tag{20}$$

with ‘count parameters’  $\alpha_m^0 = 1$ . These parameters can be thought of as corresponding to the assumption of having previously observed one instance of each model type. These parameters produce a flat prior. The posterior is approximated to also be a Dirichlet

$$p(f|\tilde{Y}) = \text{Dir}(f; \alpha) \tag{21}$$

where  $\tilde{Y}$  indicates data from all subjects. The count parameters  $\alpha_m$ , are initialised as  $\alpha_m^0$ , and then updated iteratively as follows

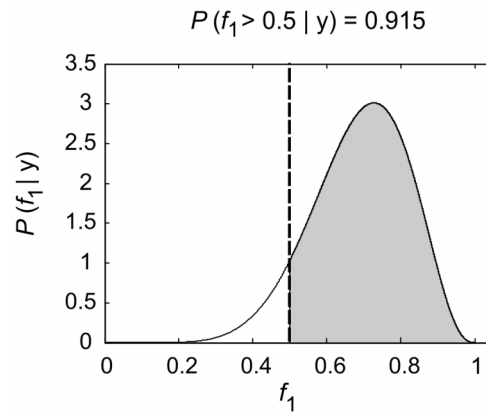
$$\begin{aligned} u_{nm} &= \exp \left[ \log p(y_n|m) + \psi(\alpha_m) - \sum_m \psi(\alpha_m) \right] \\ g_{nm} &= \frac{u_{nm}}{\sum_{m'} u_{nm'}} \\ \alpha_m &= \alpha_m^0 + \sum_n g_{nm} \end{aligned} \tag{22}$$

where  $\psi()$  is the digamma function [18]. Here  $\log p(y_n|m)$  is the entry in the log evidence table from the  $n$ th subject (row) and  $m$ th model (column). The quantity  $g_{nm}$  is the posterior probability that subject  $n$  used the  $m$ th model. This is ‘posterior’ as in after seeing the model evidence table and, implicitly, the data from all subjects  $\tilde{Y}$ .

This algorithm can also be applied to a table of log Bayes factor values, as long as the Bayes factors have all been computed with respect to the same common model. In this paper, the common model is the null CVA model with zero canonical variates.

The goal of RFX-BMS is to estimate  $f_i$  using a table of model evidence scores, or Bayes factors with respect to a common model, from  $S$  subjects and  $K$  models. Intuitively, if the scores favour model  $i$  in 9 out of 10 subjects, then  $f_i$  will be estimated to be about 0.9. However, the estimate of  $f_i$  is also influenced by the degree to which models are favoured. For example, if for the 10th subject the score is greatly in favour of a different model then the estimate of  $f_i$  will be commensurately reduced. Given data  $Y$  (in practice, a table of logBF values), a posterior distribution,  $p(f_i|Y)$ , can be estimated using the algorithm described in [9]. The mean of this distribution,  $\langle f_i|Y \rangle$  provides an estimate of the model frequencies. This is also referred to as the ‘expected frequency’.

It is also possible to compute the probability that one model frequency exceeds another. For example, when comparing just two models we can compute  $\phi_1 = p(f_1 > f_2|Y)$ . This is known as the exceedance probability for model 1 over model 2. Figure 1 illustrates the concept of an exceedance probability. If one has maps of model evidence over anatomical space, and for multiple subjects, it is possible to produce maps of expected frequencies or exceedance probabilities. In previous work [19], for example,



**Figure 1.** This graphic demonstrates random effects model selection for the case of comparing two models. These models have frequencies  $f_1$  and  $f_2$ . These frequencies refer to the population from which the subjects were drawn. The figure plots the posterior probability of  $f_1$ . The mean of this density is  $\langle f_1|Y \rangle = 0.75$ , indicating that 75% subjects use model 1. For two models, the exceedance probability  $p(f_1 > f_2|Y) = p(f_1 > 0.5|Y)$  is given by the posterior mass in excess of  $f_1 = 0.5$ . Here, the exceedance probability is 0.915. doi:10.1371/journal.pone.0071305.g001

Exceedance Probability Maps (EPMs) were plotted for univariate General Linear Models fitted to functional MRI data. In this paper we plot expected frequency maps for CVA models in MEG source space.

