

Published in final edited form as:

Neuroimage. 2010 January 1; 49(1): 282–292. doi:10.1016/j.neuroimage.2009.08.019.

Fetal MEG evoked response latency from beamformer with random field theory

J. McCubbin¹, J. Vrba^{1,3}, P. Murphy¹, J. Temple¹, H. Eswaran¹, C.L. Lowery¹, and H. Preissl^{1,2}

¹Dept. of Obstetrics and Gynecology, University of Arkansas for Medical Sciences, Little Rock, AR 72205 USA ²MEG Center, University of Tuebingen, 72076 Tuebingen, Germany ³Elekta Neuromag Oy, Helsinki, Finland

Abstract

Analysis of fetal magnetoencephalographic brain recordings is restricted by low signal to noise ratio (SNR) and non-stationarity of the sources. Beamformer techniques have been applied to improve SNR of fetal evoked responses. However, until now the effect of non-stationarity was not taken into account in detail, because the detection of evoked responses is in most cases determined by averaging a large number of trials. We applied a windowing technique to improve the stationarity of the data by using short time segments recorded during a flash evoked study. In addition, we implemented a random field theory approach for more stringent control of false positives in the statistical parametric map of the search volume for the beamformer. The search volume was based on detailed individual fetal/maternal biometrics from ultrasound scans and fetal heart localization. Average power over a sliding window within the averaged evoked response against a randomized average background power was used as the test z – statistic. The significance threshold was set at 10% over all members of a contiguous cluster of voxels. There was at least one significant response for 62% of fetal and 95% of newborn recordings with gestational age (GA) between 28 and 45 weeks from 29 subjects. We found that the latency was either substantially unchanged or decreased with increasing GA for most subjects, with a nominal rate of about -11 ms/week. These findings support the anticipated neurophysiological development, provide validation for the beamformer model search as a methodology, and may lead to a clinical test for fetal cognitive development.

Keywords

fetal magnetoencephalography; evoked response latency; statistical significance; beamformer; MEG; random field theory

Introduction

The study of fetal brain development is of interest for early detection of neurological disorders and could lead to improved treatment and outcomes. Common disorders that may benefit from perinatal detection and potential treatment include cerebral palsy, epilepsy, deafness, and

© 2009 Elsevier Inc. All rights reserved.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

blindness. In addition to genetic factors, these disorders are associated with prenatal or perinatal infection, toxic insult, hypoxia, ischemia, hemorrhage, and low birth weight. Additional information on fetal neurophysiological status could influence the decision for early delivery intervention and premature birth (Lowery et al., 2006). Monitoring of neurophysiological markers over the perinatal period could lead to a normative database for use in existing or new treatment methods. Traditional indirect measures such as heart rate variability, ultrasound observations, and behavioral changes due to external stimuli may be augmented by direct measures of fetal brain activity using functional magnetic resonance imaging (fMRI) (Gowland and Fulford 2004) or fetal magnetoencephalography (fMEG) (Lowery et al., 2006).

Fetal MEG has been applied by several groups over the past decade (Lengle et al., 2001; Eswaran et al., 2002; Schleussner and Schneider, 2004) and the major outcome measure of these studies was latency estimation of evoked responses to visual and auditory stimulation. Several studies reported a decrease in latency for auditory evoked responses (AER) with increasing gestational age (GA). Dragonova et al. (2007) found an auditory discriminative brain response (MMN) in 46% of 76 fetal recordings over 18 subjects after heart signal attenuation by orthogonal projection (OP, McCubbin et al., 2006; Robinson et al., 2002). A visual selection of peak latency was aided by comparison with averaged noise estimate and other criteria, however there was no statistically significant correlation of latency with GA. Kiefer et al. (2008) compared AER latency in growth restricted vs. normal fetuses and found that, with increasing GA, both groups had a statistically significant latency decrease of 13 ms/week. It was also noted that the growth restricted fetuses showed a delayed response. Sheridan et al. (2008) used visual flash evoked response (VEF) to test for habituation over four sequential flashes. Although they found a low fetal evoked response detection rate, 29% of 25 records, there was evidence of habituation within that group. Visual selection of peaks in the averages was the same as used by Dragonova et al. (2007) and Kiefer et al. (2008).

The detection rates and evidence for neurophysiological maturational markers exemplified by these reports is encouraging progress, however the mostly manual process of visual examination for latency peaks in averaged time courses with poor SNR is laborious and subjective. The problem is to select statistically significant latency by some objective and automated means for investigation of neurophysiological markers. This has been addressed recently in the application of cross correlation analysis (Govindan et al., 2008) and beamformer model search (McCubbin et al., 2007, Robinson and Vrba, 2004) to fetal MEG evoked response detection. In Govindan et al. (2008), the cross correlation computed between the stimulus signal and each MEG channel was subjected to a bootstrap confidence level estimate. The peaks in the correlation time course which exceeded the confidence level were then selected as latency for each MEG channel. They used data from four selected subjects with four to seven records each (including neonatal recordings) and investigated the correlation between latency and GA. The individual subject latencies did not show a statistically significant correlation with GA, but the group data did. In McCubbin et al. (2007) a spatial filter attenuating signal components that do not correspond to a selected electromagnetic source model was used (beamformer, with model as a dipole in a conducting sphere). A model grid was constructed over the plausible volume in the maternal abdomen for fetal head location and candidate averaged evoked responses corresponding to beamformer source activity time courses were subjected to a bootstrap significance test. Representative source time courses that passed the significance test were selected (based on SNR) for latency estimation. They found a broad range of latency and no clear correlation with GA in a group of 27 subjects with single records.

These promising but somewhat conflicting early results may indicate a poor selectivity of the statistical methods applied which may be related to a high false positive rate. In fact, there has been a considerable literature recently in the area of 'family-wise error rate' or 'false discovery rate' (FDR), and its control as a solution for the multiple comparison problem. The situation

arises when the p-values of multiple hypothesis tests are compared and the single hypothesis significance threshold must be increased to maintain the overall statistical significance for the experiment. A standard approach like Bonferroni correction (Bonferroni, 1936) provides an intuitive but conservative solution with limited statistical power. Recently, several different methods to control FDR have been developed which claim improved statistical power for certain applications, mostly in genetics, positron emission tomography (PET), and fMRI. A review of the literature in this area is beyond the scope of this report (see e.g. Li et al., 2005; Marchini and Presanis, 2004).

MEG statistical analysis is closely related to statistical parametric mapping (SPM) of fMRI or PET data. Fetal MEG analysis suffers from the same problem as exemplified in Govindan et al. (2008) where p-values over about 150 MEG channels are compared, and in McCubbin et al. (2007) with thousands of p-values in the search grid for comparison. We have selected an approach which seems particularly well suited to beamformer SPM analysis (Barnes and Hillebrand, 2003) and is based on PET and fMRI applications (Friston et al., 1996; Worsley et al., 1999). The approach involves ‘statistical flattening’ of a modeled 3D Gaussian random field to account for anisotropic and inhomogeneous distribution of noise and includes a correction for multiple comparisons. The general approach of random field theory (RFT) has since been advanced and refined (Taylor and Worsley, 2007). RFT has been applied to event related potential recordings in the analysis of functional connectivity of the combined sensor/source space (Carbonell et al., 2004) and was recently extended to include the time varying correlation structure (Carbonell et al., 2009).

In this report, we describe an upgrade of the scalar beamformer model search with statistical inference using RFT. In order to get best results with the more stringent statistics, we also improved the ultrasound biometric data collection to define a more plausible and limited search volume and improved the stationarity of the record by applying a sliding analysis window. We apply this procedure to 29 subjects with a total of 111 recordings and report the corresponding findings on the latency for evoked responses. We define latency here as the first peak after the stimulus trigger in the power of the response waveform as usual for evoked response analysis.

Materials and methods

Data collection

The fetal MEG data was collected using a dedicated 151 channel instrument specifically designed for measurements over a large area of the pregnant abdomen (Lowery, 2006). Subjects with normal pregnancies were recruited with informed consent and measurement protocols were approved by the institutional review board. An ultrasound (US) exam was conducted immediately prior to the MEG session according to standard procedures and the following measures were estimated from the images: fetal head-center to heart-center distance, fetal head major and minor diameters and circumference, and fetal head surface to maternal abdominal surface distance. The record also included a fetal orientation sketch as shown in the example of fig. 1. We have quantified the orientation of the fetal head relative to the fetal heart as indicated by measuring the angle from the positive x axis (relative to the heart location) to approximately the center of the head within about ± 5 degrees. The value of 5 degrees was chosen because it reflects the perceived inaccuracy of about 1 cm over a distance of about 10 cm.

Localization coils were placed on the subject’s left and right hips and spine. A fourth coil was positioned roughly over the fetal head location determined from the US exam. Coil locations were measured at the beginning of each MEG session and were used as a rough registration of the subject to the MEG array. MEG data was recorded at a sample rate of 312.5 Hz with a bandwidth of 0 – 100 Hz with either an auditory or visual stimulation. The auditory evoked

response protocol consisted of 100 ms tone bursts at 500 Hz (80%) or 1000 Hz (20%) with approximately 3 s inter stimulus intervals (randomized ± 0.5 s) delivered through a plastic tube. Visual stimulus was generated by an array of high power red LEDs coupled to a fiber optic cable terminated in a 3 cm by 5 cm woven panel that was positioned on the maternal abdomen over the fetal head location (Wilson et al. 2008). Visual stimulus duration was 0.5 s with inter stimulus interval of 3 s ± 0.5 s.

For development of the updated model search processing method, we used a series of eight MEG datasets (the test series) collected from different subjects with GA between 28 and 35 weeks. Each dataset consisted of 10 minutes of visual stimulation and 10 minutes of auditory stimulation. The developed model search was then applied to a series of 111 datasets (the study series, 90 fetal and 21 newborn records) collected from 29 subjects over multiple (from one to six) visits in the period of 28 to 45 weeks GA and including a newborn recording from 21 of the 29 subjects (newborn GA was between 39 and 45 weeks). Only a 10 minute visual stimulation was available for analysis of the later series.

Beamformer

The beamformer model search method used in this work employs a scalar minimum variance beamformer (Van Veen et al., 1997, Sekihara et al., 2004). The beamformer is a type of spatial filter where the output is a linear combination of the measurement channel data. All data is denoted as: $\mathbf{m}(n)$, $n = 1, \dots, N$, where N is the number of time samples, and n is used instead of t_n , the $\mathbf{m}(n)$ is a $M \times 1$ column vector with M being the number of channels. The set of coefficients used for the linear combination of $\mathbf{m}(n)$ is a $M \times 1$ vector of weights \mathbf{w}_θ , where θ denotes dipole location, \mathbf{r} , and orientation in the tangential plane at the location of interest, ψ , or $\theta = (\mathbf{r}, \psi)$. Source activity from the same location and orientation is computed as

$$s_\theta(n) = \mathbf{w}_\theta^T \mathbf{m}(n), \quad (1)$$

where \mathbf{w}_θ is given by

$$\mathbf{w}_\theta^T = \frac{\mathbf{B}_\theta^T \mathbf{C}^{-1}}{\mathbf{B}_\theta^T \mathbf{C}^{-1} \mathbf{B}_\theta}, \quad (2)$$

with \mathbf{B}_θ as the model forward solution vector for position and orientation θ and \mathbf{C} is the covariance matrix of the data. The method for selecting the optimal orientation involves an eigen decomposition of the vector beamformer output (Sekihara et al., 2004). The beamformer power at θ is:

$$Q_\theta = \mathbf{w}_\theta^T \mathbf{C} \mathbf{w}_\theta, \quad (3)$$

and it gives a measure of the source activation, usually normalized by some measure of noise, e.g.

$$Z_\theta = \mathbf{w}_\theta^T \mathbf{C} \mathbf{w}_\theta / \mathbf{w}_\theta^T \boldsymbol{\Sigma} \mathbf{w}_\theta, \quad (4)$$

where the noise covariance matrix is approximated by $\boldsymbol{\Sigma} = v\mathbf{I}$, v is the mean sensor noise, and \mathbf{I} is the $M \times M$ identity matrix.

