

Measures of Spatial Similarity and Response Magnitude in MEG and Scalp EEG

Xing Tian · David E. Huber

Accepted: 9 November 2007
© Springer Science+Business Media, LLC 2007

Abstract Sensor selection is typically used in magnetoencephalography (MEG) and scalp electroencephalography (EEG) studies, but this practice cannot differentiate between changes in the distribution of neural sources versus changes in the magnitude of neural sources. This problem is further complicated by (1) subject averaging despite sizable individual anatomical differences and (2) experimental designs that produce overlapping waveforms due to short latencies between stimuli. Using data from the entire spatial array of sensors, we present simple multivariate measures that (1) normalize against individual differences by comparison with each individual's standard response; (2) compare the similarity of spatial patterns in different conditions (angle test) to ascertain whether the distribution of neural sources is different; and (3) compare the response magnitude between conditions which are sufficiently similar (projection test). These claims are supported by applying the reported techniques to a short-term word priming paradigm as measured with MEG, revealing more reliable results as compared to traditional sensor selection methodology. Although precise cortical localization remains intractable, these techniques are easy to calculate, relatively assumption free, and yield the important psychological measures of similarity and response magnitude.

Keywords EEG · MEG · Individual/anatomical differences · Overlapping waveforms · Multivariate analysis · Source modeling · Repetition priming

Introduction

The human brain is a highly distributed system and many cortical areas are simultaneously active during any task. Non-invasive surface recordings, such as scalp electroencephalography (EEG) and magnetoencephalography (MEG), use many sensors to record the voltage potentials or magnetic field responses near the surface of the head that arise from the underlying mixture of cortical sources. Each sensor in these recording methods receives a mixed signal from all neural sources and the contribution of each source depends both on the distance to that source and the relative orientation of that source [12]. Given the mixture of underlying neural sources, it is inaccurate to assume a one-to-one mapping between sensors (or local groups of sensors) and underlying sources. In particular, a change in response for a particular sensor could be due to a change in the magnitude of response of the dominant neural source, or it might instead be due to a change in the distribution of the neural sources as different cortical areas are recruited in the different conditions of interest.

The inherent ambiguity between response magnitude and cortical distribution could be addressed with equivalent dipole modeling [1] or blind source separation algorithms, such as independent component analysis (ICA) [23]. However, due to the 3-dimensional nature of passive electrophysiological recordings (as opposed to 2-D manipulated cortical slices with fMRI), there exists a so-called "inverse problem", which refers to the infinite possible cortical solutions to a particular data pattern across

X. Tian (✉)
Department of Psychology, Neuroscience and Cognitive Science
Program, University of Maryland, College Park,
MD 20742, USA
e-mail: xtian@psyc.umd.edu

D. E. Huber
Department of Psychology, University of California, San Diego,
La Jolla, CA 92093, USA
e-mail: dhuber@psy.ucsd.edu

the sensors [24]. These techniques tackle the inverse problem through simplifying assumptions, such as an assumed number of dipoles/components and independence from moment to moment and trial to trial. The accuracy of these techniques strongly relies upon these assumptions, which are known to be false in many circumstances. For instance, synchronization between cortical areas during auditory or visual detection tasks [3, 10, 11] implies that different cortical sources are temporally dependent rather than independent.

The ambiguity that arises from only considering select sensors is compounded by averaging across subjects because the same sensor may reflect entirely different mixtures of cortical sources for each individual. Nevertheless, sensor selection and subject averaging remain common practices in EEG experiments (e.g., [6, 31]) and few EEG studies address this issue directly (although see [4, 30]). For instance, it has been demonstrated that anatomical differences in the cortex directly relate to EEG scalp recording differences [2]. Beyond anatomical differences, tissue conductivity volume conduction differences also play an important role in the magnitude and pattern of scalp potentials [26]. In contrast to this situation with EEG, the role of individual differences is carefully considered in the study of neuroanatomy and functional magnetic resonance imaging (fMRI). For instance, current practice in fMRI normalizes against neuroanatomical differences by employing inflation techniques that map each individual brain onto a canonical brain that is flattened such that sulci can be visualized in a 2-D plane [8]. Analogous to this canonical brain mapping with fMRI, we present a simple technique that normalizes EEG or MEG responses across all sensors against the ‘standard’ response pattern for that individual.

In light of individual differences, EEG and MEG typically go to one of two extremes. EEG analyses often average over many participants, making sure to keep scalp electrodes in the same position in relation to the external parts of the head, with the hope that this produces a systematic positioning in relation to the brain, or at least sufficient data to overcome the differences that might otherwise confound the results with a smaller number of participants. In contrast, systematic positioning of sensors in relation to the head is all but impossible with MEG. Therefore, MEG typically involves analyzing the results from a small number of participants separately (e.g., [33]), with no method for combining the results across individuals. This avoids the potential errors caused by averaging across individual differences who have different neuroanatomy (e.g., [21]), but makes it impossible to use inferential statistics to make claims about the general population.

Traditional EEG and MEG results based on single sensors or small groups of sensors can nevertheless reliably identify whether conditions are different from each other, but cannot ascertain (1) whether the observed differences reflect the addition/subtraction of new cortical sources or (2) whether the observed differences are due to increases or decreases in response magnitude of the underlying sources. Our technique can address these functionally important questions through the multivariate comparison of the entire array of sensors in experimental conditions against a standard response pattern for each individual. For EEG this makes exact placement of electrodes less important and for MEG this allows comparison across individuals.

