Movement Compensation in Magnetoencephalography Analysis

Landon Wagner, Tulaya Limpiti, *Member, IEEE,* Ronald T. Wakai, and Barry D. Van Veen, *Fellow, IEEE*

I. Introduction

EVOKED response data in magneto-encephalography (MEG) typically has a very low signal-to-noise ratio (SNR). To address this problem, in the course of an experiment stimuli are periodically applied to a subject dozens of times and averaging is typically used to improve the SNR. While the temporal response of interest to the stimulus will not be affected by subject movement with respect to the sensor array, the spatial signature of the signal of interest across the sensor array will. As a consequence of this, in the absence of any means of incorporating head movement information into MEG signal processing algorithms, in order to stimulate the subject enough times to increase SNR to necessary levels we must require the subject to remain perfectly still throughout the experiment. This requirement is not practical in conscious subjects, necessary to undergo stimulation, of age 1 to 6 years, a critical period of cognitive development. This motivates our work of incorporating subject head movement information into MEG signal processing algorithms.

de Munck, et al. [6] employ Legendre polynomials to correct MEG data for movement by the minimum norm interpolation method of Hämäläinen and Ilmoniemi [7] and extrapolation of the data to a reference MEG grid. This method requires attaching magnetic coils to the subject head which is not agreeable with the subject age group we are interested in working with. Uutela, Taulu, et al. [8] suggest using the minimum norm estimate of the cortical current in each position to recalculate the magnetic field at a standardized sensor location. Taula, et al. [4] propose a method of standardizing multiple head positions called Signal Space Separation (SSS). SSS is based on expanding the magnetic field in terms of spherical harmonics. SSS has been applied to movement compensation in localization studies. Imada, et al. [9] employ SSS to correct for movements less than 7 millimeters in an infant speech perception localization study. Medvedovsky, et al [10] compare localization performance of two variants of SSS head movement compensation with median nerve stimulation and suggest limiting movement to less than 3 centimeters yields good results. Wehner, et al. [11] evaluated SSS when performing source localization of auditory evoked responses in 8-12 year old children and found that SSS can compensate for the spatial smoothing caused by head movements. The SSS approach does not appear to have been studied in the context of recovering evoked response morphology.

A second approach to dealing with head movement is to incorporate the head movement into the forward models used in the signal processing algorithms for performing source localization or estimation. In principle, directly incorporating movement into the signal processing alhorithms offers potential improvement over two-step approaches that first rely on interpolation. Uutela et al. [8] propose averaging the minimum norm estimate of cortical currents computed using each epoch. However, this approach does not exploit the full extent of available information, i.e., find the one minimum norm solution simultaneously satisfying the sets of linear equations associated with each head postion. Consequently, as discussed above, the low SNR associated with each individual epoch limits the quality of the individual solutions.

We present two approaches for estimating the repeated component of the evoked response while relaxing the requirement that the subject remain perfectly still. One approach uses Signal Space Separation to "transform" the data from multiple distinct head positions to a reference position. At this point we apply the work of Dogandžić and Nehorai [1]. The second approach extends the same work of Dogandžić and A. Nehorai by incorporating head movement information into the source models.

Section II of this paper develops the theory of the two approaches of this work. The efficacy of each approach is demonstrated in Section III using simulated and phantom MEG data. Section IV presents a discussion of the proposed methods. Our notation uses bold lower and uppercase symbols to denote vectors and matrices, respectively.

II. Problem Formulation

Assume multiple observations or epochs of a spatio-temporal series are recorded. The signal of interest is assumed to be repeated across epochs, while the nonrepeated component is assumed to be background brain activity and sensor noise. However, this repeatability is spatially obscured in the recorded data by the source moving with respect to the sensor array. So potentially, each trial has a different spatial identity and we represent a trial of the spatiotemporal signal measured in the p^{th} head position at N sensors and T time samples as the $N \times T$ matrix $\mathbf{S}_p$

$$\mathbf{S}_p = \mathbf{H}_p(\phi)\boldsymbol{\theta}\mathbf{C}^T,\ p = 1, \ldots, P \tag{1}$$

where P is the number of head positions, $\mathbf{H}_p(\phi)$ is a known $N \times Q$ spatial basis matrix, $\mathbf{C}$ is a known $T \times L$ temporal basis matrix, and $\boldsymbol{\theta}$ is an unknown $Q \times L$ basis coefficient matrix. The spatial basis matrix $\mathbf{H}_p(\phi)$ relates neural activity at location ϕ on the cortical surface to the measured signal at the N sensors. Example MEG or EEG spatial bases include dipoles [13], multipoles [14], and cortical patches [12]. We

say that the signal is rank Q if there are Q spatial bases in the columns of $\mathbf{H}_p(\phi)$. The bases in $\mathbf{C}$ represent prior temporal knowledge of the signal such as bandwidth or typical waveform morphologies [1], [3]. Many choices of temporal bases, including Fourier bases or wavelets can be utilized depending on the temporal attributes of interest. Without loss of generality, we assume orthogonal basis functions, i.e., $\mathbf{H}_p(\phi)^T\mathbf{H}_p(\phi) = \mathbf{I}_Q$ and $\mathbf{C}^T\mathbf{C} = \mathbf{I}_L$.

The single-trial observation at the sensors is denoted by the $N \times T$ matrix $\mathbf{X}_p^j$

$$\mathbf{X}_p^j = \mathbf{S}_p + \mathbf{N}_p^j,\ j = 1, \ldots, J_p,\ p = 1, \ldots, P \tag{2}$$

where J_p is the number of trials in the p^{th} head position, the $N \times T$ matrix $\mathbf{N}_p^j$ is zero-mean Gaussian noise with unknown positive-definite spatial covariance $\mathbf{R}_p$ and known temporal covariance $\mathbf{R}_t = \mathbf{I}_T$. Hence, we write $\mathbf{N}_p^j \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_T \otimes \mathbf{R}_p)$, where $\mathbf{I}_T \otimes \mathbf{R}_p$ represents the covariance of $\text{vec}\{\mathbf{N}_p^j\}$. The noise is assumed independent and identically distributed across trials. The subsequent analysis is applicable to the case when $\mathbf{R}_t \neq \mathbf{I}_T$ if the data are temporally whitened, i.e., we replace $\mathbf{X}_p^j$ with $\tilde{\mathbf{X}}_p^j = \mathbf{X}_p^j\mathbf{R}_t^{-1\backslash 2}$.