Model search

In the present work, we assume that the sources of electrical activity are equivalent current dipoles immersed in a spherical conducting medium. In typical adult MEG applications of beamformer, the origin of the conducting volume model sphere is assumed to be known via registration markers placed on the subject's head and the volume is scanned on a regular 3D grid to map the model source activation. In the case of fetal MEG, accurate registration of the fetal head has not been available and the beamformer has been used in a model search mode (McCubbin et al., 2007; Robinson and Vrba, 2004) whereby the model sphere origin is also scanned over a regular 3D grid within the maternal abdomen. For each sphere origin, the sphere volume is scanned for source activity, and the model center position for which the maximum activity is obtained is assumed to be the "true" model. The model sphere origin search grid was established by first searching approximately the entire maternal abdominal volume for fetal heart source activity and then using the fetal biometric and orientation data from the US exam along with geometrical constraints to identify a plausible search volume.

The fetal heart location estimate and the individual biometric data were used to construct a plausible fetal MEG search grid. A spherical shell of candidate model origins with 2 cm grid spacing was formed around the fetal heart location. The inner/outer shell radii were set to the individual head to heart distance ± 3 cm since we estimated the uncertainty of the head to heart distance measurement as less than 3 cm. The shell was trimmed to a wedge by intersection of the shell with two planes perpendicular to x-y plane, intersecting at the center of the heart, C, and subtending ± 45 deg angle with C-B line (fig. 1). Finally, only the grid points within the estimated maternal abdominal volume were retained. As a modification to the procedure to accommodate the random field approach, the model origin grid was converted into a model source/model origin grid as follows. A source grid with 1 cm spacing was constructed so as to enclose the origin grid with an additional margin of one head radius. Then all plausible origins were assigned to each source in the grid (i.e. those origins for which the source was closer than the fetal head radius and excluding a 1 cm radius at the model center).

Windowing of data record

The accumulation of covariance for the beamformer analysis assumes that the noise, signal of interest, and interfering source activity are stationary over the duration of the MEG recording. This assumption may be reasonable for an adult MEG measurement but is generally not satisfied for fetal MEG. Not only is the fetal movement uncontrolled, but the general maternal discomfort in the later stages of pregnancy sometimes makes it difficult for the subject to remain still over the entire recording session. We have attempted to minimize the non-stationarity by using a shorter analysis window and sliding it along the available record. The effect may be illustrated by comparing residual averaged heart signal amplitude after OP using either an entire 30 minute record or sequential non-overlapping 2 minute windows for determination of the MCG vectors, fig. 2. The example shows a significant reduction in the residual amplitude for shorter windows and it indicates that, at least for fetal heart analysis, windowing is advantageous. We speculate that over the 30 minute time interval the data is not stationary, but the stationarity improves when shorter time intervals are considered.

For the 10 minute evoked response analysis we used nine analysis windows with 2 minute duration and 1 minute overlap. The 2 minute windows provided better performance as compared with 3, 4, 5, or 10 minute windows over the test data series. The performance was measured by comparing the number of significant records for each processing variation: for 2, 3, 4, 5, and 10 minute analysis window widths we found significance in 6, 4, 3, 4, and 0 of 8 records, respectively. Shorter windows were not investigated because of the loss of SNR due to insufficient noise averaging over fewer trials, and the minimum data requirement for stable covariance matrix (see below). The 2 minute analysis windows provided about 40 trials (40

epochs) for averaging (~3 s inter stimulus interval, ISI, averaged over (-0.5,1.5) s relative to the trigger markers).

Various windows used for the processing described in this paper are shown in fig. 3. The data is divided into the “analysis windows”, fig. 3a, to minimize the effect of non-stationarity. Within each analysis window there are number of “epochs”, typically about 40 (fig. 3b). The epochs within the analysis window are averaged (to the stimulus) to yield an “average epoch”, as in fig. 3c. The average epoch is divided into “response power windows”, as in fig. 3d.

Processing outline

The processing sequence is outlined in the flowchart of fig. 4. The processing was repeated for a number of selected analysis windows. For each analysis window, a statistical test was applied to one or more evoked response power windows in the average. Inputs to the process were the evoked response MEG record and the biometrics data described above. The record was first marked for occurrences of fetal MCG signals using a template matching algorithm (McCubbin et al., 2006; Robinson et al., 2002) and was then segmented into a number of analysis windows. Subsequent processing was repeated for each analysis window, first of which was the fMCG model search and construction of the fMEG search grid described above.

For each analysis window, the covariance matrix and the averaged evoked epoch in the window were computed. The covariance was used to determine beamformer weights at each search grid location (or voxel) θ via eq. 2. Then the averaged epoch for each channel over all MEG channels was used as $\mathbf{m}(n)$ in eq. 1 with the weights to generate an averaged time course at each voxel. A number of post-stimulus response power windows in the average epoch were selected to test for statistically significant evoked response at each voxel using a z-statistic. A clustering algorithm was then used to identify potential regions of activation and RFT was applied to determine a false positive threshold for cluster member voxels. The processing was completed for each hypothesized response power window in the average by tabulating the significant cluster parameters. The sequence was terminated after all selected analysis windows were processed.

Processing details

The covariance matrix was computed from data concatenated from time segments of all epochs within the analysis window. The time segment within an epoch, was (0, 800) ms after stimulus markers with (a) full measurement bandwidth (BW) 0 to 70 Hz, \mathbf{C}_a , and (b) narrow bandwidth 0.5 to 10 Hz, \mathbf{C}_b . This is enough data for a stable covariance provided that the number of samples is at least $3M$ (Van Veen et al., 1997), where M is the number of channels, and the effective sample rate is $f_s = 2 BW$. Then the required time interval is $T = 3 M / (2 BW)$. For $M = 148$, $BW = 9.5$ Hz, the minimum required time is $T = 23.4$ s. The beamformer performance has been shown to be marginal for this number of samples (Brookes et al., 2008) so more data is preferable. With about 40 trials in each 2 minute analysis window, we get 32 s of data for the covariance. Note however that testing with (0, 600), (0, 800), and (0, 1000) ms covariance windows on the test series indicated very little effect (as measured by comparing the number of significant records for each processing variation).

The source orientation for maximum beamformer power and beamformer weights were determined using the matrix \mathbf{C}_a and the procedure developed by Sekihara et al. (2004). The weights are obtained as

$$\mathbf{w}_\theta = \frac{\mathbf{B}_\theta (\mathbf{C}_a + \mu \boldsymbol{\Sigma})^{-1}}{\mathbf{B}_\theta^T (\mathbf{C}_a + \mu \boldsymbol{\Sigma})^{-1} \mathbf{B}_\theta}, \quad (8)$$

where the forward solution \mathbf{B}_θ corresponds to the source orientation along the maximum beamformer power determined from \mathbf{C}_a . The beamformer power is normalized by noise, computed using weights from eq.8 and the narrow band covariance matrix \mathbf{C}_b as

$$Q_\theta = \mathbf{w}_\theta^T \mathbf{C}_b \mathbf{w}_\theta / \mathbf{w}_\theta^T \Sigma \mathbf{w}_\theta. \quad (9)$$

The weights corresponding to the model origin which produced the largest beamformer power were assigned to a given source grid location. A regularization factor of 10^3 was selected from test series results over $\mu = 10^x$, $x = 2, 3, 4$, and 6. The selection was based on maximizing the number of significant records over the processing variations.

The “true average” was taken from the average epoch as the interval (0, 800) ms relative to stimulus. The markers within the analysis window were then randomized (about ± 2 s, based on minimum ISI) to create a set of 500 random averages in the selected analysis window. For clarity, the averaging is described in detail as follows.

Source activity s_θ is a $1 \times N$ matrix (row vector). Assume that the true triggers are at times n_k , $k = 1, \dots, K$, where K is the number of trials. Then the source activity within an epoch is

$$s_{\theta k}(n_t) = s_\theta(n_k + n_t) \quad (10)$$

where n_t corresponds to a point in the post-trigger interval of interest. Time averaged source activity (averaged epoch) is

$$\bar{s}_{\theta \text{ true}}(n_t) = \frac{1}{K} \sum_{k=1}^K s_\theta(n_k + n_t) = \frac{1}{K} \sum_{k=1}^K \mathbf{w}_\theta^T \mathbf{m}(n_k + n_t) = \mathbf{w}_\theta^T \bar{\mathbf{m}}_{\text{true}}(n_t) \quad (11)$$

where $\bar{\mathbf{m}}_{\text{true}}(n_t) = \frac{1}{K} \sum_{k=1}^K \mathbf{m}(n_k + n_t)$ is averaged data. Let the randomized triggers be at times n_r^i , $r = 1, \dots, R$, where $R \sim K$ is the number of randomized trials and $i = 1, \dots, N_{\text{rand}}$, with N_{rand} as the number of trigger set randomizations. Then

$$\bar{s}_{\theta \text{ rand}}(n_t, i) = \frac{1}{R} \sum_{r=1}^R s_\theta(n_r^i + n_t) = \frac{1}{R} \sum_{r=1}^R \mathbf{w}_\theta^T \mathbf{m}(n_r^i + n_t) = \mathbf{w}_\theta^T \bar{\mathbf{m}}_{\text{rand}}(n_t, i) \quad (12)$$

where $\bar{\mathbf{m}}_{\text{rand}}(n_t, i) = \frac{1}{R} \sum_{r=1}^R \mathbf{m}(n_r^i + n_t)$ is the set of random averaged data.

The test statistic, x , was taken as the mean power over a response power window in the post-trigger interval $(n_o, n_o + N_o - 1)$, with n_o as the first time point in the response power window and N_o as the response power window width,

$$x_{\text{true}} = \frac{1}{N_o} \sum_{q=1}^{N_o} \bar{s}_{\theta \text{ true}}^2(n_o + q - 1) \quad (13)$$

$$x_{\text{rand}}(i) = \frac{1}{N_o} \sum_{q=1}^{N_o} s_{\theta \text{ rand}}^2(n_o + q - 1, i), \quad i=1, \dots, N_{\text{rand}}. \quad (14)$$

Since the expected latency for a specific fetal response is unknown, we have selected $N_o = 200$ ms wide response power window sliding over the post stimulus interval from 0 to 500 ms in steps of 50 ms, $n_o = 0, 50, 100, 150, 200, 250, 300$ ms.

