# 36 Event-Related Functional MRI

R.L. Buckner, T.S. Braver

CONTENTS



## 36.1 Introduction

One of the most exciting areas in functional magnetic resonance imaging (fMRI) today involves the development and application of methods based on the transient neuronal changes associated with individual cognitive and sensory events. These methods, referred to as single trial or event-related fMRI, have broadly expanded the spectrum of task designs and analytical techniques that can be used in neuroimaging studies. In particular, event-related fMRI (ER-fMRI) allows paradigms to depart from "blocked" testing procedures, in which long periods of task performance are integrated, to paradigms that isolate individual trial events or subcomponents of trial events (Fig. 36.1). This provides much greater flexibility in experimental design, by allowing for selective averaging of stimulus events or task conditions which may be intermixed on a trial-by-trial basis. Moreover, by focusing on responses to single events rather than to extended blocks, ER-fMRI provides a means of examining questions regarding the dynamics and time-course of neural activity under various conditions. In this chapter, we provide an in-depth discussion of ER-fMRI, describing its basic principles, methodology, current applications, and future directions. Two important foci of the chapter are: (1) current methodological issues surrounding the use of ER-fMRI, and (2) how ER-fMRI methods have been (or could be) fruitfully applied to expand the range of questions that can be asked in clinical and cognitive neuroscience studies.

R.L. Buckner, PhD; T.S. Braver, PhD; Department of Psychology, Washington University, Campus Box 1125, One Brookings Drive, St. Louis, MO 63130, USA

## 36.2 The Hemodynamic Response

The development of ER-fMRI has followed the development of our understanding of the basic principles of the blood oxygenation level-dependent (BOLD) hemodynamic response (Ogawa et al. 1990, 1992; Kwong et al. 1992). Specifically, there are two key characteristics of the hemodynamic response: (1) it

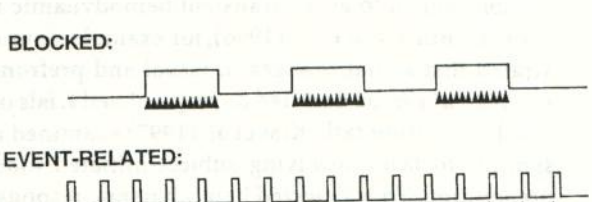


Fig. 36.1. The difference between blocked-trial and event-related paradigms. *Top,* In blocked-task paradigms, many events of the same type are presented in sequences (shown by the *closely spaced arrowheads*). Analysis proceeds by examining signal change averaged across the entire block. *Bottom,* Event-related paradigms attempt to isolate the individual trial events, or even the subcomponents of trial events. The simplest kind of event-related paradigm is illustrated, in which separate events (indicated by *arrows*) are spaced widely apart. As is discussed extensively, many kinds of event-related paradigm are possible, including those that space trials closely together (~2 s) and those that target subcomponents of individual events

can be elicited following brief periods of neuronal activity, and (2) it can be characterized by a well-behaved and reliable impulse-response function.

The first characteristic of the BOLD hemodynamic response – that it can elicited by a brief period of neuronal activity – was observed soon after BOLD contrast became available for mapping brain function. Blamire et al. (1992) presented subjects with brief visual stimuli (2 s), which presumably evoked correspondingly brief periods of neuronal activity. In response to each of the stimuli, the investigators observed transient positive changes in BOLD signal over visual cortex. Similarly, Bandettini (1993) demonstrated signal changes to even shorter task events, examining responses to finger movements for durations ranging from 0.5 s to 5 s. In all situations, including the brief 0.5 s movement duration, clear signal increases were detected in motor cortex. These early studies were followed by a number of similar investigations (Humberstone et al. 1995; Savoy et al. 1995; Boulanouar et al. 1996; Konishi et al. 1996). Savoy et al. (1995), in an extreme demonstration of the temporal limits of these methods, showed that visual stimulation as brief as 34 ms in duration will elicit small, but clearly detectable, positive signal changes.

More recently, these observations have been carried over to the realm of cognitive task paradigms. In such paradigms, the evoked neural activity is less completely understood, and the responses are considerably more subtle than in most studies using sensory stimulation. Nonetheless, isolated trial events within cognitive task paradigms have been reliably shown to evoke transient hemodynamic responses. Buckner et al. (1996), for example, demonstrated that signal changes in visual and prefrontal brain areas can be detected during isolated trials of a word generation task. Kim et al. (1997) examined responses to tasks involving subject-initiated motor preparation. They detected hemodynamic responses in motor and visual cortex. McCarthy et al. (1997) noted transient responses in parietal cortex during infrequent presentation of target letter strings (see Sect. 36.5). Taken collectively, these and other observations make clear that fMRI can detect hemodynamic responses to extremely brief neuronal events, making it possible to be utilized in a truly event-related fashion (for review see Rosen et al. 1998).

The second key characteristic of the BOLD hemodynamic response is the nature and reliability of the shape of the response to a given, brief, fixed interval of stimulation (often described as the impulse-response function). Early studies using sustained stimulation (e.g., Kwong et al. 1992) noted that the BOLD response is delayed in onset from the time of presumed neural activity by about 2–6 s. The work of Blamire et al. (1992) confirmed this finding for brief sensory events and further demonstrated that the hemodynamic response was prolonged in duration. Across a wide range of studies it has now been determined that, for a neural event that lasts a second, the robust positive deflection of the BOLD response will evolve over a 10–12 s period (see hemodynamic response for the "one-trial" condition in Fig. 36.2A). Certain, more subtle, components of the response may have considerably longer recovery periods. From this information, investigators began to

