NeuroImage **9,** 416–429 (1999)
Article ID nimg.1998.0419, available online at http://www.idealibrary.com on IDEAL®

Deconvolution of Impulse Response in Event-Related BOLD fMRI[1]

Gary H. Glover

Center for Advanced MR Technology at Stanford, Department of Diagnostic Radiology, Stanford University School of Medicine, Stanford, California 94305-5488

Received August 19, 1998

The temporal characteristics of the BOLD response in sensorimotor and auditory cortices were measured in subjects performing finger tapping while listening to metronome pacing tones. A repeated trial paradigm was used with stimulus durations of 167 ms to 16 s and intertrial times of 30 s. Both cortical systems were found to be nonlinear in that the response to a long stimulus could not be predicted by convolving the 1-s response with a rectangular function. In the short-time regime, the amplitude of the response varied only slowly with stimulus duration. It was found that this character was predicted with a modification to Buxton's balloon model. Wiener deconvolution was used to deblur the response to concatenated short episodes of finger tapping at different temporal separations and at rates from 1 to 4 Hz. While the measured response curves were distorted by overlap between the individual episodes, the deconvolved response at each rate was found to agree well with separate scans at each of the individual rates. Thus, although the impulse response cannot predict the response to fully overlapping stimuli, linear deconvolution is effective when the stimuli are separated by at least 4 s. The deconvolution filter must be measured for each subject using a short-stimulus paradigm. It is concluded that deconvolution may be effective in diminishing the hemodynamically imposed temporal blurring and may have potential applications in quantitating responses in event-related fMRI.

INTRODUCTION

Event-related functional magnetic resonance neuroimaging (fMRI) has emerged recently as a means of observing the temporal evolution of neuronal response to a stimulus (Buckner *et al.*, 1996; Rosen *et al.*, 1998). Typically the stimulus is applied for a short duration while T2*-weighted images are continuously recorded, and multiple trials separated by a relatively long intertrial interval (ITI) are employed to obtain sufficient statistical power. The resulting time series can be analyzed by various methods including time-locked averaging, which generates a single time series of duration ITI depicting the average of all trials. Often it is desired to utilize such time series data to quantitate the evolution of activity during some neuronal process, as for example during a working memory task (Cohen *et al.*, 1997).

Unfortunately, the fMRI signal is heavily filtered by the hemodynamic delay inherent in the blood oxygen level-dependent (BOLD) contrast mechanism (Buxton *et al.*, 1997), and therefore temporally evolving events are blurred. In addition to having a temporal width of about 5 s, the response function has been found to have long time constant undershoots (Krüger *et al.*, 1996), and therefore the quantitation of neuronal response from measured fMRI time series is hampered by the hemodynamic distortion. These effects result from the fact that the fMRI signal is only a secondary consequence of the neuronal activity. Regionally increased oxidative metabolism causes a transient decrease in oxyhemoglobin and increase in deoxyhemoglobin, as well as an increase in CO_2 and NO. While the process is still not fully understood, this in turn causes arterioles in the affected region to dilate and blood flow to increase, producing an overabundance of oxygen in the capillaries and venules draining the activated region. This process takes several seconds, after which the change in oxygen level results in altered magnetic state of the tissues, which modulates the MRI signal through changes in T2 or T2*. Temporal evolution of the MRI signal is therefore delayed relative to the neuronal activity because of the hemodynamic time lag in responding to the need for increased oxygen (Hu *et al.*, 1997). Buxton and colleagues have developed a phenomenological "balloon" model that can qualitatively explain both the delay and the undershoot based on transient volumetric changes in the venous outflow tract (Buxton *et al.*, 1998).

Under certain conditions, the fMRI response has

[1] Supported by NIH P41 RR 09784, the Lucas Foundation, and GE Medical Systems.
[2] The Center for Advanced MR Technology at Stanford is a program of the NIH National Center for Research Resources.

1053-8119/99 $30.00

been found to be approximately linear (Boynton *et al.*, 1996), and this has been the basis for most of the event-related fMRI analysis to date (Buckner *et al.*, 1996; Cohen, 1997; Friston *et al.*, 1998). With the assumption of linearity, the response to any stimulus can be predicted if the response to a very short stimulus (called the impulse response) is known. Friston and colleagues assumed a Poisson distribution function (Friston *et al.*, 1994), which was later found to be consistent with measurements in human visual area V1 (Boynton *et al.*, 1996). Often the impulse response is assumed to be identical for all individuals and without regard to cortical system. Cohen, for example, used the measured response from a visual experiment as the basis for quantitating sensorimotor activation with finger tapping at different rates (Cohen, 1997). However, some departure from linearity was noted in V1, especially at shorter stimulus times (Boynton *et al.*, 1996). This trend was also observed in the auditory cortex by Robson and colleagues (1998) and Binder and colleagues (1994). A study of the visual system showed strong nonlinearities for short stimuli (Vazquez *et al.*, 1998). Friston and colleagues used an extension of convolution methods with nonlinear basis functions to account for nonlinearities in auditory response to words spoken at different rates (Friston *et al.*, 1998). This modeling approach requires a priori assumptions about the impulse response in the form of preselection of basis functions and is explicitly nonlinear. Aguirre and colleagues have shown that response functions in sensorimotor cortex vary across individuals, but are relatively reproducible within scan sessions (Aguirre *et al.*, 1998). Thus, if impulse response functions are measured on an individual basis, it may be possible to use these functions for deblurring more complex responses.

The use of deblurring methods for analysis of event-related time series was reported by Hykin, who showed improved temporal resolution in a working memory task by employing deconvolution with an assumed impulse response (Hykin *et al.*, 1996). Only a binary decision was made as a function of time as to the presence or absence of a response, however, so that quantitation of the deconvolved response was not possible. Linear analysis techniques such as SPM (Friston *et al.*, 1994) can incorporate an impulse response function into the correlate; however, the intent is to improve the parametric map rather than to explicitly derive a temporally sharpened time course.

In this work we propose an alternative approach, in which deconvolution is used explicitly to remove the effect of the impulse response from the measurements in order to attempt more accurately to depict the time course of the neuronal activity from the measured time series. In order for deconvolution to be effective, it is important to understand the dynamic characteristics of the fMRI response to determine the limits of applicability of linear system theory. Therefore, the goals of this study were twofold: to measure the dynamic characteristics of motor and auditory cortical systems and to examine the use of deconvolutional deblurring methods for event-related fMRI.

Event-related experiments that allowed simultaneous observation of the BOLD signal in sensorimotor and auditory cortex were performed. Two types of experiments were developed in order to examine different aspects of linearity. The first class of experiments was designed to see whether the response to long stimuli could be predicted from the measured response to a short stimulus, by observing response to stimuli of varying duration. The second type of experiment used multiple bursts (or episodes) of activity and deconvolution to examine whether the quantitative responses at different stimulus rates and episode repetition schedules could be separately recovered from the measured time series, in the presence of overlap between the response functions from each episode. The variable rate experiment was similar to that employed by Rao *et al.* (1996).