The above discussion focused on the role of individual differences and the inherent ambiguity between pattern similarity and response magnitude. In addition to these concerns, many experimental paradigms introduce a third problem by using short inter-stimulus intervals between presentations, which results in measurements that reflect the combination of fast cortical responses to the current stimulus and slower responses from previous stimuli. However, this problem is all but unavoidable considering that many of the most widely used and informative designs in behavioral psychology rely upon short latencies between stimuli (e.g., [7, 17, 20, 27, 28]). Complicated experimental and mathematical techniques can be employed to address this issue of overlapping electrophysiological waveforms [32]. However, our simple multivariate comparison technique may also help in such situations considering that comparisons can be done separately in relation to fast (e.g., P100) and slow (e.g., N170) standard responses so as to partially untangle the combined pattern.

Similar to ICA and dipole modeling and similar to multivariate analyses in fMRI [19, 25], our technique uses the entire pattern across all the sensors. Our approach is also analogous to the use of a “localizer task” in functional magnetic resonance imaging (fMRI) studies (e.g., [18]) in that it compares responses to a standard reference. In fMRI experiments, an initial task is often used to locate a particular cortical region of interest for that individual. Subsequent experimental conditions are then analyzed in terms of this region. By analogy, we include a condition that defines a ‘standard response’ for each individual (i.e., a pattern over all sensors in some baseline condition at a chosen time following the stimulus), against which we compare the response pattern in experimental conditions. We advocate both a measure of pattern similarity to the standard response as well as the projected magnitude “in the direction of the standard response”. Next, we present the mathematical details for calculating these measures and then we demonstrate their effectiveness using data from a word priming study that recorded high density evoked MEG.

Measures of Similarity and Magnitude

Measures of similarity and magnitude can be calculated for any two patterns across the sensors, such as with two different experimental conditions, but they are most effective when there is a standard response for comparison. The standard response should measure a response that includes many of the same processes of potential interest that are employed in the experimental conditions, but remain free of overlapping waveforms and other complications. The situation that defines the standard response should be similar to the experimental conditions in terms of the stimuli, level of attention, task demands, salience, task relevance, etc. In the example to follow, the experimental conditions of interest were immediately repeated versus novel words and the standard response was generated from a highly attended prime word presented in isolation for a sufficiently long duration.

We define the entire multivariate pattern of sensors in one condition to be \bar{A} , which is an n -dimensional column vector where n is the number of sensors (i.e., all analyses are in “sensor space”). Then, the response in some other condition, \bar{B} , can be compared for its similarity as defined by the n -dimensional angle (Eq. 1) as well as its magnitude in the direction of \bar{A} through the geometric projection of \bar{B} onto \bar{A} (Eq. 2).

$$\cos \theta = \frac{\bar{A} \cdot \bar{B}}{|\bar{A}| |\bar{B}|} \quad (1)$$

According to Eq. 1, the value of cosine θ gives an index of the spatial similarity between the pattern across the sensors in two experimental conditions, A and B. We refer to use of this measure as the ‘angle test’. Similar to a correlation measure, the cosine of the angle ranges between -1 and $+1$, with -1 indicating completely opposite, $+1$ indicating completely similar, and 0 indicating dissimilar (perpendicular). This measure has the advantage that it is unaffected by the magnitude of the response. For example, if two conditions have identical response patterns across the sensors, but the response magnitude for one condition is twice as large, this comparison technique will reveal that the angle between them is 0 , and, thus, the cosine of the angle is 1.0 (i.e., perfectly similar). If the angle test reveals that there is no statistically reliable difference in the spatial similarity between two conditions (below we discuss a technique for assessing statistical reliability), this suggests that the distribution of underlying cortical sources is similar between the two conditions. The distribution of cortical sources might change in exactly the right way to produce the same pattern across the sensors, but such a coincidence is of low probability, particular when the number of sensors is large. Thus, a high value on the angle

test is likely to correspond with a similar mix of cortical responses. Conversely, if the angle test reveals that the spatial similarity of the sensors is different, then this definitely indicates that the underlying distribution of cortical responses has changed. This second conclusion is assumption free, barring confounding factors such as head position changes between conditions.

$$|\bar{B}| \cos \theta = \frac{\bar{A} \cdot \bar{B}}{|\bar{A}|} \quad (2)$$

For the ‘projection test’ based on Eq. 2, the magnitude of an experimental condition B is normalized against the standard response A, by projecting \bar{B} in the direction of \bar{A} . This projection can be used to calculate the magnitude of response for some condition in the direction of the standard response. It is not necessary that the condition of interest be similar to the standard response. However, in order to unambiguously compare two projection values in two different experimental conditions (e.g., B_1 and B_2), the conditions need to be sufficiently similar to each other. Otherwise, any apparent magnitude differences could be due to similarity differences. In other words, a critical first step is to assess whether the conditions are dissimilar with the angle test. Regardless of the observed level of similarity between the conditions, the projection values can be calculated, but if it is found that the conditions are dissimilar, then an obtained difference in the projection values does not unambiguously indicate a magnitude change and instead only indicated something has changed. Nevertheless, even in this situation, the projection test may be useful for normalizing against individual differences.

Because the projection is relative to each individual’s “clean” standard response (e.g., the M170 to a word presented in isolation), it should normalize both against individual differences (e.g., the particular pattern over sensors of the M170 for a particular individual) and against contamination from overlapping waveforms that exist with short latencies between successive stimuli (e.g., the combination of an M170 to a word presented 170 ms ago with the M400 to a prime word presented 570 ms ago). Because this technique normalizes against individual differences, the projection values can be used in traditional inferential statistical tests across the data of multiple participants.

