◆ Human Brain Mapping 5:329–340(1997) ◆

Selective Averaging of Rapidly Presented Individual Trials Using fMRI

Anders M. Dale* and Randy L. Buckner

Massachusetts General Hospital Nuclear Magnetic Resonance Center and the Department of Radiology, Harvard Medical School, Boston, Massachusetts 02129

Abstract: A major limitation in conducting functional neuroimaging studies, particularly for cognitive experiments, has been the use of blocked task paradigms. Here we explored whether selective averaging techniques similar to those applied in event-related potential (ERP) experiments could be used to demonstrate functional magnetic resonance imaging (fMRI) responses to rapidly intermixed trials. In the first two experiments, we observed that for 1-sec trials of full-field visual checkerboard stimulation, the fMRI blood oxygenation level-dependent (BOLD) signal summated in a roughly linear fashion across successive trials even at very short (2 sec and 5 sec) intertrial intervals, although subtle departures from linearity were observed. In experiments 3 and 4, we observed that it is possible to obtain robust activation maps for rapidly presented randomly mixed trial types (left- and right-hemifield visual checkerboard stimulation) spaced as little as 2 sec apart. Taken collectively, these results suggest that selective averaging may enable fMRI experimental designs identical to those used in typical behavioral and ERP studies. The ability to analyze closely spaced single-trial, or event-related, signals provides for a class of experiments which cannot be conducted using blocked designs. Trial types can be randomly intermixed, and selective averaging based upon trial type and / or subject performance is possible. *Hum. Brain Mapping 5:329–340, 1997.*

Key words: single trial; neuroimaging; visual cortex; ERP; MRI; imaging

INTRODUCTION

A major limitation in conducting functional neuroimaging studies, particularly for cognitive experiments, has been the use of blocked task paradigms that require averaging over many trials presented in close succession. Blocked task paradigms have been used because of the need to average across trials to obtain sufficient signal-to-noise ratios to generate functional activation images [Bandettini et al., 1993; Buckner et al., 1996; Frackowiak and Friston, 1995; Grabowski and Damasio, 1996; Kwong et al., 1992; Raichle, 1987]. However, such blocked trial procedures do not allow separate trials within the task blocks to be distinguished. As a result, experiments have either grouped many repetitions of the same trial type or simply accepted that the multiple, intermixed trial types within task blocks could not be separated.

Here, we present procedures which allow for selective averaging of individual trials in mixed task paradigms. Mixed trial types were presented in rapid succession to one another (as little as 2 sec apart), placing the timing domain within the range typically used during behavioral studies. The ability to analyze single-trial, or *event-related,* signals provides for a class of experiments which cannot be readily conducted using blocked designs. For example, the well-known

Contract grant sponsor: NIH; Contract grant number: DC03245-01; Contract grant sponsor: Charles A. Dana Foundation; Contract grant sponsor: Human Frontiers Science Project.

*Correspondence to: Dr. Anders M. Dale, Massachusetts General Hospital Nuclear Magnetic Resonance Center, Bldg. 149, 13th Street, Boston, MA 02129. E-mail: dale@nmr.mgh.harvard.edu

Received for publication 19 March 1997; accepted 16 May 1997

"oddball" paradigms, as studied using event-related potentials (ERPs), rely on the ability to embed a novel or rare event within a train of identical stimuli [Squires et al., 1977]. Similarly, a variety of cognitive experiments involve mixed cognitive stimuli and logged psychophysical responses, and one needs to be able to selectively bin and analyze the data based upon the subjects' performance (e.g., whether they perform correctly on a trial or not). Furthermore, random intermixing of trial types eliminates strategy effects that might otherwise confound the results in blocked task paradigms. Such designs have been extensively applied in studies based on event-related potentials or fields (measured using EEG or MEG).

Several recent demonstrations that functional magnetic resonance imaging (fMRI) can detect activation from brief periods of visual stimulation [Savoy et al., 1995; Boynton et al., 1996; Konishi et al., 1996] and individual cognitive task trials [Buckner et al., 1996; McCarthy et al., 1996] suggest that mixed trial designs are possible. However, these procedures have also shown that the fMRI hemodynamic response is delayed in relation to the neural activity and extends over many seconds, even when the individual trials last only a second or two. For example, Buckner et al. [1996] demonstrated that fMRI signal increases could be detected when averaging individual trials of a word generation task. Such signals were detected in visual areas presumably responding to the perceptual task demands (trials were visually cued) as well as in prefrontal areas involved with the higher-order task demands. The fMRI response for both visual and prefrontal areas lasted for about 10 sec.

The finding that the hemodynamic response to individual trials is temporally extended presents a potential challenge to mixed-trial fMRI studies, since the responses to temporally adjacent trials may overlap, and thus distort the selective averages for each trial type. One solution is to place trials sufficiently far apart so that their hemodynamic responses do not overlap. However, this severely limits the number of trials which can be averaged per time unit, thus limiting the achievable signal-to-noise ratio. A better solution would be to present closely spaced trials, perhaps a few seconds apart, and then apply methods for removing overlap from the selective averages for the different trial types.

In order to estimate and remove such overlap, however, we need a model for how the measured hemodynamic response to a sequence of events relates to the response to each individual event. Rugg et al. [1996] postulated that sequential trial events could diminish the ability to detect different probabilities of these events. The implicit assumption made by Rugg et al. [1996] was that separate trial events add in a nonlinear fashion and, more specifically, that this addition saturates rather quickly. Contrary to this assumption, a recent study suggests that a time-invariant linear model provides a reasonable approximation to the observed blood oxygenation level-dependent (BOLD) fMRI signal, at least for simple visual stimulation [Boynton et al., 1996]. Although Boynton et al. [1996] did observe some clear departures from linearity, it was argued that this might be due to neuronal habituation effects. Given the types of stimuli employed, particularly pulses of varying duration, neural habituation effects would in fact be expected. Similar results were obtained by Konishi et al. [1996], who demonstrated that the absolute magnitude of signal change in fixed regions of interest (ROIs) increased when comparing 200-msec, 2-sec, and 20-sec stimulus durations. In their paradigm, percent signal change increased from .2% to approximately 1% across duration times, suggesting that the signal does not saturate within that hemodynamic range.