**Experimental Data**

**Participants and Experiment.** 10 healthy young adults (6 female; on average 23 years old ( $SD=2$ )) participated in an episodic memory study. All participants gave written informed consent to participate and the study was approved by the University College London Research Ethics Committee for human-based research. All participants were financially compensated for their participation. The overall goal of this study was to examine the neural correlates of visual memory. Here, we used the data from a small part of this study in which subjects were presented with images of faces. These images were grey scaled and normalized to a mean grey value of 127 and SD of 75, of dimensions  $300 \times 300$  pixels, and shown upon a grey background (grey value of 127) subtending approximately 6 degrees of horizontal and vertical visual angle.

**MEG recordings.** MEG data were recorded with a 274 channel CTF Omega whole-head gradiometer system (VSM MedTech, Coquitlam, BC, Canada) with a 600 Hz sampling rate.

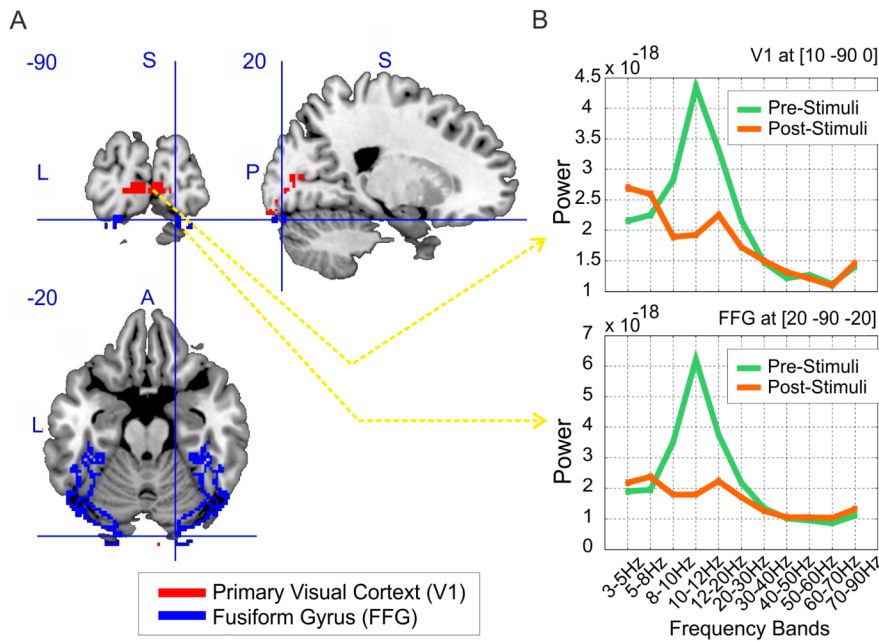
**Table 1.** Definition of feature space.

d	Frequencies(Hz)
1	3–90
3	3–10, 10–30, 30–90
5	3–8, 8–12, 12–30, 30–50, 50–90
7	3–5, 5–8, 8–12, 12–20, 20–30, 30–50, 50–90
9	3–5, 5–8, 8–10, 10–12, 12–20, 20–30, 30–40, 40–50, 50–90
11	3–5, 5–8, 8–10, 10–12, 12–20, 20–30, 30–40, 40–50, 50–60, 60–70, 70–90

This describes the fractionation of the power spectrum into  $d$  separate bands,  $d = 1, 3, 5, 7, 9$  and 11.

doi:10.1371/journal.pone.0071305.t001





**Figure 2. Regional Activity.** (A) the MNI mask for region of interest: grey matter in the primary visual cortex (V1) and fusiform gyrus (FFG), the view is from [20 -90 -20]mm. (B) The power spectrum for pre- and post-stimulus activity across 11 frequency bands for an individual participant. The top plot is the signal from a V1 voxel, [10 -90 0]mm (Talairach coordinates), and the bottom plot is the signal from a FFG voxel, [20 -90 -20]mm. doi:10.1371/journal.pone.0071305.g002

Head position inside the system was acquired via head localizer coils attached to the nasion and 1 cm anterior to the left and right pre-auricular points. Participants were seated upright and the stimuli were back-projected onto a screen approximately 1 m in front of them.

