Fusion of Magnetometer and Gradiometer Sensors of MEG in a Beamforming Framework

Hamid R. Mohseni*†, Mark W. Woolrich‡, Morten L. Kringelbach†§¶, Henry Luckhoo‡, Penny Probert Smith* and Tipu Z. Aziz§

*Institute of Biomedical Engineering, School of Engineering Science, University of Oxford, Oxford, UK

†Department of Psychiatry, University of Oxford, Warneford Hospital, UK

‡Oxford Centre for Human Brain Activity (OHBA), Department of Psychiatry, University of Oxford

§Oxford Functional Neurosurgery, Nuffield Department of Surgery, John Radcliffe Hospital, Oxford, UK

¶CFIN/MindLab, Aarhus University, Aarhus, Denmark

Abstract

Novel neuroimaging techniques have provided unprecedented information on the structure and function of the living human brain. Multimodal fusion of data from different sensors promises to radically improve this understanding, yet optimal methods have not yet been developed. Here, we demonstrate a novel method for combining multichannel signals. We show how this method can be used to fuse signals from the magnetometer and gradiometer sensors used in magnetoencephalography (MEG) and, through extensive experiments using simulation, head phantom and real MEG data, show that it is both robust and accurate. This new approach works by assuming that the lead-fields have multiplicative error. The criterion to estimate the error is given within the beamforming framework such that the estimated power is minimised in the worst case scenario. The method is compared to, and found better than, existing approaches. The closed-form solution and the conditions under which the multiplicative error can be optimally estimated are provided. This novel approach can also be employed for multimodal fusion of other multichannel signals such as MEG and EEG.

Index Terms

Magnetoencephalography, magnetometer, gradiometer, beamforming, multiplicative error.

I. Introduction

Magnetoencephalography (MEG) is a non-invasive neuroimaging technique that is showing great promise in increasing our understanding of the functional activity of the human brain. It offers excellent temporal resolution on the scale of milliseconds. However, the spatial resolution of MEG is lower than other prominent neuroimaging methods such as functional magnetic resonance imaging (fMRI), typically on the scale of around 5 mm^3, depending on both the location and the number of the MEG sensors.

The two fundamental classes of MEG sensors are *magnetometers* and *gradiometers*. Magnetometers measure the magnetic field component along the direction perpendicular to the surface of the sensor. While being very sensitive to nearby sources, the magnetometer is also sensitive to distant sources. The gradiometer is a more complex sensor which measures the spatial gradient rather than the magnitude of the field. This sensor type is less sensitive to interference located distant from the sensor, since the interference manifests itself as a homogeneous magnetic field with zero gradient, and hence also less sensitive to sources of interest located further away.

Gradiometer sensors can be arranged in two different ways to obtain different spatial derivatives of the field. One configuration, the *axial-gradiometer* arranges the sensor coils along the same radial axis. The other configuration, the *planar-gradiometer*, arranges the coils side-by-side in the same plane. The axial-gradiometer measures the sources around the sensors while the planar-gradiometer measures the sources right underneath the sensors (for more details of MEG and its sensor types, please refer to [1]).

The majority of available MEG scanners have one of these sensor types, while both sensors are included in some newer MEG scanners such as the Elekta-Neuromag MEG scanner, which is becoming the industry standard.

In the ideal situation, the gradiometers and magnetometers are measuring the same activity at the same time and should reveal the same neural activity. However, it is easy to show empirically that if the two sensor types are combined without any correction, the spatial resolution and specificity are not necessarily improved from measurements from just one type of sensor. An accurate and robust method for the fusion of magnetometer and gradiometer to improve their joint estimation is therefore an important goal in improving MEG data.

Previous research in multimodal fusion has concentrated primarily on combining one of the MEG sensor types with electroencephalography (EEG). Among these, simultaneous recording and simple combination of MEG and EEG have been well investigated [2]–[6]. In addition, there is a wide range of intelligent approaches for the fusion of MEG and EEG including using regularised linear inverse source estimation [7], independent component analysis [8], lead-field correction for dipolar sources [9], minimum l_2 norm estimation [10], mutual information [11], Bayesian estimation [12] and use of neural mass model [13]. In addition to these methods, a method was recently proposed for the fusion of EEG and two MEG sensors based on the normalisation of lead-fields and measurements [14]. This normalisation method normalises the data and the lead-fields based on their power before the source reconstruction algorithm. We compare the results from the proposed method with this previous method using data from real and phantom experiments in Section III.

Our novel method for the multimodal fusion of MEG sensors uses beamforming, a spatial filter that minimises the power of the signal while passing the activity from the location of interest [15]. We propose a beamformer based on a model in which there is a multiplicative error for one of the lead-fields. The closed-form solution of the problem is given using a Lagrange multiplier technique. Furthermore we show that if 1) the noise power tends toward zero, 2) the time courses of sources are independent, and 3) the columns of each lead-fields are mutually linearly independent, then the multiplicative error and the source covariance matrix can be exactly estimated. In this paper, the multiplicative error is estimated using a beamforming framework, but other methods for source analysis can equally be used after the lead-field modification.

The organisation of the paper is as follows. The first section states the problem formulation and the solution is presented using partitioning of the inverse of the joint covariance matrix of the magnetometer and gradiometer. A theorem is then presented revealing the assumptions for perfect estimation of the multiplicative error and covariance matrix of the source. This is followed by further theoretical discussions. Section III shows the results of simulation, phantom and real data experiments and compares these results with that of the previously published normalisation method. Finally, conclusions and future avenues for research are presented.

II. Methods

A. Problem Formulation and its Solution

Consider the following problem formulation:

$$\begin{aligned} y_g &= F_g E_g s + F_{gi} s_i + n_g \\ y_m &= F_m E_m s + F_{mi} s_i + n_m \end{aligned} \tag{1}$$

where $y_g \in \mathbb{R}^{N_g \times T}$ and $y_m \in \mathbb{R}^{N_m \times T}$ are the measurements using gradiometer and magnetometer sensors, respectively, with the associated and known lead-fields $F_g \in \mathbb{R}^{N_g \times D}$ and $F_m \in \mathbb{R}^{N_m \times D}$. Let $s \in \mathbb{R}^{D \times T}$ be the time course of the desired source and $s_i \in \mathbb{R}^{Dq \times T}$ be the time course of all other q sources which we refer to as interference. Here, D is the dimension of the source, and D_q is the sum of the dimensions of all interferences. $F_{gi} \in \mathbb{R}^{N_g \times Dq}$ and $F_{mi} \in \mathbb{R}^{N_m \times Dq}$ are the matrices that contain all lead-fields except the lead-field of the desired source. In this formulation, $E_g,\ E_m \in \mathbb{R}^{D \times D}$ are the multiplicative errors varying from location to location, and $n_g \in \mathbb{R}^{N_g \times T}$ and $n_m \in \mathbb{R}^{N_m \times T}$ are additive zero-mean Gaussian white noises.

Throughout the paper, it is assumed that y_g and y_m are stationary zero-mean processes with spatial covariance matrices R_{gg} and R_{mm} and joint covariance matrix R_y, which are symmetric and positive definite. Furthermore, we assume that these covariance matrices can be estimated using the measured data. Our aim is to provide an accurate estimation of the multiplicative errors. It is clear that the source powers may then be estimated using beamforming or any other source reconstruction method.

Note that E_g and E_m cannot be estimated simultaneously, because for a particular solution of s (or its covariance matrix), all its linear transformations are also solutions. Here, therefore, we suppose that either of E_g and E_m is known and without loss of generality suppose that $E_g = I$ and $E_m = E$, where I is the identity matrix. This approach leads to a computationally tractable, fast and closed-form solution. We will explore the impact of this assumption and the other case when $E_g = E$ and $E_m = I$ in Section III.

The estimation of the multiplicative error $\hat{E}$ can be found by considering the following modified linearly constraint minimum variance (LCMV) beamformer (for details of the LCMV please refer to [15]):

$$\begin{gathered}\arg\min_{W} \arg\max_{\hat{E}} Tr\{W^T R_y W\} \quad subject\ to: \\ W^T \begin{bmatrix} F_g \\ F_m \hat{E} \end{bmatrix} = I \end{gathered} \tag{2}$$

where $Tr\{.\}$ and $(.)^T$ are trace and transpose operators, respectively. The solution is a linear filter W^T that minimises the output power when it has been maximised by the multiplicative error $\hat{E}$. This optimisation problem ensures the output power is minimised when the error has the worst effect on it, and thus guarantees the performance of the beamformer with all range of the errors.