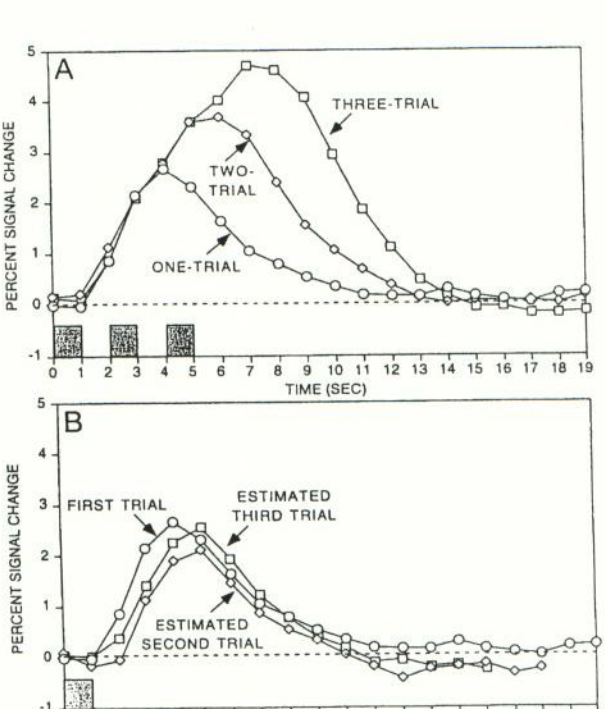


Fig. 36.2A, B. Event-related fMRI (ER-fMRI) data show approximately linear summation of the BOLD contrast signal for closely spaced (2 s apart) trials. A The raw fMRI signal intensity evoked when either one, two, or three sequential trials of a 1 s visual checkerboard stimulus are presented. The placement of each trial is indicated by *shaded bars* at the *bottom* of the graph. The response increases and is prolonged with the addition of multiple trials, indicating that the BOLD contrast signal does not saturate from one to three trials. B Estimates of the separate contributions of each trial are shown. To obtain these estimates, subtraction between trial conditions was employed. The one-trial condition was subtracted from the two-trial condition, and the two-from the three. The three estimated responses are roughly similar, although clear (but subtle) departures from linearity can be observed. These data come from early visual cortex (Adapted from Dale and Buckner 1997)

incorporate explicit models of the hemodynamic impulse-response function into analyses of time-series data, in order to account better for the delayed onset and offset properties. An important step forward in this endeavor was provided by BOYNTON et al. (1996), who demonstrated that the shape of the hemodynamic response to a number of different stimuli could be modeled as a simple function within a linear system. In particular, a function of the form:

$$h(t) = \left(\frac{(t-\delta)}{\tau}\right)^2 e^{-\frac{(t-\delta)}{\tau}} \qquad (36.1)$$

provides a good fit of activity profiles in primary visual cortex (V1) over a range of stimulus durations and intensities. h(t) refers to the signal intensity over time. Parameters d and t can be adjusted to modify the shape of the function. The advantage of having an analytical model of the BOLD response is that it provides a way of generating predictions regarding the idealized fMRI signal expected to a given neuronal event. As will be discussed below, these predictions can be incorporated into statistical procedures as reference functions in order to identify voxels responsive to specific patterns of stimulation or task-related events. Together, knowledge of the characteristics of the BOLD signal – its response to even very brief neuronal events and its shape and time-course – form the basis of ER-fMRI experimental design and data analysis.

## 36.3 Event-Related Functional MRI Data Analysis

Analytical methods have been developed in a number of laboratories to more specifically exploit the power and unique characteristics of ER-fMRI paradigms (CLARK et al. 1998; DALE and BUCKNER 1997; JOSEPHS et al. 1997; ZARAHN et al. 1997; FRISTON et al. 1998). In general, procedures used in ER-fMRI data analysis are conceptually similar to those used in EEG and MEG data analysis, which also apply an event-related approach. At the most basic level, the primary characteristics of ER-fMRI analysis are: (1) determination of the onset of individual events, and (2) selective averaging and/or explicit modeling of responses for each different event-type based on these onsets. The extraction of event-related responses is first dependent upon the image acquisition procedure. That is, images should be acquired in such a fashion that they could be aligned with the events of interest. This alignment or event-locking is most often done to the presentation of stimuli, but can also be locked to other events such as behavioral responses or even spontaneously occurring physiological events such as the beginning of seizure activity or hallucinations (see section 36.6).

Next, the full BOLD response should be extracted or modeled for each event. Of course, given that image acquisition typically occurs continuously across multiple repetitions of trials or stimulus events, this begs the question of what constitutes the appropriate time-course for a trial or stimulus event. In the evoked response potential (ERP) literature, the time-course extracted following an event is typically referred to as an epoch. The epoch chosen for a given event should encompass the duration expected for the full hemodynamic response. Based on previous studies, for transient neuronal events (i.e., less than 2 s) the duration of the hemodynamic response appears to be about 10–16 s. The duration of an epoch need not correspond to the duration between successive events. Indeed, as will be discussed below, some designs are now being explored in which trial or stimulus presentation rate is much shorter than the expected time-course of the hemodynamic response function and thus produces overlap in the epochs extracted for each event. Following extraction of the event-related response, the different event-types can be sorted into separate bins and/or modeled explicitly (including taking into account any overlap across event-types).

In the instance of selective averaging (e.g., DALE and BUCKNER 1997), the mean and variance of the fMRI time-course data can be computed separately, on a voxel-by-voxel basis, for each different trial type or task condition. Such averages yield a sufficient estimate of the central tendencies of the hemodynamic response and the associated variances for each event type. Statistical analysis procedures can then be applied to make inferences about the nature of the response. These procedures may test whether the response is significantly different from zero, or between one or more separate event types. There are a number of different approaches to this type of inferential analysis. The simplest approach is to determine whether there is a significant change in signal among the different images, which make up the epoch. This can be achieved using a one-way analysis of variance (ANOVA). A second approach is to assign fMRI images occurring with a certain delay after a trial to an "on" task state and temporally separate images to an "off" state. These two approaches,

which were used in some initial ER-fMRI studies (e.g., Buckner et al. 1996; Konishi et al. 1996; Cohen et al. 1997b), have the advantage of making few a priori assumptions about the event-related response. However, in situations in which a reasonable model of the hemodynamic response of interest is available, these approaches are likely to be less powerful than alternatives which make use of all known characteristics of the response, such as its shape and transition properties.