MATERIALS AND METHODS

Experimental Design

Subjects performed repeated trial experiments in which each trial consisted of various durations and rates of finger tapping paced by auditory cueing tones supplied through earphones. For some experiments a visual cue (an X or box) was projected on a screen 1 s before the onset of the auditory stimulus and finger tapping to alert the subject that a new trial was to begin. For the variable-rate finger-tapping experiments, an auditory cue with similar timing was provided. Image volumes that covered both motor and auditory cortex were acquired at the temporal sampling rate of 1/s. The subjects were normal volunteers from the Stanford community who gave informed consent. The cohort consisted of four males and three females, with age 36.7 $\pm$ 10.1 (mean $\pm$ SD) years.

Experiment I: Stimulus Duration

The first type of experiment used tone pulses of different durations to pace the subject to perform finger tapping at a given rate and duration. The tones (MIDI note 72, 523.25 Hz) were square-wave amplitude modulated at a 3-Hz rate for the desired duration and were repeated every 30 s (ITI) for 10 trials (Fig. 1). An additional 30 s of baseline images was collected before the first trial, for a total scan time of 330 s. The subjects were instructed to use bilateral sequential finger apposition in exact time with the tones and to remain as still as possible before and after all tones. A visual cue was used to alert the subject to the beginning of a new trial. Separate scans were obtained for each subject perform-

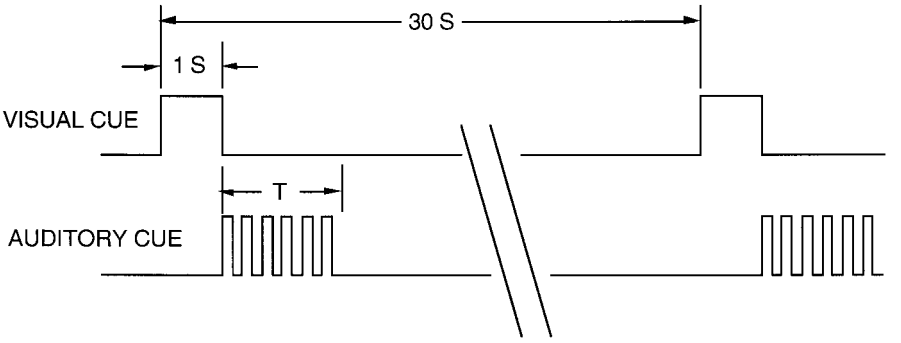

FIG. 1. Stimulus for experiment I. A visual cue (a blue box shown on screen) begins 1 s before the tones to alert subject to the onset of a finger-tapping task. The tones are square-wave modulated at 3 Hz for duration T, where T varies from ⅓ s to 16 s. Each trial is 30 s long, and there are 10 trials in each scan. Images are acquired at 1-s intervals, with $t = 0$ at the beginning of the auditory cue.

ing finger tapping with square-wave durations of ⅓, ⅔, 1, 2, 4, 8, and 16 s (1, 2, 3, 6, 12, 24, and 48 tone pulses, respectively). The exact timing of the subsecond motor movements was difficult to control because of subject response time; the ⅓ s paradigm prompted the subject with only 1 tone pulse of duration ⅙ s every 30 s, and the actual motor function was undoubtedly variable. In addition to the seven repeated trial experiments, a conventional block trial experiment was performed using the same auditory/visual cueing method, with 20 s of finger tapping repeated every 40 s for a total duration of 200 s. Five subjects performed this set of experiments.

Multiple Episode Experiments

In the second class of experiments, multiple episodes of short duration finger tapping were performed, as shown in Figs. 2A–2D. The episodes were spaced closely enough that the responses to each episode would be expected to overlap due to the slow hemodynamic function. Two experiments of a larger series of experiments are described.

Experiment IIA

In this experiment (Fig. 2A), each episode consisted of finger tapping at 3 Hz for 1 s, and each trial consisted of three concatenated episodes with variable spacing T_s between episodes ($1 \text{ s} \leq T_s \leq 5$ s). A single-episode scan was also acquired to provide an impulse response function for deconvolution of the other data sets. In all cases a 1-s visual cue was provided before the first auditory cue, but not between episodes. This set of experiments was designed to see how closely the episodes could be concatenated and resolved using deconvolution. Two subjects performed these experiments.

Experiment IIB

In a second form of the deconvolution experiments, repeated episodes were employed in each trial, as shown in Figs. 2B and 2C. The rate of finger tapping was varied between 1 and 4 Hz, and either two or four episodes were employed in an attempt to generate controlled levels of activation. Data for single episodes were also gathered to generate an impulse response function at each frequency. An auditory cue was given 1

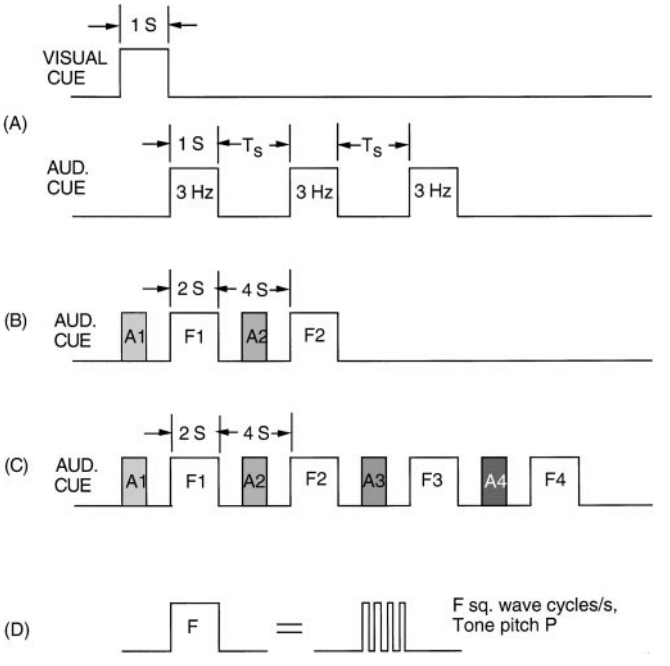

FIG. 2. Stimuli for experiment II, showing 1 trial of 10 total. The tones are square-wave modulated at rates between 1 and 4 Hz (D). (A) For experiment IIA, a visual cue begins 1 s before the tones. Separation between adjacent stimulus episodes T_s varies between 1 and 5 s, with 3-Hz tone/finger-tapping rate. Each trial is 30 s long. (B and C) For experiment IIB, a nonmodulated 1-s-long auditory alerting cue begins 1 s before the finger-tapping tones. Two or four 2-s stimuli are delivered with 4 s spacing between episodes. In some experiments the pacing rates F_i were identical (see Table 1). The MIDI pitch of both the alerting tones and the pacing tones is linearly related to F_i: for $F = 1$ Hz, $P = 66$, $A = 58$; for $F_i = 2$ Hz, $P = 70$, $A = 62$; for $F_i = 3$ Hz, $P = 72$, $A = 64$; for $F_i = 4$ Hz, $P = 74$, $A = 66$. Images are acquired at 1-s intervals, with $t = 0$ at the beginning of the finger-tapping cue.

s before each burst as shown to alert the subject to the next event. The frequency (MIDI note) of the cuing tone and the pacing tone varied with the finger-tapping rate, enabling the subject to anticipate changes in the task demand by the pitch. Because auditory alerting tones were used before the finger-tapping episodes rather than visual cues, only the sensorimotor data were analyzed. Two subjects performed these experiments, as given in Table 1. The first subject performed all scans with intertrial repetition period ITI 30 s, while the second performed only scans 1–4, 7, and 8 with ITI 35 s. Ten trials were used together with a 30-s prestimulus baseline acquisition period. An on/off block trial acquisition was also obtained. This set of experiments was designed to test the linearity of the deconvolution method with stimuli that would generate overlapping responses at different activation levels.