The single-head position observation at the sensors is denoted by the $N \times TJ_p$ matrix $\mathbf{X}_p$

$$\mathbf{X}_p = \begin{bmatrix}\mathbf{S}_p & \ldots & \mathbf{S}_p\end{bmatrix} + [\mathbf{N}_p^1\ \ldots\ \mathbf{N}_p^{J_p}],\ p = 1, \ldots, P. \tag{3}$$

Then $\mathbf{X}_p \sim \mathcal{N}(\mathbf{H}_p(\phi)\boldsymbol{\theta}\mathbf{D}_p^T, \mathbf{I}_{TJ_p} \otimes \mathbf{R}_p)$ where the $L \times TJ_p$ matrix $\mathbf{D}_p^T = \begin{bmatrix}\mathbf{C}^T & \mathbf{C}^T & \ldots & \mathbf{C}^T\end{bmatrix}$.

A. Signal Space Separation (SSS)

One modern signal processing method used to standardize different head positions in MEG signals is the Signal Space Separation (SSS) method. Using sensor geometry and the sensor array's relative position with respect to the head only, SSS provides a basis spanning all signals of magnetic origin [5].

The basis used in SSS is built out of spherical harmonic functions. An arbitrary N dimensional signal vector for a single time sample can be expressed as [4]

$$\boldsymbol{\mu} = \sum_{l=1}^{\infty}\sum_{m=-l}^{l} \alpha^{lm}\mathbf{a}_p^{lm} + \sum_{l=1}^{\infty}\sum_{m=-l}^{l} \beta^{lm}\mathbf{b}_p^{lm}, \tag{4}$$

where

$$\mathbf{a}_p^{lm} = \frac{Y_{lm}(\boldsymbol{\lambda}_p, \boldsymbol{\tau}_p)}{\boldsymbol{r}_p^{l+1}},\ \mathbf{b}_p^{lm} = \boldsymbol{r}_p^l Y_{lm}(\boldsymbol{\lambda}_p, \boldsymbol{\tau}_p)$$

$$Y_{lm}(\boldsymbol{\lambda}_p, \boldsymbol{\tau}_p) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_{lm}(\cos\boldsymbol{\lambda}_p)e^{\imath m\boldsymbol{\tau}_p}.$$

$P_{lm}(\cos\boldsymbol{\lambda}_p)$ is the associated Legendre function. As a function of the sensor coordinates in the p^{th} position, $\boldsymbol{r}_p$ is the vector specifying each sensor's radial distance from the origin, $\boldsymbol{\lambda}_p$ specifies each sensor's elevation angle and $\boldsymbol{\tau}_p$ specifies each sensor's azimuthal angle.

Note the multipliers, $\boldsymbol{r}_p^{-(l+1)}$ and $\boldsymbol{r}_p^l$ in equation (4) indicate that $\mathbf{a}_{lm}^p$ is associated with sources within the sensor array and $\mathbf{b}_{lm}^p$ is associated with sources outside the sensor array [4]. Since signals of interest lie only within the sensor array, after deconstructing the recorded data into coefficients, α_{lm}^j and β_{lm}^j we can filter out the contribution of interfering sources by leaving off the contribution of β_{lm}^j in the signal reconstruction.

The basis vectors, $\mathbf{a}_{lm}^p, \mathbf{b}_{lm}^p$, specified by $l = 1$ represent the lowest spatial frequency in the measured data while increasing l represents increasing spatial frequency. Physically, there becomes a point based on the system sensor geometry where higher l, m terms no longer represent measureable signals and instead their associated coefficients represent noise. It is above this point where we can truncate the basis. We denote the appropriate cutoff values for $\mathbf{a}_{lm}^p$ and $\mathbf{b}_{lm}^p$ to be L_{in} and L_{out}, respectively. It is desireable from a computational perspective to have the total number of basis vectors be less than the number of spatial channels in order to avoid an overdetermined system of equations. Specifically, we require $N > (L_{\text{in}}+1)^2 + (L_{\text{out}}+1)^2 - 2$ [4]. In fact, truncation of the basis results in the normal tradeoff of noise reduction versus complete signal representation. The point at which to truncate the basis for a particular sensor array can be determined experimentally through reconstructing known signals. Now equation (4) can be written as

$$\boldsymbol{\mu} = \sum_{l=1}^{L_{\text{in}}}\sum_{m=-l}^{l} \alpha_{lm}\mathbf{a}_{lm}^p + \sum_{l=1}^{L_{\text{out}}}\sum_{m=-l}^{l} \beta_{lm}\mathbf{b}_{lm}^p. \tag{5}$$

Or in matrix form one $N \times T$ epoch, S_j recorded in the p^{th} head position can be written as,

$$\mathbf{S}_p = \mathbf{M}_p\mathbf{A}_p = \begin{bmatrix}\mathbf{M}_{p,\text{in}} & \mathbf{M}_{p,\text{out}}\end{bmatrix}\begin{bmatrix}\mathbf{A}_{\text{in}} \\ \mathbf{A}_{\text{out}}\end{bmatrix} \tag{6}$$

where

$$\mathbf{M}_{p,\text{in}} = \begin{bmatrix}\mathbf{a}_{1,-1}^p & \ldots & \mathbf{a}_{L_{\text{in}}L_{\text{in}}}^p\end{bmatrix}, \tag{7}$$

$$\mathbf{M}_{p,\text{out}} = \begin{bmatrix}\mathbf{b}_{1,-1}^p & \ldots & \mathbf{b}_{L_{\text{out}}L_{\text{out}}}^p\end{bmatrix}, \tag{8}$$

$$\mathbf{A}_{\text{in}} = \begin{bmatrix}\alpha(1)_{1,-1} & \ldots & \alpha(T)_{1,-1} \\ \vdots & \ddots & \vdots \\ \alpha(1)_{L_{\text{in}},L_{\text{in}}} & \ldots & \alpha(T)_{L_{\text{in}},L_{\text{in}}}\end{bmatrix}, \tag{9}$$

$$\mathbf{A}_{\text{out}} = \begin{bmatrix}\beta(1)_{1,-1} & \ldots & \beta(T)_{1,-1} \\ \vdots & \ddots & \vdots \\ \beta(1)_{L_{\text{out}}L_{\text{out}}} & \ldots & \beta(T)_{L_{\text{out}}L_{\text{out}}}\end{bmatrix}. \tag{10}$$

Then the coefficient matrix, $\mathbf{A}_p$ can be found from equation (6).