More recently, several analyses have been developed that compare the observed event-related response in each image with an idealized model of the expected hemodynamic response function or a range of possible responses (e.g., a basis set of hemodynamic responses). This comparison is typically achieved through a regression procedure that uses the idealized response(s) as a reference function and determines the statistical significance of the correlation at each tested voxel. Within a selective averaging (event-sorting) procedure, in which the mean time-course of an event and the variance are the data of interest, statistical activation maps can be constructed via the covariance between the observed average signal and a normalized predicted impulse response function (Dale and Buckner 1997). For an extended time series which contains the full signal evolution of sequential events over time, the many separate events can be modeled using procedures more akin to typical blocked-design correlational analyses (Kim et al. 1997). Individual events can be thought of as extremely short task blocks. More recent approaches are now being explored that utilize full implementations of the general linear model (GLM) (e.g., Josephs et al. 1997; Zarahn et al. 1997; Friston et al. 1998; Clark et al. 1998). Such methods promise the most flexibility because interactions of event-types with time and performance variables can be easily coded into the design matrix. Using any of these methods, voxels that are best predicted by a particular response function in relation to the known onsets of the stimulus events will be identified.

## 36.4 Limits and Assumptions of Event-Related Functional MRI

A central question for effectively applying ER-fMRI methods is the boundary conditions for their use (e.g., how close in time can sequential events be separated and what kinds of signal-to-noise tradeoffs can be expected). It is our current belief that extremely rapid presentation rates (<2 s between sequential events) are feasible and provide a powerful means of mapping brain function. Several distinct issues directly relate to this conclusion.

### 36.4.1 Stationarity and Linearity

The first issue concerns the stationarity and linearity of the BOLD impulse-response function. Boynton et al. (1996), relying on characterization of responses in visual cortex to controlled visual stimulation, suggested that, on first approximation, changes in intensity or duration of a visual stimulus have near linear and additive effects on the evoked BOLD response. These findings appear to hold for higher cognitive processes as well, in that similar results have been observed in prefrontal cortex activity through manipulations of intensity (load) and duration within working memory (Cohen et al. 1997a). Moreover, data from visual sensory responses suggest the hemodynamic response of one neural event will summate in a roughly linear manner on top of preceding events (Dale and Buckner 1997). Figure 36.2, for example, shows the approximately linear summation for multiple presentations of a 1 s visual stimulus.

The point that summation of the hemodynamic responses is only approximately linear should be seriously considered. Subtle departures from linear summation have been observed in nearly every study that has examined response summation and, in certain studies, the nonlinearities have been quite pronounced (Friston et al. 1997; Vazquez and Noll 1998). Fortunately, using parameters typical to many studies, the nonlinearities may be subtle enough to be considered approximately linear. It has been possible to assume linearity and carry out ER-fMRI studies using presentation rates (~1 event per 2 s or less) that are much faster than the time-course of the BOLD response, or even the repetition time (TR) being applied in a study. Directly relevant to this, Dale and Buckner (1997) showed that individual responses to simple sensory stimuli could be adequately separated and analyzed by using overlap correction methods. Burock et al. (1998) pushed this limit even farther, randomly intermixing varied sensory stimuli at a rate of one stimulus every 500 ms (250 ms stimulus duration, 250 ms gap between sequential stimuli). Recent rapid ER-fMRI studies of

higher level cognitive processes, such as repetition priming (BUCKNER et al. 1998a) and face memory (CLARK et al. 1998), suggest the procedures can be effectively applied to cognitive tasks and higher-order brain regions (e.g., prefrontal cortex).

However, considerably more work must be done to define the precise limits of these "linear modeling" approaches to fMRI data analysis. One difficulty in resolving questions about how the hemodynamic signal summates over time is the fact that in many situations it is not known whether the underlying neuronal activity is itself linearly additive across time and trials. Put another way, when departures from linearity or stationarity are found, it is unclear whether they reflect an intrinsic nonlinear property of the hemodynamic response or of the underlying neuronal activity. For example, auditory word stimuli have been shown to exhibit roughly linear responses when stimuli were presented as frequently as one per 2 s or slower, but robust nonlinearities in the response are observed at higher stimulus presentation rates (FRISTON et al. 1997). It may be the case that the neuronal response to auditory words at such rapid rates is different to that of more widely spaced words; alternatively, the neuronal response to words may be constant across rates but hemodynamic response may saturate. The most parsimonious explanation is that the basic transformation between the summation of neuronal events and the BOLD response is approximately linear, at least with presentation rates typically used in fMRI studies. Instances where strong fMRI signal nonlinearities are observed, such as when very rapid stimuli are presented, may represent situations where there exist nonlinearities in the neuronal activity itself. Further investigation will clearly be needed to resolve this issue as well as develop analysis methods to take into account the nonlinearities, regardless of their origin.

### 36.4.2 Variability

A second central issue in analyzing ER-fMRI data is variability of the hemodynamic response (KIM et al. 1997). As would be expected in any real-world system, variability is present. Consequently, the relevant issues are: (1) the magnitude and practical implications of the variability; and (2) what the sources of variability can tell us about and/or do to limit our exploration of brain function. Preliminary analyses of these issues suggest the within-region hemodynamic response is reasonably stable across subjects (BUCKNER et al. 1998b), although individual subjects can clearly be shown to exhibit some variance in the timing of their response (KIM et al. 1997; RICHTER et al. 1997b). In one analysis, the hemodynamic response was examined across 13 subjects during a memory recognition task (BUCKNER et al. 1998b). Analyses were conducted separately for two different cortical regions (supplementary motor area, SMA, and extrastriate visual cortex). The response in each region was derived from many observations within a subject so that stable within-subject estimate of the responses could be obtained. The question then asked was: if the timing and shape of one subject's hemodynamic response was known for a given region, how well could it predict the other subjects' hemodynamic responses in that region? The answer was clear: the basic shape and timing of the hemodynamic response was stable across subjects (see Fig. 36.3). 72% of the variance of the shape of one subject's response could be predicted – on average – by any other subject. Moreover, the absolute range of the timing of the response was a few seconds, indicating that the standard error estimate of the mean response time was fractions of a second for the group of 13 subjects. As a further empirical dem-