Data Acquisition

All experimental data were obtained using a 1.5-T scanner equipped with high-performance gradients and receiver (GE Signa Echospeed, Rev. 5.6; Milwaukee, WI). T1-weighted FSE scans were acquired for anatomic reference (TR/TE/ETL = 68 ms/3000 ms/8). Functional acquisitions used a 2D single-shot spiral gradient-recalled sequence with linear shim corrections applied during reconstruction (Glover *et al.*, 1998). The acquisition parameters were TE/TR/FA = 40 ms/1000 ms/75°, FOV 20 cm, six 5-mm-thick axial slices, in-plane resolution 2.2 mm (matrix 90 × 90). A high-performance slice select RF pulse allowed the use of contiguous slice spacing, with three slices positioned to intersect sensorimotor cortex with no interslice gap and three positioned in auditory cortex. A small, receive-only elliptical birdcage head coil was used for all scans (Hayes *et al.*, 1996), and subjects were stabilized with foam padding packed tightly in the coil. A back-projection screen attached to the coil assembly was viewed through a mirror. Auditory and visual cues were generated by a custom program on a Macintosh computer (Mountain View, CA) and delivered by pneumatic earphones and a video projector, respectively (Resonance Technology, Van Nuys, CA). The computer also started the scan with a direct connection to the scanner's auxiliary trigger port, in order to synchronize stimulus and acquisition. Images were reconstructed into a 128 × 128 matrix with an offline computer (Sun Microsystems, Mountain View, CA) using gridding and FFTs.

TABLE 1

Experiment IIB Scan Parameters

Scan	F_1 (Hz)	F_2 (Hz)	F_3 (Hz)	F_4 (Hz)
1	1			
2	2			
3	3			
4	4			
5	3	3		
6	3	3	3	3
7	1	2	3	4
8	4	3	2	1

Image Analysis

The block trial data were analyzed using cross-correlation with a sine wave (Lee *et al.*, 1995) after linear trend removal. A 3 × 3-pixel sigma filter was used to diminish the occurrence of apparently significant single pixels. Activated pixels were identified from the cohort of surviving pixels having correlation coefficients exceeding 0.25, and the resulting activation maps were overlaid on the T1-weighted scans.

Repeated-trial data were first subjected to trend removal by fitting a second-order polynomial to the entire time series and subtracting the resulting fit. Time-locked averaging was then employed to develop an averaged series of images containing 30 or 35 time frames, depending on ITI. Pixels identified as activated (as described below) and located within a region of the brain containing cortex were averaged together for each time frame to provide a measurement of the dynamic behavior, $m(t)$. Pixels containing vessels were excluded by using a mask constructed from the time series standard deviation map, in which high standard deviation was taken to signify the presence of a vessel. In all cases the block trial correlation maps provided sufficient signal to generate reliable masks in the motor cortex for identifying activated pixels in the repeated trial data sets. In most cases the block trial maps were also used for masking the repeated-trial data sets in auditory cortex, although in some subjects the auditory response in the block trial maps was too weak, and the averaged time frame at peak signal was used instead to generate the mask. In this way, measurements $m(t)$ were obtained for both sensorimotor and auditory cortex for each experimental paradigm. Finally, the time series for all pixels was cross-correlated with $m(t)$ for the corresponding T and postfiltered as above to develop repeated-trial activation maps that were overlaid on the T1 anatomic scans as in the block trial experiments.

Linearity of the functional response was assessed in two ways. For experiment I, the $m(t)$ for a given subject measured with 1-s duration finger tapping was assumed to represent the impulse response and used to predict the responses to longer stimuli by convolving the putative impulse response with rectangular functions of duration 2, 4, 8, and 16 s. The resulting curves were then compared to the corresponding measured $m(t)$.

For the second class of experiments (IIA, IIB), measured data $m(t)$ for short (1 or 2 s) single-episode experiments were defined as the impulse response and used to generate a deconvolution filter to be applied to the multiple-episode experiments, in order to quanti-

tate the response. Let $h(t)$ be the measured impulse response, and $s(t)$ the stimulus time series. For example, the function $s(t)$ in Fig. 2A would be approximately $\delta(t) + \delta(t - \tau) + \delta(t - 2\tau)$, where δ is the Dirac delta function and $\tau = T_s + 1$ s (the 1-s rect functions are short compared to the expected characteristic width of $h(t)$). If the system is linear, the measured response for the time series $s(t)$ should be

$$m(t) = s(t) * h(t) + n(t), \quad (1)$$

where * denotes convolution and $n(t)$ is the noise in the measurement. An approximation $s'(t)$ can then be obtained from the measurements $m(t)$ using a Wiener filter $d(t)$ (Papoulis, 1977),

$$s'(t) = d(t) * m(t). \quad (2)$$

Let $H(\omega)$, $M(\omega)$, $N(\omega)$, and $D(\omega)$ be the Fourier transforms of $h(t)$, $m(t)$, $n(t)$, and $d(t)$, respectively. Then,

$$D(\omega) = \frac{H^*(\omega)}{|H(\omega)|^2 + |N(\omega)|^2}, \quad (3)$$

where * denotes complex conjugate here. The estimated response to the stimulus $s'(t)$ is then given by

$$s'(t) = \mathcal{F}^{-1}[D(\omega)\, M(\omega)], \quad (4)$$

where $\mathcal{F}^{-1}$ is the inverse Fourier transform operator. Note that in the absence of noise ($N(\omega) \rightarrow 0$), $d(t)$ becomes an inversal filter as expected, $d(t) = h^{-1}(t)$.

The noise spectrum $N(\omega)$ can be obtained from time series measurements in nonactivated cortical regions. However, for simplicity this study assumed a constant spectrum (white noise) with $|N(\omega)|^2 = N_0^2$. N_0 was estimated from measured spectra $H(\omega)$ (i.e., in the activated cortex) at high frequency, avoiding the stimulus frequency and its harmonics. Thus, we have

$$s'(t) = \mathcal{F}^{-1}\left\{\frac{H^*(\omega)M(\omega)}{|H(\omega)|^2 + N_0^2}\right\}. \quad (5)$$

Equation (5) was used to estimate the response for each of the experiments of type II. The peak height of $s'(t)$ for each episode was also plotted.

