Proceedings of the 2nd International IEEE EMBS
Conference on Neural Engineering
Arlington, Virginia · March 16 - 19, 2005

Significance tests for MEG response detection

Nayef E. Ahmar[1], Yadong Wang[2], Jonathan Z. Simon[1,3,4]

[1]Department of Electrical and Computer Engineering
[2]Cognitive Neuroscience of Language Laboratory
[3]Department of Biology
[4]Bioengineering Program
University of Maryland, College Park

Abstract-**Magnetoencephalography (MEG) is a non-invasive neurophysiological technique with high temporal resolution. Nevertheless, low signal to noise ratio may hamper its fullest capability. Many confidence tests already exist to detect strong responses for signals corrupted by noise, and we have explored their use with experimentally obtained MEG signals. We find that the tests demonstrating the most power are the F-test and Rayleigh's phase coherence test. Due to the strongly non-Gaussian nature of the MEG noise, from both neural and external perspective, a signal which is purely noise often fails the marginal tests by exceeding the number of false positive allowed. A variation of the tests is suggested that ensures the average false positive for a large number of responses, excited at frequencies different than the frequency of interest, is below any desired threshold. This is implemented for the F-test, Rayleigh's phase coherence test, and the union of the two.**

I. INTRODUCTION

Magnetoencephalography (MEG) is a noninvasive tool that measures the magnetic activity of the brain, using extremely sensitive devices such as Superconducting Quantum Interference Device (SQUID). MEG is a relatively new technique that promises good spatial resolution and extremely high temporal resolution (≤ 1ms), thus complementing other brain activity measurement techniques such as Electroencephalography (EEG) and functional Magnetic Resonance Imaging (fMRI). Because the magnetic signals emitted by the brain are on the order of a few hundred femtoteslas (10^{-13} T) shielding from external magnetic signals, including the Earth's magnetic field (~ $5x10^{-5}$ T) is necessary. Even with proper shielding, poor signal to noise ratio is a major challenge for signal processing research.

When solving the inverse problem to determine the neural sources generating the measured magnetic field, detection of significant channels is crucial to improve precision and accuracy. This can be done either by measuring for consistency across different presentations, or by contrasting the signal strength at one frequency to the noise strength in neighboring bands. Rayleigh's phase coherence test [10] satisfies the former method, while the F-test satisfies the latter. We explore both methods in addition to a family of other significance tests and suggest a joint one that exploits both phase and amplitude information and achieves best results in detecting strongest auditory responses. Although various tests have different methods, they agree with each other for most channels, with some more stringent than others.

Due to highly structured noise in the measured data, these tests typically fail to accurately control the number of false positives occurring. Therefore, we turned the problem around: for every frequency measurement, in addition to examining the responses to stimuli modulated at that frequency, we also examined the responses to all the stimuli *not* modulated at that frequency (where no signal should be found). We averaged all those known false positives to achieve the desired p-value, while tuning the marginal p-value of the most powerful tests.

II. METHODS

A. Stimuli and Data

Sinusoidally amplitude-modulated sounds [3] of 2 s duration were presented 50 times each in a random order with inter-stimulus intervals uniformly distributed between 700 and 900 ms as described in [4]. A total of 20 stimuli were generated with five modulation frequencies (1.5 Hz, 3.5 Hz, 7.5 Hz 15.5 Hz and 31.5 Hz) and four different carriers (pure tone at 707 Hz; 1/3 octave pink noise centered at 707 Hz; 1 octave noise centered at 707 Hz and 5 octave noise centered at 707 Hz). All stimuli were presented binaurally at a comfortable loudness of approximately 70 dB SPL. Eight right handed subjects (5 female) were used. Subjects were given their written informed consent for the MEG study.

The magnetic signals were recorded using a 160-channel, whole-head axial gradiometer system (KIT, Kanazawa, Japan) housed in a magnetically shielded room. Three of the 160 channels are magnetometers separated from the others and used as reference channels for noise suppression. The magnetic signals were band-passed between 1 Hz and 200 Hz, notch filtered at 60 Hz, and sampled at the rate of 500 Hz. All of 157 neural channels were de-noised [15] with a Block-LMS adaptive filter [16], using the 3 reference channels.