For each response window of interest, the z-statistic,

$$z = (x_{\text{true}} - \overline{x_{\text{rand}}}) / \sigma_{\text{rand}} \quad (15)$$

was computed where x_{true} is mean response power of the true average, $\overline{x_{\text{rand}}}$ is the mean of the mean response power over the N_{rand} random averages, and σ_{rand} is the standard deviation of the random averages mean response power, $x_{\text{rand}}(i)$, $i = 1, \dots, N_{\text{rand}}$. We justify the use of estimates for the random response power distribution mean and standard deviation in place of the population parameters by the large value of N_{rand} . For $N_{\text{rand}} = 500$, we considered the bias correction for σ_{rand} of about 0.05% to be negligible (Montgomery and Runger, 2003).

Candidate clusters of potentially significant sources were identified using the method of affinity propagation (Frey and Dueck, 2007). The clustering was based on an input ‘similarity’ matrix $\zeta(i, j)$ over all combinations of two source voxels, computed as the negative squared Euclidian distance between voxel locations (i, j) plus negative squared difference in z values (an alternative is to add negative squared difference in resels volume, defined below, but we did not find any clear advantage in the test series analysis). The distance factor was multiplied by a penalizing factor of 5 for all distances larger than one voxel cube diagonal ($\sqrt{3} \times \text{grid spacing}$). The penalizing factor value is not too important but it needs to be large enough to exclude distant voxels from the same cluster. A variety of cluster sizes were generated for evaluation by setting the ‘preferences’ or self similarities, $\zeta(i, i)$, to a common value in the set $-a \times 10^{-b}$, $a = 1, 3, 5, 7$ and $b = 1, 2, 3, 4$. Taking a maximum of 30 sources having $z > 2$, the affinity propagation method estimated a number of clusters (ranging from 1 to 30) that were then used for further evaluation in a statistical significance test. The clusters were pruned to be contiguous.

For a more stringent statistical test aimed at reduction of false positives while maintaining statistical power, we have modified a RFT false positive threshold described by Barnes and Hillebrand (2003) developed for adult MEG beamformer SPM analysis. The enabling equation, due to Friston et al. (1996), expresses the probability p of obtaining at least one cluster of a selected spatial extent, r , which would exceed an unspecified threshold, U , as

$$p = 1 - e^{-E \cdot P} \quad (16)$$

where E is the expected number of clusters to exceed U and P is the probability that the cluster size is at least ‘ r ’ resolution elements or ‘resels’. A cluster is defined as one or more contiguous source voxels in the search volume. Given a selected false positive rate for the experiment, p , we want to solve eq. 16 for U at each cluster of voxels. This provides an adjustment on p to account for multiple comparisons over the SPM and reflects the variable spatial smoothness of the statistical parameter over the volume of interest.

The method estimates local SPM smoothness by counting the number of resels, r , at each cluster and requires the transformation from a cubic grid of voxels to a tetragonal mesh with 5 tetrahedra per cube. Then changes in parameters relative to all neighbors of a voxel are referenced for use in the flattening via resel sums. First, parameters of the entire search volume geometry, called ‘resel counts’, are tabulated. We have an arbitrary shaped volume so we used the procedure for estimating resel counts over a general search volume according to Worsley et al. (1996), which results in estimates for 4 parameters of the resel volume: R_3 – total resel volume, R_2 – resel surface area, R_1 – resel diameter, and R_0 – Euler Characteristic (EC) of the volume.

We computed the resel estimate, r , (Barnes and Hillebrand (2003)eq. 15) for each cluster and for each associated tetrahedron, utilizing the clever matrix formulation with beamformer weight vectors of Barnes and Hillebrand (2003)eq. 17 – 20. Then, we summed the resels over all member tetrahedra in each cluster (Friston et al. 1996) to get r_{sum} . To get the local resel counts at each cluster, we applied the following transformation from resel space to voxel space (Worsley et al. (1996), for our case, with isotropic voxel spacing, $d, r_x = r_y = r_z = r_0 = d/$ FWHM (full width half maximum) of the Gaussian point spread function, eq. 3.2: $V(k)=R_k r_0^k$ and Barnes and Hillebrand (2003), eq. 16: FWHM= $d/(3 \sqrt{r_{sum}})$):

$$V(k)=R_k(3 \sqrt{r_{sum}})^k, k=0, \dots, 3. \quad (17)$$

E in eq. 16 above is calculated as the scalar product of the local resel count vector, $\mathbf{V} = (V(0), V(1), V(2), V(3))$, and the 4-dimensional EC density function for a Gaussian distribution (we are using a z-statistic, eq. 15, while Barnes and Hillebrand (2003) used a t-statistic and EC density functions for a t-distribution; see Worsley et al., 1996). The EC density functions are not to be confused with R_0 defined above as EC of the volume. The later is a constant to describe the geometrical shape of the search volume while the former are functions of U which describe the statistical field. In our case, the EC density functions simplify to:

$$\begin{aligned} \rho(1) &= \frac{1}{2}(1 - \operatorname{erf}(U/\sqrt{2})) \\ \rho(2) &= \frac{1}{2\pi}(4\log(2))^{1/2}e^{-U^2/2} \\ \rho(3) &= \frac{1}{(2\pi)^{3/2}}4\log(2)Ue^{-U^2/2} \\ \rho(4) &= \frac{1}{(2\pi)^2}(4\log(2))^{3/2}(U^2 - 1)e^{-U^2/2} \end{aligned} \quad (18)$$

and

$$E = \mathbf{V} \cdot \boldsymbol{\rho}. \quad (19)$$

After some simplification from the expression in Barnes and Hillebrand (2003),

$$\begin{aligned} P &= e^{-\varphi}, \\ \varphi &= \left(\frac{3}{4} \sqrt{\pi} E / \left\{ \frac{1}{2} R_3 \left[1 + \operatorname{erf} \left(-U/\sqrt{2} \right) \right] \right\} \right)^{2/3}. \end{aligned} \quad (20)$$

To get a RFT threshold value at each cluster, we numerically solved eq. 16 for the threshold, U , given the selected nominal p-value threshold, p . Statistically significant clusters were identified by $z - U > 0$ for all cluster members.

Finally, a summary of the search was saved for analysis and interpretation. The summary included: (a) summary table of all search parameters and raw output, (b) sensor array projections with passed cluster(s) source/origin locations, and (c) weights for reconstructing time course(s) of selected averaged source activity.

It should be noted that the single source time courses are statistically ambiguous with the cluster analysis. We can only say that there was significant activity somewhere within the cluster volume. The median time course over all cluster members may be plotted as a representative activity but its interpretive value is limited, especially when the SNR is poor as we show in the next section.

Latency estimation

After looping through all selected power windows for each analysis window of a dataset, the evoked response latency was estimated as the midpoint of the selected significant power window. The latency could be defined as the time between stimulus and detectable onset of the response but in neurophysiological applications it is usually taken as the time between stimulus and the first peak in the response amplitude. Therefore we took the midpoint of the significant response window which best represented the peak in the source cluster time course as a latency estimate. The error in this latency measure may be estimated as \pm one half window width. Narrower power window width would improve the latency measurement error, but at some limit there will not be enough power in the window to distinguish above background, so we trade resolution for sensitivity.

Results

A number of test analyses and results are reported in this section which may be summarized as follows. A Monte Carlo simulation provided a test of false positive rates for the processing method. An example of the fetal heart localization is presented to demonstrate the need for windowing of the data record. The detection rate for the study series is reported and the validity of the processing is demonstrated for newborn data. An example of response detection across response power windows demonstrates the latency estimation procedure. Finally, a plot of latency estimation for the study series shows the decrease in latency with GA and the graphical display is quantified by a table with correlation of latency and GA.

Monte Carlo

We conducted a Monte Carlo simulation with Gaussian noise in place of the MEG channel data to test false positive rates. The noise amplitude was independently computed for each MEG channel and for each time point in the record by selecting a random value in the zero mean Gaussian distribution with variance of 25 fT. The observed false positive rates for expected thresholds of 1, 5, and 10% were estimated from 1000 Monte Carlo iterations, summarized in the plot of fig. 5. Tests using single voxels or clusters both became overly conservative at 10%, possibly due to the method's suitability only for smaller false positive rates. This is supported by the single source rate, which seems to be approaching the ideal for smaller false positive rates, however clustering appears to remain somewhat optimistic for smaller false positive rates. Further refinements in the clustering method may be indicated since we are perhaps biasing the statistics by testing multiple clusterings (due e.g. to limitations in the cluster size parameter selection and missing condition for contiguity in the clustering algorithm). Nonetheless, our choice of a 10% false positive rate should be somewhat conservative for the results which follow.

Fetal MCG variability

For each selected analysis window, a preliminary search was conducted over approximately the entire maternal abdominal volume with the fMCG markers as trial markers. This resulted in an estimate of the fetal heart location in MEG coordinates and provided a registration with the US biometric data as a basis for search grid construction. To show the general benefit of estimating the fetal heart location for each analysis window, an example of the variation among test datasets in fetal heart movement is shown in fig. 6. In some cases (fig. 6.a) there is very little heart movement while in others (fig. 6.b) there is considerable movement.

Detection rate summary

A statistically significant response was detected for 56 of 90 fetal records (62%) and 20 of 21 newborn records (95%). The OP processed averaged evoked response over the entire 10 minute stimulated record for the one failed newborn case did show a response well above baseline, so the failure was likely due to insufficient events to average in the 2 minute analysis windows. Only 21 fetal records had more than one significant analysis window and the maximum was four (for 1 record). The newborn records indicated better but still limited stationarity with a maximum of 7 of 9 significant analysis windows (2 records) and 16 records with more than one but only 7 records with more than four.

Newborn example

An example of the model search results and comparison with OP processing for a newborn record is shown in fig. 7. The search grid construction for the newborn records was necessarily modified from the fetal procedure, however the procedure was otherwise identical to the fetal processing so the reliable detection serves as a check on the method. All successful newborn model search results had a favorable comparison with the OP processed result over the same analysis window except one for which the OP result was quite noisy. However, the OP processed average over the entire 10 minute record did provide a response so the OP failure was likely due to insufficient trials to average.

It is important to note for the interpretation of the model search time course plot of fig. 7 (and those that follow) that the individual source activity within a significant cluster has limited value. We can only say that there was statistically significant mean power over the selected response power window somewhere within the cluster and we cannot even attach meaning to the individual peaks of the median time course.

Latency estimation

The results for all statistically significant analysis windows for an individual subject and GA were reviewed in order to estimate a response latency. The latency was defined as the midpoint of the significant response power window most closely associated with the first peak in median power over all significant source clusters. The averaged time courses of all passed cluster sources for significant response power windows were computed via eq. 1 and plotted for interpretation. An example of the latency selection of 300 ms for a strong response peak is shown in fig. 8. In this example all response power windows were significant for a single analysis window (4 of 9) which is convenient for demonstration but was not generally observed. Here the peak response power window could have been selected as in fig. 8d, e, f, or g and the choice was based on the smoothed nominal waveforms (including smaller cluster medians) and neglected the sharper spikes associated with noise. This visual selection process could be automated to improve reliability, however the implementation would be complicated by large variability of the data, including noise, artifacts, and false positives.