Figure 1 illustrates the angle test and projection test, with a hypothetical example that includes spatial differences between the standard (A) and experimental conditions (B_1 and B_2), as might arise from overlapping waveforms for the experimental conditions. For demonstration purposes, only two sensors are shown (e.g., a 2-D sensor space), but the same logic applies to n sensors

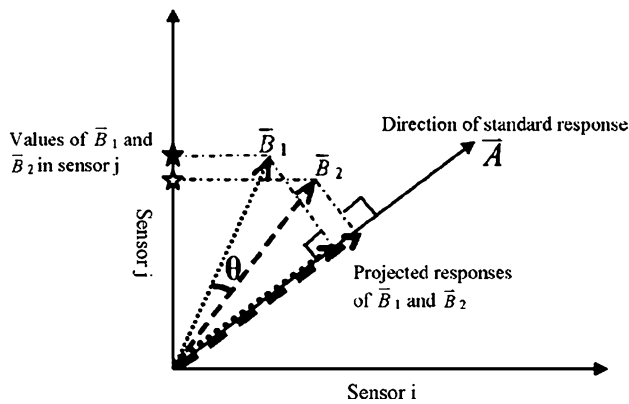


Fig. 1 Demonstration of the angle test and the projection test with a hypothetical situation involving two sensors (*i* and *j*). Results are shown for a standard response (the solid line *A*) and two experimental conditions (*B*₁, dotted line and *B*₂, dashed line). In this example, the conditions are spatially similar to each other, but are different from the standard response because, unlike the standard response, the experimental conditions include some other overlapping response. The spatial angle (θ) between the experimental conditions indicates whether the experimental conditions represent different response patterns across the two sensors. Because the conditions are sufficiently similar to each other (i.e., small angle), the projection onto the standard response indicates response magnitude that normalizes against individual differences and against overlapping waveforms (i.e., the magnitude of response in the direction of the standard response). If the sensor of largest magnitude is selected (i.e., sensor *j*) it is concluded that *B*₁ (filled star on vertical axis) is greater than *B*₂ (open star on vertical axis) but, in contrast, if the projection onto the standard response is analyzed, it is concluded that *B*₂ (length of dashed arrow) is greater than *B*₁ (length of dotted arrow)

defining *n* dimensions (i.e., a multivariate situation). The angle test between *B*₁ and *B*₂ indicates that they are sufficiently similar and, therefore, likely due to the same mixture of underlying cortical sources. Thus, the projection onto the standard response provides a pure measure of magnitude that normalizes against individual differences (i.e., different patterns in the standard response for different individuals) and also normalizes against overlapping waveforms (i.e., extraction of that component of the experimental condition that is in the direction of the standard response, rather than in the direction of the overlapping response). The bold lines along the direction of standard response are the projected normalized responses of experimental conditions. A traditional analysis based on the sensor with the largest response compares sensor *j* in the two conditions, and concludes that condition *B*₁ produced a larger response than condition *B*₂. In comparison, projection of the experimental conditions onto the standard response concludes that condition *B*₁ produced a smaller response magnitude than condition *B*₂.

The general procedure for using these measures includes the following steps. (1) identify an appropriate standard response in the experimental design; (2) use the angle test

to assess the similarity of conditions of interest; (3) project the experimental conditions onto the standard response to normalize against individual differences and overlapping waveforms. If the answer to step 2 concludes that the conditions are dissimilar, the projection of step 3 is still useful for normalizing, but it does not unambiguously indicate magnitude versus similarity. However, if conditions are not found to be similar according to the angle test, then the projection values unambiguously indicate magnitude and increases versus decreases can be taken to correspondingly indicate increases versus decreases of cortical response.

In order to statistically test each step, a null hypothesis distribution is needed. There are several techniques that could be applied to define a null distribution, although we take the relatively simple approach of comparing the first half versus second half of trials within the experiment for a given condition (i.e., cross-validation over time), versus the same first/second half of trials comparison between conditions. The selected trials could be determined by odd versus even trials, or through repetitive non-parametric bootstrap samples, but first versus second half of the experiment is simple to calculate and includes trends over time in the null distribution. In this manner, the angle and projection tests become simple t-test of between versus within conditions.

Next, we provide an example where these measures were used to assess M170 responses in a short-term repetition paradigm with visually presented words. The response of interest was to the briefly flashed target word. The first word (i.e., the long prime) was used to define a standard response for the M170 in terms of the spatial pattern associated with a single visually presented word. This event was chosen for the standard response because it presented a single highly attended word. First, the angle test was implemented to statistically test the existence of individual differences in the standard response (otherwise there's no need for a standard response to obtain normalization). Next, the angle test was used to assess the similarity across the 3 different target conditions (*novel*, a target word that is different than both prime words; *short*, a target word that repeats the second prime, which was presented 150 ms prior the target; and *long*, a target word that repeats the first prime, which was presented 2,000 ms prior to the target). Because the target conditions were found to be sufficiently similar to each other, the projection test was used to unambiguously indicate response magnitude. The projection measure normalized the target conditions against the standard response and the results were analyzed across individuals to determine if, in general, target response magnitude varied across the conditions. Essentially, we asked "how much of an M170" occurred in each condition for each individual by using all

the data across the entire sensor array as compared each individual's standard M170 to visually presented words.

An Example with MEG: Immediate Word Repetitions

The example implementation of these techniques, reported next, may seem overly complex. However, this complexity proves to be instructive. A major advantage of these measures is that they can be applied to high density sensor data that varies greatly across individuals (as is the case with MEG data in this example) and that they can be applied to isolate a small response (e.g., the target word presented for just 50 ms) that overlaps greatly with a previous stimulus (e.g., the prime word presented 150 ms before) or overlaps greatly with a subsequent stimulus (e.g., the pattern mask presented immediately after the target word). The chosen paradigm is a classic threshold word identification paradigm (e.g., [16]), which requires immediately preceding primes, brief targets, and subsequent masks. Therefore, this task is ideal for demonstrating the effectiveness of these measures for untangling the otherwise confusing combination of responses.