Responses to each stimulus were taken on each channel from 300 to 2300 ms post-stimulus and concatenated, resulting in 20 responses (of 2 ms resolution and 100 s duration) for

0-7803-8709-0/05/$20.00©2005 IEEE

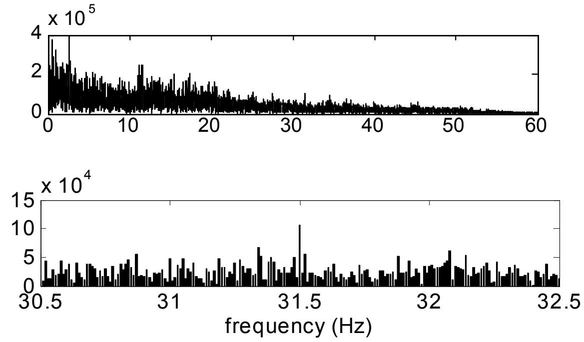


Fig. 1. (top panel) Magnitude squared of the Fourier transform of magnetic field response from one channel to a sinusoidal amplitude modulated tone with modulation frequency 31.5 Hz and carrier frequency 707 Hz. (bottom panel) Same FFT magnified to show the frequencies nearest to 31.5 Hz.

each of the 157 channels. Each response was transformed by the Fast Fourier Transform (FFT), resulting in 20 complex frequency responses (of 0.01 Hz resolution and 250 Hz extent) for each of the 157 channels. See Figure 1 for the magnitude squared of the FFT of the response of a single channel to the 31.5 Hz amplitude modulated sinusoid tone. The SSR peak at 31.5 Hz is stereotypically narrow with a width of 0.01 Hz. Also, as seen in Fig.1, background responses became noisier with decreasing frequency.

B. *Signal Detection*

These twenty different stimuli consisting of 5 modulation frequencies and four carriers were used to test the different tests introduced.

1) *Rayleigh's Phase Coherence Test*

For each of the 2 s stimulus responses of the number of presentations (N=50), an FFT was performed and the phase at the stimulus frequency was measured. We then take the projection onto the real and imaginary axes, and sum individual projection for all presentations. The phase coherence, denoted *R*, ranges between 0 and 1 where 0 is uniformly random and 1 significant [10,5,12,2]. The phase coherence is formally:

$$R_P = \frac{1}{N}\sqrt{(\sum_{i=1}^{N}\cos\theta_i)^2 + (\sum_{i=1}^{N}\sin\theta_i)^2} \quad (1)$$

The significance of the result was assessed using approximation formula suggested by [1]:

$$P = \mathrm{e}^{-NR_P^2} \quad (2)$$

An improvement to the phase coherence method was suggested by [2] where all phase measurements are projected onto an expected phase. For our MEG data, we used neighboring channels to compute the expected phase; however, this weighting method did not improve on the simple phase coherence test, and typically had less power. This is consistent with noise contamination whose phase is spatially coherent.

2) *F-Test for Hidden Periodicity*

As explained in [11,9,8,2], this test examines the signal to noise ratio for the signal at stimulus frequency to the background noise at neighboring frequencies. After taking the FFT of the concatenated 50 presentations, the average power of a total of 120 frequency bins separated by 0.01Hz, (60 below and 60 above the stimulus frequency) is measured, denoting background noise. Total noise bandwidth is 1.2 Hz. The formula to compute such ratio is given by:

$$R_F = \frac{120\left|a_{sf}\right|^2}{\sum_{i=sf-60, i\neq sf}^{i=sf+60}\left|a_i\right|^2} \quad (3)$$

The significance of this ratio is evaluated through the F distribution with (2 and 240) degrees of freedom [9,13].

An improvement to the F-test was suggested by [2] where complex values were projected onto an expected phase, creating a t-test. For our MEG data, we used neighboring channels to compute the expected phase. As in the hoped-for improvement to the phase coherence method, this weighting method did not improve on the simple F-test, and typically had less power. Again, this is consistent with noise contamination whose phase is spatially coherent.