**Data Analysis**

At each point in source space we generated a *d*-dimensional feature vector of power in *d* frequency bands. Here we used 6 different features. The frequency bands are as defined in Table 1. The average power in the frequency band was computed separately across 1 second before (-1000 to 0) and 1 second after (0 to 1000 ms) onset of visual stimulus epochs.

We analysed data in two regions of interest, V1 and FFG, defined using the MNI grey matter masks shown in Figure 2A. We excluded any voxels which overlapped in the low-resolution source localization grid space (10 millimetres resolution). The FFG mask included 1600 voxels and the V1 mask included 574 voxels.

In this paper, our *Y* variable contains class labels with a scalar +1 indicating post-stimulus, and a -1 indicating pre-stimulus. We have  $\nu=1$ . Therefore our CVA model has at most a single canonical component,  $h=1$ . Our *X* variable contains the *d*-dimensional power spectra. Thus, the number of parameters in the CVA model is  $k=d+1$ .

The matrix *Y* was prepared in the following way. Let each *j* index Fourier bins within one of the pre-defined spectral bands ranging from  $F^{lower}$  to  $F^{upper}$  frequency. Then at each source location *s* the activity at trial *n* is formulated as  $y_n = \sum_{j=F^{lower}}^{F^{upper}} (w_s B_{nj})(w_s B_{nj})^*$  based on equation 2, where \* signifies the complex conjugate. In order to give equal weighting to all frequency bands (some of which have markedly less power) for each band we removed the mean value (across all conditions) and normalized the variance (in power) to unity.

For each subject we computed the  $\log BF(d)$  maps for  $d=1,3,5,7,9$  and 11. Each  $\log BF(d)$  map is the log-evidence for a model with a single canonical vector, and *d* signal features, minus the log-evidence of the null model (no canonical vectors).

**Results**

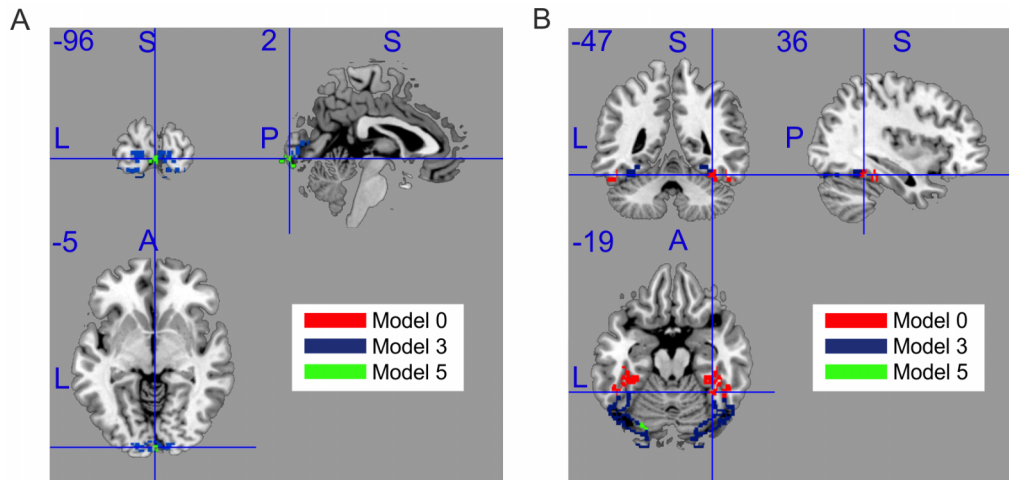
We studied 3–90 Hz oscillatory activity in pre- and post-visual stimulus presentation (an unfamiliar face) in source space, specifically in the primary visual cortex (V1) and fusiform gyrus (FFG). The aim of our analysis was to decode the differences in pre versus post stimulus activity, based on different feature sets (see Table 1 in the Methods and Material section). We hypothesised that there will be differences elicited by onset of the stimuli in the regions of interest. These regions were defined using the

**Table 2. Pairwise model comparisons versus the null.**

ROI	Model 1	Model 3	Model 5	Model 7	Model 9	Model 11
FFG (1600 voxels)	64 (4.0%)	615 (38.4%)	438 (27.3%)	239 (14.9%)	148 (9.2%)	103 (6.4%)
V1 (574 voxels)	32 (5.5%)	510 (88.8%)	466 (81.1%)	362 (63.0%)	205 (35.7%)	58 (10.1%)