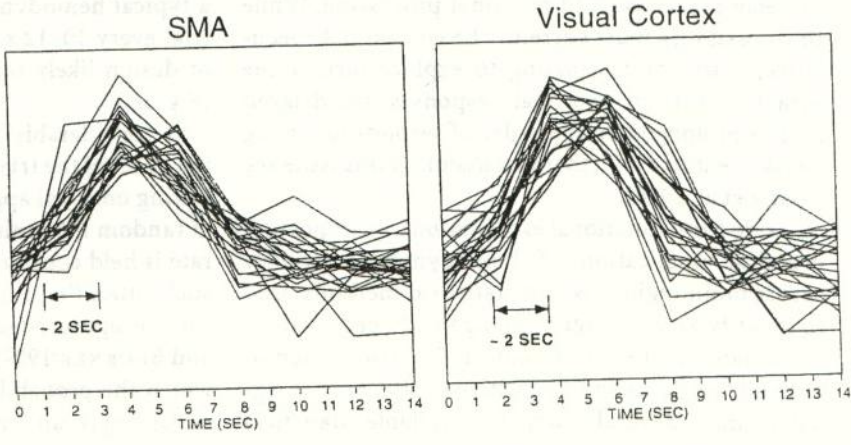


**Fig. 36.3.** The hemodynamic response from 13 separate subjects (two conditions per subject) are plotted in arbitrary unit for each of two brain regions: supplementary motor area (SMA) and visual cortex. As can be seen, variance of the timing of the response has a range of about 2 s for each region, suggesting strong central tendencies in the data and enabling signal averaging across subjects. (Adapted from BUCKNER et al. 1998b)

onstration, the response of one group of six subjects was compared to a second independent group of seven subjects. They nearly overlapped. Given this demonstration of the degree of reliability of the BOLD response within given regions, it should be possible to interpret changes in the onset or shape of the response to different event-types as reflecting changes in neural processing in that region (e.g., Luknowsky et al. 1998), even when central tendencies across averaged groups of subjects are considered.

Information about variation in the delay and shape of the hemodynamic response across brain regions has implications for questions regarding the latency and cascade of neural processing. In this regard, it is noteworthy that marked variations in the timing and shape of responses have been observed across regions even within the same subjects. For example, delays have been noted between visual and prefrontal regions on the order of 0.5–1 s during both word generation and memory retrieval tasks (Buckner et al. 1996; Schacter et al. 1997). In addition, more extreme delays between separate prefrontal regions (about 2–4 s) were noted during the memory retrieval task. While the source of this variation is currently unknown, several possibilities exist. One possible interpretation is that various delays result from differences in the underlying vasculature being sampled across regions (Lee et al. 1995). Vasculature explanations likely account for much of the pixel-by-pixel variance seen within a single functionally specialized region such as V1. However, this possibility seems unlikely to account for all of the broader regional findings observed so far, considering the longest delays have been noted in spatially averaged data and specifically in anterior prefrontal cortex, where large vasculature is minimal (Schacter et al. 1997). A second, more intriguing, possibility is that the observed regional differences in delay reflect delayed neuronal processing. While this possibility must currently be considered speculative, it will be fascinating to explore further the idea that certain neuronal responses are delayed and/or prolonged on the order of seconds following a task event (for a further discussion of this issue see Rosen et al. 1998).

Aside from functional explanations, an important practical implication of hemodynamic variance across brain regions is that statistical methods used to identify areas of signal change will need to allow for variance in the timing and/or shape of the hemodynamic response to be sensitive to all forms of signal change. Several currently available statistical methods either possess this feature (e.g., Schacter et al. 1997; Friston et al. 1998) or make no assumption about the shape (Cohen et al. 1997b).

### 36.4.3 Power

A nontrivial issue that confronts ER-fMRI is determining the relative tradeoffs in power when considering among several possible paradigm designs. For the purposes of this chapter, we reduce the issue of power to two separate questions: (1) What is the ideal mean rate of stimulus presentation? (2) What is the ideal temporal sequencing of the events about that mean rate? The two issues will interact and each has a large affect on the power of an ER-fMRI paradigm. It is perhaps easiest to start a discussion with the simplest example, in which the interstimulus interval is held constant. For purposes of this chapter, we also leave out discussion of interactions between paradigm design and colored noise properties of fMRI studies.

When the interstimulus interval is held constant (e.g., Buckner et al. 1996), the power of the design will begin to rapidly decrease as the intertrial interval becomes much less than the width of the hemodynamic response. It is easy to envision why this occurs: at fast rates, the hemodynamic response from a preceding event is decaying as the hemodynamic response of a subsequent is increasing. Within a linear model, the two will largely cancel each other and the observable response will diminish. Because sequential events are always positioned at a constant distance from each other in time, no deconvolution of the response is possible. Mathematical modeling (Cox and Bandettini 1998) and empirical studies (Bandettini and Cox 1998) suggest that, within a fixed interval design of this sort, the optimal rate for a typical hemodynamic response will be about one trial every 10–12 s. Fortunately, this particular kind of design likely represents the worst case scenario (Fig. 36.4).

A considerably more powerful design is to jitter the onset of the trials in time. For example, instead of having one trial appear every 8 s, trials could appear at random intervals either 6 or 10 s apart. The mean rate is held constant, but the timing is jittered. With such jitter, the underlying hemodynamic response can be appreciated through deconvolution (Dale and Buckner 1997). Considering the specific case in which the probability of an event occurring or not occurring is random and fixed, the obtainable power

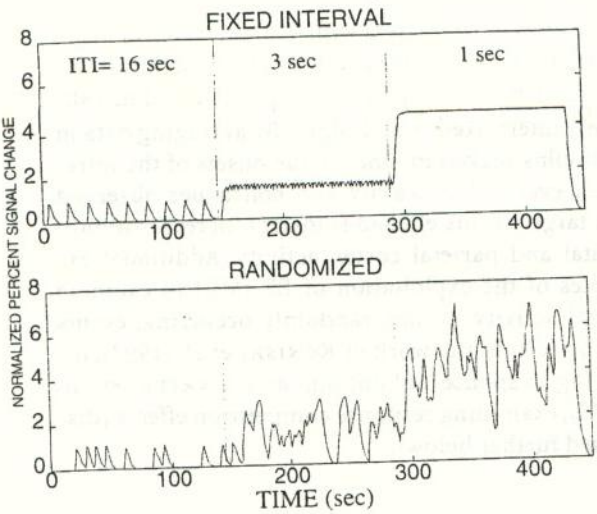


Fig. 36.4. Comparison of fixed interval spacing and randomized spacing of trials for three mean intertrial rates (16 s, 3 s, 1 s). A clear interaction between the kind of ER-fMRI paradigm and rate can be observed. *Top*, For fixed interval paradigms, as the rate of stimulus presentation is increased, the deflection from the mean decreases: sequential events cancel each other out (e.g., the signal goes flat in the 1 s condition). This behavior does not reflect saturation but rather the cumulative effect of summed responses within a linear system (see BUROCK et al., 1998 for more details). *Bottom*, For randomized paradigms, in which trial spacing is jittered in time, the deflection about the mean actually increases at faster rates. This deflection represents signal that can be deconvolved and used as the basis for time-course and statistical activation map generation. The increased mean deflection as stimulus presentation rate increases indicates increased power. For randomized designs, which are appropriate for a number of applications, extremely rapid presentation rates are possible. (Modified from BUROCK et al. 1998)