To find the solution, first suppose that $\hat{E}$ is known. Following methods used in beamforming [15], W can be estimated using the Lagrange multiplier method:

$$W^T = (\begin{bmatrix} F_g^T & \hat{E}^T F_m^T \end{bmatrix} R_y^{-1} \begin{bmatrix} F_g \\ F_m \hat{E} \end{bmatrix})^{-1} \begin{bmatrix} F_g^T & \hat{E}^T F_m^T \end{bmatrix} R_y^{-1} \tag{3}$$

By putting (3) into the constraints in (2), it is clear that this estimate of W^T always satisfies the constraint for any choice of $\hat{E}$. As a result, $\hat{E}$ can be estimated by only maximising the power $\mathcal{P} = Tr\{W^T R_y W\}$.

Using (3), and some algebraic manipulation, the power is expressed as:

$$\mathcal{P} = Tr\{\left(\begin{bmatrix} F_g^T & \hat{E}^T F_m^T \end{bmatrix} R_y^{-1} \begin{bmatrix} F_g \\ F_m \hat{E} \end{bmatrix}\right)^{-1}\} \tag{4}$$

As an example, the behaviour of $\mathcal{P}$ with respect to $\hat{E}$ has been plotted in Fig. 1. In this figure, $D = 1$ and the lead-fields and the covariance matrices have been randomly selected. This figure shows that all plots are convex and have one global maximum.

In equation 4, if F_g is full column rank, then it is straightforward to show that $\begin{bmatrix} F_g \\ F_m \hat{E} \end{bmatrix}$ is also full column rank regardless of $\hat{E}$ and F_m. As we already assumed that R_y is positive definite and thus invertible, the expression inside of the above bracket is also invertible and therefore exists.

By partitioning the inverse of joint covariance matrix into $R_y^{-1} = \begin{bmatrix} R_{gg}^{-} & R_{gm}^{-} \\ R_{mg}^{-} & R_{mm}^{-} \end{bmatrix}$, the first line of (2) using (4) is expressed as:

$$\arg\max_{\hat{E}} \mathcal{P} = \arg\max_{\hat{E}} Tr\{\left(\hat{E}^T F_m^T R_{mm}^{-} F_m \hat{E} + \hat{E}^T F_m^T R_{mg}^{-} F_g + F_g^T R_{gm}^{-} F_m \hat{E} + F_g^T R_{gg}^{-} F_g\right)^{-1}\} \tag{5}$$

The derivative of the above expression with respect to $\hat{E}$ is equal to (see Appendix A):

$$\begin{aligned} \frac{\partial \mathcal{P}}{\partial \hat{E}} = &-\left(2\hat{E}^T F_m^T R_{mm}^{-} F_m + (F_m^T R_{mg}^{-} F_g)^T + F_g^T R_{gm}^{-} F_m\right) \\ &\left(\hat{E}^T F_m^T R_{mm}^{-} F_m \hat{E} + \hat{E}^T F_m^T R_{mg}^{-} F_g + F_g^T R_{gm}^{-} F_m \hat{E} + F_g^T R_{gg}^{-} F_g\right)^{-2} \end{aligned} \tag{6}$$

where $X^{-2} = X^{-1} X^{-1}$. Since the expression inside the second bracket is invertible, its inverse is full rank and the above expression is only zero if the expression inside the first bracket is zero. Hence, the problem in (4) is convex and the global maximum, using the fact that $F_m^T R_{mg}^{-} F_g = (F_g^T R_{gm}^{-} F_m)^T$, is given by:

$$\hat{E} = -(F_m^T R_{mm}^{-} F_m)^{-1} (F_m^T R_{mg}^{-} F_g) \tag{7}$$

Finally, using block matrix inversion [16], R_{mm}^{-} and R_{mg}^{-} are given by:

$$\begin{aligned} R_{mm}^{-} &= (R_{mm} - R_{mg} R_{gg}^{-1} R_{gm})^{-1} \\ R_{mg}^{-} &= -R_{mm}^{-1} R_{mg} (R_g g - R_{gm} R_{mm}^{-1} R_{mg})^{-1} \end{aligned} \tag{8}$$

where R_{mm} is the covariance matrix of the magnetometer and $R_{mg} = R_{gm}^T$ is the cross-covariance matrix between magnetometer and gradiometer. Equation (7) also can be rewritten as:

$$\hat{E} = -(F_{om}^T R_y^{-1} F_{om})^{-1} (F_{om}^T R_y^{-1} F_{go}) \tag{9}$$

where $F_{om} = \begin{bmatrix} 0 \\ F_m \end{bmatrix}$ and $F_{go} = \begin{bmatrix} F_g \\ 0 \end{bmatrix}$. In the beamformer, the inverse of the joint covariance matrix R_y^{-1} is required and should be estimated, thus $\hat{E}$ can be calculated faster using (9) than (7). In other applications, if there is no estimation of the joint covariance matrix R_y^{-1}, estimation of $\hat{E}$ equation (7) using (8) is faster than (9).

B. *Performance Analysis*

In equation (7), $(F_m^T R_{mm}^{-} F_m)^{-1}$ can be considered as the covariance matrix of the source using just the magnetometer data, and $F_m^T R_{mg}^{-} F_g$ can be considered as the spatiotemporal correlation between the two sensor types. Hence, the method normalises the lead-field using the power of the source estimated by the magnetometer only and its correlation with the source estimated by the gradiometer. It is also clear from (7) and (8) that $\hat{E}$ is proportional to the cross-covariance matrix of the magnetometer and gradiometer measurements R_{mg}. This implies that if there is no correlation between two sensor types $R_{mg} = 0$, then $\hat{E} = 0$ and the source reconstruction result is based only on the gradiometer which was assumed to have the correct lead-field. This is reasonable based on the proposed model, because if the two modalities show completely different reconstructions, one or none of them has to be selected as the correct reconstruction based on some prior information.

In general, it may be useful to know the conditions for optimality of the method, which can help us to evaluate if the method is suitable for a given data-set. The following theorem states when the error E and the source covariance matrix C_s can be perfectly estimated. However, we shall see in the results that, in practise, this choice of the correct modality is not critical.

Theorem 1: Suppose i) the column of F_g are linearly independent from column of F_{gi} as well as F_m from F_{mi}, ii) the time course of source s is statistically independent from the interferences s_i, and iii) the noises n_g and n_m are zero-mean white Gaussian and with power tending to zero.

Under the above conditions, we can infer that $\hat{E} = E$ using (7) or (9), and $\hat{C}_s = C_s$ using $\hat{C}_s = W^T R_y W$.

Proof: See Appendix B.

Remark 1: Not practically but theoretically important, any zero-mean noise with covariance matrix C_n, where $|C_n| \to 0$, can be used instead of white Gaussian noise in *Theorem 1* ($|C_n|$ is the Frobenius norm of C_n). This can be shown by modifying to the proof of *Theorem 1*.

Corollary 1: If in equation (1) $E_g = E$ and $E_m = I$ (contrary to what we have assumed already), but one uses equation (7) or (9) to estimate E, then under the conditions stated in *Theorem 1*, $\hat{E} = E^{-1}$ and $C_s = \hat{E}^T \hat{C}_s \hat{E}$.

Proof: Trivial from *Theorem 1*.

For a perfect reconstruction, conditions (i) and (ii) in *Theorem 1* are necessary for any source analysis method. In other words, if the sources are correlated or the source and the interference lead-field matrices are dependent, no algorithm can perfectly estimate the multiplicative error and reconstruct the sources. Furthermore, it can be concluded from *Theorem 1* and *Corollary 1* that for moderately noisy data, selecting the modality which has the correct lead-field does not have much impact on the results and always s or a linear transformation of s is estimated. However, in a noisy environment, selecting a modality which has the correct lead-field is important for source estimation.

In some applications, we may need to fuse three or more sensor types. For instance, one may consider magnetometer, planar- and axial-gradiometer as three different sensors. Fusion of EEG and MEG gradiometer and magnetometer sensors would be another case—similar to the approach in [14]. For the convenience of the reader, we have provided the solution of this generalised problem in Appendix C.

III. Method Validation and Discussion

A. Computer Simulation Experiment

For the first experiment, MEG data with Gaussian sources were simulated. The number of sources were set to 15 and the number of sensors for each sensor type to 102. The sources were non-correlated and their lead-fields were chosen randomly in a Monte Carlo like simulation; i.e. the location of sensors and sources were chosen randomly. The multiplicative error $E_m = E$ was also chosen randomly in each simulation according to a uniform random distribution in the range of 0 to 1, and we set $E_g = I$. Gaussian white noise was added to the signal after applying the lead-field to the time series of the sources. The following results were produced as the average of 1000 Monte Carlo simulations. The error also is defined as the mean squared error between original and estimated time course which were normalised by their power.