3) *Multitaper DPSS*

A multitaper method based on windows from the discrete prolate spheroid sequences (DPSS) is also used to detect sinusoids embedded in noise based on their amplitude [14]. It is very similar to the F-test, but due to the DPSS windows, it averages over neighboring frequency bins. For our data it had less power than the simpler F-test. This is consistent with our experimental design which puts all the power of the signal into a single frequency bin, with no spectral splatter or frequency widening. In this design, smoothing in the frequency domain serves little purpose and only allows additional noise into the signal's frequency bin.

C. *"F- or Phase coherence" Balanced test*

Identifying the amplitude F-test for hidden periodicity, and phase coherence test as most prominent tests suggests a joint test since each conveys different and complementary information from the other. This was done by looking at the union of the sets of channels selected by each of the two tests and modifying the *p* values (the measure of how much evidence we have against the null hypothesis) appropriately.

All tests used were theoretically consistent. Nevertheless, the false positive rate was not achieved in measurements known to contain no signal. This should not be surprising, since all the above tests use Gaussian white noise as the null-hypothesis, and typical MEG noise is non-Gaussian. Rather than set a fixed false positive (α) value that correspond to a fixed *p* value, we computed a real estimate of false positives at the frequency of interest by averaging large number of observations of false positives (α_{avg}) for responses of different stimulus frequency. Accordingly, we tune α values to achieve the desired α_{avg} .Finally, to ensure equal contribution from either of the 2 tests, we balance the tuning equally by enforcing a constant ratio throughout. On average, the tuning reduced α by 30%.

III. Results

The joint balanced "F- or Phase coherence" test outperformed all other tests including the randomization test described in [1]. Our result agrees that information from both phase and amplitude better estimate signals than using only one as found in [2].

This is demonstrated in Fig. 2a which shows the complex field distribution at 3.5 Hz for a stimulus modulated at 3.5 Hz. Arrows represent the magnetic field response, at each of 157 channels, as phasors: the length of the arrow denotes amplitude and the orientation denotes phase. Circles mark those channels identified as significant by the joint balanced test ($p < 1/157$). It is clear that many of the channels strong in magnitude are not significant (see especially the frontal channels in the right hemisphere). Fig. 2b shows the response at the same frequency (3.5 Hz) but from a stimulus whose modulation frequency was 1.5 Hz, and so only noise is expected. One significant channel is found, which is consistent with $p < 1/157$ for 157 channels. Notice further the apparent spatial coherence of the phase structure, despite the fact that this is a pure noise response. Analogous maps for the 7.5 Hz (and 15.5 Hz) cases are shown in Fig. 3. The significant channels may be found higher numbers in either hemisphere. In the case of Fig. 3b, there are 2 false positives. Recall that the test is designed so that there is, on average, one false positive for all responses in which there is no signal expected.

Fig. 4a demonstrates the best fit dipole to the head map shown in Fig. 3a using only the 30 significant channels from right hemisphere (the dipole was fit to the phase of the complex magnetic field corresponding to maximum spatial variance, resulting in a real field and a real dipole). For comparison, another dipole fitted same head map using all channels from the right hemisphere Fig. 4b. All quantitative measures improved: dipole location, Goodness of Fit (GOF), and cross correlation between measured and theoretical values. Improvement varied from mild to major based on how many channels are significant and distribution of strong signals in the neighborhood that falsely bias the dipole position, strength, and orientation.