will increase as the interstimulus interval is decreased (BUROCK et al. 1998). Such behavior in statistical power is nearly opposite to that observed for the case of a fixed interval and considerably more amenable to cognitive and behavioral task design (Fig. 36.4). When the interval between separate events of the same type are jittered in time, short intertrial intervals (~2 s) typical of behavioral and ERP studies provide considerably more power than analogous designs with long intertrial intervals (see BUROCK et al. 1998 for a discussion of the nature of the distribution from which these random intervals can be best selected).

## 36.5 Implications for Human Brain Mapping

The observation that fMRI is sensitive to transient neuronal events has a number of important implications for cognitive neuroscientific and clinical applications. First, fMRI studies can now achieve the same flexibility in experimental design found in behavioral and ERP studies. In particular, these studies typically use paradigms in which individual trial events are elicited rapidly, with distinct event types randomly intermixed. Second, classification of events often cannot be determined a priori – such as when the subject's response to the event is a factor. Being able to isolate fMRI signals allows the BOLD response to be sorted by subject performance (e.g., by whether an event is performed correctly or not, or based on how long an event takes to complete). Third, certain kinds of neuronal responses may not be stable over time and are therefore best addressed through trial-by-trial measurements. Paradigms exploring novelty effects and learning provide prime examples. Fourth, rare and/or unpredictable events can be the target event of interest in which signal change from sequential or common events is the basis. Finally, ER-fMRI provides a means for resolving the activity and dynamics of within-trial events, which may aid many questions in attention and working memory that require the separation of cue- vs probe-related activity, or in examining the dynamics of activity during the delay period between these events.

Thus, the flexibility afforded to experimental design by ER-fMRI allows new classes of behavioral paradigms to be tested. However, as with any methodological advancement, the utility of ER-fMRI will be evaluated by the new scientific questions that are elucidated by their application. Below, we discuss several examples which both demonstrate the impact ER-fMRI is having on the field of cognitive neuroscience and illustrate how ER-fMRI already has, or could be, exploited along several of the dimensions described above.

### 36.5.1 Mixed Trial Designs

The most basic, and possibly most general, advance of ER-fMRI is the increased flexibility it affords to experimental design. In particular, the ability to separate out fMRI responses in an event-specific manner allows the possibility of intermixed event

types. Mixed-trial designs provide a control for many of the strategic effects that can occur when conditions are blocked. For example, many findings in cognitive psychology will be influenced by practice, asymmetrical transfer between conditions, and/or fatigue. Mixed-trial designs avoid many of these effects by equally distributing conditions throughout the experimental session.

An additional issue associated with blocked-trial designs is that they may cause differences in the processing strategies adopted by subjects during task performance, which may result in differential patterns of neural activity. Subjects may anticipate or implicitly adjust performance when they are able to predict trials of a certain type. An illustration of this issue comes from blocked studies of recognition memory in which debate arose as to the role of certain prefrontal areas in memory retrieval (RUGG et al. 1996; BUCKNER et al. 1998b, c; WAGNER et al., 1998a). RUGG et al. (1996), in an influential study, suggested that certain prefrontal areas were most active during trials in which subjects successfully retrieved information. The manner in which this conclusion was derived, however, may have influenced the interpretation. In the aforementioned study and all those that proceeded, blocks of trials that had fewer or greater numbers of successfully retrieved items were the basis of exploration. RUGG and colleagues noted that blocks with the most successfully retrieved items activated the prefrontal areas to the greatest degree. However, by comparing an ER-fMRI study and a blocked-trial study similar to that of RUGG et al. (1996), BUCKNER et al. (1998b, c) were able to show that modulation of the prefrontal areas in relation to retrieval success was closely tied to the use of a blocked-trial procedure; ER-fMRI separation of successfully retrieved and correctly rejected items indicated equal levels of prefrontal signal change. The preliminary conclusion is that the blocked-trial paradigm, in which item events are predictable, may have encouraged subjects to adjust their strategy (implicitly or explicitly) and influenced prefrontal participation. ER-fMRI paradigms can circumvent this issue by presenting trials randomly and contrasting different trial types under conditions in which the specific upcoming event type cannot be predicted.

Another benefit of mixed-trial designs in fMRI studies occurs when experimental effects are necessitated upon the use of low probability events within trains of stimuli. Along these lines, MCCARTHY et al. (1997) explored the ERP P300 (which is evoked by rare events) using ER-fMRI. They presented continuous strings of characters with a target string appearing unpredictably once every 20 or so trials – a procedure that can only be accomplished within randomly intermixed trial designs. By averaging data in a stimulus-locked manner to the onsets of the infrequent events, MCCARTHY and colleagues observed that target events elicited transient increases in prefrontal and parietal cortex activity. Additional examples of the exploitation of ER-fMRI to examine neural activity to rare, randomly occurring, events can be seen in the work of KONISHI et al. (1997), involving response inhibition, and CARTER et al. (1998), examining response competition effects (discussed further below).