Fig.2 (a) shows the error of the three methods versus signal to noise ratio (SNR). SNR is defined in the sensor space as the ratio of the mean power of the signal of interest (applying the lead-field only to the source of interest) to the mean power of the added noise across the sensors. In this figure, the 'all' method means when two sensors were used without any correction (assuming both $E_g = I$ and $E_m = I$), ‘grad’ means when only gradiometer was used, and ’fuse-grad’ means when the proposed method were used with the gradiometer having the correct lead-field—Table I provides a reference to the different methods used in the following experiments. We have not presented the results of the magnetometer, as it clearly gives worse results compared to the gradiometer which has a perfect lead-field.

Fig.2 (a) demonstrates that the ’fuse-grad’ method has the best performance compared to other methods in various ranges of SNRs. Note that the sensitivity of the three methods to the additive Gaussian noise is almost similar.

Fig. 2(b) shows the error in the three methods versus signal to interference ratio (SIR). SIR is defined as the ratio of the mean power of the desired source to the mean power of the other interfering sources (after applying the lead-field to their time series). Interference has no impact on the accuracy of the ’fuse-grad’ method. The reason is that 102 sensors are enough to estimate the power of the desired source as well as to cancel out the other sources. However, the ‘grad’ method is struggling to reject the interference in the presence of the noise. Furthermore, because of the presence of multiplicative error, the ‘all’ method shows worse performance compared to the others.

Next we investiagte the impact of the assumption that which modality has the correct lead-field. We set $Eg = 1$ and assume that $E_m = E$ which is uniformly increasing. In this particular simulation to be able to plot E_g, scalar beamformer was used ($D = 1$). The results from ‘fuse-grad’ and ‘fuse-mag’ (assume that magnetometer has the correct lead-field while the contrary is true), ’mag’, ’grad’ and ’all’ methods are presented in Fig. 3. As we mentioned before, in moderate noise environment this assumption has not large impact on the results (ignoring the amplitude of the reconstructed time courses). Comparing to when only one sensor type is used, the ‘fuse’ methods well outperform the other methods which are using only one sensor types.

Fig. 4 presents the error versus the number of sensors. The ’fuse-grad’ method has a better performance compared to the other methods. According to *Theorem 1* the lead-field of the sources should be mutually independent to perfectly reconstruct the sources. In other words, the number of sensors should be greater than the total number of lead-fields columns, which means at least 45 sensors for 15 sources is needed. When the number of sensors are greater than 100, both methods have similar and perfect performance. This suggests that, for this example, approximately 100 sensors are enough to reconstruct one source within 15 other interferences and more sensors may be needed if the data are noisier. Moreover, the error of the ‘all’ method is larger than the other two methods since the multiplicative error has a big impact on the results.

It may be concluded that if the number of sensors is sufficiently large compared to the number of sources, one modality would be enough to reconstruct the sources. In human brain imaging, however, the number of sources are much larger than the number of sensors, and the proposed fusion method should therefore improve source reconstruction.

B. Phantom Experiment

In addition to the simulated data, the Elekta-Neuromag phantom, designed to calibrate the MEG scanner, was imaged to validate the method. The phantom contains 32 dipoles with fixed locations that produce two cycles of a sine wave with frequency of 20Hz. The phantom is shown in Fig. 5(a) and its dipole locations are plotted in Fig. 5(b). In this phantom, only one dipole can be active at a time. The onset of the sine waves were recorded and the data were then low-pass filtered with cut-off frequency of 40Hz. MaxFilter which is a commercial software and recommended by the Elekta company were applied to the continiuos data [17]. After epoching, the base-line was corrected using 100ms segment of the trial before the onset of the sine wave.

We compared the results of the new 'fuse' method with the 'normalisation' method for fusion of MEG and EEG data proposed by Henson et al. [14]. In their method the lead-field of the matrix and measurements for each modality was normalised by the quantities: $\sqrt{\frac{1}{N}Tr\{F^TF\}}$ and $\sqrt{\frac{1}{T}Tr\{y^Ty\}}$, respectively. This method places approximately the same weighing on each modality rather than weighting them according to their multiplicative error.

Table I: Methods used in the phantom and real experiments:

'mag'	only magnetometer were used
'grad'	only gradiometer were used
'fuse-grad'	fusion method that assumes gradiometer has the correct lead-field
'fuse-mag'	fusion method that assumes magnetometer has the correct lead-field
'normalisation'	previous method proposed by Henson et al [14]
'all'	both magnetometer and gradiometer without any correction were used

The phantom data has high SNR, and there was little difference in the source localisation between the different methods. Therefore, the success of each method is judged by comparing the true time series at the source with the estimation from the measurements. The results are presented in Fig. 6. The result labelled 'fuse-grad' and 'fuse-mag' are the results of the 'fuse' methods with a correct lead field assumed for gradiometer and magnetometer respectively. The results labelled 'grad' and 'mag' show the results when only one sensor type was used: gradiometer and magnetometer, respectively. The result labelled 'normalisation' shows Henson's method [14] and 'all' is when both sensors without any correction are used.

Figs. 6(a) and (b) indicate that the fusion methods have estimated more accurately the time series in comparison to using only one sensor modality. Fig. 6(c) shows that the normalisation methods is better than the 'all' method but none of them outperforms the 'fuse-grad' and 'fuse-mag' methods. It is also notable that mean (variance) of the trace of estimated E_m in 'fuse-grad' was 0.069 (0.0052) and estimated E_g in 'mag-grad' was 0.234 (0.4506). This means that the gradiometer was assigned bigger weight compared to the magnetometer.

The range of errors for each modality was plotted in Fig. 7. The error of the fusion methods are significantly smaller compared to those using only one sensor type ($F = 17.3, p < 0.001$)[1]. There is no significant differences between the two methods of fusion ($F = 1.2, p > 0.1$). It is also clear that the normalisation method did not significantly improved the results compared to the single sensor methods ($F = 3.08, p > 0.05$) and that the 'all' method actually had a significantly larger error compared to other methods ($F = 134.23, p < 0.001$). Hence, the phantom experiment confirms the advantages of the new method in the cases where SNR is high.

In order to investigate the spatial resolution, we also calculate the Full Width Half Maximum (FWHM). The FWHM is defined by fitting a Gaussian function to the 1-D profile (taken in the x-direction, i.e. left to right, passing through the estimated dipole location) of the corresponding spatial map and setting FWHM $= 2.35\sigma_{FWHM}$, where σ_{FWHM} is the standard deviation of the fitted Gaussian function. The results are shown in Fig. 8 in which 'fuse-grad' and 'fuse-mag' clearly show less FWHM in each individual trial and therefore have better spatial resolutions at the same time as having increased temporal accuracy as demonstrated in Fig. 7.

C. Real Data Experiment

In this section, we compare the results of the methods using MEG data obtained from a face recognition experiment designed to locate deep sources in the brain. This is clearly a much more challenging environment than the simulation, with potential for multiple sources and low SNR.

Lesion studies, single-unit recordings and neuroimaging techniques including positron emission tomography (PET), fMRI and MEG have implicated a distributed network of brain regions in decoding face stimuli. Some studies suggest that a specific region of the brain, the fusiform face area (FFA), is the primary face processing region in the human brain [18]–[20]. The FFA is a part of the human visual system and located within the fusiform cortex, just above the cerebellum. Various studies using MEG

[1] F-statistic is the ratio of the mean squares for error, and the p value is derived from the cdf of F. MATLAB R2010b were used to calculate this analysis using `anova1` function.

and EEG have demonstrated that face-specific components in the brain signals peaks at approximately 170ms after presentation the face [21], [22]. This face-specific components has become known as the M170 in MEG studies and appears to be significantly earlier and larger for faces than other objects in the FFA.

In this experiment a series of animal and human faces were presented to the participants. Each image were presented for 300 ms and the time interval between images were 1500 ms. Intra-individual head movement was kept to a minimum, and head position was localised immediately before the start of the experiment. The sampling rate was 1Khz and after linearly filtering data in the range of 1–40Hz the recorded data were epoched and averaged for animal and human faces. Same as the phantom experiment the MaxFilter were used to further denoise the data. The trials were visually inspected and few of the trials with large variations were removed. The difference between MEG signals for human and animal faces from 150ms to 190ms after stimuli were used to estimate the covariance matrix. The covariance matrix was diagonally regularised with 5% of its trace. A single layer realistic head model was used in which the brain was divided into a number of cells. The distance between adjacent cells was 5mm. The power of the beamformer at each cell were computed and normalised with those of the noise estimated from the pre-stimulus trials.