RESULTS

Experiment I: Stimulus Duration

The measured responses averaged over five subjects performing the task in Fig. 1 are given in Figs. 3A and 3B, for motor and auditory cortex, respectively. The subsecond (⅓ s, ⅔ s) motor responses do not differ significantly from each other, although the peak amplitude for the auditory measurements in a monotone increasing function for $T \leq 8$ s. The predicted responses for 2 s $\leq T \leq$ 16 s are shown in Figs. 3C and 3D, using the 1-s data as the impulse response function. It is apparent from these curves that neither cortical system is linear. In particular, linearity predicts that for short stimuli the peak amplitude should be proportional to the stimulus duration T, as shown in Figs. 3C and 3D. In addition, the short duration measurements include an undershoot that is suppressed in the longer T scans. This is clear in the predicted results, which have much greater undershoots following the positive response than the measurements. The $T = 1$ s average data are shown in Figs. 3E and 3F, together with the standard deviations, which indicate moderately reproducible characteristics across subjects performing the same task. The data were fit to analytic functions of the form

$$y(t) = c_1 t^{n_1} e^{-t/t_1} - a_2 c_2 t^{n_2} e^{-t/t_2}$$

$$c_i = \max(t^{n_i} e^{-t/t_i}),$$

where for the motor response $n_1 = 5.0$, $t_1 = 1.1$ s, $n_2 = 12.0$, $t_2 = 0.9$ s, and $a_2 = 0.4$ and for the auditory response, $n_1 = 6.0$, $t_1 = 0.9$ s, $n_2 = 12.0$, $t_2 = 0.9$ s, and $a_2 = 0.35$. The full width half maximum (FWHM) of the positive peak for the sensorimotor impulse response is about 5.0 s, while that for the auditory system is about 3.5 s. The greater FWHM for the sensorimotor response may be explained in part by physiological response time and imprecision of subjects performing the finger exercise.

Measured and predicted time series results for the motor cortex of one of the five subjects are shown in Fig. 4. These results confirm that the nonlinearity conclusions observed in the average responses remain valid in individual cases, and this was true for all five subjects. The conclusions are similar for the auditory characteristics, which are accordingly not shown. The motor and auditory activation maps for the same subject at all stimulus durations are shown in Fig. 5. The sites of activation did not vary substantially as the task duration within the trial was varied. Note that the sensorimotor activation is greatest for the long task, while the auditory response is largest at intermediate T.

Experiment IIA: Temporal Separability

Experimental results for two subjects performing the three-episode task (Fig. 2A) are given in Fig. 6. The measured time series are shown in the first and third rows, and the corresponding deconvolved time series are given in the second and fourth rows for each of the time separations, T_s. The impulse responses were measured on the same subjects in separate single 1-s episode scans using 10 trials of 30 s each and were similar to those in Fig. 3E. The deconvolution is successful when the interepisode spacing is ≥4 s, allowing much better separation of the individual

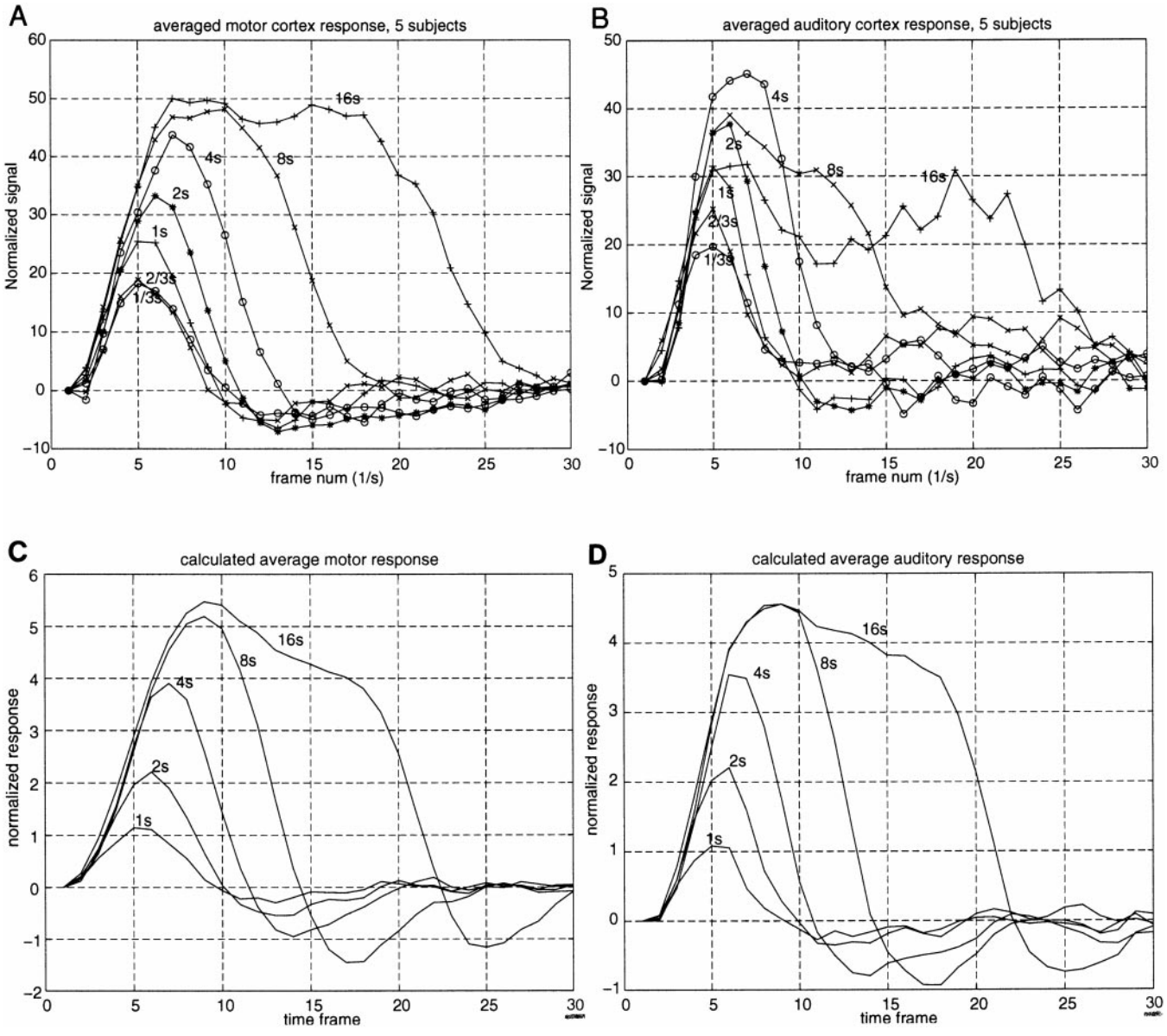

FIG. 3. Measured average activation response for five subjects performing stimulus duration experiments in (A) motor cortex and (B) auditory cortex. Scaling of ordinate uses 50 units = 2% of mean image intensity in brain. (C and D) Corresponding curves predicted by using measured 1-s data as impulse response function and convolving with rectangular function of specified width. Data have been normalized so that all curves start at 0. Averaged motor (E) and auditory (F) response (mean ± std dev over the 5 subjects) for the 1-s stimulus condition, overlaid on analytic functions fit to the data (solid lines, see text for parameters).

events than is obtained in the raw data. Ideally the deconvolved results would be approximations to $s(t) \sim \delta(t) + \delta(t - T_s - 1\text{ s}) + \delta(t - 2T_s - 2\text{ s})$, but the peaks are blurred by noise and departure from linearity, i.e., concordance between the assumed impulse response measured in the single episode scan and the responses in the three-episode scans. With greater signal-to-noise ratio the process could be expected to perform better with smaller T_s, but in practice it is not likely that separation with T_s substantially less than the FWHM of the impulse response could be recovered.