This experiment is only summarized here, and is reported in full elsewhere [15]. As seen in Fig. 2, the task of the participants was to identify the briefly flashed target word (e.g., PATCH) presented in the center of the screen immediately after the prime words. First, a single prime word appeared above the midline, which provided the standard response (as well as a long duration prime). Next, a second prime word appeared 1,850 ms later, below the midline. Finally, these two prime words were replaced by a single target word, which provided the evoked MEG responses for the three conditions of interest (a *novel* target, a target that repeated the *short* duration prime, and a target that repeated the *long* duration prime). Target word durations were set at the perceptual threshold for each participant such that accuracy was approximately 75% in forced choice identification (e.g., a choice between HURRY and PATCH).

Humphreys et al. [16] used a similar paradigm and found differences in the magnitude of immediate repetition priming with masked versus unmasked words. The current paradigm tested these effects by controlling for response bias with two-alternative forced choice testing, rather than naming the briefly flashed target word. With just a single prime, as in the Humphreys et al. studies, comparison of the MEG response to the target following a brief prime versus following a long duration prime would be problematic because only the short duration prime condition involves overlapping waveforms. Furthermore, short versus long duration primes might involve different degrees of alertness. For instance, the abrupt onset of a word may result in a transient attentional response with a different

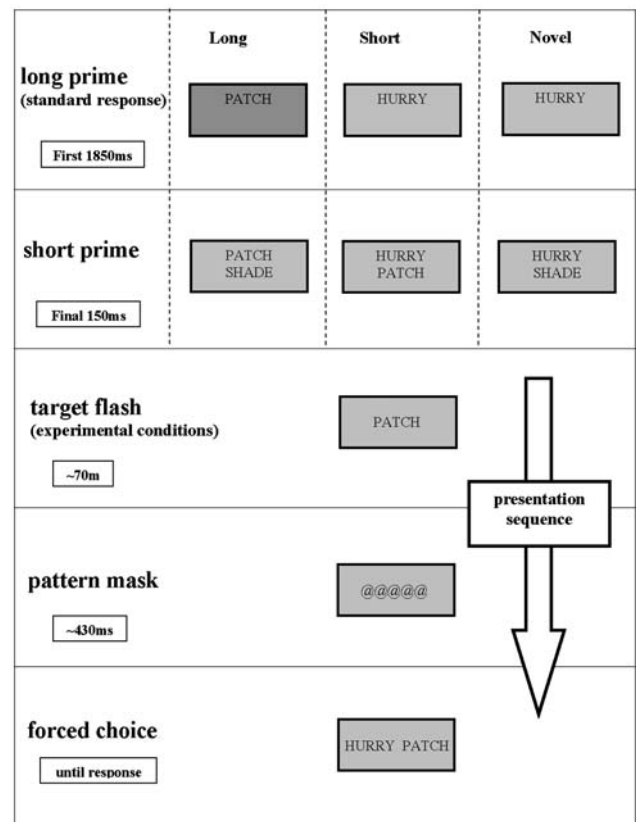
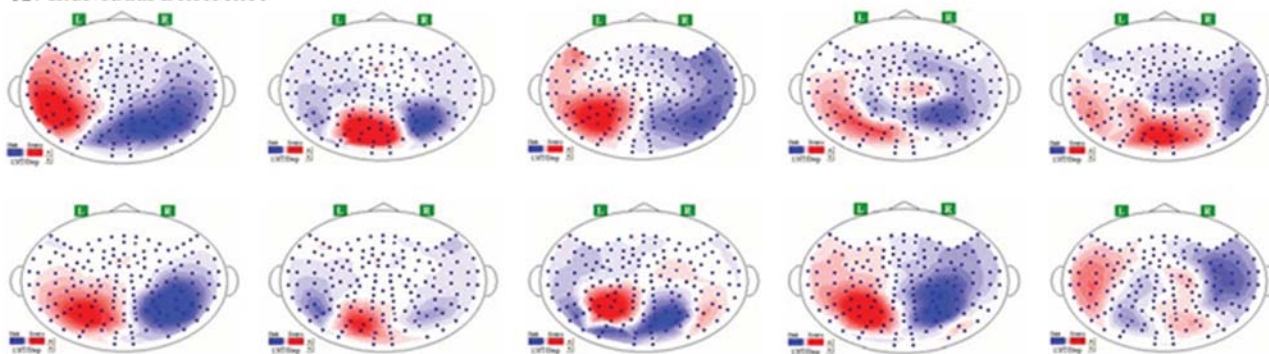
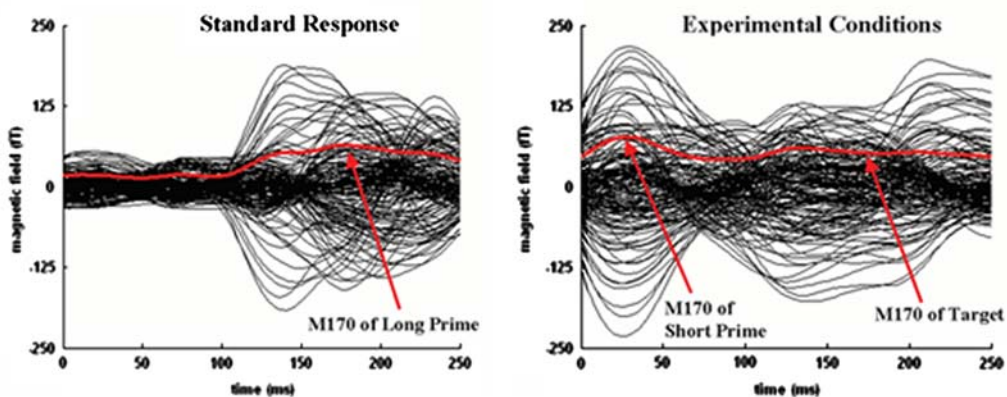
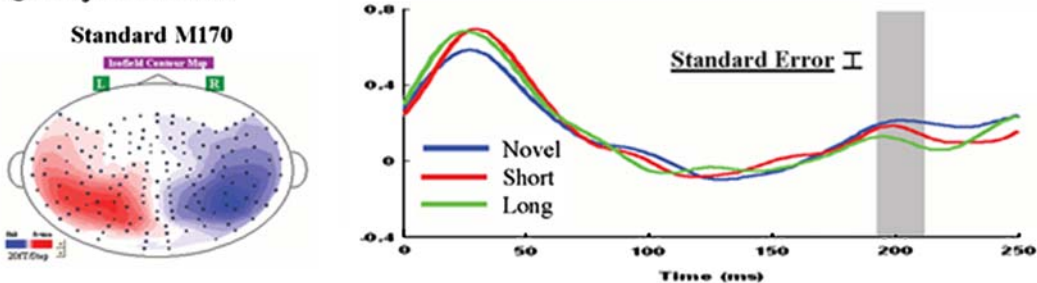
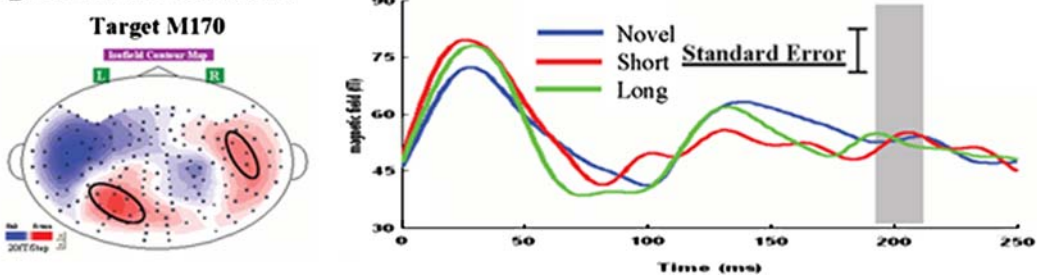