Fig. 9 shows the output power of beamformer normalised by norm of the associated lead-fields using different methods; i.e. normalised by the power of projected Gaussian white noise (see [15]). The reconstructed activity was thresholded using the normalised threshold of 0.9. The same anatomical plane is displayed for each method. In all the figures the solid white volume is the mask representing the full fusifom cortex region depicted using the automated anatomical labelling (AAL) atlas, within which the FFA can be found [23].

Fig. 9(a) shows the results using only the gradiometer measurements. The peak power spectrum is close to the FFA area but it is biased towards the superficial source. Fig. 9(b) shows the results using only magnetometers. The results were improved with larger overlap with the white volume, in line with the magnetometer's greater sensitivity to deep sources such as the FFA.

Figs. 9(c) and (d) show the results from the fusion methods with true lead-field in gradiometer and magnetometer, respectively. Both methods show better performance compared to the results in Figs. 9(a) and (b), seen from the greater overlap between the source and FFA, but show some difference in the source activity detected in the visual cortex (Fig. 9(c) axial view). It is notable that there is similarity between the results of 'fuse-grad' and 'grad' only, and between 'mag-fuse' and 'mag' method. The results suggest that to reconstruct deep sources, a choice of magnetometer as the base would seem more appropriate, whereas for superficial sources the gradiometer is better. Another issue for choosing a modality as the base is the amount of noise present. In a noisy environment, using the gradiometer as the base can be expected to outperform the magnetometer.

Table II shows the coordinates in the standard MNI (Montreal Neurological Institute) space of the peak of the power spectrum reconstructed using the various methods. In this table, the 'peak to peak error' is the distance of the peak of each power spectrum from the coordinate of the FFA ($x = 40$, $y = -55$, $z = -10$ [19]). It should be noted that the maximal peak of this activity is not the best indicator of the localisation of face responses given that they can be found within a larger area in the fusiform cortex with considerable variability between people. We therefore also include another measure, 'Correlation with the fusiform mask' which is the correlation of top 10 percentile of the power spectrum with the full fusiform cortex mask [23]. Furthermore, this table included the mean (variance) of the estimated multiplicative errors across the voxels. 'Fusion' methods have bigger variances compared to the 'normalisation' method means they are adaptive method in terms of assigning weight to lead-field of different voxels.

Fig. 9(e) shows the results using the 'normalisation' method, which weights both measurement sets such that to have a same lead-field power. It can be seen that although the reconstructed activity has overlap with the FFA, it has a lower resolution than the other methods and the peak is displaced to the visual cortex away from the FFA. Finally, Fig. 9(f) shows the results when using the two sensors with no correction. This method is the worst, showing no activity within FFA and is thus the worst result. As in the other experiments, the new fusion methods are seen to outperform the others.

To gain further insight, Fig. 10 shows the power of the beamformer weights for each sensor. By partitioning the weight vector in equation (3) as $W^T = [W_g^T \quad W_m^T]$, the power assigned to the gradiometer and the magnetometer, W_g and W_m, can be determined. The left column of Fig. 10 is the weight power of the gradiometer and the right column is the weight power of the magnetometer.

Fig. 10(a) and (b) show the results from 'grad' and 'mag', respectively. The results are similar to the reconstructed powers. Fig. 10(c) and (d) show the results from the 'fuse-grad' method. Since it is assumed that the gradiometer has the correct lead-field, it generally assigns larger weights to the gradiometer compared to the magnetometer. Similar results are seen for the 'mag-fuse' method in Figs. 10(e) and (f), in which larger weights are assigned to the magnetometer. Figs. 10(g) and (h) show the contribution of magnetometers and gradiometers using the 'normalisation' method. This figure demonstrates that the gradiometers have the same weights for all locations, but the magnetometers have greater weights around the FFA for which their measurements are relatively larger. Figs. 10(i) and (j) show the results for the 'all' method showing that the gradiometer generally have very small weights compared to the magnetometer, and hence a small contribution to the final results. Roughly speaking, these figures show the contribution of each sensor type in the final results which depends simply on magnitude in the 'all' method, on the average magnitude in the 'normalisation method' and on the explicit estimation of multiplicative errors in the new fusion method.

Table II: Coordinates and errors (in mm) of the reconstructed FFA

method	**MNI coordinates** x	y	z	**peak to peak error**	**correlation with the fusiform mask**	**estimated E_g and E_m**
'grad'	20	-48	-20	23.4	0.22	$E_g = 1$, $E_m = 0$
'mag'	32	-86	-14	32.3	0.40	$E_g = 0$, $E_m = 1$
'fuse-grad'	34	-40	-6	16.6	0.56	$E_g = 1$, $E_m = -0.0439(0.001)$
'fuse-mag'	40	-42	-2	15.3	0.65	$E_g = -0.2316(0.040)$, $E_m = 1$
'normalisation'	2	-90	2	53.0	0.12	$E_g = 0.0159(1.2e-4)$, $E_m = 0.0459(4.6e-4)$
'all'	-2	-100	2	62.7	0	$E_g = 1$, $E_m = 1$

IV. Conclusions

We have proposed a novel method for the multimodal fusion of magnetometer and gradiometer in MEG imaging. The method is formulated within the beamforming framework, and its closed-form solution is presented through partitioning the inverse of the joint covariance matrix. The conditions for optimality of the method have also been presented, and illustrate that the method does not introduce any new assumptions to those underlying the beamformer framework. Simulation, phantom and real experiments show that combining the novel method with beamforming increases the reliability of source localisation compared to other methods. The implementation of the proposed method is straightforward, and with the beamforming source reconstruction does improve the results.

In forthcoming studies, we will further validate the method in other basic experiments such as sensory-motor and auditory paradigms in a group of participants. The method, therefore, promises to be useful when employed for psychology and neuroscience experiments investigating the spatial and temporal sources of brain activity in different paradigms.

We are also interested in the estimation of the multiplicative error for correction of the lead-field within other source reconstruction techniques (e.g. minimum norm estimation). Moreover, the method can be used for optimally combining information from simultaneous EEG and MEG signals and thus become a viable strategy for further enhancing our understanding of human brain activity.

APPENDIX A

Here, we show that

$$\frac{d}{dX}Tr\{(X^TAX+BX+C)^{-1}\} = -(2X^TA+B)(X^TAX+BX+C)^{-2} \tag{10}$$

suppose x_{ij} is the entry of matrix X that lies in the ith row and the jth column, and e_i and e_j are the ith and jth standard bases. Then, using the facts that $d(Tr\{X\}) = Tr\{dX\}$, $dX/dx_{ij} = e_ie_j^T$ and $d(X^{-1}) = X^{-1}(dX)X^{-1}$ (see [16]), we have

$$\begin{aligned}
&\frac{d}{dx_{ij}}Tr\{(X^TAX+BX+C)^{-1}\} \\
&= -Tr\{(X^TAX+BX+C)^{-1}(2X^TAe_ie_j^T+Be_ie_j^T)(X^TAX+BX+C)^{-1}\} \\
&= -Tr\{e_j^T(2X^TA+B)(X^TAX+BX+C)^{-2}e_i^T\} = -[(2X^TA+B)(X^TAX+BX+C)^{-2}]_{ij}
\end{aligned} \tag{11}$$

which results equation (10).

APPENDIX B

PROOF OF *Theorem 1*

First note that the covariance and cross-covariance matrices are given using (1) by:

$$\begin{aligned}
R_{gg} &= F_gC_sF_g^T + F_{gi}C_iF_{gi}^T + \sigma_gI \\
R_{mm} &= F_mEC_sE^TF_m^T + F_{mi}C_iF_{mi}^T + \sigma_mI \\
R_{gm} &= R_{mg}^T = F_gC_sE^TF_m^T + F_{gi}C_iE^TF_{mi}^T
\end{aligned} \tag{12}$$

Lemma 1: (A General Singular Value Decomposition) If we have F and F_i, then there exist unitary matrices U and V, and invertible X such that:

$$\begin{aligned}
XFV &= \begin{bmatrix} \Gamma \\ 0 \end{bmatrix} \\
XF_iU &= \begin{bmatrix} 0 \\ \Sigma \end{bmatrix}
\end{aligned} \tag{13}$$

where Γ and S are diagonal matrices.

proof: This is a special case of the general singular value decomposition given in [24].