These results suggest that despite the lack of linearity shown by the duration experiments (Figs. 3 and 4), deconvolution can recover concatenated events even when the impulse response is strongly overlapping, as long as the events are separated by at least 4 s. This conclusion motivated the next series of experiments.

Experiment IIB: Variable Activation Levels

The motor results for two subjects performing the tasks shown in Figs. 2B and 2C are given in Figs. 7 and 8. The auditory responses were not analyzed because the alerting tone distorted the task.

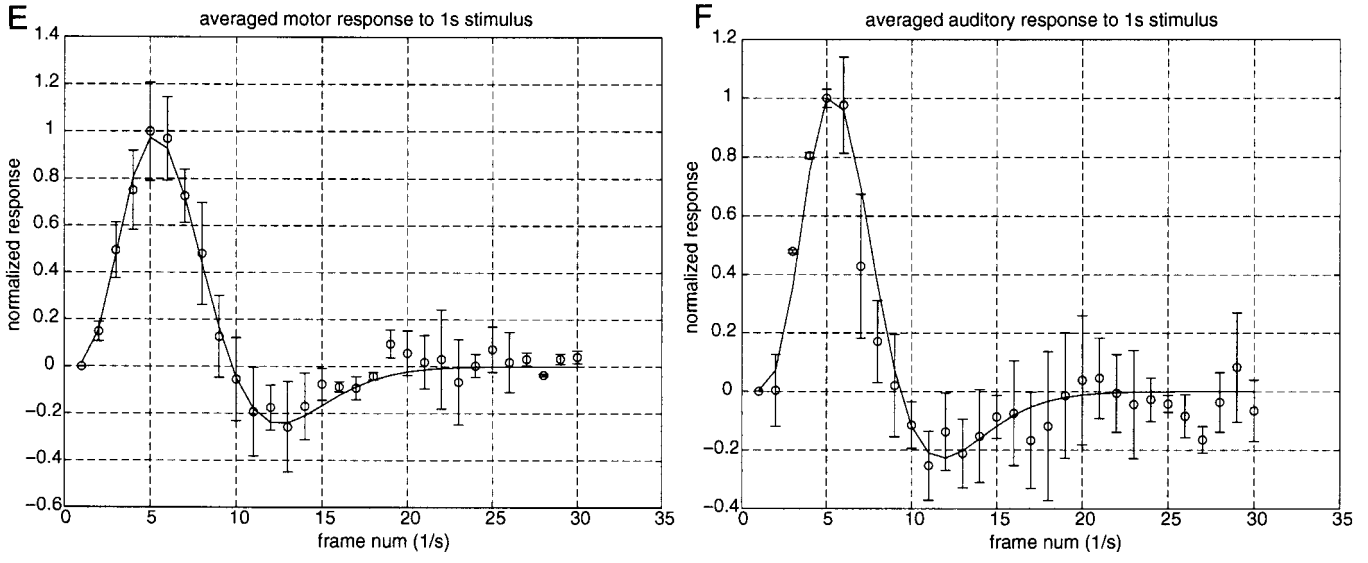

FIG. 3—*Continued.*

For the first subject, the impulse response nominally increases in amplitude as the finger-tapping frequency increases (Fig. 7A), in agreement with others (Rao *et al.*, 1996; Cohen, 1997). There is substantial undershoot after the main positive lobe, and this characteristic is reflected in the measured data for all the other tasks (Figs. 7B–7E). Despite this, the deconvolution process is able to restore reasonable approximations to the stimulus, $s'(t)$, as shown in the second row. The 3-Hz response in Fig. 7A was used for $h(t)$. The distortion and its reduction by deconvolution is particularly apparent in the two-episode task, Fig. 7B. If used without correction, the measured $m(t)$ data would indicate substantial reduction in activity for the second episode and negative postepisode activity that could be falsely interpreted as neuronal in nature. The amplitudes of the peak response in $s'(t)$ for each stimulus frequency in scans 1–4, 7, and 8 are shown in Fig. 8A.

For the second subject, the shape of the impulse response does not vary greatly between 1 and 4 Hz (Fig. 7F), but differences are nevertheless apparent in the variable frequency tasks (Figs. 7G and 7H). Again, significant distortion and its reduction by deconvolution using the 3-Hz data in Fig. 7F for $h(t)$ is seen in Fig. 7H. The amplitudes of the peak response in $s'(t)$ for each stimulus frequency are shown in Fig. 8B. Better

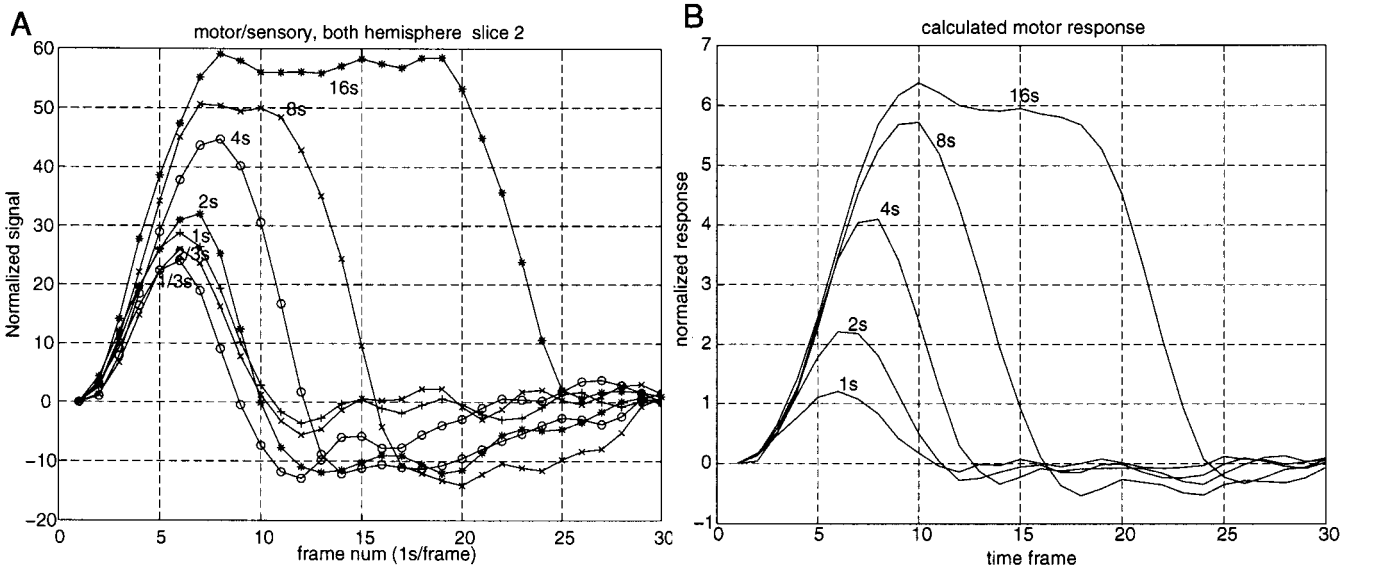

FIG. 4. (A) Measured motor activation response for a single subject. (B) Corresponding curve calculated by convolving measured 1-s data for that subject in (A) as impulse response function with rectangular function of specified width. Data have been normalized so that all curves start at 0.