We consider the initial head position, $p = 1$, to be the position of interest and we apply SSS to data recorded from head positions following the first. Then for epochs recorded from head positions $p = 2, \ldots, P$ the "corrected" data is

$$\mathbf{S}_j = \mathbf{M}_{1,\text{in}}\mathbf{A}_{\text{in}} \tag{11}$$

which is the data we would observe had there been no head movement, plus noise.

If we define the $N \times TJ_p$ noise observation matrix for the p^{th} head position as $\mathbf{N}_p = \begin{bmatrix}\mathbf{N}_p^1 & \ldots \mathbf{N}_p^{J_p}\end{bmatrix}$ then the whole $N \times TJ$ observation matrix can be written as

$$\tilde{\mathbf{X}} = \mathbf{H}_1(\phi)\boldsymbol{\theta}\mathbf{F}^T + \begin{bmatrix}\mathbf{N}_1 & \tilde{\mathbf{N}}_2 & \ldots & \tilde{\mathbf{N}}_P\end{bmatrix} \tag{12}$$

where the $L \times JT$ matrix $\mathbf{F}^T = \begin{bmatrix}\mathbf{D}_1^T & \mathbf{D}_2^T & \dots & \mathbf{D}_P^T\end{bmatrix}$ and noise matrices, $\mathbf{N}_p$ for head positions $p = 2, \dots, P$ have been transformed to $\tilde{\mathbf{N}}_p$ by the SSS operation.

B. Maximum Likelihood Estimation - SSS (MLESSS)

If we were to assume that after applying the SSS operation to data within head positions $p = 2, \dots, P$ the covariance matrix of each head position was identical then the probability density function for the data matrix $\tilde{\mathbf{X}}$ can be written as

$$f(\tilde{\mathbf{X}}; \mathbf{R}, \mathbf{H}_1(\phi), \boldsymbol{\theta}) = (2\pi)^{-(NTJ/2)}(\det\mathbf{R})^{-(TJ/2)} \times \exp\Big\{ -\frac{1}{2}\mathrm{tr}(\mathbf{R}^{-1}\mathbf{Q}(\phi))\Big\} \quad (13)$$

where

$$\mathbf{Q}(\phi) = (\tilde{\mathbf{X}} - \mathbf{H}_1(\phi)\boldsymbol{\theta}\mathbf{F}^T)(\tilde{\mathbf{X}} - \mathbf{H}_1(\phi)\boldsymbol{\theta}\mathbf{F}^T)^T. \quad (14)$$

Note that our assumption above does not hold. The application of SSS will result in a different noise covariance for different head positions [5]. However, we argue the assumption is within reason and apply it within this section.

Following the SSS operation the problem setup is the same as in [1], [3]. It can be shown that the MLE for $\mathbf{R}$ is given as a function of the unknown $\boldsymbol{\theta}$ by [2, Theorem 3.1.5]

$$\begin{aligned}\hat{\mathbf{R}} &= \frac{1}{TJ}(\tilde{\mathbf{X}} - \mathbf{H}_1(\phi)\boldsymbol{\theta}\mathbf{F}^T) && (15)\\ &\times\ (\tilde{\mathbf{X}} - \mathbf{H}_1(\phi)\boldsymbol{\theta}\mathbf{F}^T)^T && (16)\\ &= \frac{1}{TJ}\mathbf{Q}(\phi). && (17)\end{aligned}$$

Then the MLE of the signal amplitude $\boldsymbol{\theta}$ is found by minimizing [1]

$$\min_{\boldsymbol{\theta}} |(\tilde{\mathbf{X}} - \mathbf{H}_1(\phi)\boldsymbol{\theta}\mathbf{F}^T)(\tilde{\mathbf{X}} - \mathbf{H}_1(\phi)\boldsymbol{\theta}\mathbf{F}^T)|. \quad (18)$$

Following [1] we define the following quantities:

$$\begin{aligned}\mathbf{X}_{\mathbf{F}} &= \mathbf{X}\mathbf{F}(\mathbf{F}^T\mathbf{F})^{-1/2} && (19)\\ \mathbf{X}_{\bar{\mathbf{F}}} &= \mathbf{X}\bar{\mathbf{F}} && (20)\\ \mathbf{Q}_{\bar{\mathbf{F}}} &= \mathbf{X}_{\bar{\mathbf{F}}}\mathbf{X}_{\bar{\mathbf{F}}}^T && (21)\\ \tilde{\boldsymbol{\theta}} &= \boldsymbol{\theta}(\mathbf{F}^T\mathbf{F})^{1/2}. && (22)\end{aligned}$$

It was shown in [1] that the estimate for $\tilde{\boldsymbol{\theta}}$ is

$$\hat{\tilde{\boldsymbol{\theta}}} = \big(\mathbf{H}_1^T(\phi)\mathbf{Q}_{\bar{\mathbf{F}}}^{-1}\mathbf{H}_1(\phi)\big)^{-1}\mathbf{H}_1^T(\phi)\mathbf{Q}_{\bar{\mathbf{F}}}^{-1}\mathbf{X}_{\mathbf{F}}. \quad (23)$$

At this point, we search over ϕ to find the matrices, $\mathbf{H}_1(\phi)$, $\boldsymbol{\theta}$ that maximize the log-likelihood function

$$\log\big(f(\tilde{\mathbf{X}}; \mathbf{R}, \mathbf{H}_1(\phi), \boldsymbol{\theta})\big) \sim -\frac{1}{2}\mathrm{tr}(\mathbf{R}^{-1}\mathbf{Q}(\phi)) \quad (24)$$

in order to find the source location, $\hat{\phi}$, that best represents the recorded data.

C. Approximate Multiple Position Maximum-Likelihood (AMPML)

At this point we introduce our second approach. Instead of the two-step approach of transforming the data prior to processing as in the MLESSS method we propose another method that extends the work of [1] in order to accomodate multiple head positions. We call this approach Approximate Multiple Position Maximum-Likelihood (AMPML).