Our goal in the experiments reported here was to explicitly determine whether the hemodynamic response to rapidly presented isolated trials adds in a linear fashion and whether individual trials could serve as the basis for mixed trial task designs. We avoided issues surrounding neuronal habituation effects by using single trials of fixed duration, sufficiently separated in time to minimize neuronal habituation effects between adjacent events. This allowed us to more rigorously test the linear hemodynamic response model in regimes relevant to single-trial experiments.

These tests were made across four experiments where we explored the hemodynamic response function derived from closely spaced trials. Our main finding was that the BOLD fMRI response to multiple trials added in a roughly linear fashion, although some departures from linearity were observed. We were further able to use a relatively simple analysis procedure based on selective averaging to create activation maps contrasting separate trial types spaced as little as 2 sec apart. These results have broad implications for the field of functional neuroimaging. They establish that it is possible to conduct studies that isolate individual trials using experimental paradigms identical to those used in typical behavioral and ERP experiments.

GENERAL METHODS

Overview

Our goal in conducting these studies was to explore how the fMRI response to closely spaced trials super-

imposes temporally. Stimuli known to produce robust activations in known areas of the visual cortex were selected. Trials consisted of either full-field counterphased 8-Hz flickering checkerboards (experiments 1 and 2) or right- and left-hemifield counterphased flickering checkerboards (experiments 3 and 4). Activation along the calcarine cortex was explored in each experiment. General methods pertaining to all the experiments are presented first, followed by the specific methods and results for each of the individual experiments.

Subjects

Seven subjects between ages 18–35 volunteered for participation. Each subject was normal or corrected-to-normal in visual acuity. Informed consent was obtained prior to scanning in the manner approved by the Human Studies Committee of the Massachusetts General Hospital.

General magnetic resonance (MR) procedures

Imaging was performed on a 1.5 T General Electric scanner with an echo-planar imaging upgrade (Advanced NMR Systems, Wilmington, MA). A custom-designed bilateral quadrature surface coil was used. Visual stimuli were presented to the subject using a PowerMacintosh (Apple Computer) connected to a Sharp 2000 color LCD projector. Images were projected through a collimating lens (Buhl Optical) onto a screen, mounted within the magnet bore, which could be viewed through mirrors.

Subjects lay on the flat scanner bed with their heads immobilized, using a bite-bar as a means of reducing motion. For each subject, conventional structural images as well as echo-planar functional images were acquired over a 2-hr session. High-resolution anatomic images were acquired (conventional T1-weighted spoiled GRASS (SPGR), 60-slice sagittal, 2.8-mm thickness). An automated echo-planar shim procedure was run to improve B_0 magnetic field homogeneity [Reese et al., 1995]. Slices were selected for all remaining echo-planar acquisitions such that five 7-mm slices were positioned perpendicular to the calcarine cortex. A T1-weighted inversion recovery echo-planar image was acquired for anatomic alignment (TR = 20 sec, TI = 1,100 msec, 1.5625 mm in plane resolution). Finally, T2*-weighted functional images were acquired using an gradient echo sequence sensitive to BOLD contrast (TR = 1 sec, TE = 50 msec, $\alpha = 90°$, 3.125-mm inplane resolution). Functional images were acquired within runs of 240 timepoints. Eight discarded timepoints were acquired prior to each run to allow T1 stabilization.

Selective averaging

Functional runs contained mixed trial types that were between 2–20 sec long, depending on the experiment. Analysis procedures were developed to sort the functional runs by trial type and to compute the averaged fMRI time-courses, along with associated variance estimates, for each trial type on a voxel-by-voxel basis. More precisely, the average fMRI response $y_{i,j}$ for trial type i at time j was computed for each functional run as

$$y_{i,j} = \frac{\sum_{k}^{n_i} s_{t_{i,k}+j}}{n_i}$$

and the associated variance estimate $\sigma^2_{i,j}$ as

$$\sigma^2_{i,j} = \frac{\sum_{k}^{n_i} (s_{t_{i,k}+j} - y_{i,j})^2}{n_i - 1}$$

where n_i is the number of trials of type i in the particular run, $t_{i,k}$ is the start-time (in discrete samples) of the k^{th} trial of type i, and $s_{t_{i,k}} + j$ refers to the $(t_{i,k} + j)^{th}$ sample of the fMRI signal. Across multiple runs, cumulative estimates of $y_{i,j}$ and $\sigma^2_{i,j}$ were computed by simple weighted averaging, with the average and variance for each run weighted by their corresponding number of degrees of freedom (n_i and $n_i - 1$, respectively).

Statistical analysis

Statistical activation maps were generated by calculating the covariance between the observed average signal (or signal difference) and a normalized predicted impulse-response function. More precisely, let

$$q = \frac{\sum_{i}^{m} (y_i - \bar{y})(h_i - \bar{h})}{\sqrt{\frac{\sum_{i}^{m} \sigma_i^2}{m}} \sqrt{\sum_{i}^{m} (h_i - \bar{h})^2}},$$

$$\bar{y} = \frac{\sum_{i}^{m} y_i}{m}, \quad \text{and } \bar{h} = \frac{\sum_{i}^{m} h_i}{m}$$

where m is the number of time points in the averaging window, y_i is the i^{th} time point of the selectively averaged fMRI signal (or signal difference), σ_i is the estimated standard deviation of y_i, and h_i is the i^{th} time point of the predicted fMRI response. The assumed fMRI impulse-response function (within an arbitrary scaling factor) was given by

$$h(t) = \left(\frac{(t-\delta)}{\tau}\right)^2 e^{(t-\sigma)/\tau}$$

with parameters $\delta = 2.5$ sec and $\tau = 1.25$ sec [see Boynton et al., 1996].

Under the assumption of stationary, additive white noise, the random variable q is Student's t distributed with number of degrees of freedom equal to the total number of degrees of freedom of σ_i^2. The statistical significance of the activation at each voxel can thus be calculated using a simple t-test [Press et al., 1992]. When computing the statistical significance of the contrast between two trial types, the mean difference between trial types at each timepoint and associated variance was used in the analysis. Statistical activation images containing *P*-values associated with the results of the t-test were displayed on a pseudocolor scale on a voxel-by-voxel basis.

Linear hemodynamic response model

One of the goals of this study was to test whether the fMRI responses to distinct trials add linearly in time. More precisely, using the notation established above, we wished to test the time-invariant linear hemodynamic response model, which states that the observed fMRI signal s_t at time t is given by

$$s_t = \sum_{i}^{N} \sum_{k}^{n_i} y_{i,t-t_{i,k}} + \eta_t$$

where N is the number of trial types, and η_t represents additive noise. If this model is correct, it would greatly simplify the analysis and interpretation of fMRI data. For instance, the fMRI response waveforms to each trial type can be estimated from the observed fMRI signal by deconvolution, even when the individual trials are presented at a rapid rate [Ganis et al., 1997].