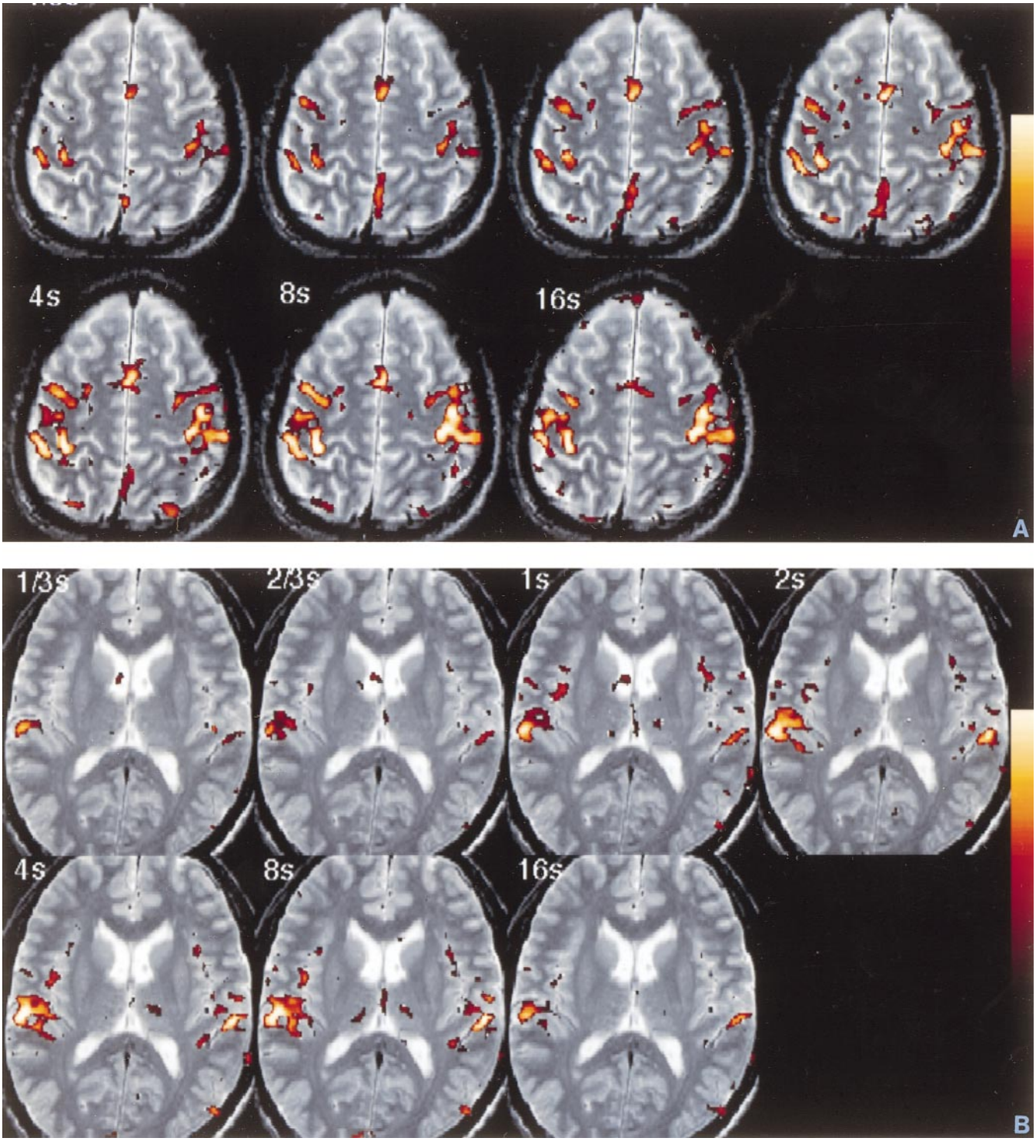

FIG. 5. Activation response for subject in Fig. 4 for (A) motor and (B) auditory cortex. Color scale shows correlation coefficients between 0.55 and 0.95.