Total (and percentage) of voxels with posterior expected frequencies  $\langle f_i | Y \rangle$  greater than 0.9. These are voxels where higher (than zeroth) order models are favoured in more than 90% of the population.

doi:10.1371/journal.pone.0071305.t002



**Figure 3. Multi-way model comparison maps.** The maps show voxels where pre- versus post-stimulus activity is best discriminated by each model, for V1 (left) and FFG (right). (A) V1 view from [2 -96 -5]mm (B) FFG view from [36 -47 -19]mm (Talairach coordinates). Models 1, 7, 9 and 11 were not best at any voxel. Note that the null model is best for anterior FFG, but model 3 is best for posterior FFG. doi:10.1371/journal.pone.0071305.g003

anatomical masks based on MNI brain (Figure 2A). In the first step, the MEG data for each individual participant was source localized and the features were defined as average power within the specified frequency bands. Figure 2B illustrates the average power across the 11 frequency bands (Model 11) in selected voxels from V1 and FFG. In the next step we studied which feature space best decoded the signals in each ROI based on the experimental design.

Firstly, we separately compared each of the models  $d = 1, 3, 5, 7, 9$  and 11 to the null model. Table 2 summarizes the results of these pairwise comparisons for each ROI. We compute the percentage of voxels with posterior expected frequencies greater than 0.9. For V1, the clearest result is that model 3 is better than the null in 88.8% voxels. For FFG, the findings are less clear cut, with both models 3 and 5 being better than the null in 38.4% and 27.3% voxels, respectively. These findings imply that the null model is better than the alternative models at a large proportion of voxels in FFG. We investigated this further with a set of multi-way model comparisons.

For the multi-way comparisons we first created a model-comparison map with voxels colored to show which model has the highest frequency  $\langle f_i | Y \rangle$  in the population. Figure 3 shows the results. Note that models 1, 7, 9 and 11 do not have the highest frequency at any voxel. The figure shows that model 3 is favoured in posterior FFG whereas the null model is favoured in anterior FFG. We can therefore infer that, in FFG, stimulus-induced changes in power spectra are restricted to posterior regions.

We also report summary statistics collapsed across all voxels in each region. Table 3, for example, shows the percentage of voxels in which the different models are favoured. For FFG, the null

model is favoured in half of the voxels and, as we have seen in Figure 3, these voxels are in the anterior region. For V1 model 3 is favoured in 90.8% voxels.

A subtlety with random effects model selection is that the expected frequencies depend on the models in the comparison set. For example, in a pairwise (two-way) comparison model 3 beats model 0 at 510 voxels (see Table 2). Whereas in a multi-way comparison model 3 wins at 521 voxels (Table 3). The 11 extra voxels reflect spectral differences that were previously attributed to model 0, but given a wider comparison set are attributed to model 5. This is analogous to voting in elections where the addition of an extra option can ‘split the vote’ [20].

Finally, we report expected model frequencies averaged over all voxels in each ROI. This is shown in Figure 4. Models 3 and 5 are, on average, the most frequently selected in V1, and models 0 and 3 are the most frequently selected in FFG.

## Discussion

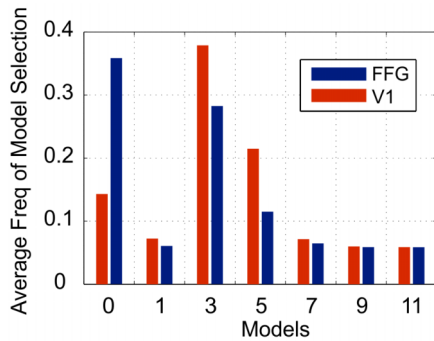
Here we proposed a solution to the problem of group-level inference from multivariate modelling of MEG data. The combination of scoring CVA models using BIC and assessing consistency across a group using Bayesian random effects model inference provides a principled solution. We applied this approach to source space power spectra in regions of interest to decode pre- versus post-stimulus epochs, using various feature sets (models). These feature sets differed in the degree to which the power spectra were fractionated (number of frequency bands). We were able to show that although all features sets provided some degree of discrimination between experimental conditions; the optimal