In some cases, an unexpectedly early peak response was detected, as exemplified in fig. 9, with a latency around 100 ms or earlier. These peaks were identified as artifact, as discussed below, and excluded from the latency selection.

In other cases, a statistically significant response power was associated with a time course which consisted of narrow spikes rather than an expected broader morphology. An example is shown in fig. 10 and these were rejected as false positive responses.

The selected response power window results for each subject over all GA sessions were then plotted as shown in the example of fig. 11 and latency estimates were tabulated. As the example demonstrates, the response time course was not always robust and sometimes there was only one single voxel cluster available to estimate latency (fig. 11c and 11d). The example of fig. 11e and 11f shows the latency selected from the earlier peak even though it was not produced by the largest cluster. After rejection of artifacts and false positives, many subjects in the study series were reduced to one or two GAs for latency estimates (including newborn records). In fact only 6 subjects had more than two GAs for tabulation.

Response latency vs. GA

Latency estimation for the study series is summarized in the plot of fig. 12 for 27 of 29 subjects including 8 with only one selected latency estimate. The remaining two subjects in the series had no significant GA records. It was observed in fig. 12a that several early GA latency estimates were unusually short and a few late GA (newborn) latency estimates were surprisingly long. Two of the 19 subjects with more than one latency estimate had a GA record with latency that was substantially longer (outside the ± 100 ms latency error estimate, fig. 12b) than the first measured GA record latency and were marked as a contrary sub-group. Of the remaining 17 subjects (89%), 9 demonstrated latency reduction with GA as expected for normal development. The other eight subjects showed no meaningful change of latency with GA (remaining within the ± 100 ms latency error estimate, fig. 12b).

Latency vs. GA correlation

For those subjects with more than two GA records, the least squares fit slope and associated correlation coefficients are listed in Table 1. The mean slope was -15 ms/week. Considering the ± 100 ms error bars on each latency estimate for an individual subject, it can be simply shown that the average slope is equal to the slope determined directly from the data (column 2). Standard deviation of the slope, column 4, was computed by simulation using large number (30,000) of various error realizations, assuming that the data errors are uniformly distributed with zero mean. Taking all data of fig. 12a together, the fit slope was -5 ms/week with a correlation coefficient of only 0.3. However, when a few outliers were discarded, the fit slope increased to -11 ms/week with correlation coefficient of 0.6 (plotted as a dotted line in fig. 12a). This analysis provides a numerical value for the general trend observed in the plots of fig. 12.

Discussion

We have implemented a statistical procedure based on RFT to control false positives in the SPM from a fetal MEG beamformer model search. The performance of the algorithm was tested for false positives by Monte Carlo method and found to be somewhat optimistic at lower thresholds, but conservative for a higher p-value threshold that we used (10%). The effect of the RFT procedure was to reduce the nominal p-value threshold for significance so we attempted to optimize the model search parameters to provide best SNR data. The search volume was individually constructed for each analysis window of interest in the record based on ultrasound biometrics. The search volume was referenced to the estimated fetal heart

location averaged over the analysis window of interest. In previous work an ‘actogram’ was used (McCubbin et al., 2007) to identify fetal movement based on MCG variations and then select segments with relatively stable MCG. However, there is not a clear correlation between torso, limb, and head movements nor were maternal movements monitored. Furthermore, there undoubtedly exist additional contributors to non-stationarity within the maternal abdomen and even within the fetal head (e.g. eye and mouth movements, spontaneous burst activity).

As a best effort to deal with this situation, we chose an analysis window width of 2 minutes and stepped it along the 10 minute record in 1 minute increments. In general, wider windows produced lower detection rates and narrower windows were not considered because there would not be enough data for construction of the beamformer covariance. Under the assumption that fetal or maternal movements are a major source of the non-stationarity, the analysis window length could alternatively be determined by some movement measure such as fetal heart position variability. However the correlation between fetal heart movement and fetal head movement might not always be strong. Additionally, there may be other unidentified contributors to non-stationarity. A possible improvement would be an automated step-down testing of window widths from e.g. 5 minutes to optimize for those records in a series with better stationarity. An additional enhancement to the model search was the variable width and location of the post-stimulus response power window. This was considered important for fetal evoked response detection because of the unknown and variable latency or duration.

It’s interesting to note that, even with the relatively high SNR conditions of the newborn measurement, detection is not assured. This may be partly due to inappropriate search grid selection (newborn positioning was not too well known, grid was too coarse for high SNR application) or insufficient spatial sampling of the sensor array, but also likely includes the nonstationarity due to newborn movements as well as the attending mother’s movements. The evidence from tabulation of passed analysis window counts indicates a maximum stationarity of four windows for fetal data and seven for newborn data. Nearly half of the fetal records had only a single significant 2 minute window while only about half of the newborn records had more than 3 significant windows.

The observation of statistically significant response power at less than about 100 ms latency was a complication in the latency estimation. We justify the identification of this component as artifact by noting that there are no known reports of neonate newborn flash evoked response latency below about 200 ms (Kræmer et al., 1999; Tsuneishi et al., 1995). The origin of the artifact could be an event related eye blink or limb twitch, however this will have to be the subject of a future investigation.

The most convincing validation of fetal evoked response detection is the demonstration of some neurophysiological marker using a standardized processing over a large group of records. We selected the processing parameters by detection performance over a small test series and then processed a large series of different records as a batch with the same processing. The detection rate was 56 of 90 fetal records (62%) and 20 of 21 newborn records (95%). The number of fetal latency estimates were further limited by rejection of artifacts and false positives, nonetheless the correlation between estimated latency and GA revealed the expected latency decrease with GA over the group in general at -11 ms/ week and over individual subjects for most subjects in rough agreement with Kiefer et al. (2008). Individually, the latency reduction was quite variable. Over subjects with more than 2 GA latency estimates, the slope ranged from -30 to $+5$ ms/week with a mean of -15 ms/week, which is in agreement with the general trend. The results of the fetal evoked response analysis may not seem too impressive to those with experience in postnatal or adult electrophysiology, however we believe that this report represents the most convincing evidence for reliable detection of evoked fetal brain activity to date.

While there are many improvements that could be made to optimize the performance, we have shown that a beamformer model search with appropriate parameters and careful control of false positives in the search volume has potential for reliable and routine latency detection of fetal evoked responses. The clustering method was somewhat cumbersome and could be replaced by a more general implementation of the RFT which would identify statistically significant activation clusters over the whole search space without the need for the separate clustering step. We intend to explore the parameter space for optimizations, consolidate the processing software, and apply the method to larger fetal data series. The averaging of evoked response over many trials reduces the noise but also may diminish the response amplitude if the latency and duration varies over trials. In future work it may be useful to devise an experiment for a paired sample statistical test for comparison with these results.

Acknowledgements

The work was supported by NIH grant R01NS036277.

References

- Barnes GR, Hillebrand A. Statistical flattening of MEG beamformer images. *Hum. Brain Map* 2003;18:1–12.
- Benjamini Y, Heller R. Screening for parital conjunction hypotheses. *Biometrics* 2008;64(4):1215–1222. [PubMed: 18261164]
- Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilit. *Publicazioni del R. Istituto Superiore di Scienze Economiche e Commerciali di Firenze* 1936;8:3–62.
- Brookes MJ, Vrba J, Robinson SE, Stevenson CM, Peters AM, Barnes GR, Hillebrand A, Morris PG. Optimising experimental design for MEG beamformer imaging”. *NeuroImage* 2008;39:1788–1802. [PubMed: 18155612]
- Carbonell F, Galán L, Valdés P, Worsley KJ, Biscay RJ, Díaz-Comas L, Bobes MA, Parra M. Random field-union intersection tests for linear electrophysiological imaging. *Neuroimage* 2004;22:268–276. [PubMed: 15110017]
- Carbonell, F.; Worsley, KJ.; Trujillo-Barreto, N.; Sotero, RC. *Human Brain Mapping*. 2009. Random field-union intersection tests for detecting functional connectivity in EEG/MEG imaging. in press
- Chumbley JR, Friston KJ. False discovery rate revisited: FDR and topological inference using Gaussian random fields. *NeuroImage* 2009;44:62–70. [PubMed: 18603449]
- Dragonova R, Eswaran H, Murphy P, Lowery C, Preissl H. Serial magnetoencephalographic study of fetal and newborn auditory discriminative evoked responses. *Early Hum. Dev* 2007;83:199–207. [PubMed: 16863685]
- Frey BJ, Dueck D. Clustering by passing messages between data points. *Science* 2007;315:972–976. [PubMed: 17218491]
- Friston KJ, Holmes A, Poline J-B, Price CJ, Frith CD. Detecting activations in PET and fMRI: levels of inference and power. *NeuroImage* 1996;40:223–235. [PubMed: 9345513]
- Eswaran H, Wilson JD, Preissl H, Robinson SE, Vrba J, Murphy P, Rose DF, Lowery CL. Magnetoencephalographic recordings of visual evoked brain activity in the human fetus. *The Lancet* 2002;360:779–780.
- Govindan RB, Wilson JD, Preissl H, Murphy P, Lowery CL, Eswaran H. An objective assessment of fetal and neonatal auditory evoked responses. *NeuroImage* 2008;43:512–527.
- Gowland P, Fulford J. Initial experiences of performing fetal fMRI. *Exp. Neurol* 2004;190:S22–S27. [PubMed: 15498538]
- Kiefer I, Siegel E, Preissl H, Ware M, Schauf B, Lowery C, Eswaran H. Delayed maturation of auditory-evoked response in growth-restricted fetuses revealed by magnetoencephalographic recordings. *Am. J. Obstet. Gynecol* 2008;199:503.e1–503.e7. [PubMed: 18533116]
- Krämer M, Abrahamsson M, Sjöström A. The neonatal development of the light flash visual evoked potential. *Doc. Ophthalmologica* 1999;99:21–39.