However, given the complex physical and physiological mechanisms underlying the observed fMRI signal, it is highly unlikely that the linear model is precisely correct. For instance, the fact that an fMRI signal cannot increase or decrease without bound implies that the signal can only be linear within a certain range, although the dynamic range of the fMRI signal is thought to be considerably greater than the signal changes observed in typical functional single-trial activation studies. Several past studies suggest boundary conditions for how much expansion of the hemodynamic signal is possible under single-trial task conditions. Konishi et al. [1996] observed a fivefold increase from .2% to 1% in peak signal magnitude as the duration of a visual flickering checkerboard was increased from 200 msec to 20 sec. Boynton et al. [1996] reported an approximate threefold increase as the contrast of fixed duration visual stimuli were manipulated. Finally, Buckner et al. [1996] showed a two- to threefold increase between prefrontal activation observed with single-trial task conditions as compared to blocked task conditions for a word-generation paradigm. Findings such as these suggest that there is sufficient reserve to accommodate hemodynamic summation when multiple, separate single trials are considered. However, this prediction needs to be tested explicitly. Other potential violations of linearity (or time-invariance) assumptions could arise from hemodynamic refractoriness or habituation effects, i.e., the hemodynamic processes might require some time to recover between events.

EXPERIMENT 1

Methods

The goal of this experiment was to determine if the hemodynamic response adds linearly across closely spaced fMRI trials. To accomplish this, clusters of two closely spaced trials (5 sec apart) were compared with single trials (we call these one-trial clusters) (Fig. 1). The logic of this manipulation was that the one-trial clusters could be subtracted from the two-trial clusters and thereby give an estimate of the added hemodynamic response contribution of the second trial. This ''estimated'' second-trial response could then be compared to the hemodynamic response from the one-trial clusters as a simple test of the linear hemodynamic response model. Each trial consisted of a 1-sec full-field 8-Hz counterphased flickering visual checkerboard. Thus, for the two-trial clusters, two 1-sec visual checkerboards were presented separated by 5 sec, while in the one-trial clusters only a single 1-sec checkerboard was presented. Twenty-second separated clusters were presented to allow the effects of the visual stimulation to decay before another set of trials. In this way, the one-trial and two-trial clusters could be compared without confounds due to overlap across clusters. Runs contained 12 clusters (six one-trial and

six two-trial clusters), and eight runs were performed with each subject. Furthermore, one- and two-trials clusters followed each other randomly to avoid order effects. Odd and even trials of each type were separated and analyzed independently to allow reliability to be assessed within-subject.

Results

A clear hemodynamic response could be observed to both the one-trial and two-trial clusters in each of the subjects. Consistent with previous work [Buckner et al., 1996; Boynton et al., 1996; DeYoe et al., 1994; Konishi et al., 1996; Savoy et al., 1995], the hemodynamic response began after about 2 sec and lasted for about 8–10 sec. On visual inspection, the hemodynamic response to the two-trial clusters was significantly increased beyond that of the one-trial clusters. Furthermore, for the one-trial clusters a peak around 5–6 sec could be observed as well as a noticeable second peak around 9–11 sec for the two-trial clusters, presumably reflecting the contributions of the overlapping trials. These observations were highly reliable across odd and even trials in each subject (Fig. 2A,C).

Of central importance, when the one-trial clusters were subtracted from the two-trial clusters, a hemodynamic response highly similar to the one-trial response was observed. This response, when shifted by 5 sec to account for the delay between trials, overlapped considerably with the one-trial hemodynamic response on the rise portion of the curve and decayed at a slightly faster rate, undershooting the hemodynamic response of the one-trial clusters (Fig. 2B,D). This undershoot was observed for both subjects. The similarity between the single-trial hemodynamic response and the estimated second-trial response suggests that the hemodynamic responses from individual trials add roughly linearly.

To test whether activation maps could be produced solely from the hemodynamic response contribution of the second trials, statistical activation maps (t-statistic, see above) were calculated based on the estimated second-event hemodynamic response. Figure 3 shows the results. Robust activations were observed for the second-trial responses, which were highly similar to those observed for the first-trial responses.

These results demonstrate the feasibility of separating hemodynamic responses to trials spaced at least 5 sec apart. In the next experiment, we investigated the feasibility of separating the hemodynamic responses to even more closely spaced trials.

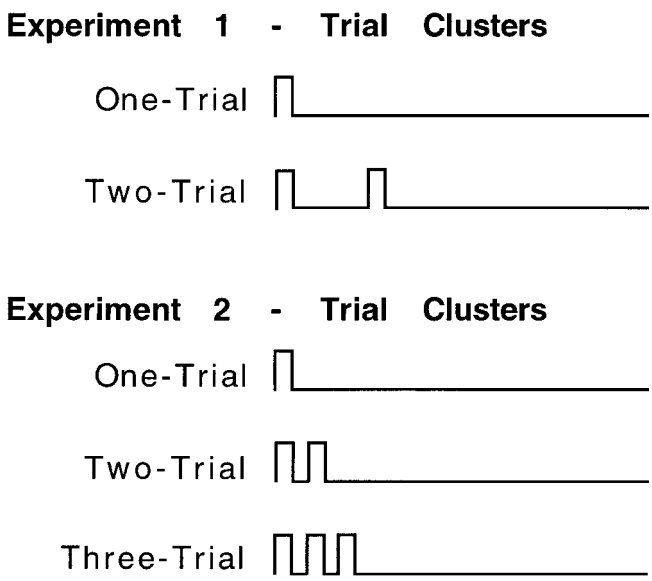

Figure 1.
Diagram of trial clusters for experiments 1 and 2. The total length for trial clusters was kept constant (20 sec) across experiments. The delay between trials and the number of trials was varied across experiments and cluster types as shown.

EXPERIMENT 2

Methods

The goal of this experiment was to determine if the hemodynamic response adds linearly across trials spaced 2 sec apart. Procedures were similar to those of experiment 1, except that three types of trial clusters were examined: one-trial clusters, two-trial clusters, and three-trial clusters (Fig. 1). Thus, experiment 2 differed from experiment 1 by shortening the intertrial interval to 2 sec and exploring trial clusters with up to three trials. Nine runs were collected, with 12 trial clusters per run (four of each type).