### 36.5.2 Post-hoc Sorting

Another dimension in which ER-fMRI increases the flexibility of functional neuroimaging studies is in the ability to sort trials on a post-hoc basis. The most practical aspect of this capability is in the area of artifact rejection. As is done in behavioral research, there are many instances in which it may be desirable to exclude responses on certain trials, based on performance measures such as outlier responses in terms of response time, incorrect responses, or instances in which subjects either do not fixate appropriately or have high motion-related artifact. ER-fMRI task designs make this possible by allowing the investigator to determine post-hoc which trials contain artifacts and exclude these from further analysis. One might even imagine situations in which the "good" trials are few and far between, such as in patients and children in whom motion is a serious issue. So long as enough artifact-free events can be collected, imaging protocols based on infrequently kept events can provide a powerful means of examining subject groups for whom data loss is considerable. In another example, BIRN et al. (in press) addressed the issue of speech-related motion artifact, which is a challenge to fMRI. They showed that modeling the hemodynamic response function to weight images occurring several seconds after the actual event of speaking allows movement artifacts occurring during speech to be minimized (remember, the hemodynamic response to an event is delayed by a few seconds). While not explicitly post-hoc, the study of BIRN et al. (in press) illustrates another instance in which throwing out, or more specifically in their instance, weighting, the inclusion of certain data differentially, may benefit

data analysis.

A more substantive advantage conveyed by post-hoc sorting of fMRI data is for studies that ask questions regarding changes in brain activity associated with certain characteristics of subject performance. For example, in the recognition memory paradigm of BUCKNER et al. (1998b), discussed above, individual trial responses were sorted based on subjects' recognition judgments to previously studied items. ER-fMRI studies using performance-based sorting can also be applied to answer questions that link neural activity with subsequently measured behavioral effects, rather than behavioral effects acquired at the time of the activity. A prime example of such a procedure occurs when examining neural mechanisms underlying memory encoding. Signal change measured during the original presentation of an item can be later sorted according to whether subjects correctly remember or forget the item on a subsequent memory test. WAGNER et al. (1998b) used such a procedure to show that to-be-remembered words activate prefrontal and temporal regions more than to-be-forgotten words. In their study, words were presented rapidly, and subject performance after the imaging session dictated how the word items would be sorted. Such paradigm flexibility previously available to other modalities (e.g., ERP; PALLER et al. 1988) is now afforded to fMRI via event-locked sorting.

A third example of how ER-fMRI can be used to examine neural activity based on subject performance comes from a study by CARTER et al. (1998). In this study, correct vs error trials were sorted separately and compared during performance of a variant of a working memory task called the Continuous Performance Test. Anterior cingulate activity was found to be significantly increased during error trials, consistent with previous reports from the ERP literature of error-related activity stemming from a medial frontal generator (GEHRING et al. 1993; DEHAENE et al. 1994). In addition, CARTER et al. (1998) were also able to examine activity in this same anterior cingulate region during correct performance of low-frequency trials thought to produce differing degrees of response competition. This analysis found that anterior cingulate activity was increased in high-response competition trials relative to trials in which competition was low. These results were interpreted as suggesting that the region monitors the presence of processing conflict, rather than errors per se.

ER-fMRI can also be used in experimental designs in which it is desirable to sort trials based on other aspects of subject performance, such as reaction time (RT). For example, it may be of interest to compare activity in fast response vs slow response trials. An extension of this approach would be to identify regions whose activity correlated with RT on a trial-by-trial basis. This type of approach is complementary to the averaging procedures discussed in Sect. 36.3. In particular, averaging across trials allows the central tendencies in the modal hemodynamic response to be observed, while examination of single-trial hemodynamic responses may provide information about the variability in responses across individual events. Of course, this type of analysis is dependent upon having adequate signal-to-noise characteristics in the fMRI signal. RICHTER et al. (1997a) have recently provided evidence (using a high-field strength magnet) that a single trial of a mental rotation task is sufficient to detect parietal cortex activation. Such a demonstration underscores the kind of analysis that is possible; if individual trial events can provide detectable signals then we should be readily able to correlate with and sort between different item types in fMRI data analysis. As our understanding of the characteristics and limitations of event-related fMRI studies improves, it is likely that techniques will increasingly take advantage of this property in the future.

### 36.5.3 Within-Trial Components and Dynamics

Many clinical and cognitive neuroscientific questions involve resolving the activity and dynamics of the component neural events that may occur within a trial. For example, at the most basic level, in a sensorimotor task, sensory processing should occur in different neural regions and involve different dynamics than does motor processing. However, within a blocked-trial paradigm it would not be possible to dissociate these two components of processing. D'Esposito and colleagues (ZARAHN et al. 1997) recently examined whether ER-fMRI could be used to observe such dissociation. They had subjects make comparisons about the locations of pairs of objects either presented simultaneously (no-delay condition) or separated by 12 s delays (delay condition). Comparison of activation time-courses across the two conditions clearly showed temporal separation of the fMRI signal. The no-delay condition exhibited activation in regions of motor cortex well before similar activation in the delay condition. A similar pattern of results was found in a study by

KIM et al. (1997), which used a visual precuing procedure occurring at variable delays before sequential finger movement.

ER-fMRI can also be used to answer questions about the dynamics of higher level cognitive processes. COURTNEY et al. (1997) provided one of the earliest illustrations of this type of study. They explored a working memory paradigm and organized intermixed trials such that the event of encoding a stimulus could be temporally separated from the act of maintaining the stimulus in working memory. By using analysis procedures that separated the within-trial components related to encoding from those related to maintenance, they were able to show a pathway of brain areas active during both encoding and maintenance but showed differential participation between the two kinds of processes. Posterior visual areas contributed proportionately more to perceptual encoding operations, while prefrontal areas made their greatest contribution to maintenance operations. COHEN et al. (1997b) reported a similar finding by examining the within-trial time courses for a letter working memory task (called the n-back task). They looked for effects of time within the trials and its relation to a second factor, memory load. The logic was that load-sensitive sustained activity across the trial would indicate areas involved in maintenance or other sustained memory-related processes such as rehearsal. In contrast, transient activity (i.e., showing an effect of time) that was load-insensitive would not be related to memory processes. Consistent with the work of COURTNEY et al. (1997), they found effects of the latter type within certain prefrontal and parietal areas and effects of the former type within posterior visual areas and sensorimotor cortex.