- Lengle JM, Chen M, Wakai RT. Improved neuromagnetic detection of fetal and neonatal auditory evoked responses. *Clin. Neurophysiol* 2001;112:785–792. [PubMed: 11336893]
- Li SS, Bigler J, Lampe JW, Potter JD, Feng Z. FDR-controlling testing procedures and sample size determination for microarrays. *Statist. Med* 2005;24:2267–2280.
- Lowery CL, Eswaran H, Murphy P, Preissl H. Fetal Magnetoencephalography. *Sem.Fetal Neonatal Med* 2006;11:430–436.
- Marchini J, Presanis A. Comparing methods of analyzing fMRI statistical parametric maps. *NeuroImage* 2004;22:1203–1213. [PubMed: 15219592]
- McCubbin J, Robinson SE, Cropp R, Moiseev A, Vrba J, Murphy P, Priessl H, Eswaran H. Optimal reduction of MCG in fetal MEG recordings. *IEEE Trans. Biomed. Eng* 2006;53:1720–1724. [PubMed: 16916111]
- McCubbin J, Murphy P, Eswaran H, Preissl H, Yee T, Robinson SE, Vrba J. Validation of the flash-evoked response from fetal MEG. *Phys. Med. Biol* 2007;52:5803–5813. [PubMed: 17881801]
- Montgomery, DC.; Runger, GC. *Applied Statistics and Probability for Engineers*. Vol. 3rd ed. New York: Wiley and sons; 2003.
- Porcaro C, Zappasodi F, Barbati G, Salustri C, Pizella V, Rossini PM, Tecchio F. Fetal auditory responses to external sounds and mother's heart beat: detection improved by independent component analysis. *Brain Res* 2006;1101:51–58. [PubMed: 16784726]
- Robinson, SE.; Vrba, J.; McCubbin, J. Separating fetal MEG signals from the noise. In: Nowak, H., et al., editors. *Biomag 2002*. Berlin: VDE Verlag GmbH; 2002. p. 665-667.
- Robinson, SE.; Vrba, J. In: Halgren, E.; Ahlfors, S.; Hamalainen, M.; Cohen, D., editors. *Cleaning fetal MEG using a beamformer search for the optimal forward model; Proceedings of the 14th International Conference on Biomagnetism; 2004*. p. 339-340.
- Schleussner E, Schneider U. Developmental changes of auditory-evoked fields in fetuses. *Exp. Neurol* 2004;190(Suppl 1):S59–S64. [PubMed: 15498543]
- Schwartzman A, Dougherty RF, Lee J, Ghahremani D, Taylor JE. Empirical null and false discovery rate analysis in neuroimaging. *NeuroImage* 2009;44:71–82. [PubMed: 18547821]
- Sekihara K, Nagarajan S, Poeppel D, Marantz A. Asymptotic SNR of scalar and vector minimum-variance beamformers for neuromagnetic source reconstruction. *IEEE Trans. Biomed. Eng* 2004;51:1726–1734. [PubMed: 15490820]
- Sheridan CJ, Preissl H, Siegel E, Murphy P, Ware M, Lowery CL, Eswaran H. Neonatal and fetal response decrement of evoked responses: a MEG study. *Clin. Neurophysiol* 2008;119:796–804. [PubMed: 18226946]
- Taylor JE, Worsley KJ. Detecting sparse signal in random fields, with an application to brain mapping. *Journal of the American Statistical Association* 2007;102:913–928.
- Tsuneishi S, Casaer P, Fock JM, Hirano S. Establishment of normal values for flash evoked potentials (VEPs) in preterm infants: a longitudinal study with special reference to two components of the N1 wave. *Electroenceph. Clin. Neurophysiol* 1995;96:291–299. [PubMed: 7635074]
- Van Veen BD, van Drongelen W, Yuchtman M, Suzuki A. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans. Biomed. Eng* 1997;44(9):867–880. [PubMed: 9282479]
- Vrba J, Robinson SE, McCubbin J, Lowery L, Eswaran H, Murphy P, Preissl H. Searching for the best model: ambiguity of inverse solutions and application to fetal magnetoencephalography. *Phys. Med. Biol* 2007;52:757–776. [PubMed: 17228119]
- Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum. Brain Map* 1996;4:58–73.
- Worsley KJ, Andermann M, Koulis T, MacDonald D, Evans AC. Detecting changes in nonisotropic images. *Hum. Brain Map* 1999;8:98–101.
- Wilson JD, Adams AJ, Murphy P, Eswaran H, Preissl H. Design of a light stimulator for fetal and neonatal magnetoencephalography. *Physiol. Meas* 2008;30(1):N1–N10. [PubMed: 19104135]

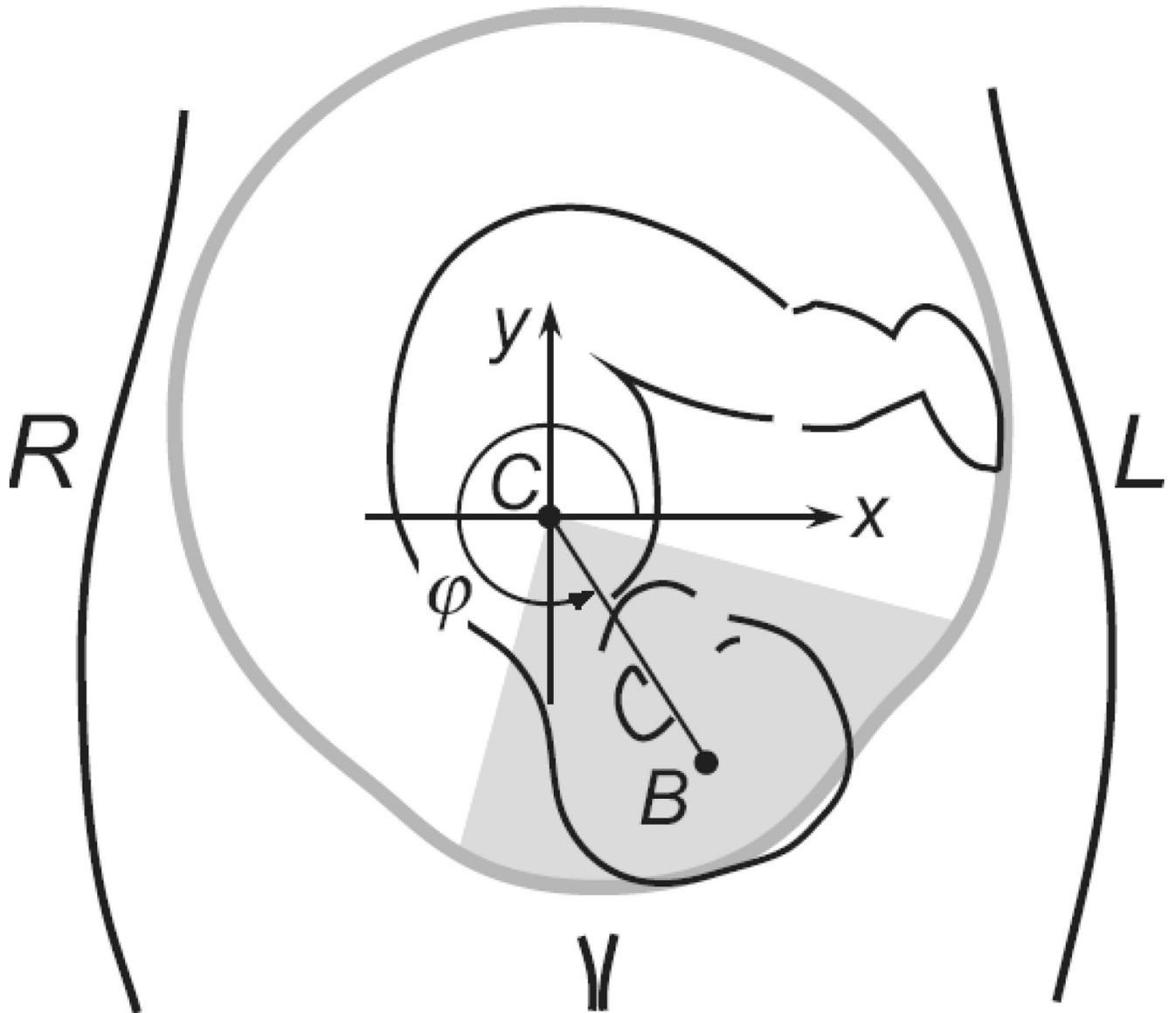


Figure 1. Fetal orientation sketch example with orientation angle (ϕ) measurement. The x and y coordinates are indicated in the figure, z coordinate is perpendicular to the figure plane. Shaded region indicates a ± 45 deg wedge used to limit the search grid. C – heart center, B – head center.

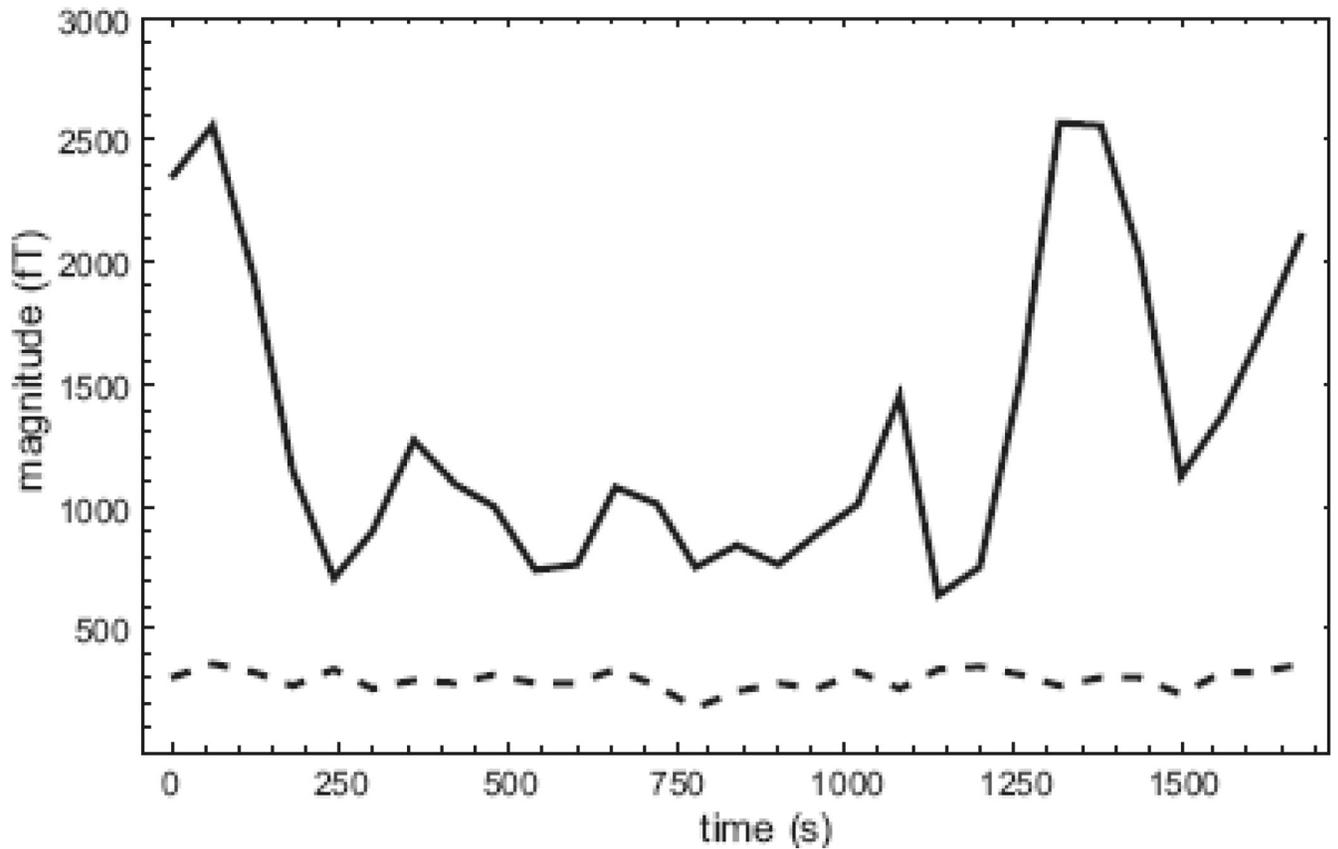


Figure 2. Residual average mMCG peak vector magnitude vs. time over a 30 minute fMEG record with bandwidth 0 – 70 Hz. The residual was computed for each of 15 non-overlapping 2 minute analysis windows after OP using either vectors determined once over the entire record without windowing (solid line), or over each 2 minute window (dotted line).