Results

Much as in experiment 1, a robust hemodynamic response was detected in relation to each of the trial types, with temporally more extended activation associated with increased numbers of trials (Fig. 4A). By subtracting the time-course of clusters containing one trial from those containing two trials, and those with two trials from those with three trials, the fMRI signal for each trial was estimated. Figure 4B shows the results. Consistent with the data from the 5-sec intertrial intervals, the similarity between the first, second, and third trial responses supports the linear hemody-

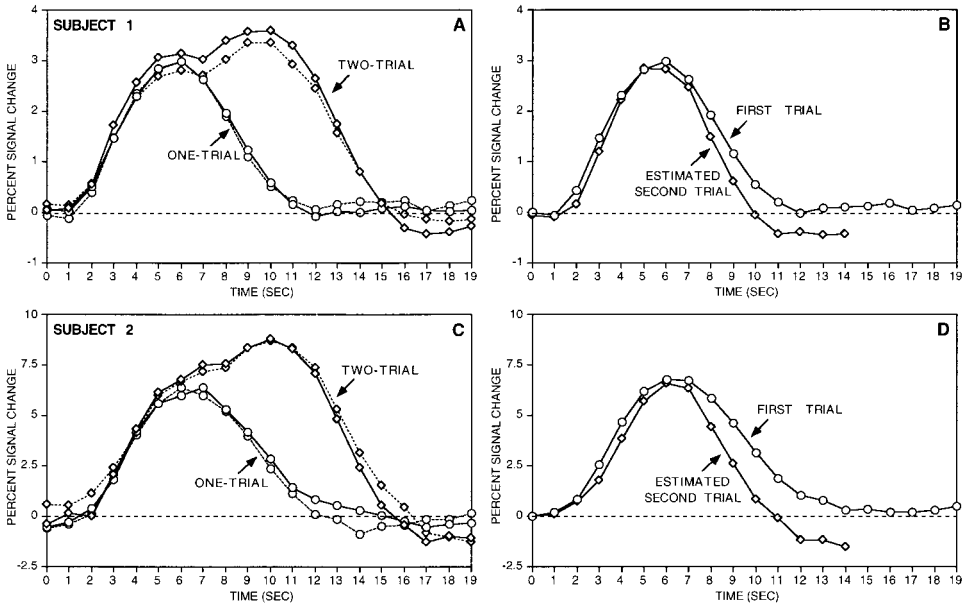

Figure 2.

Time-course data for selectively averaged trials from experiment 1 are shown for two subjects. **A, C:** Raw data for the one-trial and two-trial clusters. Solid and dashed lines show separately the odd and even trials within a cluster type to demonstrate the reliability of the time-courses. Two-trial clusters are consistently longer in duration and reach a higher peak than the one-trial clusters. **B, D:** Estimated contributions of the first trial from the one-trial clusters and the second trial from the two-trial clusters (see Experiment 1 results). Both trials show similar hemodynamic response functions, although some divergence is observed in the declining phase of the response.

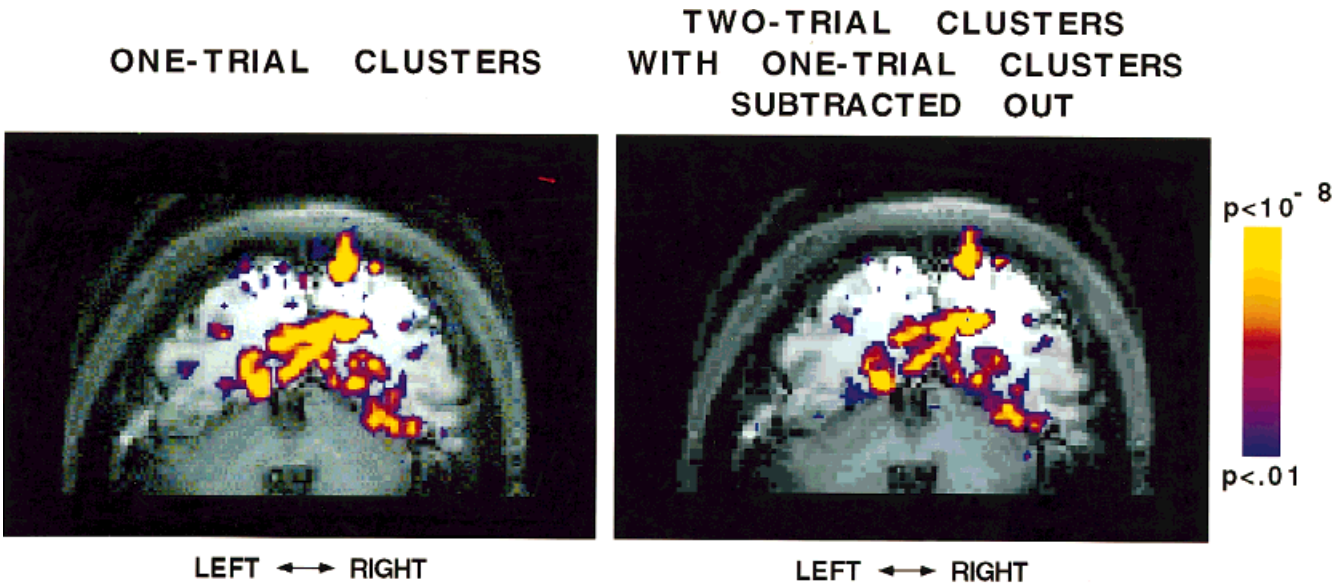

Figure 3.

Statistical maps of visual cortex activation for the first trial from the one-trial clusters and the second trial from the two-trial clusters are shown superimposed on top of echo-planar anatomic images (see Experiment 1 results). The images come from a slice through the occipital lobe, perpendicular to the calcarine sulcus. Brighter colors represent more significant activation, as indicated by the color bar. The images are nearly identical between the two trials, suggesting that activation maps can be made exclusively from the added contribution of subsequent trials presented in a series and that these images are comparable to those produced from isolated single trials.