Fig. 2 Presentation sequence for the reported experiment. The task of participants was to identify the briefly flashed target word by selecting between two choice words at the end of the trial sequence. Target flash durations were set separately for each individual to achieve threshold performance of 75% correct. The upper prime word appeared first (long prime), remaining onscreen for 1850 ms in isolation, thus providing a standard response. During the final 150 ms prior the target word, the lower prime word appeared (short prime). MEGs to the target flash provided three experimental conditions, depending on whether the target was different than both primes, a repeat of the long duration prime, or a repeat of the short duration prime

MEG signature that would exist following a brief prime but not following a long duration prime. To address these concerns, the current paradigm presented on every trial a first prime for 2,000 ms (the long prime) as well as a second prime for the final 150 ms (the short prime) before the target. Thus, all conditions are identical up until the target, and any attentional effects should be equivalent.

Individual Differences

The response to the long prime was used as the standard response because the long prime was seen in isolation for 1,850 ms as part of the highly attended sequence of events. Figure 3a shows the M170 standard response results from

A: Individual Difference**B: Grand Average Waveforms****C: Projection Results****D: Sensor Selection Results**

all 10 participants, demonstrating different spatial patterns for the M170 to visually presented words.

In order to statistically assess reliability of these individual differences (i.e., are the differences in Fig. 3a reliable individual differences or just due to sampling noise), the experiment was separated into two halves, and

M170 patterns for each half were determined separately in order to obtain a null hypothesis measure of spatial pattern variability. There were 400 trials in the experiment and so standard M170 responses to the first 200 trials were calculated separate from the last 200 trials. The angle test was performed for the 10 within subject comparisons for the

◀ **Fig. 3** (a) Spatial patterns of the M170 standard responses for all 10 individuals. In general, individual differences in the similarity of these patterns were found to be reliable as indicated by a statistical test of the angle between the standard responses across individuals. (b) Comparison between the grand averaged waveforms of standard response to the long prime (left), versus the grand averaged experimental response to the target word (right), which immediately followed the second prime. Both figures represented waveforms in 157 channels of grand-average results. The bold red lines in both figures is the Root-mean-square (RMS) of the 157 channels. For the experimental conditions, there is no clear M170 peak, possibly because of individual differences, overlapping waveforms, or because the target word is only presented briefly. (c) Projection results. The topographic map is the grand average of the standard M170 responses. The graph shows the average projection measures at each moment in time for three different target conditions, with projection calculated separately for each individual according the standard responses in a. (d) Traditional sensor selection and Root Mean Square results. The topographic map is the grand average target M170. The

10 sensors with the largest positive magnitude are circled and these were selected to produce traditional sensor selection analyses. The graph shows average Root Mean Square for the selected sensors at each moment in time. The topographic maps in c and d were taken at 176 ms after word onset (the long prime onset for c and the target word onset for d). A comparison of these topographic maps reveals large differences, which suggests that the target response is contaminated by overlapping responses from the second prime, which occurred just 150 ms before the target. The shaded areas in the graphs of both c and d indicate the 22 ms average window used for statistical analyses. The error bar in each graph indicates one standard error of the mean difference between novel and repeated words, averaged over the short and long conditions as calculated for the M170 time window. There was no difference among M170 responses of the three target conditions using sensor selection. However, the projection measure uncovered the small target M170 peak and revealed that the target word produced less of a cortical response when it repeated the long duration prime

data of the 10 participants (i.e., angle between first half and second half for each individual), versus the 45 between subject comparisons (10 choose two combinations of first half versus second half when these halves are for different individuals). Using an independent samples t-test, the cosine angle for the between subjects comparison was significantly lower (more dissimilar) than the within subjects comparison, $t(21.5) = 7.812$, $P < .001$. Because the assumption of equal variances was violated for the independent samples t test ($F = 4.26$, $P < .05$), the degrees of freedom has been appropriately adjusted. This indicates that different people have different spatial patterns for the M170 responses to visually presented words. This result highlights the need to normalize against these individual differences.