For P head positions, the probability density function for the $N \times TJ$ observation matrix $\mathbf{X}$ may be written

$$f(\mathbf{X}, \mathbf{R}_p; \mathbf{H}_p(\phi), \boldsymbol{\theta}) = (2\pi)^{-\frac{NTJ}{2}}\Big(\prod_{p=1}^{P}|\mathbf{R}_p|^{-\frac{TJ_p}{2}}\Big) \times \exp\Big\{ -\frac{1}{2}\sum_{p=1}^{P}\mathrm{tr}\big(\mathbf{R}_p^{-1}\mathbf{Q}_p(\phi)\big)\Big\}. \quad (25)$$

where $\mathbf{Q}_p(\phi) = (\mathbf{X}_p - \mathbf{H}_p(\phi)\boldsymbol{\theta}\mathbf{D}_p^T)(\mathbf{X}_p - \mathbf{H}_p(\phi)\boldsymbol{\theta}\mathbf{D}_p^T)^T$. In this problem a closed form solution for the maximum-likelihood estimator of $\mathbf{R}_p$ is not available. So instead we estimate $\mathbf{R}_p$ in the following way. Define the $J_pT \times J_pT - L$ matrix $\bar{\mathbf{D}}_p$ whose columns are an orthonormal basis for the null-space of the columns of $\mathbf{D}_p$. That is $\mathbf{D}_p^T\bar{\mathbf{D}}_p = \mathbf{0}$ and $\bar{\mathbf{D}}_p^T\bar{\mathbf{D}}_p = \mathbf{I}$. Then define

$$\mathbf{X}_{\bar{\mathbf{D}}p} = \mathbf{X}_p\bar{\mathbf{D}}_p. \quad (26)$$

Note that since the rows of $\mathbf{S}_p$ lie outside of the space spanned by the rows of $\bar{\mathbf{D}}_p$ so that $\mathbf{X}_{\bar{\mathbf{D}}p}$ contains noise alone. We estimate $\mathbf{R}_p$ by using $\mathbf{X}_{\bar{\mathbf{D}}p}$ to construct the sample noise covariance matrix. That is,

$$\hat{\mathbf{R}}_p = \frac{1}{J_pT - L}\mathbf{X}_{\bar{\mathbf{D}}p}\mathbf{X}_{\bar{\mathbf{D}}p}^T \quad (27)$$

Because we assume $\mathbf{R}_p$ is known our task is to maximize the exponential term in equation (25) with respect to $\boldsymbol{\theta}$ and ϕ. Equivalently, our task is to find the solution of the following expression;

$$\min_{\boldsymbol{\theta},\phi}\Bigg[\sum_{p=1}^{P}\mathrm{tr}\big\{\mathbf{R}_p^{-1}(\mathbf{X}_p - \mathbf{H}_p(\phi)\boldsymbol{\theta}\mathbf{D}_p^T) \times (\mathbf{X}_p - \mathbf{H}_p(\phi)\boldsymbol{\theta}\mathbf{D}_p^T)^T\big\}\Bigg]$$

$$= \min_{\boldsymbol{\theta},\phi}\Bigg[\mathrm{tr}\Big\{\sum_{p=1}^{P}\big(\mathbf{R}_p^{-1}\mathbf{X}_p\mathbf{X}_p^T - \mathbf{R}_p^{-1}\mathbf{X}_p\mathbf{D}_p\boldsymbol{\theta}^T\mathbf{H}_p^T(\phi) - \mathbf{R}_p^{-1}\mathbf{H}_p(\phi)\boldsymbol{\theta}\mathbf{D}_p^T\mathbf{X}_p^T + \mathbf{R}_p^{-1}\mathbf{H}_p(\phi)\boldsymbol{\theta}\mathbf{D}_p^T\mathbf{D}_p\boldsymbol{\theta}^T\mathbf{H}_p^T(\phi)\big)\Big\}\Bigg] \quad (28)$$

Since $\mathbf{D}_p^T\mathbf{D}_p = J_p\mathbf{I}$ the last term in the previous equation becomes $J_p\mathbf{R}_p^{-1}\mathbf{H}_p(\phi)\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{H}_p^T(\phi)$. Using linearity and invariance under cyclic permutations properties of the trace operator

we can write

$$\min_{\boldsymbol{\theta},\phi}\Bigg[\mathrm{tr}\Big\{\boldsymbol{\theta}^T\Big(\sum_{p=1}^{P}J_p\mathbf{H}_p^T(\phi)\mathbf{R}_{Np}^{-1}\mathbf{H}_p(\phi)\Big)\boldsymbol{\theta} - \boldsymbol{\theta}^T\Big(\sum_{p=1}^{P}\mathbf{H}_p^T(\phi)\mathbf{R}_{Np}^{-1}\mathbf{X}_p\mathbf{D}_p\Big) - \Big(\sum_{p=1}^{P}\mathbf{D}_p^T\mathbf{X}_p^T\mathbf{R}_{Np}^{-1}\mathbf{H}_p(\phi)\Big)\boldsymbol{\theta} + \sum_{p=1}^{P}\mathbf{R}_{Np}^{-1}\mathbf{X}_p\mathbf{X}_p^T\Big\}\Bigg]. \quad (29)$$

Let

$$\mathbf{F}(\phi) = \sum_{p=1}^{P}J_p\mathbf{H}_p^T(\phi)\mathbf{R}_{Np}^{-1}\mathbf{H}_p(\phi) \quad (30)$$

$$\mathbf{G}(\phi) = \sum_{p=1}^{P}\mathbf{H}_p^T(\phi)\mathbf{R}_{Np}^{-1}\mathbf{X}_p\mathbf{D}_p \quad (31)$$

$$\mathbf{K} = \sum_{p=1}^{P}\mathbf{R}_{Np}^{-1}\mathbf{X}_p\mathbf{X}_p^T. \quad (32)$$

Then equation (29) can be written more clearly as

$$\min_{\boldsymbol{\theta},\phi}\Big[\mathrm{tr}\big\{\boldsymbol{\theta}^T\mathbf{F}(\phi)\boldsymbol{\theta} - \boldsymbol{\theta}^T\mathbf{G}(\phi) - \mathbf{G}^T(\phi)\boldsymbol{\theta} + \mathbf{K}\big\}\Big] \quad (33)$$