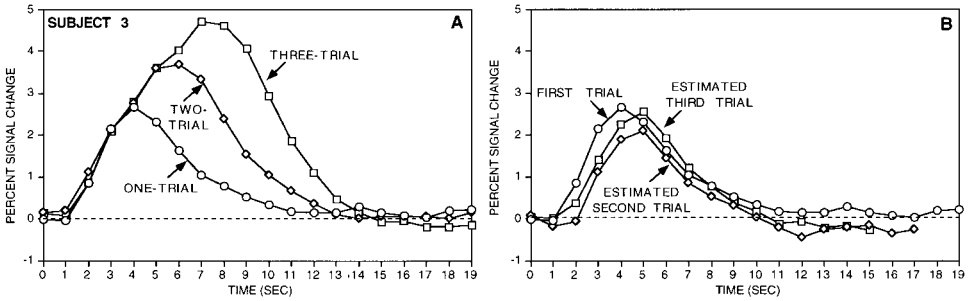

Figure 4.

Time-course data for selectively averaged trials from experiment 2 are shown in a format similar to that of Figure 2. **A:** Raw data from the one-, two-, and three-trial clusters are shown separately and reveal a larger and prolonged hemodynamic response as the number of trials increase. **B:** The estimated contribution for each trial event is shown and, as in experiment 1, the hemodynamic response added in a roughly linear fashion, although a slight departure from linearity is again suggested by the undershoot in the second and third trial responses, not present in the first.

namic response model. Having established that the hemodynamic responses to individual trials add roughly linearly for brief full-field visual stimuli, we next sought to determine if selective averaging methods could be used to distinguish the fMRI responses to different conditions in mixed, randomized experimental designs.

EXPERIMENT 3

Methods

The goal of this experiment was to produce activation images for randomly intermixed trial types. Two visual trial types were examined: 1 sec of left-hemifield stimulation, and 1 sec of right-hemifield stimulation. These two trial types were randomly presented to subjects on a continuous basis for 240 sec. Across runs, the mean intertrial interval was either 5 sec or 10 sec.[1] The intertrial time was not fixed but rather jittered within a 4-sec range centered on the mean intertrial interval. Thus, for the 5-sec mean intertrial interval, presentations occurred randomly at 3 sec, 5 sec, or 7 sec, and for the 10-sec mean intertrial intervals, presentations occurred randomly at 8 sec, 10 sec, or 12 sec. The main reason for using jittered intertrial intervals was that this potentially allows overlap from adjacent trials to be estimated and removed [Hansen, 1983; Woldorff, 1993; Ganis et al., 1997], although such overlap-correction procedures did not need to be applied in the current study.[2] Counterbalancing included assuring that right- and left-hemifield trials occurred randomly, as well as assuring that the jittered intertrial times randomly occurred equally often for both trial types.

Analyses proceeded by selectively averaging across each of the two trial types and directly subtracting the time-courses from each trial type on a voxel-by-voxel basis. Statistical activation images were then constructed by calculating the covariance of the resulting difference time-courses with the normalized predicted hemodynamic response function, as in experiment 1. The result was an activation image that showed signal increases for right- > left-hemifield trials when one direction of the subtraction was considered, and left- > right-hemifield trials when the other direction of the subtraction was considered.

Results

Contrasting the two trial types revealed robust hemisphere-specific activation in the visual cortex for both the 5-sec and 10-sec mean intertrial intervals (Fig. 5). It should be noted, however, that these activations represent *differences* between the two trial types and not simply those areas activated by each trial type vs. a neutral baseline (e.g., fixation). Thus, areas activated in common by the two trial types are not highlighted by this analysis.

The observation that clear signal changes could be detected in each hemisphere demonstrates that selec-

[1]Note that in this context the term ''intertrial interval'' refers to the interval between the onset of adjacent trials, regardless of trial types (sometimes referred to as *stimulus onset asynchrony,* or SOA). Hence, given two trial types presented in randomized order, the mean interval between trials of any given type is twice the stated intertrial interval.

[2]Due to the counterbalancing of conditions, there should be no differential overlap between right- and left-hemifield trials, and thus no overlap-correction had to be applied to the difference time-course.

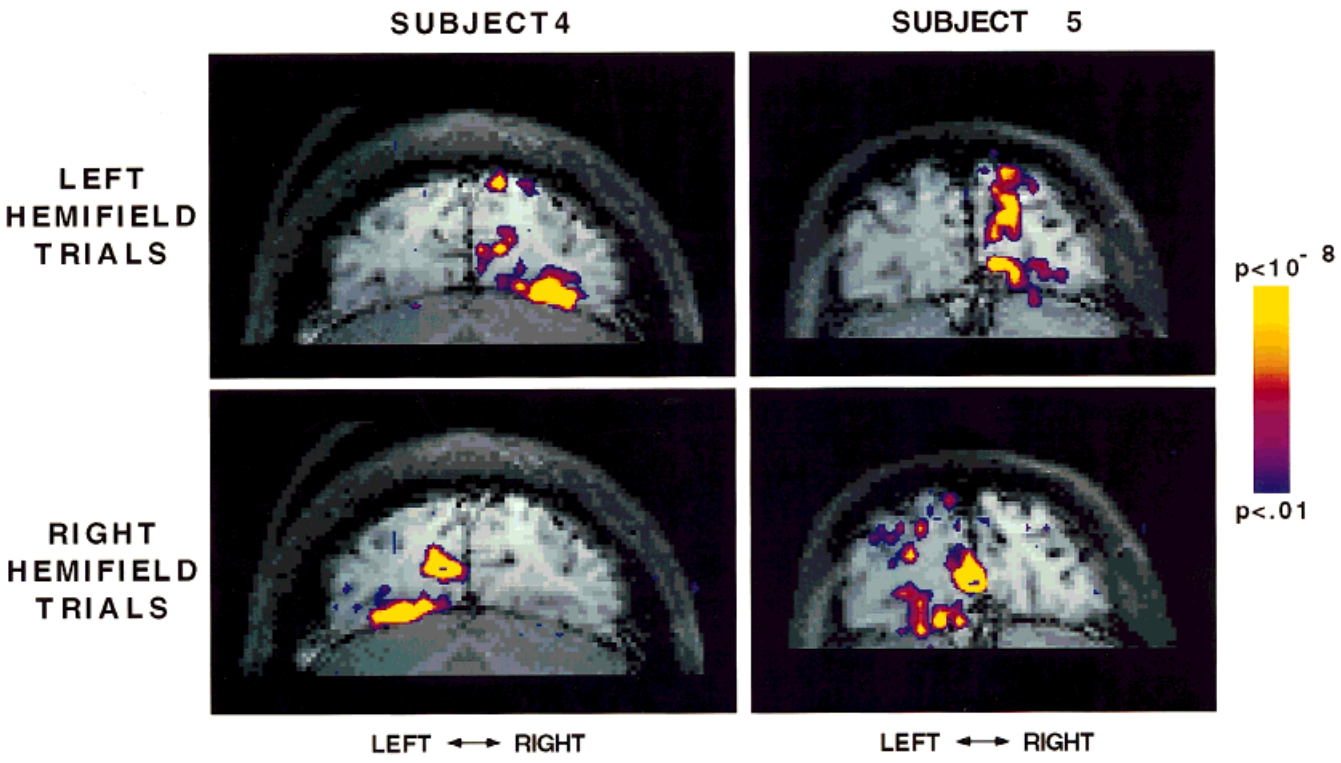

Figure 5.