**Table 3. Multi-way between model comparison.**

ROI	Model 0	Model 1	Model 3	Model 5	Model 7	Model 9	Model 11
FFG (1600 voxels)	812 (50.7%)	0	769 (48.1%)	19 (1.2%)	0	0	0
V1 (574 voxels)	1 (0.1%)	0	521 (90.8%)	52 (9.1%)	0	0	0

Total (and percentage) of voxels where model  $i$  is the most favoured model (ie. has the largest posterior expected frequency  $\langle f_i | Y \rangle$ ).

doi:10.1371/journal.pone.0071305.t003



**Figure 4. Average frequency of model selection.** For each model we average the posterior expected frequencies  $\langle f_i | Y \rangle$  across all voxels in each ROI (FFG in blue and V1 in red). Models 3 and 5 are, on average, the most frequently selected in V1, and models 0 and 3 are the most frequently selected in FFG. doi:10.1371/journal.pone.0071305.g004

feature set in general consisted of three approximately classical bands (3–10, 10–30, 30–90 Hz). Although our features sets were selected simply to illustrate the new approach, it is clear that the multivariate models outperformed univariate models even for this simple decoding task.

An optimal statistical model provides a balance between model fit and model complexity [21] and this general principle naturally applies to the CVA models we have employed. In this paper model complexity is assessed by the number of model parameters (see equation 16 where  $k$  is the number of parameters). Thus, the univariate model (model 1) has low complexity ( $k=1$ ) but is suboptimal at most voxels because it has a poor model fit (it has a low correlation with the experimental design). Conversely, model 11 has a good model fit but is suboptimal because it has a large complexity ( $k=11$ ); the improvement in the canonical correlation is not justified by the larger number of parameters.

Our data analyses focussed on inferring the optimal feature dimension across various exemplar feature sets. But there are also other ways in which random effects Bayesian model selection can be used for finding the optimal feature set. These include, for instance, fixing the number of features, but changing the feature set (eg. by breaking up spectra in a different way, or using phase/amplitude or more exotic nonlinear features). In our study the multivariate features were power spectra at single source voxels in regions of interest. But one could also apply the approach to data from local regions of voxels as in ‘searchlight’ approaches [22].

A perhaps subtle aspect of the RFX-BMS approach is that it is concerned only as to whether, for example, more subjects use model A than model B. This does not require that the parameters of the winning model are consistent over that group. For example,

## References

- Newman EL, Norman KA (2010) Moderate excitation leads to weakening of perceptual representations. *Cerebral cortex* (New York, NY: 1991) 20: 2760–2770.
- Duncan KK, Hadjipapas A, Li S, Kourtzi Z, Bagshaw A, et al. (2010) Identifying spatially overlapping local cortical networks with MEG. *Human Brain Mapping* 31: 1003–1016.
- Fuentemilla L, Penny W, Cashdollar N, Bunzeck N, Duzel E (2010) Theta-coupled periodic replay in working memory. *Current Biology* 20: 1–7.
- Jafarpour A, Horner AJ, Fuentemilla L, Penny WD, Duzel E (2012) Decoding oscillatory representations and mechanisms in memory. *Neuropsychologia*.
- Zhou J (2008) Dimension reduction based on constrained canonical correlation and variable filtering. *The Annals of Statistics* 36: 1649–1668.
- Chalise P, Fridley B (2012) Comparison of penalty functions for sparse canonical correlation analysis. *Computational Statistics and Data Analysis* 56: 2452–254.
- Barnes G, Litvak V, Brookes M, Friston K (2011) Controlling false positive rates in massmultivariate tests for electromagnetic responses. *Neuroimage* 56: 10721081.
- Onton J, Makeig S (2006) Information-based modeling of event-related brain dynamics. *Progress in brain research* 159: 99–120.
- Stephan K, Penny W, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. *Neuroimage* 46: 1004–17.
- Robinson S, Vrba J (1999) Functional neuroimaging by synthetic aperture magnetometry (SAM). In: *Recent Advances in Biomagnetism*. Sendai, Japan: Tohoku University Press.
- Barnes GR, Hillebrand A (2003) Statistical flattening of MEG beamformer images. *Human brain mapping* 18: 1–12.
- Sekihara K, Nagarajan S, Poeppel D, Marantz A, Miyashita Y (2002) Application of an MEG eigenspace beamformer to reconstructing spatio-temporal activities of neural sources. *Hum Brain Mapp* 15: 199–215.