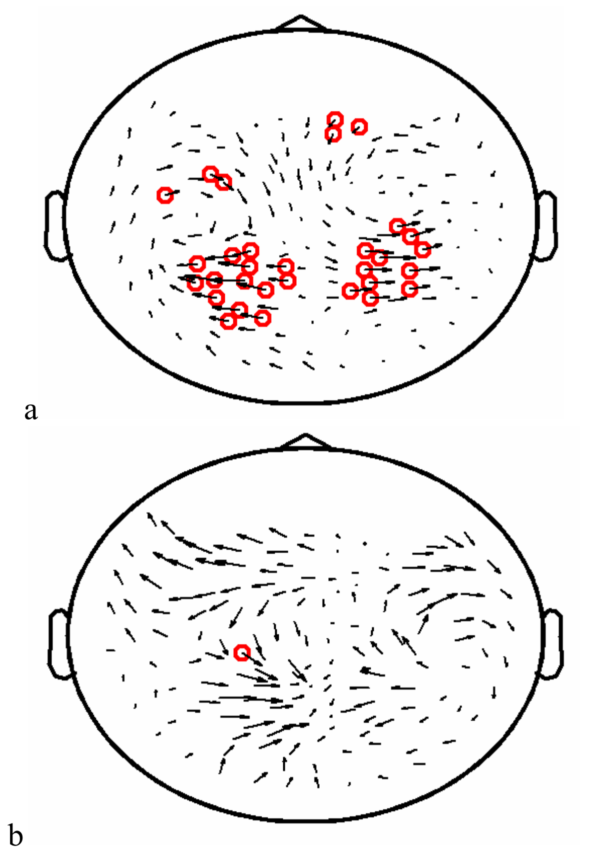


Fig. 2. (a) Response at 3.5 Hz for a 3.5 Hz modulated stimulus. 31 significant channels out of a total of 157. Arrows represent amplitude and phase of signals at the stimulus frequency. The circles denotes significant channels. (b) Response at 3.5 Hz for a 1.5 Hz modulated stimulus. The circle denotes a "significant" channel which must be a false positive. Note that the strongest responding channels were not found to be significant.

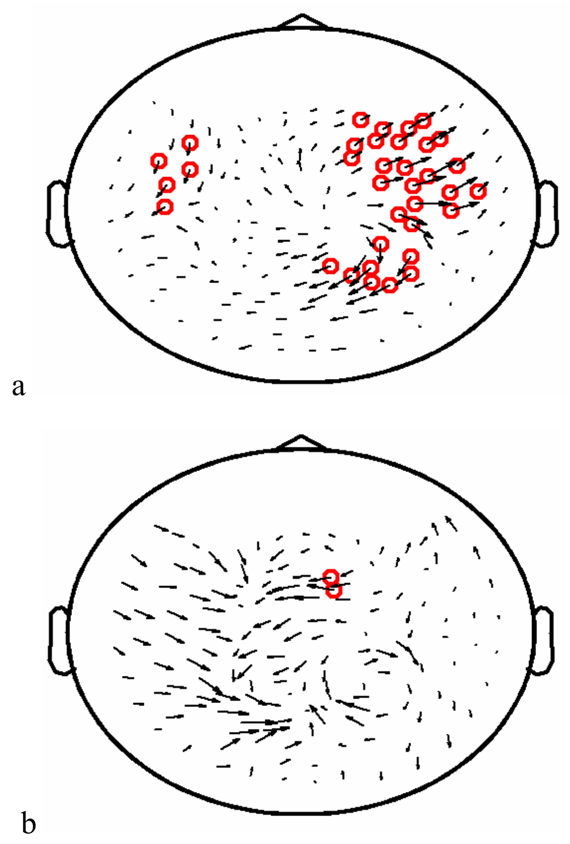
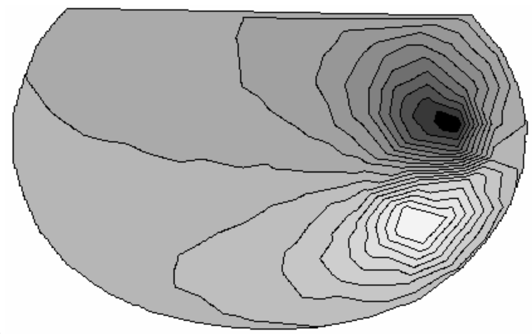


Fig. 3. (a). Response at 7.5 Hz for a 7.5 Hz modulated stimulus. 35 significant channels out of a total of 157. Arrows and circles as in Fig. 2. (b) Response at 7.5 Hz for a 15.5 Hz modulated stimulus. Two false positives.