Statistical activation maps of visual cortex activation for selectively averaged right- and left-hemifield trials from experiment 3 are shown for two subjects. The images come from runs where the two trial types were randomly intermixed with a mean interstimulus interval of 5 sec. The pseudocolor scale represents increased statistical significance (t-statistic) for the direct comparison between trial types, and is overlaid on top of an echo-planar anatomic image. Clear right- and left-hemifield activation is present, as expected based on visual hemifield stimulation, demonstrating that rapidly presented mixed-trial types can be used to construct robust statistical activation maps.

tive averaging procedures can be used to contrast intermixed trial types using fMRI. We next sought to determine the temporal limits of such a procedure in experiment 4.

EXPERIMENT 4

Methods

The goal of this experiment was to determine whether selective averaging of intermixed fMRI trials could be implemented for trials spaced as little as 2 sec apart. Procedures similar to experiment 3 were used, except that three intertrial intervals were tested: 2 sec, 5 sec, and 10 sec. These intervals were fixed, as it was impossible to jitter the 2-sec interval, which was the condition of interest in this experiment. Analyses proceeded in a manner identical to experiment 3, with a specific focus on determining whether activations could be detected for trials randomly presented 2 sec apart.

Results

Clear hemisphere-specific activations were observed at all intertrial intervals, with spatially similar activation patterns (Fig. 6). Highly consistent results were obtained in both subjects. Activations were most robust at the 2-sec intertrial interval, probably due to the fact that more trials were present per unit of time. The observation that signal changes were robust and most significant at the 2-sec intertrial interval suggests that selective averaging with fMRI can examine continuous trials spaced as close in time as typical behavioral and ERP studies.

DISCUSSION

Using selective averaging techniques similar to those applied in ERP experiments, we have demonstrated fMRI responses to single trials spaced as little as 2 sec apart (experiments 1 and 2). When more than one trial

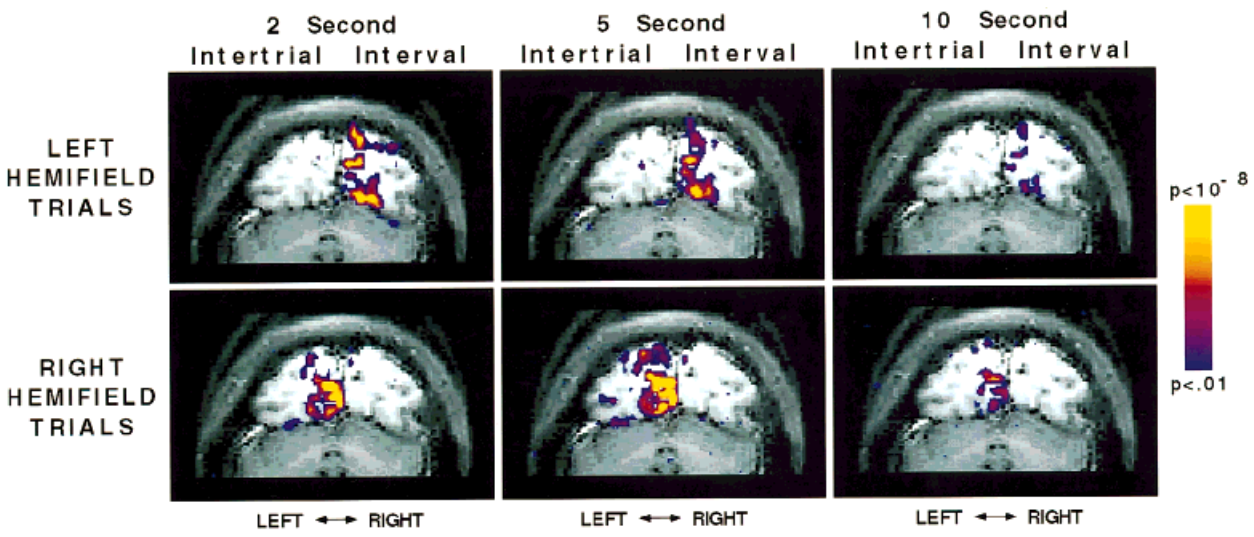

Figure 6.

Statistical activation maps of visual cortex activation for selectively averaged right- and left-hemifield trials from experiment 4 are shown for one of the two subjects. The pseudocolor scale is identical to that in Figure 5. Three separate mean interstimulus intervals are shown: 2 sec, 5 sec, and 10 sec. Similar activation patterns are observed across the three interstimulus intervals, with the most robust maps in the 2- and 5-sec conditions.

was presented in rapid succession, the BOLD fMRI response appeared to add in a roughly linear fashion, although subtle but consistent departures from linearity were observed.

We have further demonstrated that selective averaging procedures could be used to construct statistical activation maps that directly compare two trial types presented within a randomly intermixed format. Such activation maps were shown to be robust even when trials were presented within a few seconds of each other (experiment 4), thereby placing the allowed timing of trials within the range commonly used by behavioral and ERP studies.

Implications

These procedures allow for several new kinds of paradigm design and analyses that were not previously available using blocked designs. Most importantly, different trial types can be randomly intermixed. In experiments 3 and 4 we demonstrated how such a procedure can be used in its simplest form by contrasting two trial types. The procedure should generalize straightforwardly to situations where more than two trial types are considered.