our model comparisons generally showed multiple voxels in which post-stimulus activity was better discriminated from pre-stimulus activity when the spectrum was described using a triplet (power in low, medium and high frequencies) rather than a scalar (power across all frequencies). This does not necessarily mean that the pattern of frequency responses was consistent over subjects. For example, half the subjects may have increases in low frequency power post-stimulus, and the other half decreases. One way to directly test for this scenario using the same scheme would be to see if a model using a fixed canonical vector over subjects (essentially a univariate test with data projected onto a single canonical vector) has more evidence than a model in which the canonical vector is allowed to vary (as here). Similar model comparison approaches could be used to test for differences (e.g. in the feature set) between different study groups, such as a patient group and a control group.

A large amount of neuroimaging research implements multivariate analysis using pattern recognition approaches based on artificial neural networks or support vector machines. A further benefit of the approach described in this paper is that model optimality is assessed using Bayes Factors, whereas the optimality of most pattern recognition approaches is assessed using cross validation [23]. Our approach is therefore more computationally efficient. For example, in the case that the assessment of model optimality is based on 10 fold cross-validation, our approach is 10 times faster.

We have applied our approach to beamformer reconstructed MEG data in source space. As noted in the introduction, it could also be applied to data in sensor space or data projected onto independent components [8]. As one of our goals has been to find data features that are consistent across a group of subjects, it would therefore also be necessary to use independent components that are consistent across the group. Fortunately, there are already established methods for doing this based on group-wise significance testing [24] or clustering [25].

This paper was based on MEG data; however, the combination of scoring CVA models and random effects model inference is not limited to MEG data and can be applied to any neuroimaging modality. In fMRI, for example, CVA can be applied to fMRI time series from a region of interest [26]. That said, the main advantage we see of this approach is that it provides principled population level inference on optimal feature space and dimension which could be particularly useful for data-rich neuroimaging techniques (like M/EEG).

## Author Contributions

Conceived and designed the experiments: LF ED. Performed the experiments: LF AJ. Analyzed the data: AJ GB WDP. Contributed reagents/materials/analysis tools: AJ GB WDP. Wrote the paper: AJ GB WDP. Designed the software used in analysis: GB WDP.

13. Penny W, Flandin G (2005) Bayesian analysis of fMRI data with spatial priors. In: Proceedings of the Joint Statistical Meeting (JSM). American Statistical Association.
14. Litvak V, Mattout J, Kiebel S, Phillips C, Henson R, et al. (2011) EEG and MEG data analysis in SPM8. *Comput Intell Neurosci* 2011: 852961.
15. Chatfield C, Collins AJ (1980) *Introduction to Multivariate Analysis*. Chapman and Hall, London.
16. Wang C (2007) Variational bayesian approach to canonical correlation analysis. *IEEE Trans Neur Networks* 18: 905910.
17. Friston KJ, Ashburner J, Kiebel SJ, Nichols TE, Penny WD, editors(2007) *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press.
18. Press W, Flannery B, Teukolsky S, Vetterling W (1991) *Numerical Recipes in C*. Cambridge University Press.
19. Rosa M, Bestmann S, Harrison L, Penny W (2010) Bayesian model selection maps for group studies. *Neuroimage* 49: 217–224.
20. Penny WD, Stephan KE, Daunizeau J, Rosa MJ, Friston KJ, et al. (2010) Comparing families of dynamic causal models. *PLoS Comput Biol* 6: e1000709.
21. Pitt MA, Myung IJ, Zhang S (2002) Toward a method of selecting among computational models of cognition. *Psychological review* 109: 472–491.
22. Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America* 103: 3863–3868.
23. Bishop CM (2006) *Pattern Recognition and Machine Learning*. Springer.
24. Hyvriinen A (2011) Testing the ICA mixing matrix based on inter-subject or inter-session consistency. *NeuroImage* 58: 122–136.
25. Spadone S, de Pasquale F, Mantini D, Della Penna S (2012) A k-means multivariate approach for clustering independent components from magnetoencephalographic data. *NeuroImage* 62: 1912–1923.
26. Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage* 19: 261–270.