Although not central to application of these measures, we note that the reported analyses also normalized for individual differences in the timing of the M170 responses. This was done by determining peak M170 times in the standard response waveform to find appropriate M170 temporal offsets for each individual. These individually appropriate times were then used for subsequent tests (i.e., we assumed that these same peak times were applicable to the briefly flashed target words). Twenty-two milliseconds windows were placed around these individually determined M170 peak responses for data averaging purposes, both for the standard responses as well as the target condition responses. Separate statistical tests validated the reliability of these timing differences.

Target Repetition Effects

The three target conditions were first compared to each other using the angle test to check if they produced different spatial patterns. Such a finding would indicate that

one or more of the conditions involved recruitment of a cortical response not present in the other conditions. Statistical reliability was again determined by dividing the experiment into trials from the first half of the experiment versus trials from the second half of the experiment. The patterns for each half were again compared to each other with the angle test, with this occurring for first/second half angle measures from the same condition (the within values for null hypotheses) versus first/second half angle measures from different conditions (the between values for experimental conditions). Because this was a comparison of conditions, rather than individuals, these comparisons were calculated within subject (i.e., repeated measures). Because there were three conditions, this entailed three different possible comparisons between conditions, which were then averaged in comparison to the average of the three within condition comparisons, yielding one between conditions measure and one within conditions measure for each participant. In a dependent samples test across the 10 participants, there was no significant difference in the similarity (angle test) of the between conditions comparisons versus the within condition comparisons, $t(9) = -1.216$, $P = 0.255$. This suggests that the same distribution of cortical responses was involved in the three target conditions. Thus, the subsequent projection test was taken to indicate magnitude differences rather than pattern differences.

After confirming that the three conditions were not dissimilar, magnitude changes were determined by projecting each condition onto the standard response (i.e., the M170 to the long prime). In doing so, the target word's response was reduced to a single magnitude measure that normalized against individual differences. Besides allowing statistical tests across individuals, this normalization also helped reduce contamination from the ongoing later components (e.g., M400) in response the short prime,

which was presented just 150 ms prior to the target. By projecting the target response onto the standard M170, the resultant magnitude more cleanly reflects the M170 component, with the overlapping M400 response to the short prime partially factored out. The degree of success in this decontamination depends on the spatial similarity between the standard response and the unwanted overlapping response. The possible contamination from the M400 to the short prime is highlighted in Fig. 3b and the topographic maps of 3c and 3d, which show the grand averaged (i.e., across individuals) standard response 170 ms after the long prime (the waveforms are shown in the first graph of 3b and the topographic map is shown in 3c). This contamination is seen by comparing these standard responses to the grand average target response 170 ms after presentation of the target (the waveforms are shown in the second graph of 3B and the topographic map is shown in 3d). As seen in the figures, these two topographic patterns, which include 4,000 trials, are very different and the M170 waveform to the target appears to be missing as indicated by the Root Mean Square (RMS) shown in the red line.

First, we report the results from the projection measure and, next, we compare these to the results with traditional sensor selection. A repeated measures one-way ANOVA across the three priming conditions as applied to the projection test values averaged over a 22 ms window (see Table 1 and Fig. 3c), revealed significant differences for the M170, $F(2,18) = 5.202$, $P < .025$. After Geisser-Greenhouse correction to the degrees of freedom, there were still significant differences for the M170, $F(1.75,15.72) = 5.202$, $P < .025$. Subsequent contrasts revealed that the M170 to a repeated word was smaller than the novel condition, but only following a long duration prime, $t(9) = 3.916$, $P < .01$, with no priming effect following a short duration prime, $t(9) = .919$, $P = .382$. This finding replicated the same result found for the N170 in an ERP experiment with identical design [15]. Figure 3c portrays the projection test results, first showing the standard M170 topographic map and then the projection values at each moment in time for the three conditions. The topographic map shows the grand average standard response for illustrative purposes, but the separate standard

responses shown in Fig. 3a were used for the projections prior to averaging across subjects. The shaded region of the waveform graph indicates the 22 ms window used for statistical analyses and the error bar is the standard error of the mean repetition priming difference averaged across short and long duration priming. Unlike the grand average waveforms in the second graph of Fig. 3b, which reveal no apparent target M170 response, there is now a definite peak for the M170 in response to the target (as well as an earlier large peak for the M170 to the short prime). Thus, this normalization technique extracted the M170 component from the overlapping waveforms and did so with a separate normalization for each individual, thereby recovering the small M170 peak to the briefly presented and masked target word.

Next, we compare these results to a traditional technique based on sensor selection. We first calculated the grand average response to the experimental conditions at 170 ms (shown in the topographic map of Fig. 3d) and then selected the 10 sensors with the largest positive magnitude across the three conditions (these sensors are circled in Fig. 3d). Subsequent analyses were performed only for these sensors. Using these sensors, an average M170 response for each individual in each condition was calculated based on the 22 ms time window. Comparing across the three priming conditions with this sensor selection measure, a repeated measures one-way ANOVA found no differences, $F(2,18) = .591$, $P = .565$. After Geisser-Greenhouse correction to the degrees of freedom, there was still no significant difference for the M170 using sensor selection, $F(1.54,12.36) = .591$, $P = .527$ (see Table 1 for the raw values and measures of standard error for each analysis technique). The corresponding waveforms in Fig. 3d may appear to indicate differences between the three conditions, but perhaps the most important result in Fig. 3d is the height of the standard error bar, which is more than twice as large as compared to the projection results in Fig. 3c. Furthermore, even with the maximal M170 sensors selected, there is no obvious M170 peak to the target, unlike the graph in Fig. 3c. Thus, the sensor selection results are too unreliable to conclude that there were any effects of priming condition or even that there was an M170 to the briefly presented target.