Moreover, more subtle trial types that examine cognitive events can be studied using single-trial procedures. Buckner et al. [1996] demonstrated that prefrontal responses to single trials of word generation could be detected when trials were spaced widely apart (16 sec). McCarthy et al. [1996] and Schacter et al. [1997] further demonstrated the use of single-trial procedures for small (<.5%) signal changes that were detected during ''oddball'' and memory recognition tasks, respectively. They used averaging across subjects to obtain their results and spaced their trials widely apart, as in Buckner et al. [1996]. We anticipate that the present procedures, which have demonstrated the feasibility of selectively averaging closely spaced visual stimulation trials, will generalize to more complex trial types within cognitive task paradigms.

The use of mixed trial types in cognitive paradigms will eliminate strategy effects that might occur in blocked trial paradigms. For example, in studies of memory retrieval, a focus of recent research has been to explore situations in which subjects recognize items as compared to situations where they do not. Because these studies were conducted within blocked trial paradigms, the percentage of recognized items was varied [e.g., Tulving et al., 1994; Schacter et al., 1996; Rugg et al., 1996]. However, a concern has been that subjects adopt different strategies when trial types are blocked [see Rugg et al., 1996 for discussion]. Mixed-trial paradigms allow trials to be randomly presented and sorted post hoc, circumventing many issues related to how target probabilities influence subject strategies.

A second benefit is the ability to sort and contrast trial types based on subject performance. For example,

trial sorting might occur based on whether trials are correctly performed or not, or if reaction times fall within a given range, or if subjects perceive or do not perceive an object at threshold, or based on how the trials bias performance at a later time [e.g., differences based on subsequent memory performance, Paller, 1990]. All of these possibilities can be explored within mixed-trial procedures that allow sorting of trials post hoc.

Caveats

There are a few potential caveats in interpreting our data. First, while we found that trials added in a ''roughly'' linear manner, we did observe subtle departures from linearity. The first event appeared to show a hemodynamic response function that was not identical to the estimated subsequent trial responses: later trials consistently showed a slight undershoot below baseline relative to the first trial. It should be pointed out, however, that experiments 1 and 2, in a way, represent ''worst-case'' tests of the linear model, since the intertrial interval varies greatly between the first and subsequent trials within the clusters (between 5–20 sec in experiment 1, and between 2–20 sec in experiment 2). Thus, any hemodynamic refractoriness or habituation effects would be exaggerated relative to what would be expected in a typical activation study, where the intertrial intervals vary within a much smaller range. Further studies will be needed to test the linear model in regimes more similar to those used in typical activation studies, as well as to determine the optimal stimulus rate given potential nonlinearities.

It should also be noted that the experiments reported here were explicitly designed to eliminate the need for more advanced overlap correction methods. By counterbalancing the trial types, i.e., where each trial type is followed and preceded by each trial type equally often, the overlap between adjacent responses cancels out by simple subtraction, when differences between trial types are considered [e.g., see Woldorff, 1993]. In some cases it may be impossible or undesirable to counterbalance the trial types in this manner. For instance, when trials are sorted based on a subject's performance on each trial, different trial types are not likely to be perfectly counterbalanced. In such cases, explicit overlap correction methods, as described elsewhere [Ganis et al., 1997; Woldorff, 1993; Hansen, 1983], will generally be required. Without such overlap correction, it would seem conservative to space trials sufficiently far apart (perhaps about 10 sec or more) so that contributions from overlapping hemodynamic response functions are minimized.

The statistical analyses used here (see General Methods) are based on covariance with an assumed hemodynamic response function. These methods can be applied with the present parameters specifying one hemodynamic response function, or a set of functions representing a range of parameters. However, in contrast with the multiple regression method recently proposed by Courtney et al. [1996] and Maisog et al. [1996] and the fixed time point selection applied by Buckner et al. [1996], the selective averaging method described here, and the overlap correction method described by Ganis et al. [1997], do not inherently depend on a predefined shape or latency of a hemodynamic response function. It should be possible to use statistical tests, based, for instance, on resampling and cross-validation, which do not require that the response function be known a priori. This is an important advantage, as the shape and latency of the single-trial hemodynamic response is known to vary considerably from region to region in the human brain [Binder et al., 1995; Buckner et al., 1996; Schacter et al., 1997].

Another important caveat is that although the physiological mechanisms underlying the hemodynamic response to regional neuronal activity are thought to be similar across the brain, the results reported here apply only to activation along the calcarine sulcus to simple visual stimulation. Our present results were based on task trials chosen to produce 1-sec epochs of stimulation that would add linearly across trials. Such a procedure allowed us to explore how the hemodynamic response behaved in relation to neural activity without having to factor in large nonlinearity in the summation of the neuronal activity itself. The use of higher-order task trials and examination of regions outside the visual cortex [e.g., those examined by Aguirre et al., 1997; Buckner et al., 1996; Schacter et al., 1997; McCarthy et al., 1996] may produce nonlinearity due to the underlying neural activity above and beyond those due to the transformation of neural activity into a hemodynamic response. Further studies will be needed to test the linearity of the hemodynamic response in other brain regions, and in higher-level cognitive tasks.

Finally, it is important to note that although these results demonstrate that fMRI is capable of detecting changes in cerebral activation evoked by single trials, the relatively slow time-course of the hemodynamic response remains a limitation on the temporal resolution of the technique. However, by combining fMRI with recordings of EEG and/or MEG it may be possible to dramatically improve the temporal resolution of the estimated spatiotemporal activity maps

[Belliveau, 1993; Dale and Sereno, 1993; Dale et al., 1995; George et al., 1995]. Using the selective averaging methods described here, it will be possible to use the exact same experimental designs for both EEG/MEG and fMRI experiments, thus facilitating such combined studies.

CONCLUSIONS

We have demonstrated that the BOLD fMRI responses to individual trials of visual stimulation add roughly linearly, even when the trials are spaced as little as 2 sec apart. We have further shown that selective averaging methods can also be used to distinguish the fMRI response to different trial types, even when these are randomly intermixed and rapidly presented. These results suggest that selective averaging combined with overlap correction methods may enable fMRI experimental designs to be used that are identical to those used in typical behavioral, ERP, and neurophysiological studies.

ACKNOWLEDGMENTS

We thank Bruce Rosen and Jack Belliveau for helpful discussions and Bruce Rosen for resource support; Mary Foley, Terrance Campbell, and Tetsuya Takahashi for help with data collection; and Robert Weisskoff for help with the MRI procedures. R.L.B. was supported by NIH grant DC03245-01 and the Charles A. Dana Foundation. A.M.D. was supported by a grant from the Human Frontiers Science Project.