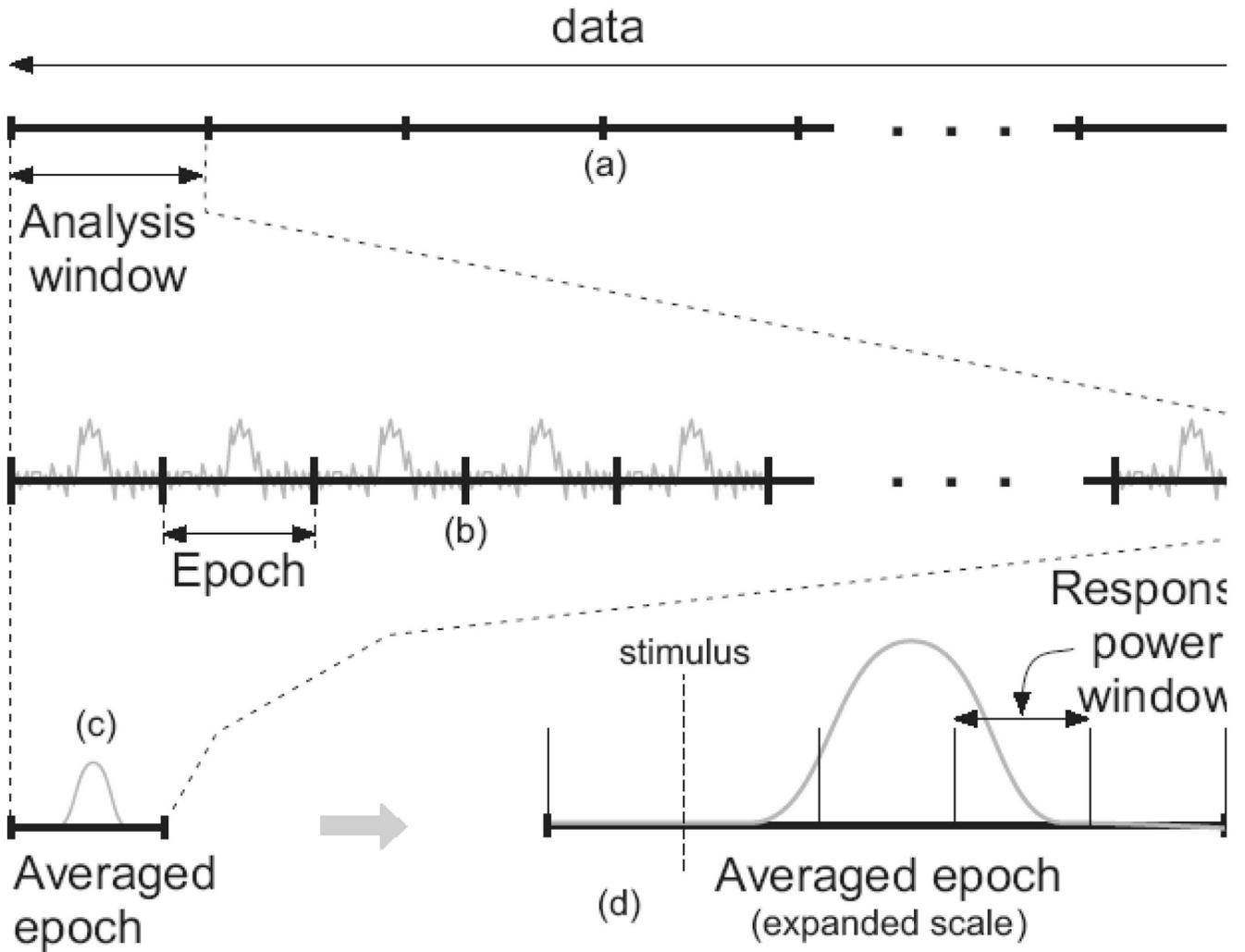


Figure 3. Schematic diagram of various windows and their naming used in this paper. (a) data record segmented into analysis windows, (b) evoked response epochs in one analysis window, (c) averaged epoch from all epochs in an analysis window, and (d) averaged epoch segmented into response power windows. Analysis windows and response power windows may overlap.

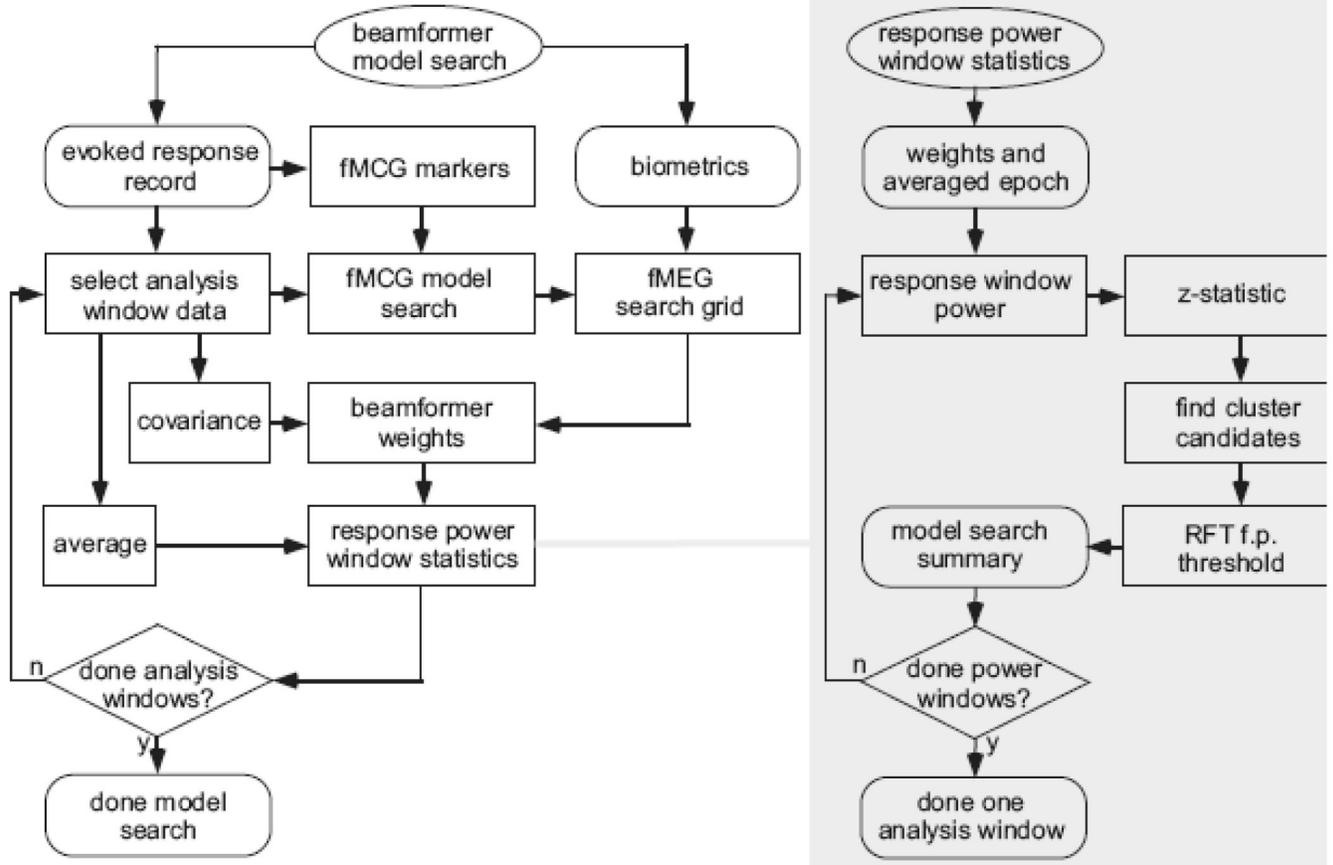


Figure 4. Beamformer model search processing flowchart. Detail flowchart of the process ‘response power window statistics’ is highlighted in gray.

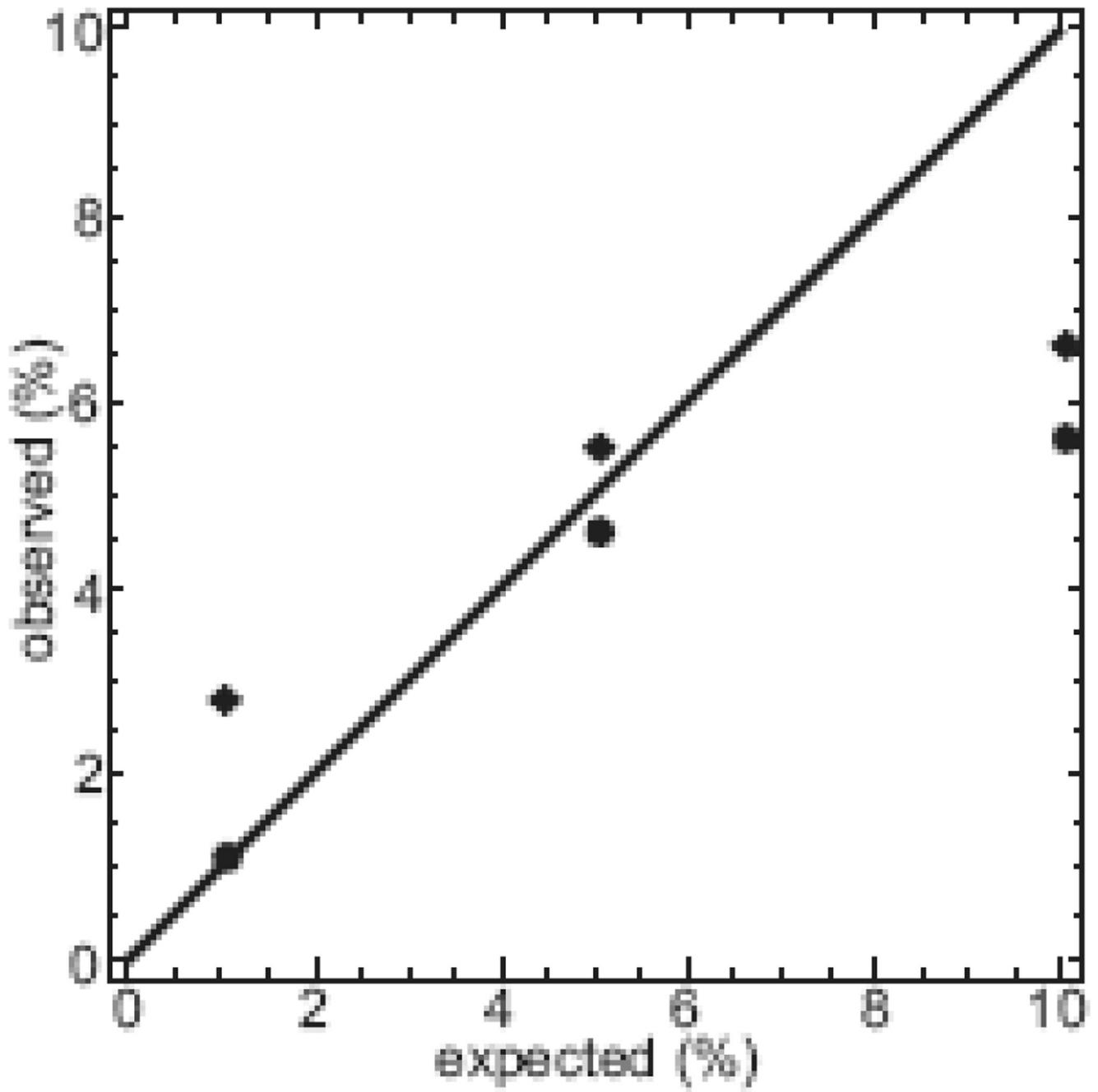


Figure 5. False positive rate (%), expected vs. observed single voxel (circle) and clusters (diamond). Solid line represents observed = expected.