We complete the square to write equation (33) as

$$\min_{\boldsymbol{\theta},\phi}\Big[\mathrm{tr}\big\{\big(\boldsymbol{\theta} - \mathbf{F}^{-1}(\phi)\mathbf{G}(\phi)\big)^T\mathbf{F}(\phi)\big(\boldsymbol{\theta} - \mathbf{F}^{-1}(\phi)\mathbf{G}(\phi)\big) + \mathbf{K} - \mathbf{G}^T(\phi)\mathbf{F}^{-1}(\phi)\mathbf{G}(\phi)\big\}\Big]. \quad (34)$$

Because $\mathbf{F}$ is positive-definite, the minimum satisfies $\boldsymbol{\theta} - \mathbf{F}^{-1}(\phi)\mathbf{G}(\phi) = \mathbf{0}$ so

$$\boldsymbol{\theta}(\phi) = \mathbf{F}^{-1}(\phi)\mathbf{G}(\phi) \quad (35)$$

or

$$\boldsymbol{\theta}(\phi) = \Big(\sum_{p=1}^{P}J_p\mathbf{H}_p(\phi)^T\mathbf{R}_{Np}^{-1}\mathbf{H}_p(\phi)\Big)^{-1} \times \Big(\sum_{p=1}^{P}\mathbf{H}_p(\phi)^T\mathbf{R}_{Np}^{-1}\mathbf{X}_p\mathbf{D}_p\Big) \quad (36)$$

Substituting $\boldsymbol{\theta}(\phi)$ into equation (34), we are left with the problem,

$$\min_{\phi}\Big[\mathrm{tr}\big\{\mathbf{K} - \mathbf{G}^T(\phi)\mathbf{F}^{-1}(\phi)\mathbf{G}(\phi)\big\}\Big] \quad (37)$$

in order to find the source location, $\hat{\phi}$, that best represents the recorded data.

III. Results

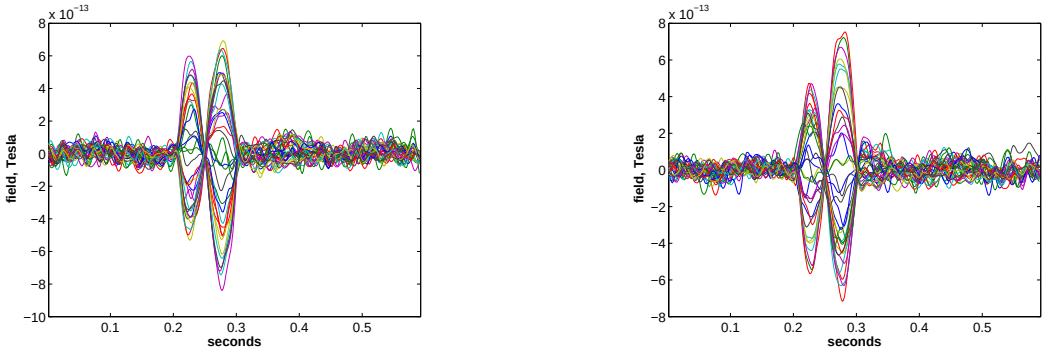
A. Phantom Generated Waveforms

The successful application of Signal Space Separation relies on accurate knowledge of sensor geometry [4]. Much of the work in the literature is carried out on a sensor array of hundreds of sensors allowing whole-head coverage and gradiometers precisely constructed through a precise thin film process [5]. The MEG system available for this research is a Magnes-II dual-dewar system employing either one or two parabolic sensor arrays of 35 functioning hand-wound gradiometers each. The effects of imprecisely constructed sensors (mismatch) is a topic studied in the literature but the effects of employing a relatively small number of sensors as well as not having whole head coverage is not. Both are issues with our available hardware. To study these effects we used the Magnes-II MEG system to collect data generated from a single dipole source within a phantom. Sources of noise in the experiment is inherent noise from the function generator and MEG sensors; no artificial noise was added. A waveform mimicking a typical evoked sinusoidal response was generated from the initial phantom position and after rotating the phantom about its vertical axis which resulted in a dipolar source translation of 8.6 millimeters. Waveforms from each phantom position are shown in Figures 1a and 1b. Figure 2a shows the effect of applying SSS to interpolate the data from the second phantom position to the first. We use $L_{\text{in}} = 4$ to perform the SSS operation as we must have $(L_{\text{in}} + 1)^2 < N = 35$. Figure 2b is the average of 64 epochs after applying SSS interpolation to each. As a comparison of averaged waveforms, Figure 2c is the average of the 64 epochs from the first phantom position. Comparing Figures 2a and 2b we can clearly see that by increasing the SNR through averaging familiar signal characteristics emerge. In other words applying SSS on our hardware decreases the SNR.

We demonstrate the effects of AMPML and SSSML on high-SNR phantom data used to show the effects of SSS on the Magnes-II dual-dewar MEG hardware. We use four phantom positions and record 64 epochs in each position. There is approximately 4.5 degrees of rotation between each position and the dipolar source is located such that a total rotation of 13.5 degrees results in a total source translation of approximately 8.6 millimeters. When applying AMPML and SSSML to the phantom dataset, each yields a localization error of 2.4 millimeters. The source dipole chosen is the dipole nearest the physical source location. i.e., the 2.4 millimeter localization error is the minimum. Since the estimated waveform is reconstructed as $\mathbf{S} = \mathbf{H}_1(\hat{\phi})\boldsymbol{\theta}\mathbf{C}^T$ and with the true waveform unknown we propose the average of the sixty-four epochs from the first head position to be a reference signal in place of the unknown true signal. This reference signal is shown in Figure 3. Based on this reference signal, the root mean squared error of the estimated signals under recursive maximum-likelihood and SSS compared to the reference signal are 26.351 and 36.728 femto-Tesla, respectively. Reconstructed waveforms using recursive maximum-likelihood and SSS are shown in Figures 4 and 5, respectively.