These initial studies regarding the neural activity dynamics in working memory pave the way for more detailed explorations using ER-fMRI. In particular, it will now be possible to ask questions regarding how task factors affect multiple different aspects of the neural response such as its latency and duration, in addition to its amplitude. This aspect of ER-fMRI is critical because for many cognitive neuroscience questions the effect of task variables may be hypothesized not to affect the peak amplitude of response. An example of this type of question comes from the recent work of BRAVER et al. (1998) on interference effects in working memory. They presented interfering stimuli during the delay periods of a working memory task, a manipulation that was previously shown to produce maintenance-specific decrements in behavioral performance. Cue-related responses observed in dorsolateral prefrontal cortex were the same amplitude under both baseline and interference conditions. In contrast, the activity decayed more quickly during the delay period under interference, suggesting that the manipulation selectively affected the duration rather than the amplitude of the response in prefrontal cortex.

The advent of ER-fMRI has also paved the way for future studies asking related questions about the onset latency of neural responses under various task conditions. For example, it might be hypothesized that the response of a particular brain region would be delayed as a task increased in difficulty, or when an additional processing stage must occur prior to its activation. Methodological developments by Menon and colleagues (LUKNOWSKY et al. 1998) suggest onset latency differences of 125 ms or less may be possible to resolve.

## 36.6 Future Directions

Currently available ER-fMRI methods offer an array of benefits to human brain mapping research as outlined above. One broad area of growing interest is the integration of ER-fMRI with information from other imaging modalities. This interest arises because the ultimate limits of temporal resolution with fMRI will likely be imposed by the underlying changes in physiology that they measure, which are (for BOLD contrast) fairly indirect, temporally blurred, and affected by differences in regional vasculature (LEE et al. 1995). Transient coordination of neuronal activity occurring across segments of cortex on the time scale of tens to a few hundred milliseconds will likely remain the domain of other techniques (EEG, MEG, and perhaps certain optical imaging methods; GRATTON and FABIANI (in press). These techniques are capable of meaningful, rapid measurements of brain activity but provide relatively coarse spatial resolution. In order to overcome this limitation, methods for combining the temporal resolution of EEG and MEG with the spatial resolution of functional MRI are being developed (e.g., DALE et al. 1995, 1997). Using such methods, it is now becoming possible to study the precise spatiotemporal orchestration of neuronal activity associated with perceptual and cognitive events. Event-related fMRI allows a further refinement of such integration by affording the ability to study the same exact paradigms in fMRI settings and during EEG and MEG sessions.

Another direction of future research concerns the further evolution of statistical analyses for mapping event-related hemodynamic responses. Methods that make few or no assumptions about the shape of the hemodynamic response as well as those that consider the possibility of nonlinear summation of the hemodynamic response are on the horizon. Moreover, as new and more powerful analysis methods provide greater understanding of the true sensitivity and specificity of fMRI data, our ability to address increasingly sophisticated hypotheses about the underlying mechanisms of brain action will continue to improve. In addition, we have yet to come to fully understand the relative signal-to-noise tradeoffs between blocked-trial and event-related paradigms, and for event-related paradigms that space or position trials in fundamentally different ways. It has been empirically demonstrated that randomly intermixed trial events spaced just a few seconds apart can be used to generate activity maps as well as time-course information for the hemodynamic responses associated with these intermixed event types (DALE and BUCKNER 1997; BUCKNER et al. 1998a; CLARK et al.1998; WAGNER et al. 1998b). However, the limitations of this approach are still yet to be determined and may involve constraints imposed by cognitive and neuronal factors that may interact with trial spacing.

Along these lines, it may be important to examine how neural activity is modulated by the local sequential structure of trials even if these are spaced widely enough in time to resolve without overlap correction. Findings of modulation of brain activity by local sequential structure have already been observed in ERP research (SQUIRES et al. 1976), and this provides another arena in which ER-fMRI can provide both convergent and complementary information regarding an as yet unexplained neural phenomena. Another area of experimental design that should be explored in the future concerns pushing the definition of an event. Currently, almost all ER-fMRI studies treat the external presentation of a stimulus as the event of interest. Yet, here again, EEG and MEG research provide demonstration that measurements of brain activity can be aligned to other task-related events such as behavioral responses, eye movements, onset of seizure activity or other non-stimulus locked occurrences. Of course, there are methodological challenges that need to be faced before such techniques can be successfully incorporated, but the basic and clinical research applications of ER-fMRI make the avenues highly worth pursuing.

*Acknowledgments.* We thank Amy Sanders and Deanna Barch for thoughtful comments on an earlier version of this manuscript. Anders Dale, Marc Burock, and Peter Bandettini provided valuable discussion. Support for the work presented in this chapter comes from grants by the National Institute of Mental Health (MH57506-01), the National Institute on Deafness and Other Communication Disorders (DC03245-02), and the McDonnell Center for Higher Brain Function.

## References

Bandettini PA (1993) MRI studies of brain activation: Dynamic characteristics. In: Functional MRI of the brain. Society of Magnetic Resonance in Medicine, Berkeley

Bandettini PA, Cox RW (1998) Contrast in single trial fMRI: Interstimulus interval dependency and comparison with blocked strategies. Proceedings of the International Society for Magnetic Resonance in Medicine, 6th meeting 1:161

Birn RM, Bandettini PA, Cox RW, Shaker R et al. (in press) Event-related fMRI of tasks involving brief motion. Hum Brain Map

Blamire AM, Ogawa S, Ugurbil K et al. (1992) Dynamic mapping of the human visual cortex by high-speed magnetic resonance imaging. Proc Natl Acad Sci USA 89:11069–11073

Boulanouar K, Demonet JF, Berry I, Chollet F, Manelfe C, Celsis P (1996) Study of the spatiotemporal dynamics of the motor system with fMRI using the evoked response of activated pixels: a deconvolutional approach. Proceedings of the International Society for Magnetic Resonance in Medicine, 4th meeting, 3

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

Braver TS, Cohen JD, Barch DM, Noll DC et al. (1998) Effects of interference in working memory on prefrontal cortex activity: a test of a computational model. Neuroimage 7:515

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

Buckner RL, Goodman J, Burock M et al. (1998a) Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. Neuron 20:285–296

Buckner RL, Koutstaal W, Schacter DL, Dale AM, Rotte MR, Rosen BR (1998b) Functional-anatomic study of episodic retrieval: II. Selective averaging of event-related fMRI trials to test the retrieval success hypothesis. Neuroimage 7:163–175

Buckner RL, Koutstaal W, Schacter DL, Wagner AD, Rosen BR (1998c) Functional-anatomic study of episodic retrieval using fMRI: I. Retrieval effort versus retrieval success. Neuroimage 7: 151–162