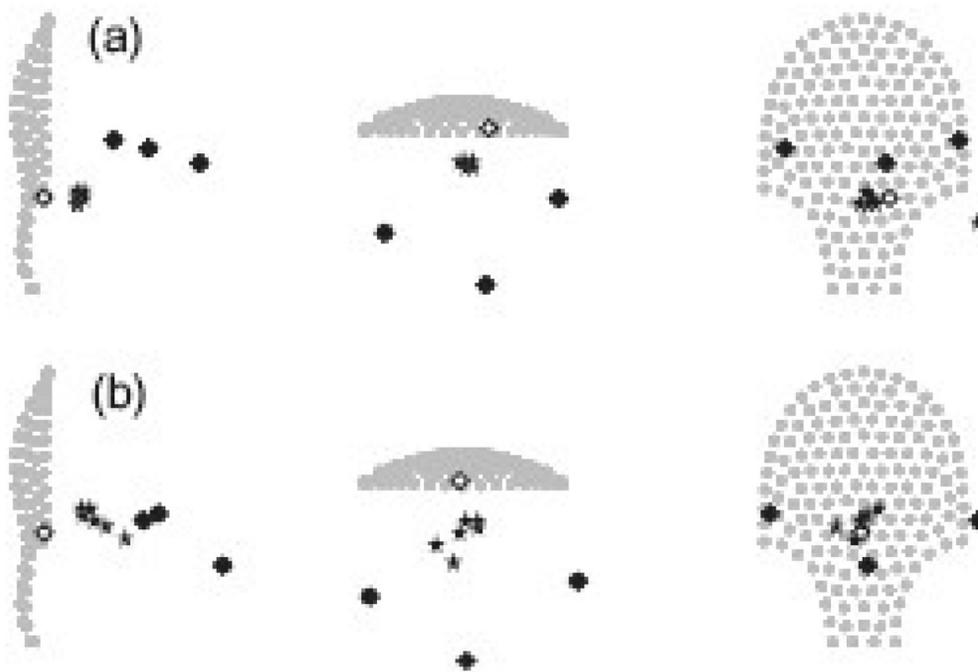


Figure 6. Heart location estimates (stars) over 9 visual stimulus windows with 2 minute duration. The quietest (a) and most active (b) examples over the test series are superimposed on projections of the fMEG sensor array (gray dots) with maternal abdominal markers (open and filled diamonds).

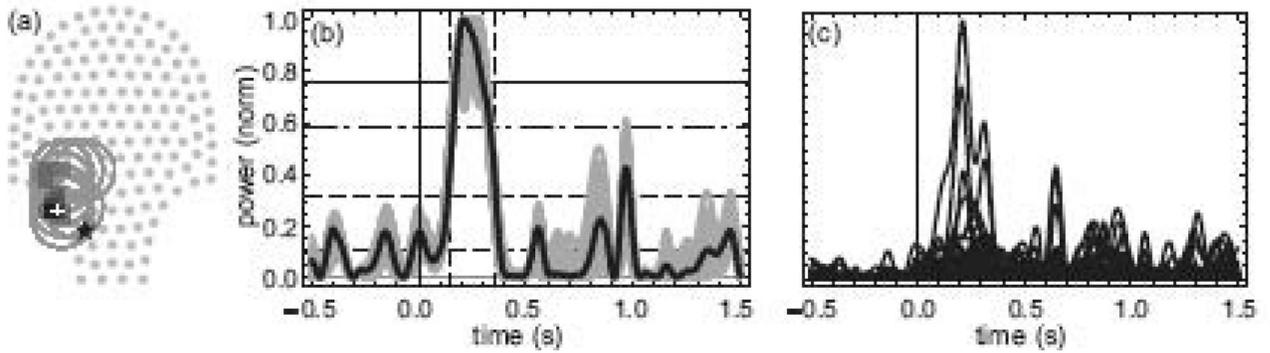


Figure 7.

Example result for a newborn record. (a) The projection onto the x-y plane of the sensor array (gray dots) with newborn heart location estimate (star), all members of all source clusters which passed statistical false positive threshold of 0.10 (gray squares, gray level is a function of source metric value ($z - U$) with white as zero level so that sources with largest metric values are darkest; overlaid with largest values on top), and associated model spheres (gray circles), one largest cluster center location (white cross); (b) The median time course for the cluster with the largest number of members (black) and smaller clusters median (gray) time courses are overlaid (normalized power vs. time in sec from trigger), dashed vertical lines: response power window limits. Horizontal lines: dashed lines are $\overline{x_{\text{rand}}} - \sigma_{\text{rand}}$, $\overline{x_{\text{rand}}} + \sigma_{\text{rand}}$ (random averages power \pm one standard deviation), dash-dotted line is $x_{\text{v}} = U\sigma_{\text{rand}} + \overline{x_{\text{rand}}}$ (RFT threshold power, please refer to eq. 15), and solid line is x_{true} (mean power over response power window). The statistical parameters were taken from the one cluster member with best metric value ($z - U$); (c) Superposition of squared MEG channels after OP.

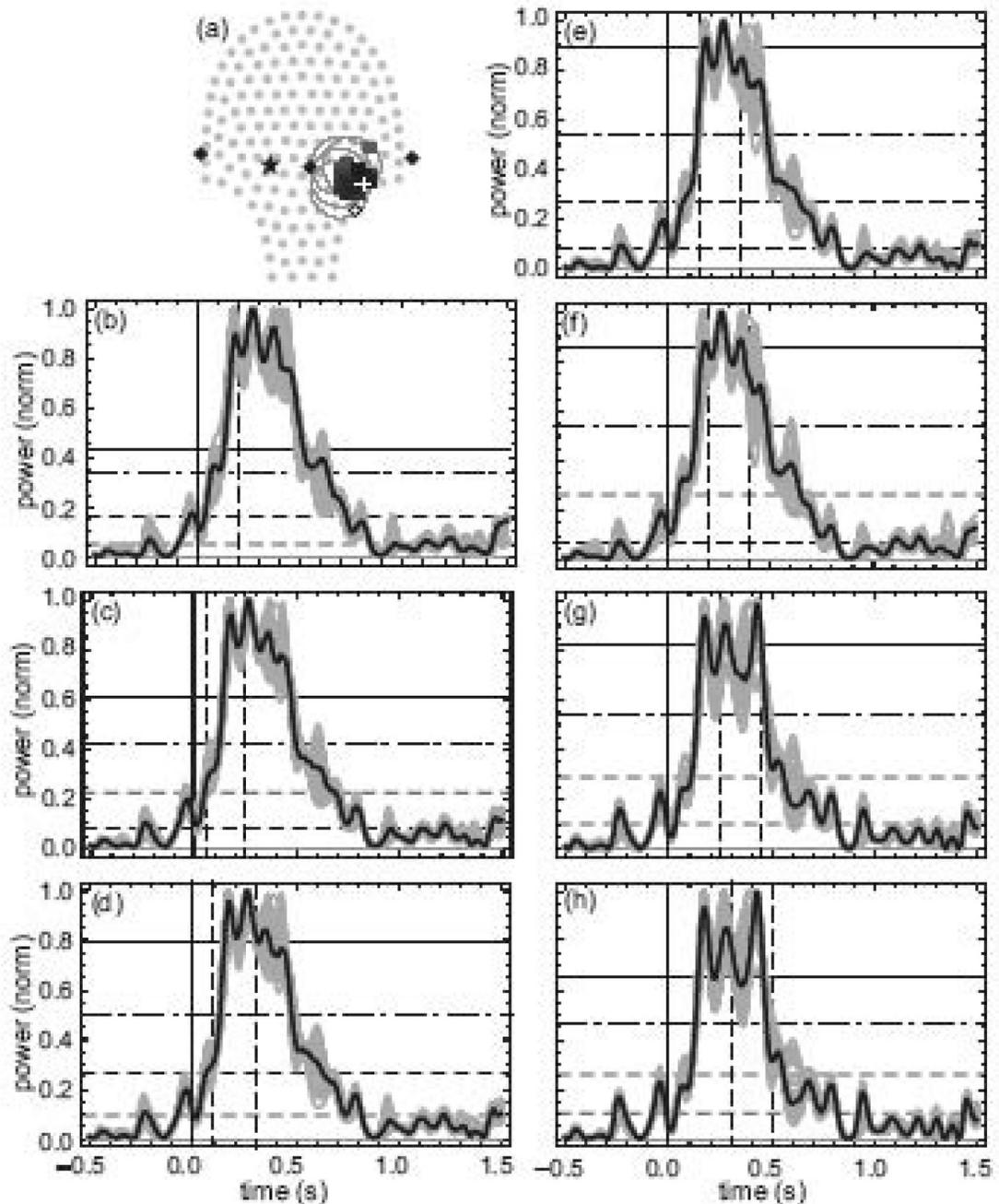


Figure 8.

Example of latency selection for subject NT19 at 33 weeks GA; (a) typical projection and (b – h) time course plots for significant clusters over each analysed response power window: (b) 0 – 200 ms, (c) 50 – 250 ms, (d) 100 – 300 ms, (e) 150 – 350 ms, (f) 200 – 400 ms, (g) 250 – 450 ms, and (h) 300 – 500 ms. Legend is the same as for fig. 7 except for the addition of maternal abdominal coils (solid diamonds) and 4th coil (open diamond) locations in the projection. The projection (a) is for the selected latency response power window (f) however in this example all were very similar as are the median time courses of significant clusters.