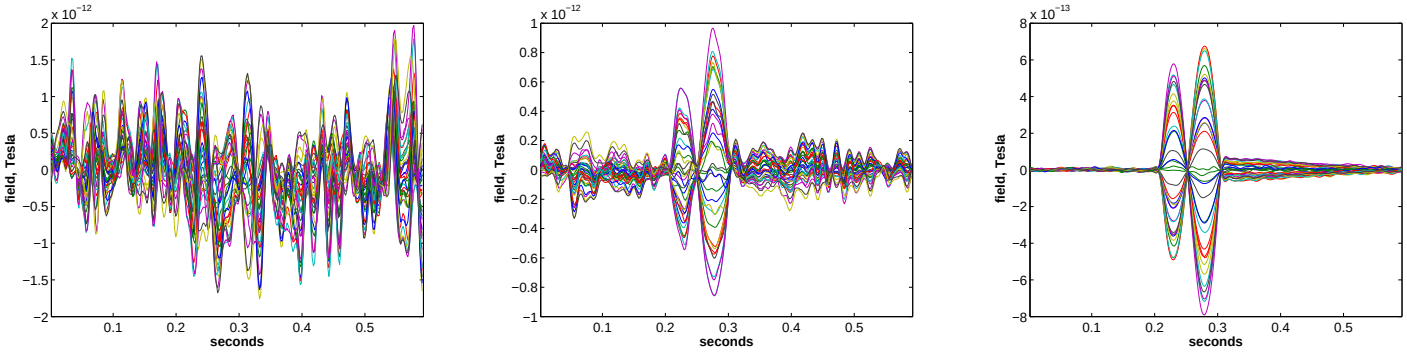
B. Simulated Waveforms

In order to compare results on simulated data with results on phantom generated data we generate simulated data using both the 37 and 74 sensor configurations of the Magnes II system (Biomagnetic Technologies, Inc.) in our lab. We also utilize both sensor configurations in order to observe gains in both algorithms due to an increased number of sensors. Each simulation run consists of 100 600 millisecond long epochs.



(a) (b)

Fig. 1. Data from two phantom positions with a common source waveform. (a) Single unprocessed epoch from the initial phantom position. (b) Single unprocessed epoch after the phantom was rotated on its axis approximately 13.5°.



(a) (b) (c)

Fig. 2. (a) Result of applying SSS to the epoch shown in Fig 1b. Note how SSS enhances the noise to a level that obscures a once visible waveform. (b) Average of the 64 epochs in the second phantom position. Note how familiar signal characteristics reappear. (c) As a reference for Fig 2b, average of the 64 epochs in the first phantom position.

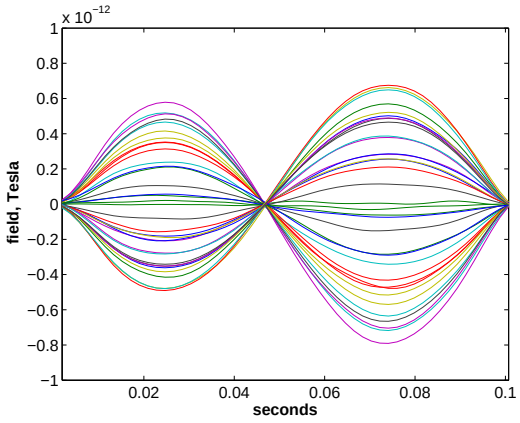


Fig. 3. Reference signal (average of data from the first head position)

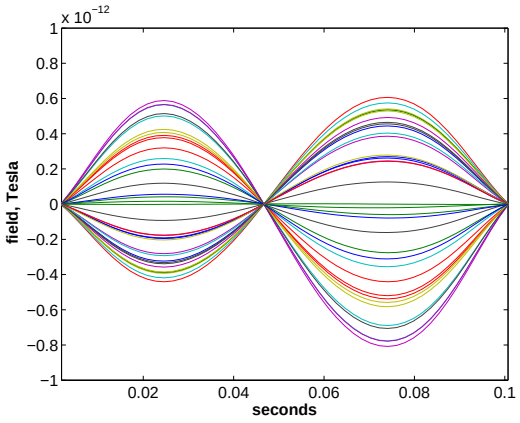


Fig. 4. AMPML reconstruction of phantom data

The signal of interest represents a typical evoked response and is 100 milliseconds long. Fig. 6 depicts the noiseless waveform used in simulation. To generate the localization-error vs. SNR and RMSE vs. SNR curves we generate J unity-power $N \times T$ noise matrices $\mathbf{N}_j$ once and apply a scalar to the matrices to achieve the desired SNR per epoch values in the curves. We compare algorithm performance under two noise structures. The first is to add spatially and temporally white noise to the data. The second is to add spatially colored, temporally white noise to the data. We make this comparison in order to observe how well each algorithm exploits structure in the noise. To construct a colored noise realization we first generate 30 dipoles whose coordinates are uniformly distributed within the head model. Thirty independent $1 \times T$ Gaussian distributed

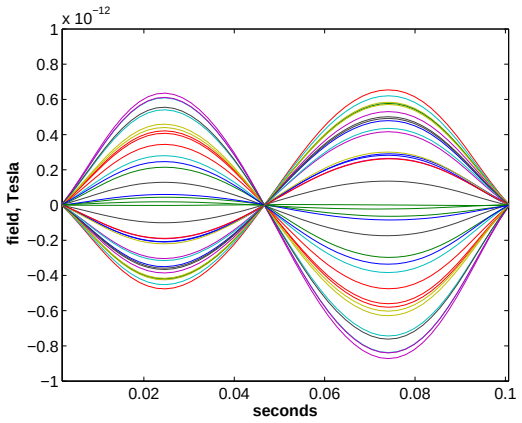


Fig. 5. SSSML reconstruction of phantom data

vectors with mean zero and a random variance uniformly distributed on the interval $[0.1, 1]$ are generated to construct the time sequence of each dipole. Finally, white noise with one hundred times less power than the colored noise is added to the colored noise matrix. Fig. $7a$ and Fig. $7b$ compare the localization error and root mean squared error (RMSE), respectively, using 35 channels. The error metrics are computed as the average over 20 independent simulations. Clearly, at realistic operating SNR values and considering white and colored noise scenarios separately, AMPML algorithm offers gains in both source localization and waveform estimation performance. This shift to the right of the AMPML curve can be explained by the observations made from Figs. $2a$ and $2b$ - applying SSS to a dataset under our sensor configuration has the effect of lowering the SNR. So while using SSSML we are able to localize the source and estimate the waveform well at a certain point, it requires a higher SNR than the one step AMPML algorithm. As expected, each algorithm performs better under the colored noise scenario due to the ability of the maximum-likelihood algorithms to exploit the colored noise structure.