Burock MA, Buckner RL, Woldorff MG, Rosen BR, Dale AM (1998) Randomized event-related experimental designs

allow for extremely rapid presentation rates using functional MRI. NeuroReport 9:3735–3739

Carter CS, Braver TS, Barch DM et al. (1998) Anterior cingulate cortex, error detection, and the online monitoring of performance. Science 280:747–749

Clark VP, Maisog JM, Haxby JV An fMRI study of face perception and memory using random stimulus sequences. J Neurophysiol 79:3257–3265

Cohen JD, Nystrom LE, Sabb FW, Braver TS, Noll DC (1997a) Tracking the dynamics of FMRI activation in humans under manipulations of duration and intensity of working memory processes. Soc Neurosci Abstr 23:1678

Cohen JD, Perlstein WM, Braver TS, Nystrom LE, Noll DC, Jonides J, Smith EE (1997b) Temporal dynamics of brain activation during a working memory task. Nature 386:604–607

Courtney SM, Ungerleider LG, Keil K, Haxby JV (1997) Transient and sustained activity in a distributed neural system for human working memory. Nature 386:608–611

Cox RW, Bandettini PA (1998) Single trial fMRI: the optimal inter-stimulus interval. Proceedings of the International Society for Magnetic Resonance in Medicine, 6th meeting 1:244

Dale AM, Buckner RL (1997) Selective averaging of rapidly presented individual trials using fMRI. Hum Brain Map 5:329–340

Dale A, Ahlfors SP, Aronen H, et al. (1995) Spatiotemporal imaging of coherent motion selective areas in human cortex. Soc Neurosci Abstr 21:1275

Dale AM, Halgren E, Lewine JD, Buckner RL, Paulson K, Marinkovic K, Rosen BR (1997) Spatio-temporal localization of cortical word repetition-effects in a size judgement task using combined fMRI/MEG. Neuroimage 5: S592

Dehaene S, Posner MI, Tucker DM (1994) Localization of a neural system for error detection and compensation. Psychol Sci 5:303–306

Friston KJ, Josephs O, Rees G, Turner R (1997) Nonlinear event-related responses in fMRI. Magn Reson Med 39:41–52

Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R (1998) Event-related fMRI: characterizing differential responses. Neuroimage 7:30–40

Gehring WJ, Goss B, Coles MGH, Meyer DE, Donchin E (1993) A neural system for error detection and compensation. Psychol Sci 4:385–390

Gratton G, Fabiani M Dynamic brain imaging: the study of the time course of activity in localized brain areas. Psychonom Bull Rev (in press)

Humberstone M, Barlow M, Clare S et al. (1995) Functional magnetic resonance imaging of single motor events with echo planar imaging at 3 T, using a signal averaging technique. Proc Soc Magn Reson 3rd Sci Meet Exhib 2:858

Josephs O, Turner R, Friston K (1997) Event-related fMRI. Hum Brain Map 5:243–248

Kim SG, Richter W, Ugurbil K (1997) Limitations of temporal resolution in functional MRI. Magn Reson Med 37.631–636

Konishi S, Yoneyama R, Itagaki H et al. (1996) Transient brain activity used in magnetic resonance imaging to detect functional areas. Neuroreport 8:19–23

Konishi S, Nakajima K, Uchida I, Sekihara K, Miyashita Y (1997) Temporally resolved no-go dominant brain activity in the prefrontal cortex revealed by functional magnetic resonance imaging. Neuroimage 5:S120

Kwong KK, Belliveau JW, Chesler DA et al. (1992) Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc Natl Acad Sci USA 89:5675–5679

Lee AT, Glover GH, Meyer CH (1995) Discrimination of large venous vessels in time-course spiral blood-oxygen-level-dependent magnetic-resonance functional neuroimaging. Magn Reson Med 33:745–754

Luknowsky DC, Gati JS, Menon RS (1998) Mental chronometry using single trials and EPI at 4 T. Proc Int Soc Magn Reson Med 6th Meet 1:167

McCarthy G, Luby M, Gore J, Goldman-Rakic P (1997) Infrequent events transiently activate human prefrontal and parietal cortex as measured by functional MRI. J Neurophysiol 77:1630–1634

Ogawa S, Lee T, Nayak A, Glynn P (1990) Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magn Reson Med 14:68–78

Ogawa S, Tank DW, Menon R, Ellerman JM, Kim SG, Merkle H, Ugurbil K (1992) Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proc Natl Acad Sci USA 89:5951–5955

Paller KA, McCarthy G, Wood CC (1988) ERPs predictive of subsequent recall and recognition performance. Biol Psychol 26:269–276

Richter W, Georgopoulos AP, Ugurbil K, Kim SG (1997a) Detection of brain activity during mental rotation in a single trial by fMRI. Neuroimage 5:S49

Richter W, Ugurbil K, Georgopolous A, Kim SG (1997b) Time-resolved fMRI of mental rotation. Neuroreport 8:3697–3702

Rosen BR, Buckner RL, Dale AM (1998) Event related fMRI: Past, present, and future. Proc Natl Acad Sci USA 95:773–780

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

Savoy RL, Bandettini PA, O'Craven KM et al. (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 3rd Sci Meet Exhib 2:450

Schacter DL, Buckner RL, Koutstaal W, Dale AM, Rosen BR (1997b) Late onset of anterior prefrontal activity during true and false recognition: an event-related fMRI study. Neuroimage 6:259–269

Squires KC, Wickens C, Squires NK, Donchin E (1976) The effect of stimulus sequence on the waveform of the cortical event-related potential. Science 193:1142–1146

Vazquez AL, Noll DC (1998) Nonlinear aspects of the BOLD response in functional MRI. Neuroimage 7:108–118

Wagner AD, Desmond JE, Glover GH, Gabrieli JDE (1998a)Prefrontal cortex and recognition memory: fMRI evidence for context-dependent retrieval processes. Brain 121:1985–2002

Wagner AD, Schacter DL, Rotte M et al. (1998b) Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. Science 281:188–191

Zarahn E, Aguirre G, D'Esposito M (1997) A trial-based experimental design for fMRI. Neuroimage 6:122–138