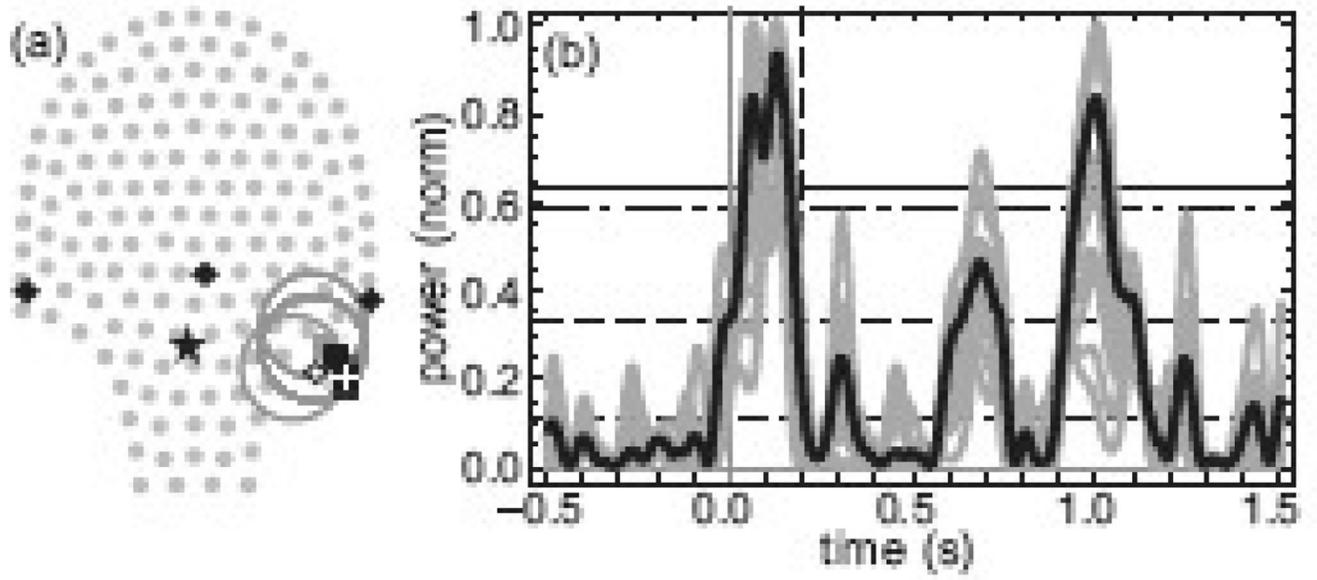


Figure 9.
Example of a response identified as artifact; (a) projection and (b) time course plot. Legend is the same as for fig. 8.

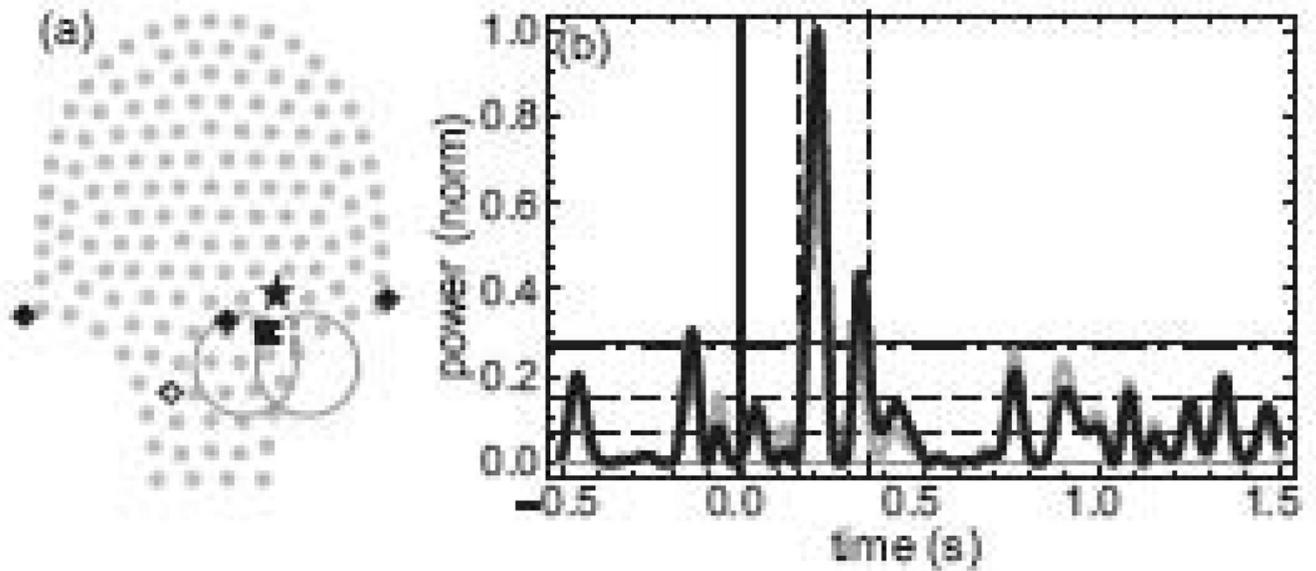


Figure 10.

Example of a response identified as false positive; (a) projection and (b) time course plot.

Legend is the same as for fig. 8. Note that in this case the RFT threshold power and mean power over the response power window are nearly coincident, indicating that this cluster was of very low significance.

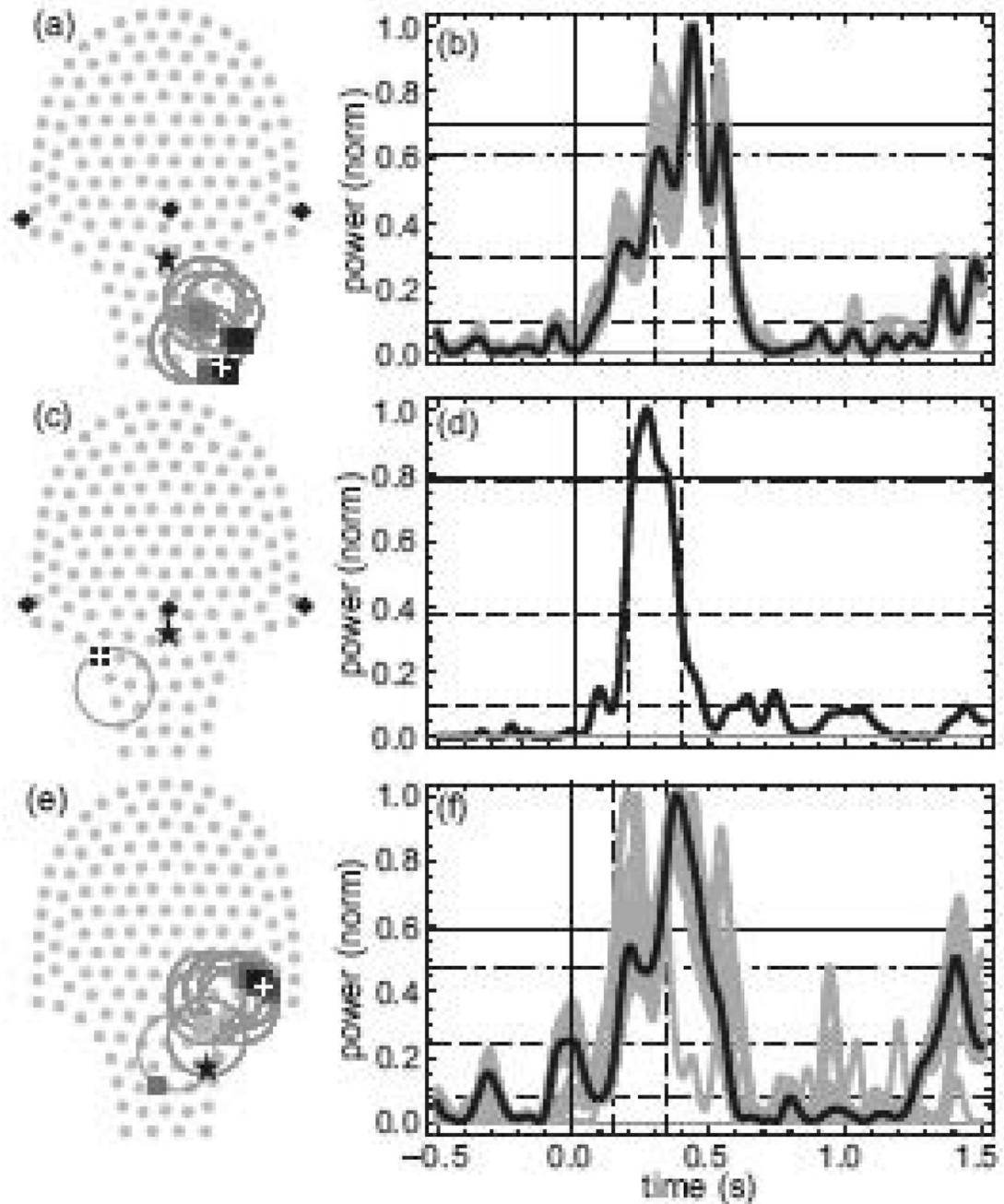


Figure 11. Example of selected significant response power windows for all GA records of one subject (NT02); (a, b) 34 weeks GA, (c, d) 36 weeks GA, and (e, f) 39 weeks GA newborn. Legend is the same as for fig. 8. Note that the RFT threshold power and mean power over response power window are nearly coincident in (d), indicating that this cluster was of very low significance.

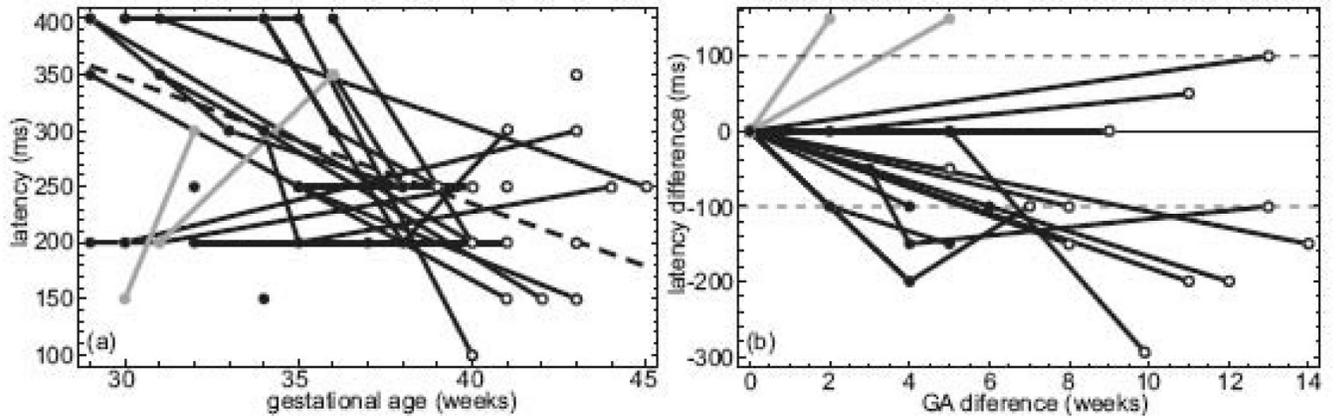


Figure 12.

(a) Response power window latency vs. GA for two subjects with latency which did substantially increase with GA (gray) and 17 subjects for which it did not (black). Dashed line is the least squares fit to all data after discarding outliers (latency < 250 ms for GA < 32 weeks or latency > 300 for GA > 40 weeks) (b) same data normalized for each subject with more than one GA latency estimate to latency difference (latency[GA] – latency [first measured GA]) vs. GA difference (GA – first measured GA.) to accentuate the key observation of latency reduction with increasing GA. Latency error bars are estimated as ± 100 ms, marked as horizontal dotted lines associated with the first measured GA. Power window width was 200 ms with center points from 100 to 400 ms in 50 ms steps. Newborn points are marked with open circles.

Table 1

Slope, correlation coefficient, and standard deviation for subjects with more than 2 GA records. Last column reports the number of GA records used in the calculation.

Subject	Slope (ms/wk)	Corr. Coeff.	s.d.	Points to fit
NT15	-29	0.85	8	3
NT02	-27	0.96	16	3
NT14	-17	1.00	7	3
NT27	-13	0.52	11	3
NT03	-6	0.52	6	4
NT16	+5	0.99	7	3