Fig. $8a$ and Fig. $8b$ compare the localization error and RMSE, respectively, using 70 channels. The same general observations can be made as in the 35 channel simulations but now we can also observe how the SSSML curves better fit the AMPML curves in each noise scenario compared to the 35 channel simulations. In the RMSE plots, we see a shift to the left in the AMPML SNR curves when moving between 35 and 70 channels but we see a larger shift in the SSSML performance. This can be explained by referencing [5] - increasing the number of sensors has the effect of improving the effect of SSS on noise enhancement.

C. Simulated Waveforms - Sensor Mismatch

It is stressed in the literature that SSS is a method sensitive to the accuracy of the calibration of the measurement device [5]. The sensors in our MEG system are hand wound, where each is understood to be wound within one percent of the nominal specification. This makes mismatch an important effect when interpreting real data where the true signal and estimation errors are unknown. We simulate and analyze sensor mismatch in this section. This is done by assuming perfect sensor construction in the algorithms but generating test data with sensor error. Based on the knowledge of one percent construction accuracy we generate independent normal scalars for each coil with mean one and variance 0.01 and apply each to corresponding coils in the forward model when generating data. The same set of random scalars is used across SNR values within an experiment but a new set of scalars is generated for each of the twenty experiments shown averaged in the SNR curves. The resulting localization error and waveform RMSE for 35 and 70 channels can be seen in Figs. $9a$ - $10b$. From Figs. $9a$ and $10a$, localization error performance of 35 and 70 channel configurations, respectively, we see that AMPML is not as sensitive to sensor mismatch as SSS.

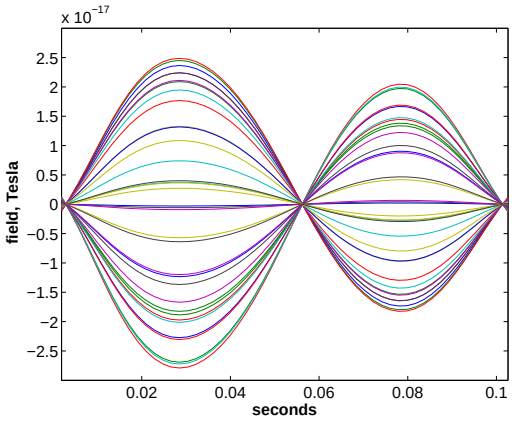
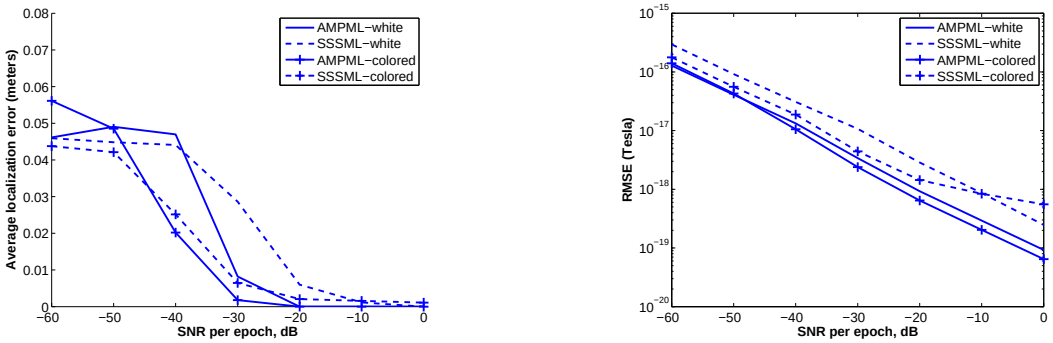


Fig. 6. Single epoch of the noiseless waveform used in simulation.

IV. Discussion

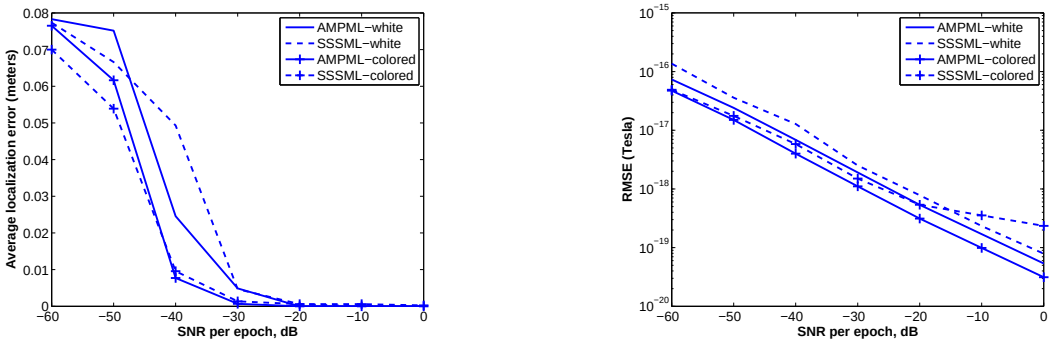
We proposed two methods for estimating evoked dipole responses when the recorded data is corrupted by a known subject head trajectory. The SSSML method is a two step algorithm that through the SSS transformation standardizes all head positions to a reference head position. Following the SSS transformation the maximum-likelihood methods of [1] for estimating dipole responses of stationary subjects is applied. The AMPML method is a one step algorithm that extends the work of [1] to incorporate known head movement information directly into the source models. In the absence of a closed form solution for the maximum-likelihood estimator of the covariance matrix we propose a substitute that is based on projecting the recorded data onto the null space of the known temporal basis matrix for the signal of interest and using the resulting signal-free data to compute a noise sample covariance matrix.

We presented numerical simulations demonstrating the performance of the SSSML and AMPML methods. The SSSML and AMPML methods were compared where under the available MEG hardware the AMPML method was shown to be a more accurate estimation algorithm and more robust to imperfect sensor construction.