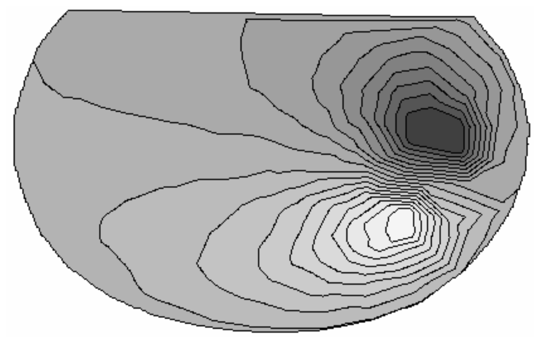


Fig. 4. Interpolated intensity contour maps for the magnetic fields shown in Fig. 3a. (a) Interpolated intensity contour map for the most significant channels: Goodness of fit (GOF) 84% Correlation Coefficient. 92%. (b) Interpolated intensity contour map for all right hemisphere channels: GOF 78% Correlation Coefficient 90%.

To test how well the algorithm performed against other choices of significant channels, we applied a permutation test to the magnetic distribution shown in Fig. 3a. From the right hemisphere, 30 channels were chosen at random and labeled “significant”, and a dipole was fit to those channels, and its GOF calculated. Then that step was repeated 1000 times to compute a cumulative distribution function of the GOF. The GOF of dipole for the significant channels based on the joint test, 84%, was not achieved in any of the 1000 permutations (i.e. $p \leq 0.1\%$)

IV. Discussion

F-test and Rayleigh’s phase coherence tests, outperformed other tests in detecting significant channels for MEG responses, but a joint test using both phase and amplitude performed better jointly than marginal ones. Expected-phase weighted tests fared more poorly, presumably because the expected phase used was the local spatial average, which was typically coherent even when no signal was present. For this reason and similar properties of the noise, the null hypothesis of Gaussian noise, independent across channels, was not appropriate, leading us to rescale the probability distributions in order to match the measured false positive rate. For the purpose of dipole fitting of auditory responses due to sinusoidal amplitude modulated tones, using significant channels determined by the joint tests yielded better goodness of fit. Using all channels per hemisphere, including those corrupted by noise could yield a dipole fit with high GOF even when there are few (or no) significant channels present, but the experimental relevance of such a fit is dubious at best.

References

[1] Fisher NI. Statistical analysis of circular data, Cambridge, MA: Cambridge University Press, 1993.
[2] Picton, T.W., Dimitrijevic, A., John, M.S. & Van Roon, P. 2001. The use of phase in the detection of auditory steady-state responses. Clin Neurophysiol, 112, 1698-1711.
[3] Ross B, Borgmann C, Draganova R, Roberts LE, Pantev C (2000) A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. J Acoust Soc Am 108:679-691.
[4] Simon J.Z., Y. Wang, D. Poeppel, J. Xiang , N. Ahmar (2004) MEG Steady State Responses To Auditory Stimuli Of Varying Complexity. Society for Neuroscience abstracts.
[5] Cohen LT, Rickards FW, Clark GM. A comparison of steady-state evoked potentials to modulated tones in awake and sleeping humans. J Acoust Soc Am 1991;90:2467-2479.
[6] Dobie RA, Wilson MJ. Analysis of auditory evoked potentials by magnitude-squared coherence. Ear Hear 1989;10:2-13.
[7] Dobie RA, Wilson MJ. Objective detection in the frequency domain. Electroencephalogr Clin Neurophysiol 1993;88:516-524.
[8] Fisher RA. Tests of significance in harmonic analysis. Proc R Soc Lond Ser A 1929;125:54-59.
[9] John MS, Picton TW. MASTER: a Windows program for recording multiple auditory steady-state responses. Comp Methods Prog. Biomed 2000a;61:125-150.
[10] Lord Rayleigh. On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. Philos Mag 1880;10:73-78.
[11] Schuster A. On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. Terrestr Magnet Atmos Electr 1898;3:13-41.
[12] Zar JH. Biostatistical analysis, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 1999.
[13] Ross, B., Picton, T. W., and Pantev, C. (2002). "Temporal integration in the human auditory cortex as represented by the development of the steady-state magnetic field," Hear. Res. 165(1-2), 68-84.
[14] Percival, D.B., and A.T. Walden. Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques. Cambridge: Cambridge University Press, 1993.
[15] Ahmar N., Simon J., MEG Adaptive Noise Suppression using Fast LMS, unpublished.
[16] Haykin, S. S. (2002). Adaptive filter theory. Upper Saddle River, N.J., Prentice Hall.