Lemma 2: If the column vectors of F and F_i are independent and $C \in \mathbb{R}^{D\times D}$ is a full rank matrix, then:

$$\lim_{\sigma\to 0} F^T(FCF^T + F_iC_iF_i^T + \sigma I)^{-1}F \approx (C+\sigma H)^{-1} \tag{14}$$

$$\lim_{\sigma\to 0} F^T(FCF^T + F_iC_iF_i^T + \sigma I)^{-1}F \approx F^T(FCF^T + \sigma I)^{-1}F \tag{15}$$

$$\lim_{\sigma\to 0} F_i^T(FCF^T + \sigma I)^{-1}F = 0 \tag{16}$$

where H is a square matrix independent from σ and C. Also, Γ and V are diagonal and unitary matrices, respectively.

proof: Suppose $F = X^{-1}\begin{bmatrix}\Gamma \\ 0\end{bmatrix}V^T$ and $F_i = X^{-1}\begin{bmatrix}0 \\ \Sigma\end{bmatrix}U^T$ according to *lemma 1*, and let $XX^T = \begin{bmatrix}X_{11}^2 & X_{12}^2 \\ X_{21}^2 & X_{22}^2\end{bmatrix}$. Since F and F_i are independent, the left hand side of (14) is expressed as:

$$\begin{aligned}
&\lim_{\sigma\to 0} V[\Gamma \quad 0](X^{-1})^T(X^{-1}\begin{bmatrix}\Gamma VCV^T\Gamma + \sigma X_{11}^2 & \sigma X_{12}^2 \\ \sigma X_{21}^2 & \Sigma UC_iU^T\Sigma + \sigma X_{22}^2\end{bmatrix}(X^{-1})^T)^{-1}X^{-1}\begin{bmatrix}\Gamma \\ 0\end{bmatrix}V^T \\
&= \lim_{\sigma\to 0} V[\Gamma \quad 0]\begin{bmatrix}\Gamma VCV^T\Gamma + \sigma X_{11}^2 & \sigma X_{12}^2 \\ \sigma X_{21}^2 & \Sigma UC_iU^T\Sigma + \sigma X_{22}^2\end{bmatrix}^{-1}\begin{bmatrix}\Gamma \\ 0\end{bmatrix}V^T \\
&= \lim_{\sigma\to 0} V\Gamma(\Gamma VCV^T\Gamma + \sigma X_{11}^2 - \sigma^2 X_{12}^2(\Sigma UC_iU^T\Sigma + \sigma X_{22}^2)^{-1}X_{21}^2)^{-1}\Gamma V^T \\
&\approx \lim_{\sigma\to 0} V\Gamma(\Gamma VCV^T\Gamma + \sigma X_{11}^2)^{-1}\Gamma V^T = \lim_{\sigma\to 0}(C + \sigma V^T\Gamma X_{11}^2\Gamma V)^{-1}
\end{aligned} \tag{17}$$

in which we used block matrix inversion to convert the second line into the third line of the above equation. By defining $H = V\Gamma^{-1}X_{11}^2\Gamma^{-1}V^T$ and considering V and Γ are invertible, equation (14) is obtained. Equations (15) and (16) can also be proved in a similar way.

To prove that $\hat{E}$ is equal E when the noise power tends to zero, first suppose E is invertible. We express (7) using (12) as:

$$\hat{E} = (F_m^T(R_{mm} - R_{mg}R_{gg}^{-1}R_{gm})^{-1}F_m)^{-1}F_m^TR_{mm}^{-1}R_{mg}(R_{gg} - R_{gm}R_{mm}^{-1}R_{mg}^{-1})F_g \tag{18}$$

The middle term $F_m^T R_{mm}^{-1} R_{mg}$ using (14)–(16) and $\sigma_m \to 0$ can be rewritten as:

$$\begin{aligned} F_m^T R_{mm}^{-1} R_{mg} &= F_m^T (F_m E C_s E^T F_m^T + F_{mi} C_i F_{mi}^T + \sigma_m I)^{-1} (F_m E C_s F_g^T + F_{mi} E C_i F_{gi}^T) \\ &= (E C_s E^T)^{-1} E C_s F_g^T = (E^{-1})^T F_g^T \end{aligned} \tag{19}$$

By putting (19) into (18), $\hat{E}$ is expressed as:

$$\hat{E} = (F_m^T (R_{mm} - R_{mg} R_{gg}^{-1} R_{gm})^{-1} F_m)^{-1} (E^{-1})^T F_g^T (R_{gg} - R_{gm} R_{mm}^{-1} R_{mg}^{-1}) F_g \tag{20}$$

The first term of the above equation using (14)–(16) is rewriten as:

$$\begin{aligned} &(F_m^T (R_{mm} - R_{mg} R_{gg}^{-1} R_{gm})^{-1} F_m)^{-1} = \Big(F_m \{ F_m E C_s E^T F_m^T + F_{mi} C_i F_{mi}^T + \sigma_m I \\ &- (F_m E C_s F_g^T + F_{mi} C_i F_{gi}^T)(F_g C_s F_g^T + F_{gi} C_i F_{gi}^T + \sigma_g I)^{-1} (F_g C_s E^T F_m^T + F_{gi} C_i F_{mi}^T) \} F_m \Big)^{-1} \\ &= \Big(F_m \{ F_m E C_s E^T F_m^T + \sigma_m I - F_m E C_s F_g^T (F_g C_s F_g^T + \sigma_g I)^{-1} F_g C_s E^T F_m^T \}^{-1} F_m \Big)^{-1} \end{aligned} \tag{21}$$

By approximating $F_g^T (F_g C_s F^T + \sigma_g I)^{-1} F_g \approx (C_s + \sigma_g H_g)^{-1}$ using (14), the above equation is equivalent to:

$$\begin{aligned} &(F_m^T (R_{mm} - R_{mg} R_{gg}^{-1} R_{gm})^{-1} F_m)^{-1} \\ &\approx (F_m^T \{ F_m E C_s E^T F_m^T + \sigma_m I - F_m E C_s (C_s + \sigma_g H_g)^{-1} C_s E^T F_m^T \}^{-1} F_m)^{-1} \\ &= E \{ C_s - C_s (C_s + \sigma_g H_g)^{-1} C_s + \sigma_m E^{-1} H_m (E^{-1})^T \} E^T \end{aligned} \tag{22}$$

The last line of the above equation was obtained by rearranging the second line and applying (14) (H was replaced by H_m). Using Woodbury matrix identity $(A+B)^{-1} = A^{-1} - A^{-1}(A^{-1} + B^{-1})A^{-1}$, (22) is simplified and re-expressed as

$$\begin{aligned} &E \{ (C_s^{-1} + (\sigma_g H_g)^{-1})^{-1} + \sigma_m E^{-1} H_m (E^{-1})^T \} E^T \\ &= E \{ \sigma_g H_g - \sigma_g^2 H_g (C_s + \sigma_g H_g)^{-1} H_g + \sigma_m E^{-1} H_m (E^{-1})^T \} E^T \end{aligned} \tag{23}$$

Using (22) and (23) for $\sigma_g \to 0$, we have:

$$(F_m^T (R_{mm} - R_{mg} R_{gg}^{-1} R_{gm})^{-1} F_m)^{-1} \approx E (\sigma_g H_g + \sigma_m E^{-1} H_m (E^{-1})^T) E^T \tag{24}$$

Similarly, for $\sigma_m \to 0$, we can show that:

$$F_g^T (R_{gg} - R_{gm} R_{mm}^{-1} R_{gm})^{-1} F_g \approx (\sigma_g H_g + \sigma_m E^{-1} H_m (E^{-1})^T)^{-1} \tag{25}$$

By putting (24) and (25) into (20), the desired result $\hat{E} = E$ is obtained. The case where E is not invertible one may consider $E + \delta I$, and after $\sigma_g, \sigma_m \to 0$ tends δ to zero.

Showing the second part of the theorem stating that the covariance matrix of the source can be perfectly estimated is straightforward. By using E instead of $\hat{E}$ and defining $F_a = \begin{bmatrix} F_g \\ F_m E \end{bmatrix}$ and $F_{ai} = \begin{bmatrix} F_{gi} \\ F_{mi} \end{bmatrix}$, and using equation (1), R_y is given by:

$$R_y = F_a C F_a + F_{ai} C_i F_{ai} + \begin{bmatrix} \sigma_g I & 0 \\ 0 & \sigma_m I \end{bmatrix} \tag{26}$$

Using (4) the covariance matrix of the source is expressed as:

$$\hat{C}_s = W^T R_y W = (F_a^T (F_a C F_a^T + F_{ai} C_i F_{ai}^T + \begin{bmatrix} \sigma_g I & 0 \\ 0 & \sigma_m I \end{bmatrix})^{-1} F_a)^{-1} \tag{27}$$

Similar to proof of (14), it is clear that $\hat{C}_s = C$ when $\sigma_m, \sigma_g \to 0$.