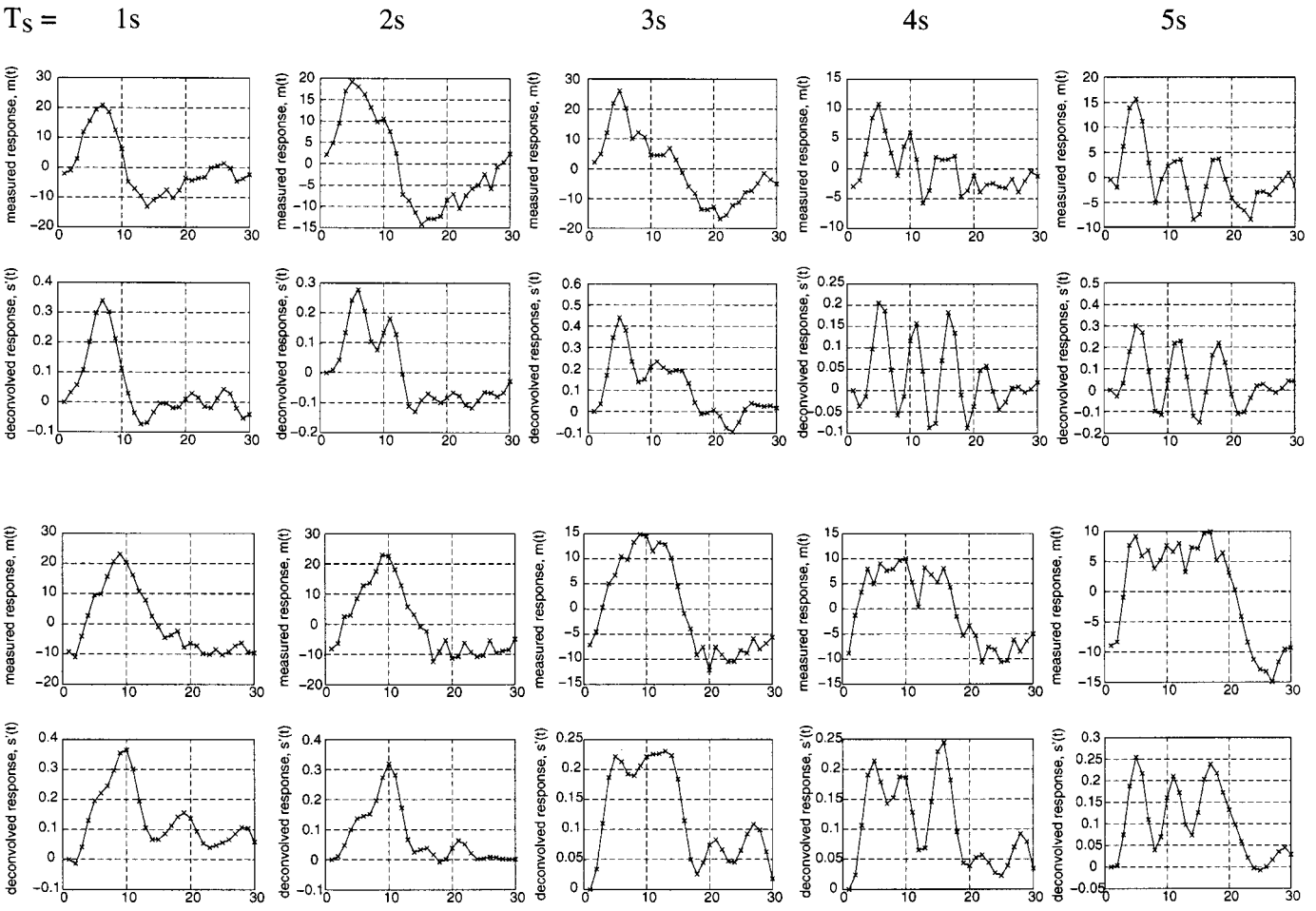

FIG. 6. Top row: Measured motor activation response for one subject performing temporal separation experiment IIA (Fig. 2A), for 1 s $\leq T_s \leq$ 5 s, ITI = 30 s. Second row: Corresponding deconvolved response curves. The plots have been shifted five frames (5 s) in order to allow easier comparison with the raw measurements. The three individual 1-s episodes are not resolved unless the spacing between episodes (T_s) is 4 s or more. Third row: Measured motor response for second subject. Fourth row: Corresponding deconvolved response curves. Note that for this subject, the response for short T_s appears to be broader and the individual episodes are less well resolved for T_s = 4 s.

agreement is observed between the single- and the multiple-episode tasks than for subject AS.

Deconvolution Parameters

The performance of the Wiener deconvolution was explored using $m(t)$ data from the multiple-episode experiments and $h(t)$ from the same subject and experimental series (Figs. 9A–9D) in comparison with $h(t)$ taken from the average response obtained in experiment I (Figs. 9E–9H). The noise parameter N_0 was varied between 10 and 40 for both sets of calculations. Using the subject's $h(t)$, the deconvolution is well behaved when N_0 is chosen appropriate to the noise level in the spectrum $|H(\omega)|$ (Fig. 9C). A value of N_0 too small (=10) yields substantial ringing, while too large a value (=40) damps the response, leading to incomplete deconvolution of the impulse response from $m(t)$. A value of 25 was used in Figs. 7G and 7H. However, the average impulse response, while less noisy (Fig. 9G), did not provide as good a deconvolved result for any value of N_0, as shown in Fig. 9H. From these results and others not shown, it is apparent that a canonical $h(t)$ does not suffice, and the impulse response must be chosen for the individual subject in agreement with Aguirre *et al.* (1998).

DISCUSSION

In general, the sensorimotor response to a short burst of finger-tapping activity is relatively strong and has a prolonged undershoot lasting up to 30 or more seconds, as shown both in averaged and in single-subject data. The character of the response to short stimuli in both auditory and motor cortex is profoundly nonlinear. With reference to the single-subject data in Fig. 4A, even very brief episodes of finger tapping