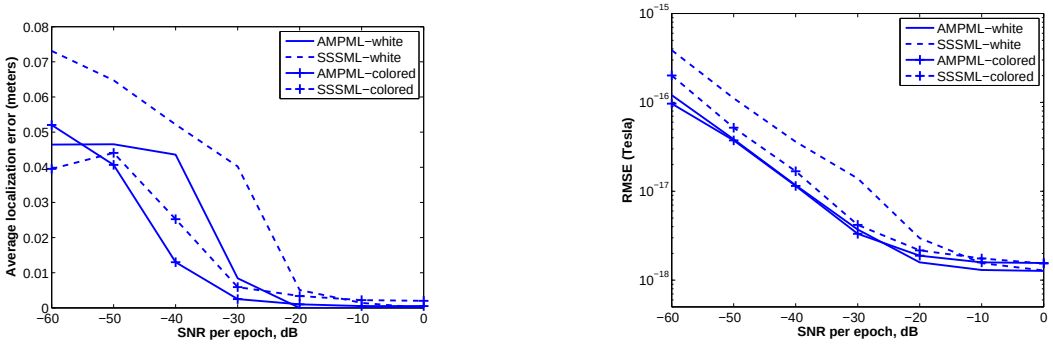
(a) (b)

Fig. 7. Error metrics as a function of SNR per epoch using 35 channels, 100 epochs per experiment, average of 20 experiments. (a) Localization error. (b) Estimated waveform RMSE.



(a) (b)

Fig. 8. Error metrics as a function of SNR per epoch using 70 channels, 100 epochs per experiment, average of 20 experiments. (a) Localization error. (b) Estimated waveform RMSE.



(a) (b)

Fig. 9. Simulation of the effects of mismatch between hardware and software calibration. Error metrics as a function of SNR per epoch using 35 channels, 100 epochs per experiment, average of 20 experiments. (a) Localization error. (b) Estimated waveform RMSE.

REFERENCES

[1] A. Dogandžić and A. Nehorai, "Estimating evoked dipole responses in unknown spatially correlated noise with EEG/MEG arrays," *IEEE Trans. Signal Processing,* vol. 48, pp. 13-25, Jan. 2004.

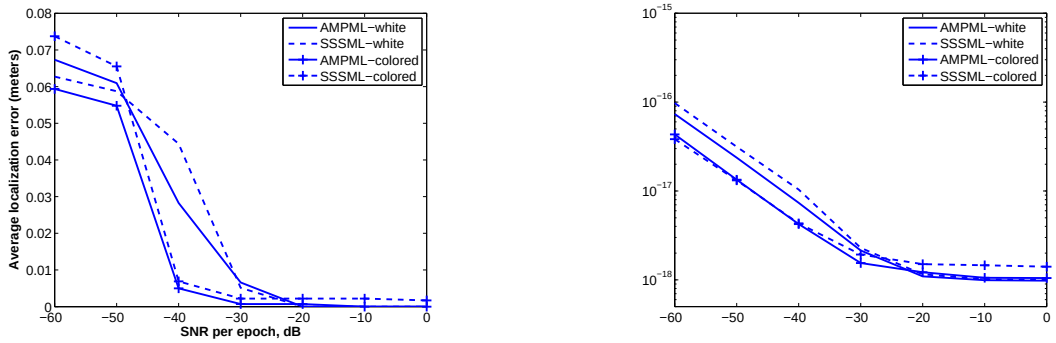
[2] R. Muirhead, *Aspects of Multivariate Statistical Theory.* New York: Wiley, 1982.

[3] B. V. Baryshnikov, B. D. Van Veen, and R. T. Wakai, "Maximum-Likelihood Estimation of Low-Rank Signals for Multiepoch MEG/EEG Analysis," *IEEE Trans. Biomed. Eng.,* vol. 51, pp. 1981-1993, Nov. 2004.

[4] S. Taulu and M. Kajola, "Presentation of electromagnetic multichannel data: The signal space separation method," *J. Appl. Phys.,* vol. 97, Jun. 2005.

[5] S. Taulu, J. Simola, and M. Kajola, "Applications of the Signal Space Separation Method," *IEEE Trans. Signal Processing,* vol. 53, pp. 3359-3372, Sept. 2005.

[6] J. C. de Munck, J. P. A. Verbundt, D. Van't Ent, and B. W. Van Dijk, "The use of an MEG device as 3D digitizer and motion monitoring system," *Phys. Med. Biol.,* vol. 46, pp. 2041-2052,



(a) (b)

Fig. 10. Simulation of the effects of mismatch between hardware and software calibration. Error metrics as a function of SNR per epoch using 70 channels, 100 epochs per experiment, average of 20 experiments. Fig. 10a is the localization error. Fig. 10b is the estimated waveform RMSE.

July 2001.

[7] M. S. Hämäläinen and R. J. Ilmoneimi, "Interpreting magnetic fields of the brain: minimum norm estimates," *Med. Biol. Eng. Comput.,* vol. 32, pp. 35-52, 1994.

[8] K. Uutela, S. Taulu, and M. Hämäläinen, "Detecting and Correcting for Head Movements in Neuromagnetic Measurements," *NeuroImage,* vol. 14, pp. 1424-1431, 2001.

[9] T. Imada, Y. Zhang, M. Cheour, S. Taulu, A. Ahonen, and P. K. Kuhl, "Infant speech perception activates Broca's area: a developmental magnetoencephalography study," *NeuroReport,* vol. 17, pp. 957-962, 2006.

[10] M. Medvedovsky, S. Taulu, R. Bikmullina, and R. Paetau, "Artifact and head movement compensation in MEG," *Neurology, Neurophysiology and Neuroscience,* vol. 4, 2007.

[11] D. T. Wehner, M. S. Hämäläinen, M. Mody, and S. P. Ahlfors, "Head movements of children in MEG: Quantification, effects on source estimation, and compensation," *NeuroImage,* vol. 40, pp. 541-550, 2008.

[12] T. Limpiti, B. D. Van Veen, and R. T. Wakai, "Cortical patch basis model for spatially extended neural activity," *IEEE Trans. Biomed. Eng.,* vol. 53, no. 9, pp. 1740-1754, Sep. 2006.

[13] M. Scherg and D. von Cramon, "Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model," *Electroenceph. Clin. Neurophysiol.,* vol. 62, pp. 32-44, 1985.

[14] J. Mosher, R. Leahy, D. Shattuck, and S. Baillet, "MEG source imaging using multipolar expansions," in *Proc. 16th Conf. Inf. Process. Med. Imag. (IPMI 1999),"* LNCS, vol. 1613, Visegrád, Hungary: Springer-Verlag, Jun./Jul. 1999, pp. 15-28.