APPENDIX C

In this appendix, we present the generalised solution of the problem in equation (1) which is stated as:

$$\begin{aligned} y_g &= F_g s + F_{gi} s_i + n_g \\ y_m &= F_m E_m s + F_{mi} s_i + n_m \\ y_e &= F_e E_e s + F_{ei} s_i + n_e \end{aligned} \tag{28}$$

Here, $F_e \in \mathbb{R}^{N_e \times D}$ is the forward matrix of the third sensor (e.g. EEG) and n_e is the additive white Gaussian noise. By writing the constraint of equation (2) for this problem as $W^T \begin{bmatrix} F_g \\ F_{me} \hat{E}_{me} \end{bmatrix} = I$, where $F_{me} = \begin{bmatrix} F_m & 0 \\ 0 & F_e \end{bmatrix} \in \mathbb{R}^{N_m + N_e \times 2D}$ and $\hat{E}_{me} = \begin{bmatrix} \hat{E}_m \\ \hat{E}_e \end{bmatrix} \in \mathbb{R}^{2D \times D}$, the solution therefore is:

$$\hat{E}_{me} = -(F_{ome}^T R_y^{-1} F_{ome})^{-1} (F_{ome}^T R_y^{-1} F_{goo}) \tag{29}$$

where $F_{ome} = \begin{bmatrix} 0 & 0 \\ F_m & 0 \\ 0 & F_e \end{bmatrix} \in \mathbb{R}^{(N_g + N_m + N_e) \times 2D}$ and $F_{goo} = \begin{bmatrix} F_g \\ 0 \\ 0 \end{bmatrix} \in \mathbb{R}^{(N_g + N_m + N_e) \times D}$.

ACKNOWLEDGMENT

This work is funded by the Wellcome Trust and EPSRC under grant number WT 088877/Z/09/Z.

References

[1] P. Hansen, M. Kringelbach, and R. Salmelin, Eds., *MEG: an introduction to methods*. Oxford University Press, 2010.

[2] C. Wood, D. Cohen, B. Cuffin, M. Yarita, and T. Allison, "Electrical sources in human somatosensory cortex: identification by combined magnetic and potential recordings," *Science*, vol. 227, no. 4690, pp. 1051–1053, 1985.

[3] W. W. Sutherling, P. H. Crandall, T. M. Darcey, D. P. Becker, M. F. Levesque, and D. S. Barth, "The magnetic and electric fields agree with intracranial localizations of somatosensory cortex," *Neurol.*, vol. 38, pp. 1705–1714, 1988.

[4] A. M. Dale and M. I. Sereno, "Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach," *J. Cogn. Neurosci*, vol. 5, pp. 162–176, 1993.

[5] J. Phillips, R. Leahy, J. Mosher, and B. Timsari, "Imaging neural activity using MEG and EEG," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 16, no. 3, pp. 34–42, 1997.

[6] F. Babiloni, D. Mattia, C. Babiloni, L. Astolfi, S. Salinari, A. Basilisco, P. M. Rossini, M. G. Marciani, and F. Cincotti, "Multimodal integration of EEG, MEG and fMRI data for the solution of the neuroimage puzzle," *Magnetic Resonance Imaging*, vol. 22, no. 10, pp. 1471–1476, 2004.

[7] F. Babiloni, F. Carducci, F. Cincotti, C. Del Gratta, V. Pizzella, G. L. Romani, P. M. Rossini, F. Tecchio, and C. Babiloni, "Linear inverse source estimate of combined EEG and MEG data related to voluntary movements," *Human Brain Mapping*, vol. 14, no. 4, pp. 197–209, 2001.

[8] H. Zavala-Fernandez, T. H. Sander, M. Burghoff, R. Orglmeister, and L. Trahms, "Multi-modal ICA exemplified on simultaneously measured MEG and EEG data," in *Proceedings of the 7th international conference on Independent component analysis and signal separation*, ser. ICA'07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 673–680.

[9] M.-X. Huang, T. Song, D. J. H. Jr., I. Podgorny, V. Jousmaki, L. Cui, K. Gaa, D. L. Harrington, A. M. Dale, R. R. Lee, J. Elman, and E. Halgren, "A novel integrated MEG and EEG analysis method for dipolar sources," *NeuroImage*, vol. 37, no. 3, pp. 731–748, 2007.

[10] A. Molins, S. M. Stufflebeam, E. N. Brown, and M. S. Hamalainen, "Quantification of the benefit from integrating MEG and EEG data in minimum l2-norm estimation," *NeuroImage*, vol. 42, no. 3, pp. 1069–1077, 2008.

[11] S. Baillet, L. Garnero, G. Marin, and J.-P. Hugonin, "Combined MEG and EEG source imaging by minimization of mutual information," *Biomedical Engineering, IEEE Transactions on*, vol. 46, no. 5, pp. 522–534, 1999.

[12] S. C. Jun, "MEG and EEG fusion in bayesian frame," in *Electronics and Information Engineering (ICEIE), 2010 International Conference On*, vol. 2, 2010, pp. V2–295–V2–299.

[13] A. Babajani-Feremi and H. Soltanian-Zadeh, "Multi-area neural mass modeling of EEG and MEG signals," *NeuroImage*, vol. 52, no. 3, pp. 793–811, 2010.

[14] R. N. Henson, E. Mouchlianitis, and K. J. Friston, "MEG and EEG data fusion: Simultaneous localisation of face-evoked responses," *NeuroImage*, vol. 47, no. 2, pp. 581 – 589, 2009.

[15] B. Van Veen, W. Van Drongelen, M. Yuchtman, and A. Suzuki, "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering," *Biomedical Engineering, IEEE Transactions on*, vol. 44, no. 9, pp. 867 –880, 1997.

[16] M. Brookes, *The Matrix Reference Manual*, 2011. [Online]. Available: http://www.ee.imperial.ac.uk/hp/staff/dmb/matrix/intro.html

[17] J. Vrba, S. Taulu, J. Nenonen, and A. Ahonen, "Signal space separation beamformer," *Brain Topogr*, 2009.

[18] J. SERGENT, "Functional neuroanatomy of face and object processing," *Brain*, vol. 115, pp. 15–36, 1992.

[19] N. Kanwisher, J. McDermott, and M. M. Chun, "The fusiform face area: A module in human extrastriate cortex specialized for face perception," *The Journal of Neuroscience*, vol. 17, no. 11, pp. 4302–4311, 1997.

[20] E. Halgren, T. Raij, K. Marinkovic, V. Jousmaki, and R. Hari, "Cognitive response profile of the human fusiform face area as determined by MEG," *Cerebral Cortex*, vol. 10, no. 13, pp. 69–81, 2000.

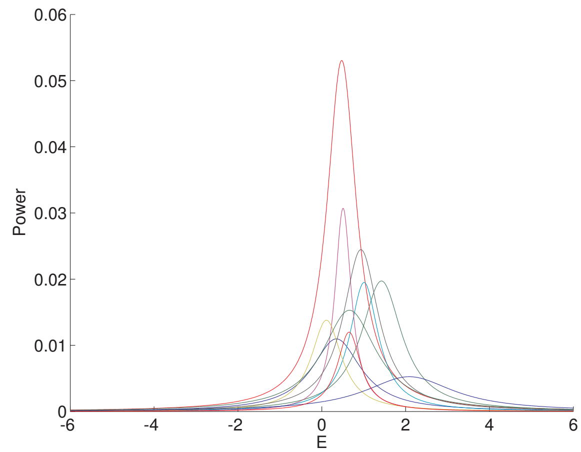
[21] I. Deffke, T. Sander, J. Heidenreich, W. Sommer, G. Curio, L. Trahms, and A. Lueschow, "MEG/EEG sources of the 170-ms response to faces are co-localized in the fusiform gyrus," *NeuroImage*, vol. 35, no. 4, pp. 1495 – 1501, 2007.

[22] J. Liu, A. Harris, and N. Kanwisher, "Stages of processing in face perception: an MEG study," *Nature Neuroscience*, vol. 5, pp. 910–916, 2002.

[23] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain," *NeuroImage*, vol. 15, no. 1, pp. 273–289, 2002.