Although traditional sensor selection failed to find any reliable results across individuals, the projection measure not only found reliable results, but, furthermore, these results replicated N170 ERP results with the same experimental design [15]. Furthermore, such ‘repetition suppression’ effects immediately following primes are in agreement with several other published results. For instance, the M170 to a face is likewise smaller in magnitude when presented immediately after a face as compared to objects from other categories [13]. Similarly,

Table 1 Target M170 results with traditional sensor selection versus the project measure

	M170 (sensor selection)		M170 (projection test)	
	Mean (fT)	Standard error	Mean	Standard error
Novel	44.829	—	.207	—
Short	40.384	2.896	.176	.034
Long	43.598	4.545	.113**	.024

** Repetition effects significant at the .01 level

with a slightly different paradigm using visual words, early repetition effects have been documented for the P150 [14]. However, unlike these previous results, which were based on sensor selection, the current technique unambiguously indicates that smaller MEG values with priming correspond to smaller cortical responses, and are not due to a change in the pattern across sensors.

Discussion

Using MEG responses to short-term repetition priming of visually presented words as an example data set, we demonstrated the effectiveness of the proposed similarity and magnitude measures. These techniques use all the data from all recorded sensors (i.e., these are multivariate measures), to determine reliable measures of spatial similarity and response magnitude. Calculation of the angle between patterns determined whether conditions produced different patterns, such as would be the case if they involved a different mix of underlying neural sources. Because no differences were found, the projection test indicated response magnitude without worry that apparent differences were due to recruitment of different cortical responses in some conditions but not others. Calculation of the projection between each condition and a standard response provided a magnitude measure that normalized against individual differences and reduced contamination from overlapping responses. We found that (1) individual spatial differences were large and reliable; (2) the conditions of interest were similar to each other in terms of the topographic pattern; and (3) priming reduced the magnitude of the distributed cortical response for the M170 to a target that repeated a long duration prime. In contrast, a traditional analysis based on sensor selection failed to find any reliable effects.

Similar to these measures, registration methods have been developed to handle individual differences [22]. For instance, intersubject registration involves mapping one participant's imaging data onto another participant's imaging data and atlas registration involves the registration of each participant's imaging data onto canonical imaging data. Registration has been successfully employed with MEG data to overcome individual difference (e.g., [5]. However, the angle and projection measures that we employed are more easily implemented and do not require atlas or between subject registration because they instead normalize each person's data against their own standard response.

One limitation of these measures is their inability to localize specific cortical sources. As discussed in the introduction, source localization techniques suffer from limitations due to possibly erroneous simplifying

assumptions. Considering the limitations of these source localization algorithms, alternative analysis methods have been proposed that attempt to delineate separate processes rather than separate cortical areas. Similar to our proposed techniques, Haig and Gordon [9] used spatial projection to provide a measure of response magnitude in different conditions. However, their technique was applied to the subject averaged data rather than separately for each individual. First, they found the spatial pattern corresponding to the average difference between the conditions of interest. Next, the results for each individual were projected onto this difference pattern and inferential statistics were applied to these projection scores from each individual. This is similar to our technique except that we advocate normalizing each individual by their own standard response. Additionally, we suggest that the similarity between conditions needs to be checked before implementing projection in order to rule out changes in cortical recruitment between conditions.

An important statistical concern with scalp EEG and MEG analyses based on individual sensors or groups of sensors is the need to correct for multiple comparisons (i.e., one could continue to select different sensors until a desired result is found). A conservative correction for the pair-wise type I error rate in order to achieve a desired family-wise type I error rate (i.e., the probability that one or more comparisons incorrectly indicates a significant difference), is to use the Bonferroni inequality, in which case the significance level is the desired type I error rate (e.g., .05) divided by the number of sensors or groups of sensors (e.g., 157 for case of the MEG data we reported, requiring that any particular result achieve a probability of .00032 to reach significance). In truth, it is rare that researchers perform this conservative correction because it would require more data collection than is typically feasible. However, multivariate measures across all sensors avoid the problem of multiple comparisons by distilling the data to a single measure of spatial similarity or response magnitude.

Besides spatial similarity and response magnitude, cortical processes can be delineated in their timing. For instance, a study performed by Uhl et al. [29] examined spatiotemporal patterns in EEG, revealing components at different phases at a given temporal frequency (i.e., different modes). In this manner, the evoked response at a sensor was explained by changes in the distribution of the temporal modes. This again points out the dangers of assuming a one-to-one mapping between sensors and responses of interest; not only does an evoked response reflect a distribution of cortical sources, but the underlying sources may each contain a distribution of temporal patterns. Our technique of spatial similarity and response magnitude is not intended as a competitor to such analyses

and is instead complementary. When working with a theory based on the precise timing of cortical processes, such as with spike timing models, or oscillator models and power spectrum effects, then these timing analyses are appropriate. However, when working with a theory based on degree of activation, such as with average firing rate or hemodynamic effects, then spatial similarity and response magnitude are appropriate.

We suggest that these measures provide several distinct advantages over traditional analyses based on individual sensors or groups of sensors. First, they indicate whether the distribution of cortical sources has changed or whether the response magnitude of cortical sources has changed; such a conclusion cannot be made with single sensors or groups of sensors, which can only indicate whether there is a difference. Second, they normalize against individual differences, allowing inferential statistics across participants. Third, by projecting onto a standard response, they reduce the problem of overlapping waveforms that arises in experiments with multiple rapid presentations. Fourth, they distill multi-sensor data to a single number of similarity or magnitude, thus avoiding the statistical problem of multiple comparisons. Fifth, they do not require collection of anatomical information for each individual. Thus, one can calculate these measures without costly structural MRI data. Finally, and perhaps most importantly, these measures are easy to implement and do not require sophisticated software.