REFERENCES

Aguirre G, Zarahn E, D'Esposito M (1997): A test of the relationship between hippocampal activity and correct word recognition with trial-based FMRI. Cogn Neurosci Soc 4th Annu Meet 4:63.

Bandettini PA, Jesmanowicz AJ, Wong EC, Hyde JS (1993): Processing strategies for time-course data sets in functional MRI of the human brain. Magn Reson Med 30:161–173.

Belliveau JW (1993): MRI techniques for functional mapping of the human brain: Integration with PET, EEG/MEG and infrared spectroscopy. In: Uemura K, Lassen NA, Jones T, Kanno I (eds): Quantification of Brain Function. Tracer Kinetics and Image Analysis in Brain PET. Amsterdam: Elsevier Science, pp 639–667.

Binder JR, Rao SM, Hammeke TA, Frost JA, Bandettini PA, Jesmanowicz A, Hyde JS (1995): Lateralized human brain language systems demonstrated by task subtraction functional magnetic resonance imaging. Arch Neurol 52:593–601.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996): Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 16:4207–4221.

Buckner RL, Bandettini PA, O'Craven KM, Savoy RL, Petersen SE, Raichle ME, Rosen BR (1996): Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic imaging. Proc Natl Acad Sci USA 93:14878–14883.

Courtney SM, Maisog JM, Ungerleider LG, Haxby JV (1996): Extrastriate and frontal contributions to face and location working memory: An fMRI study. Soc Neurosci Abstr 22:968.

Dale AM, Sereno MI (1993): Improved localization of cortical activity by combining EEG and MEG with MRI cortical reconstruction: A linear approach. J Cogn Neurosci 5:162–176.

Dale AM, Ahlfors SP, Aronen HJ, Belliveau JW, Huotilainen M, Ilmoniemi RJ, Kennedy WA, Korvenoja A, Liv AK, Reppas JB, Rosen BR, Sereno MI, Simpson, GV, Standertskjold-Nordenstam C-G, Virtranen J, Tootell RBH (1995): Spatiotemporal imaging of coherent motion selective areas in human cortex. Soc Neurosci Abstr 21:1275.

DeYoe EA, Bandettini P, Neitz J, Miller D, Winans P (1994): Functional magnetic resonance imaging (FMRI) of the human brain. J Neurosci Methods 54:171–187.

Frackowiak RSJ, Friston KJ (1995): Methodology of activation paradigms. In: Boller F, Grafman J (eds): Handbook of Neuropsychology. Amsterdam: Elsevier Science, pp 369–382.

Ganis G, Kutas M, Schendan ME, Dale AM (1997): DOC: A direct method for temporal overlap correction of ERP and fMRI responses. Cog Neurosci 4th Ann. Meeting, 4:42.

George JS, Aine CJ, Mosher JC, Schmidt DM, Ranken DM, Schlitt HA, Wood CC, Lewine JD, Sanders JA, Belliveau JW (1995): Mapping function in the human brain with MEG, anatomical MRI and functional MRI. J Clin Neurophysiol 12:406–431.

Grabowski TJ, Damasio AR (1996): Improved functional imaging techniques: The dream of a single image for a single mental event. Proc Natl Acad Sci USA 93:14302–14303.

Hansen JC (1983): Separation of overlapping waveforms having known temporal distributions. J Neurosci Methods 9:127–139.

Konishi S, Yoneyama R, Itagaki H, Uchida I, Nakajima K, Kato H, Okajima K, Koizumi H, Miyashita Y (1996): Transient brain activity used in magnetic resonance imaging to detect functional areas. NeuroReport 8:19–23.

Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, Kennedy DN, Hoppel BE, Cohen MS, Turner R, Cheng H-M, Brady TJ, Rosen BR (1992): Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc Natl Acad Sci USA 89:5675–5679.

Maisog JM, Courtney S, Van Horn JD, Haxby JV (1996): Multivariate multiple regression on fMRI data to map functionally distinct areas. NeuroImage 3:79.

McCarthy G, Luby M, Gore JC, Goldman-Rakic P (1996): Functional magnetic resonance imaging in a visual oddball task. NeuroImage 3:548.

Paller KA (1990): Recall and stem-completion priming have different electrophysiological correlates and are modified differentially by directed forgetting. J Exp Psychol [Learn Mem Cogn] 16:1021–1032.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992): Numerical Recipes in C: The Art of Scientific Computing, 2nd ed. Cambridge: Cambridge University Press.

Raichle ME (1987): Circulatory and metabolic correlates of brain function in normal humans. In: Plum F (ed): Handbook of Physiology: The Nervous System, Vol V. New York: Oxford University Press, pp 643–674.

Reese TG, Davis TL, Weisskoff RM (1995): Automated shimming at 1.5 T using echo-planar image frequency maps. J Magn Reson Imaging 5:739–745.

Rugg MD, Fletcher PC, Frith CD, Frackowiak RSJ, Dolan RJ (1996): Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval. Brain 119:2073–2083.

Savoy RL, Bandettini PA, O'Craven KM, Kwong KK, Davis TL, Baker JR, Weisskoff RM, Rosen BR (1995): Pushing the temporal resolution of fMRI: Studies of very brief visual stimuli, onset variability and asynchrony, and stimulus-correlated changes in noise. Proc Soc Magn Reson Med 3:450.

Schacter DL, Alpert NM, Savage CR, Rauch SL, Albert MS (1996): Conscious recollection and the human hippocampal formation: Evidence from positron emission tomography. Proc Natl Acad Sci USA 93:321–325.

Schacter DL, Buckner RL, Koutstaal W, Dale AM, Rosen BR (1997): Brain regions associated with true and false recognition: A single-trial fMRI study. Cogn Neurosci Soc 4th Annu Meet 4:68.

Squires KC, Donchin E, Herning RI, McCarthy G (1977): On the influence of task relevance and stimulus probability on event-related-potential components. Electroencephalogr Clin Neurophysiol 42:1–14.

Tulving E, Kapur S, Markowitsch HJ, Craik FIM, Habib R, Houle S (1994): Neuroanatomical correlates of retrieval in episodic memory: Auditory sentence recognition. Proc Natl Acad Sci USA 91:1989–1991.

Woldorff MG (1993): Distortion of ERP averages due to overlap from temporally adjacent ERPs: Analysis and correction. Psychophysiology 30:98–119.