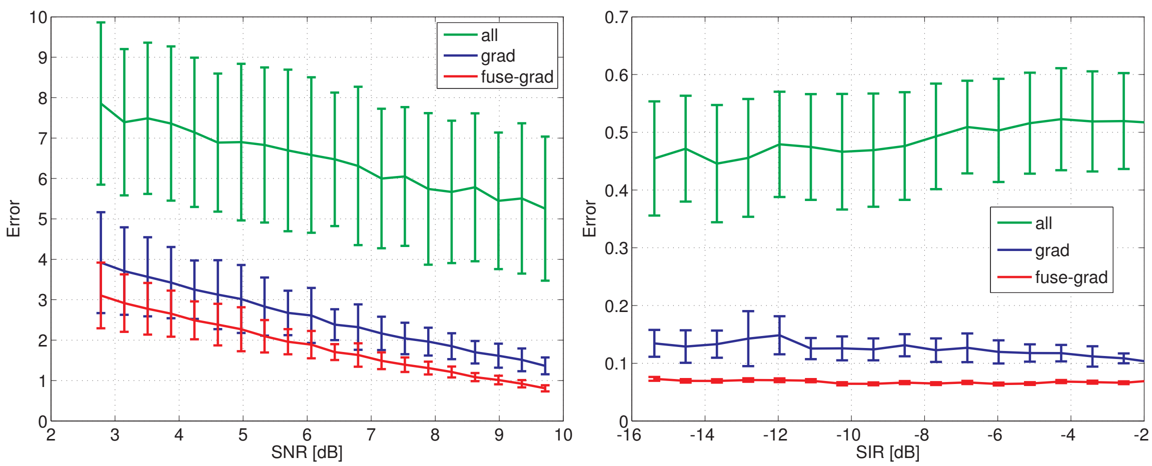
[24] C. C. Paige and M. A. Saunders, "Towards a generalized singular value decomposition," *SIAM Journal on Numerical Analysis*, vol. 18, no. 3, pp. 398–405, 1981.

LIST OF FIGURES



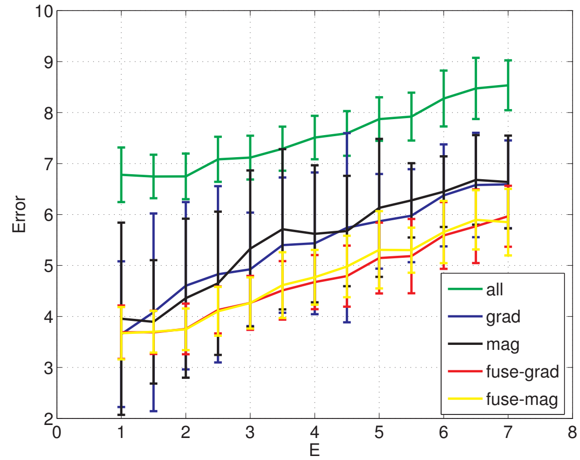
(a)

Figure 1: Examples of the power as a function of E (see equation (4)). All other parameters were selected randomly to generate 10 graphs. This function is convex and therfore has only one maximum.



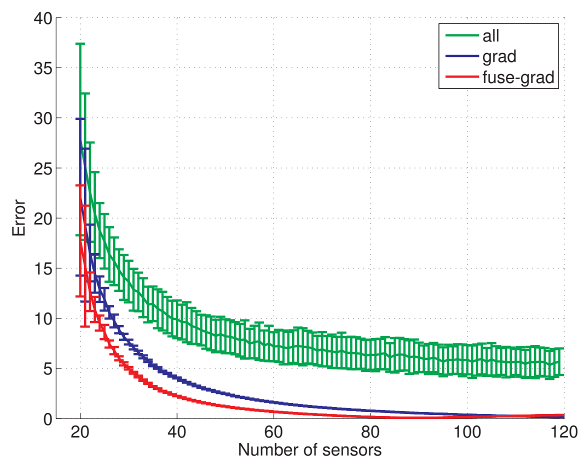
(a) (b)

Figure 2: Average of 1000 Monte Carlo simulations for (a) SNR versus error while SIR is set to approximately 5dB and (b) SIR versus error while SNR is set to approximately 5dB. Error is defined as the norm of the difference between the normalised reconstructed and original time courses. The multiplicative error E_g was set to the identity matrix and E_m was randomly selected in each Monte Carlo simulation.



(a)

Figure 3: The output error versus the true multiplicative error $E_m = E$. Here $E_g = 1$ and scalar beamformer was used while SNR and SBR were approximately fixed to 5dB. 'Fusion' methods outperforms the other methods meaning that it can reconstruct the shape of the signal (ignoring the amplitude) more accurately compared to other methods.

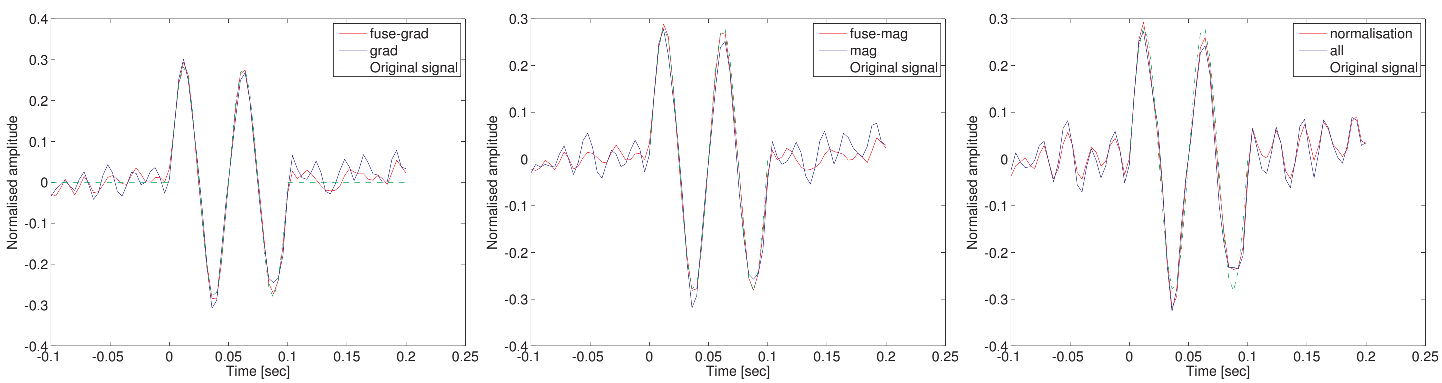


(a)

Figure 4: Effect of the number of sensors on the output error. Number of sensors can have a large impact on the results when it is small compared to the number of sources.

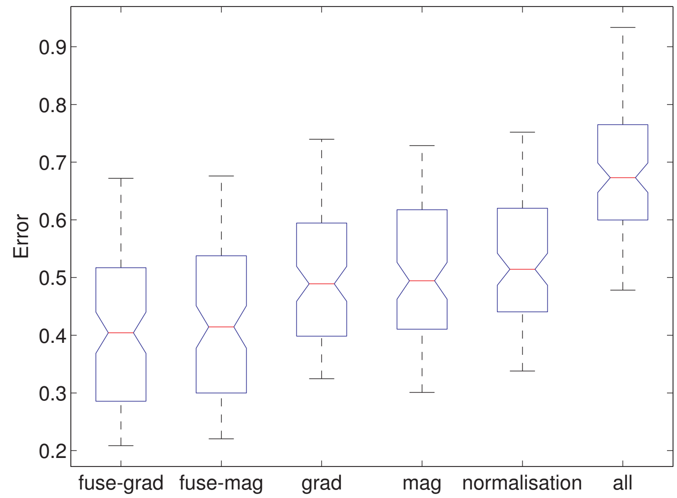
(a) (b)

Figure 5: (a) A schematic diagram of the phantom used in our experiment, (b) location of dipoles (green stars) inside the phantom which was approximated by red the markers.



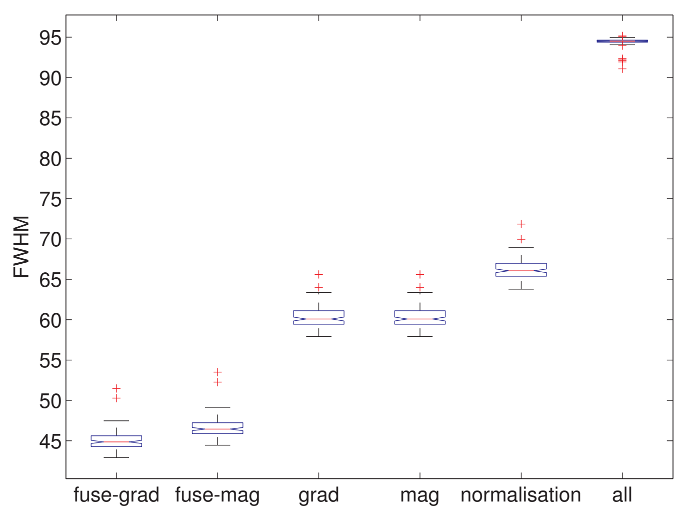
(a) (b) (c)

Figure 6: An example of estimated time-courses for only one dipole and one trial. The real waveform generated by phantom (green dotted) and its approximation using the methods explained in the text. 'Fuse' methods show better performance compared to others.