Acknowledgment We thank T. Crowley, and V. Hagan for research assistance. This research was supported by NIMH grant MH063993-04.

References

- Baillet S, Garnero L. A Bayesian approach to introducing anatomical-functional priors in the EEG/MEG inverse problem. *Biomed Eng IEEE Trans* 1997;44:374–85.
- Basile LFH, Brunetti EP, Pereira Jr JF, Ballester G, Amaro Jr E, Anghinah R, Ribeiro P, Piedade R, Gattaz WF. Complex slow potential generators in a simplified attention paradigm. *Int J Psychophysiol* 2006;61:149–57.
- Baudena P, Halgren E, Heit G, Clarke JM. Intracerebral potentials to rare target and distractor auditory and visual stimuli. III. Frontal cortex. *Electroencephalogr Clin Neurophysiol* 1995;94:251–64.
- Campanella S, Gomez C, Rossion B, Liard L, Debatisse D, Dubois S, Delinte A, Bruyer R, Crommelinck M, Guerit JM. A comparison between group-average and individual evoked potential analysis. *Neurophysiol Clin* 1999;29:325–38.
- Corouge I, Hellier P, Gibaud B, Barillot C. Interindividual functional mapping: a nonlinear local approach. *Neuroimage* 2003;19:1337–48.
- Dehaene S, Naccache L. Imaging unconscious semantic priming. *Nature* 1998;395:597.
- Di Lollo V, Enns J, Rensink R. Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *J Exp Psychol Gen* 2000;131:590–3.
- Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. *Neuroimage* 1999;9:195–207.
- Haig AR, Gordon E. Projection onto centroids difference vectors: a new approach to determine between group topographical differences, applied to P3 amplitude in Schizophrenia. *Brain Topogr* 1995;8:67–73.
- Halgren E, Baudena P, Clarke JM, Heit G, Liegeois C, Chauvel P, Musolino A. Intracerebral potentials to rare target and distractor auditory and visual stimuli. I. Superior temporal plane and parietal lobe. *Electroencephalogr Clin Neurophysiol* 1995a;94:191–220.
- Halgren E, Baudena P, Clarke JM, Heit G, Marinkovic K, Devaux B, Vignal JP, Biraben A. Intracerebral potentials to rare target and distractor auditory and visual stimuli. II. Medial, lateral and posterior temporal lobe. *Electroencephalogr Clin Neurophysiol* 1995b;94:229–50.
- Hämäläinen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev Mod Phys* 1993;65:413.
- Harris A, Nakayama K. Rapid face-selective adaptation of an early extrastriate component in MEG. *Cereb Cortex* 2007;17:63–70.
- Holcomb PJ, Grainger J. On the time course of visual word recognition: an event-related potential investigation using masked repetition priming. *J Cogn Neurosci* 2006;18:1631–43.
- Huber DE, Tian X, Curran T, O'Reilly RC, Woroch B. The dynamics of integration and separation: ERP, MEG, and neural network studies of immediate repetition effects, submitted.
- Humphreys GW, Besner D, Quinlan PT. Event perception and the word repetition effect. *J Exp Psychol Gen* 1988;117:51–67.
- Kanwisher NG. Repetition blindness: type recognition without token individuation. *Cognition* 1987;27:117–43.
- Kanwisher N, Tong F, Nakayama K. The effect of face inversion on the human fusiform face area. *Cognition* 1998;68:B1–B11.
- Kriegeskorte N, Goebel R, Bandettini P. Information-based functional brain mapping. *Proc Nat Acad Sci* 2006;103:3863–8.
- Levin DT, Simons DJ. Failure to detect changes to attended objects in motion pictures. *Psychon Bull Rev* 1997;4:501–6.
- Liu L, Ioannides AA. A correlation study of averaged and single trial MEG signals: the average describes multiple histories each in a different set of single trials. *Brain Topogr* 1996;8:385–96.
- Maintz JBA, Viergever MA. A survey of medical image registration. *Med Image Anal* 1998;2:1–36.
- Makeig S, Jung T-P, Bell AJ, Ghahremani D, Sejnowski TJ. Blind separation of auditory event-related brain responses into independent components. *Proc Nat Acad Sci* 1997;94:10979–84.
- Mosher JC, Leahy RM, Lewis PS. EEG, MEG: forward solutions for inverse methods. *Biomed Eng IEEE Trans* 1999;46:245–59.
- Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* 2006;10:424–30.
- Nunez PL, Srinivasan R. *Electric fields of the brain: the neurophysics of EEG*. New York: Oxford University Press; 2006.
- Posner M, Snyder C, Davidson BJ. Attention and the detection of signals. *J Exp Psychol* 1980;109:160–74.
- Raymond J, Shapiro K, Arnell K. Temporary suppression of visual processing in an RSVP task: an attentional blink? *J Exp Psychol Hum Percept Perform* 1992;18:849–60.

29. Uhl C, Kruggel F, Opitz B, Yves von Cramon D. A new concept for EEG/MEG signal analysis: detection of interacting spatial modes. *Hum Brain Mapp* 1998;6:137–49.
30. Whittingstall K, Stroink G, Dick B. Dipole localization accuracy using grand-average EEG data sets. *Clin Neurophysiol* 2004;115:2108–12.
31. Wilding EL, Rugg MD. An event-related potential study of recognition memory with and without retrieval of source. *Brain* 1996;119:889–905.
32. Woldorff MG. Distortion of ERP averages due to overlap from temporally adjacent ERPs: analysis and correction. *Psychophysiology* 1993;30:98–119.
33. Xu Y, Liu J, Kanwisher N. The M170 is selective for faces, not for expertise. *Neuropsychologia* 2005;43:588–97.