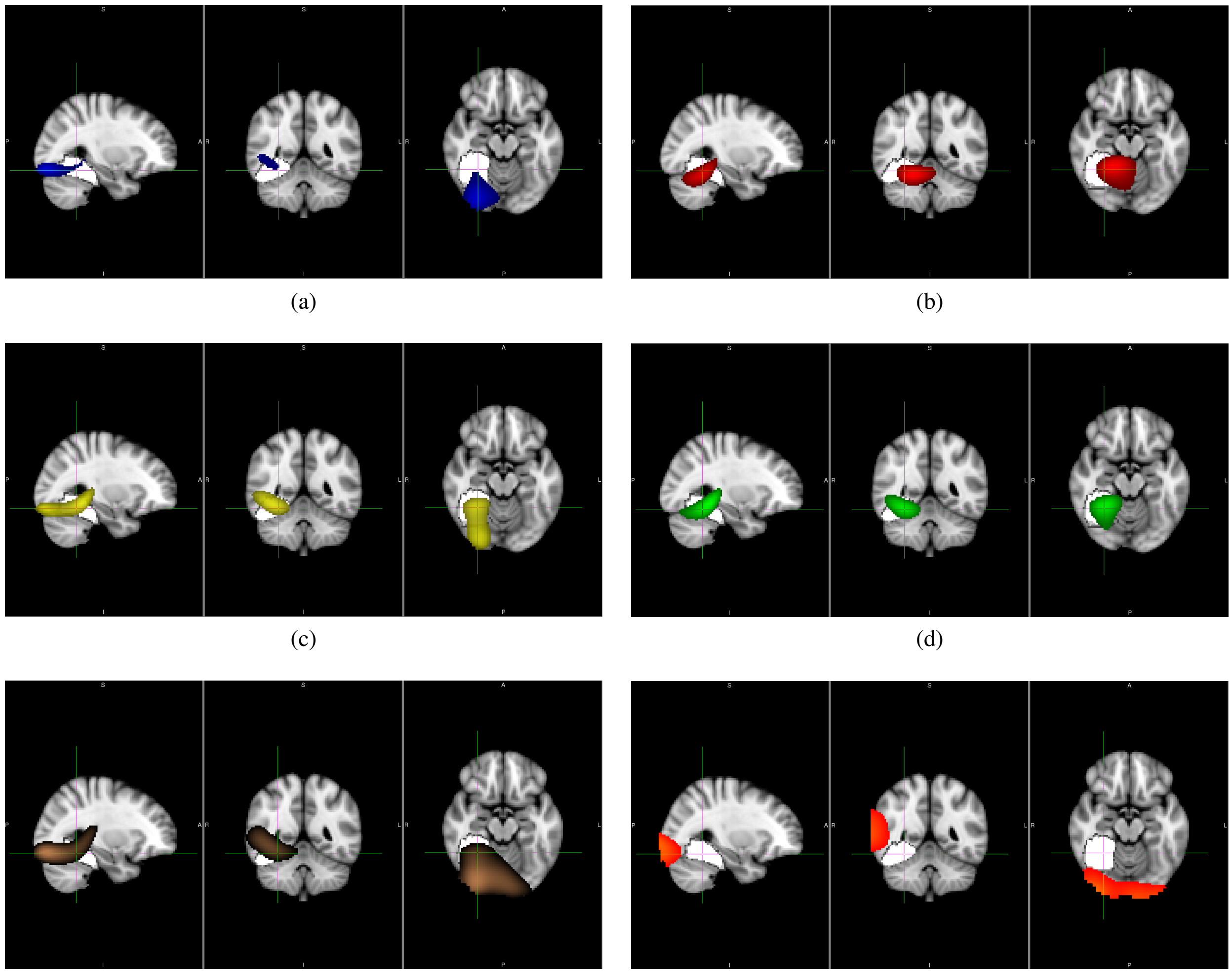
(a)

Figure 7: Box plot of the errors for different methods in phantom experiment. The central red mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points.



(a)

Figure 8: Box plot of the FWHM for different methods in phantom experiment. The fusion methods clearly show less FWHM and thus higher spatial resolution.



(a) (b)

(c) (d)

(e) (f)

Figure 9: Localising the face response showing by the normalised estimated output power of beamformer using (a) ‘grad’, (b) ‘mag’, (c) ’fuse-grad’, (d) ‘fuse-mag’, (e) ‘normalisation’ and (f) ‘all’ method. The solid white volume is the extent of the fusiform cortex within which face specific responses are found and specifically the fusiform face area (FFA).

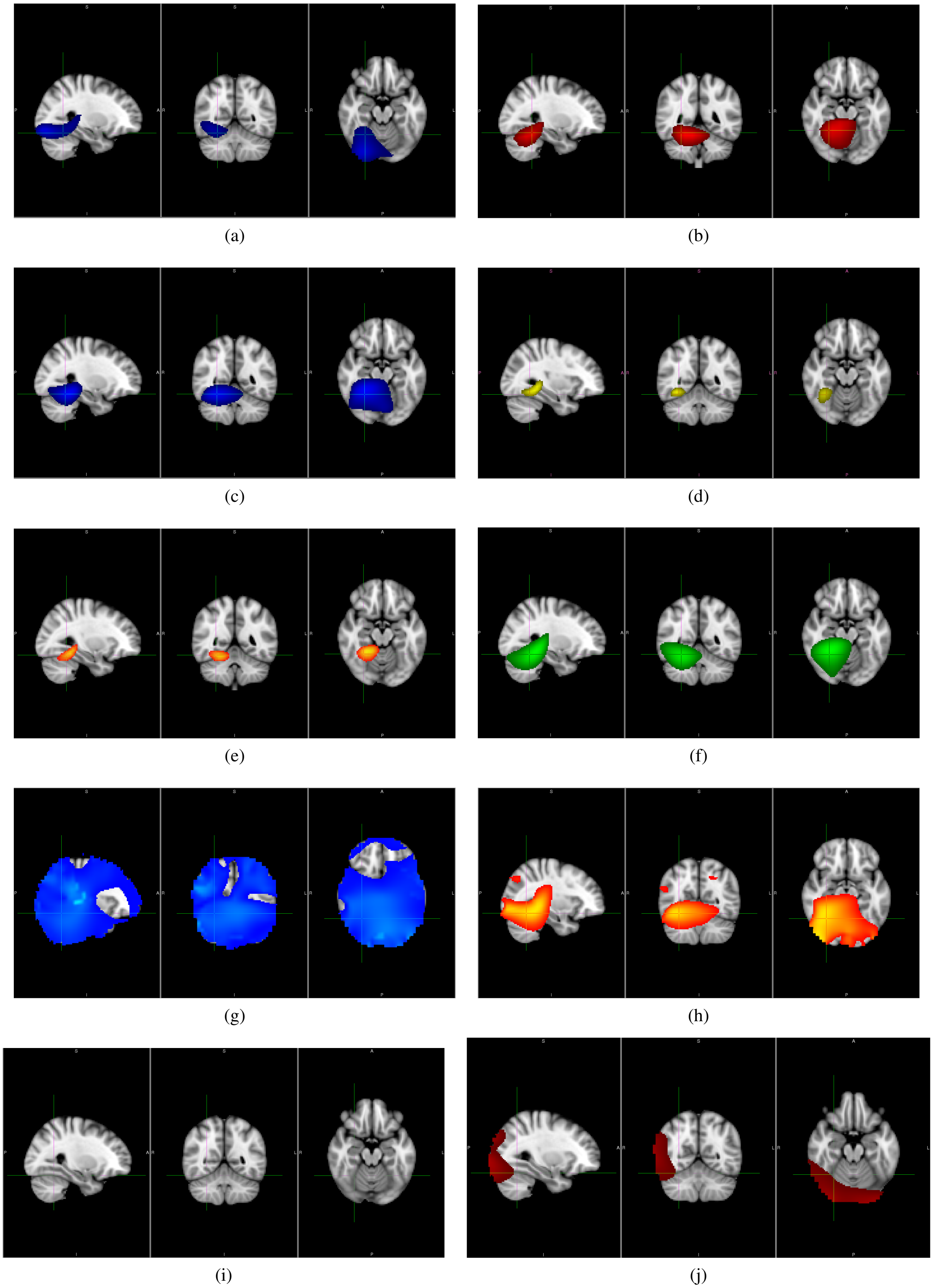


Figure 10: Power of the beamformer weights which were applied to gradiometer (left column) and magnetometer (right column) signals. (a) ‘grad’, (b) ‘mag’, (c and d) ’fuse-grad’, (e and f) ‘fuse-mag’, (g and h) ‘normalisation’ and (i and j) ‘all’ method. ’Fuse’ methods assign adaptive weights to each location according to the spatiotemporal correlation between magnetometer and gradiometer.