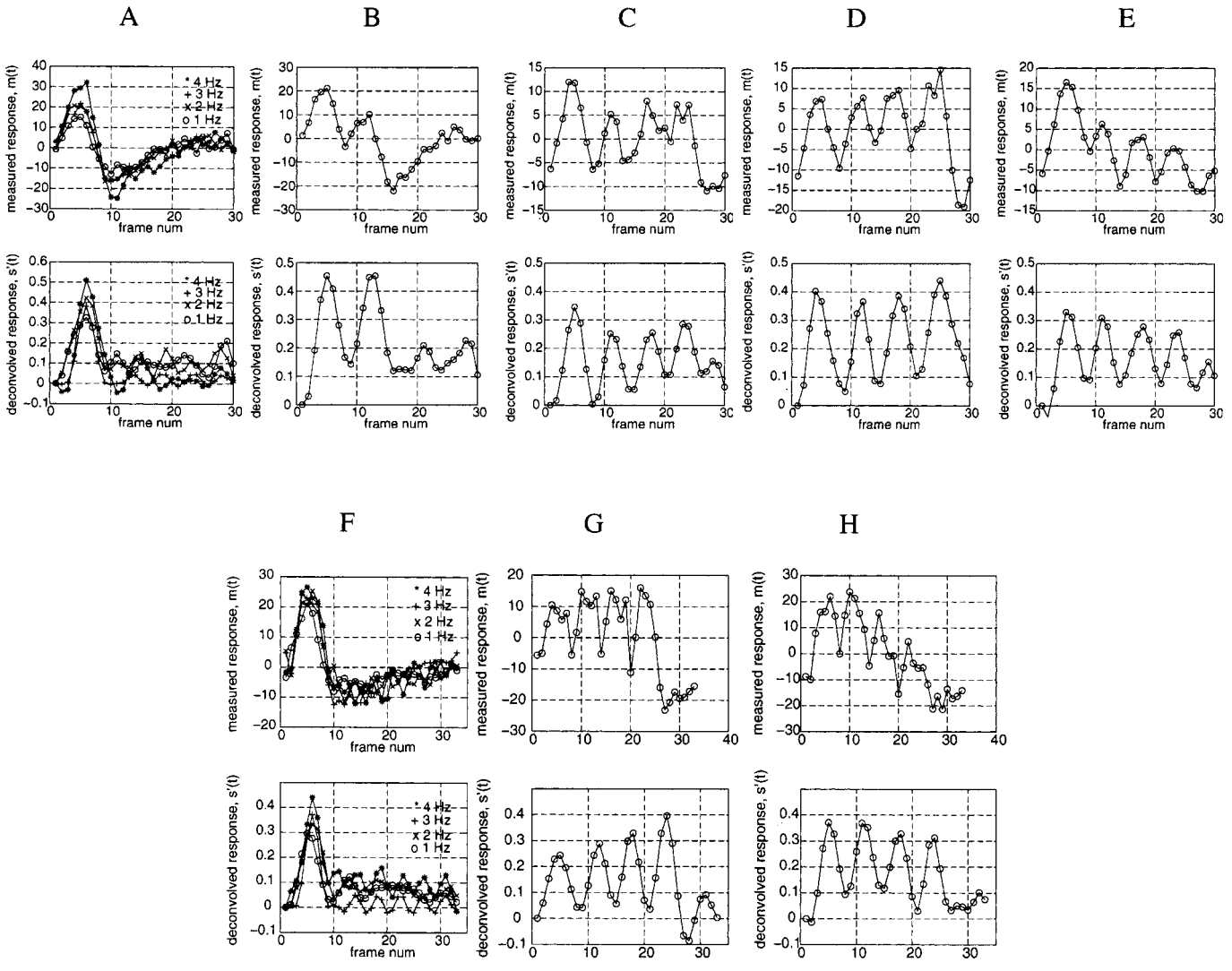

FIG. 7. Two subjects (top and bottom rows) performing activation-level experiment IIB (Figs. 2B and 2C). The 3-Hz single-episode data were used for $h(t)$. Top row: Subject AS: Measured and deconvolved motor responses for first subject with ITI = 30 s and (A) 1- to 4-Hz 2-s task, (B) two-episode 3-Hz task, (C) four-episode 3-Hz task, (D) 1-Hz/2-Hz/3-Hz/4-Hz task, (E) 4-Hz/3-Hz/2-Hz/1-Hz task. Bottom row: Subject JF: Measured and deconvolved motor responses for second subject with ITI = 35 s and (F) 1- to 4-Hz 2-s task, (G) 1-Hz/2-Hz/3-Hz/4-Hz task, (H) 4-Hz/3-Hz/2-Hz/1-Hz task.

trigger a nearly constant amplitude response for durations up to 1 or 2 s. In contrast, if the system was linear the amplitude would be directly proportional to the duration for episodes with duration ≪FWHM of the impulse response, as shown in Fig. 4B. As mentioned previously, it is possible that individual subjects' performance of the subsecond tasks may be compromised by physiological ability. However, even the multisubject average data show a lack of linearity for short duration stimuli (Fig. 3A).

A possible explanation for the relative insensitivity of the BOLD response to stimulus duration can be postulated to lie with the mechanism that governs the dilation of the arterioles feeding the affected capillary bed. We hypothesize that when a threshold increase in CO_2 or NO occurs in the venous tract, a biochemical message is sent to the arteriolar system for increased flow, but there is a minimum amount that can be achieved that is nearly independent of the change in oxy- and deoxyhemoglobin levels. It is as if the dilation "switch" can only open for a minimum time to provide a minimum bolus. With larger changes in these levels, the switch remains open longer and the increased flow continues until the task terminates and the oxygen demand returns to baseline values.

Using Buxton's balloon model, the calculated response amplitude for short-duration stimuli (modeled by proportionately short-duration triangular waveform inflow to the venous balloon) is found to be linear with duration, in disagreement with the measured data

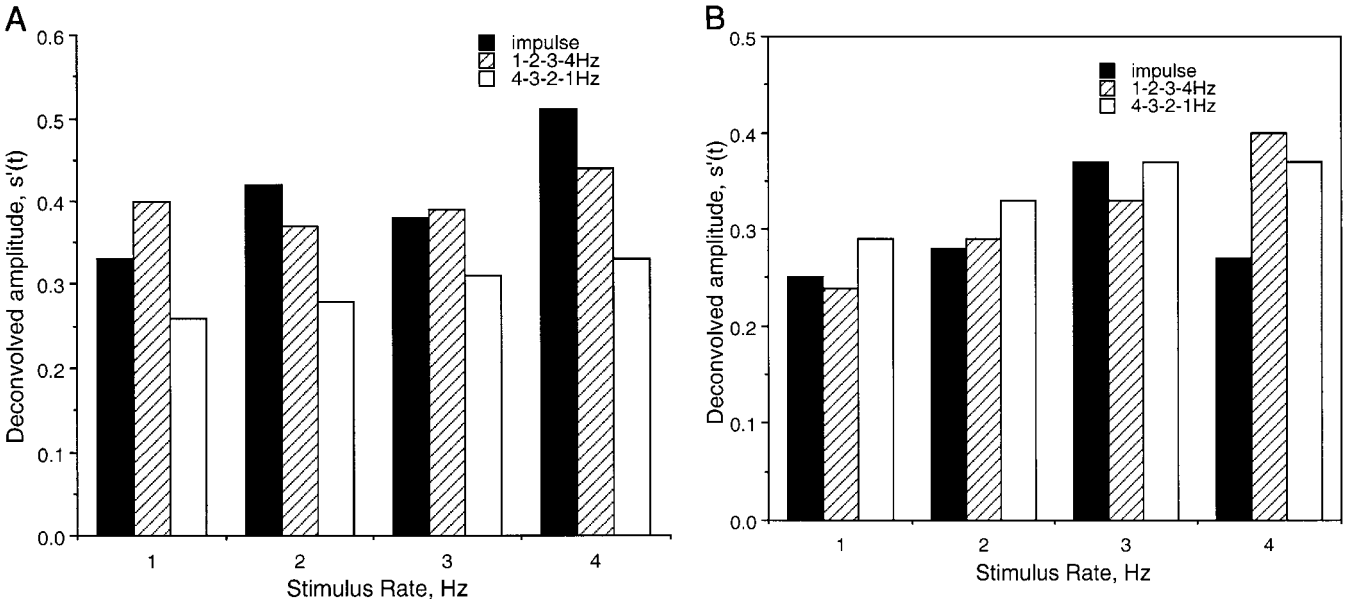

FIG. 8. Peak amplitudes of deconvolved motor response for experiment IIB for tasks in Fig. 7. Shown are results for single-episode tasks (dark) and ascending and descending frequency four-episode tasks. (A) Subject AS, (B) subject JF.

(Buxton *et al.*, 1998). Therefore, and in accordance with our hypothesis, the balloon model was modified so that the flow into the balloon has a trapezoidal form with ramp up/down times independent of the stimulus duration t_s and set to 3 s and a flat top duration t_f that is a Fermi function of t_s,

$$t_f = t_s e^{(t_s - t_0)/w}/(1 + e^{(t_S - t_0)/w}), \quad (7)$$

where t_0 and w are parameters set to 1 and 3 s, respectively. The f_{out} function was chosen as

$$f_{out} = 1 + \lambda_1(v - 1) + \lambda_2(v - 1)^{\alpha}, \quad (8)$$

where $\lambda_1 = 0.2$, $\lambda_2 = 4.0$, $\alpha = 0.47$. In addition, the f_{in} maximum value was set to 1.9 normalized units, and all other parameters were the same as in Buxton's work. The modified model calculations are shown in Fig. 10 together with the sensorimotor data from Fig. 3A. The agreement is satisfactory, lending credence to the hypothesized minimum bolus mechanism.

The auditory response shows a somewhat more linear response for short-duration stimuli, but is highly nonlinear for longer stimuli (Fig. 3B). This is due in part to the hemodynamics, but also because the auditory neuronal response tends to be enhanced by transitions in acoustic input, i.e., onset and termination of sounds, and depressed by constant sounds (Kandel *et al.*, 1991).

A consequence of the nonlinearity of the observed BOLD response is that linear methods fail to accurately predict the response to long stimuli using the measured impulse response for a short-duration stimulus. This means that simulations based on correlation (as in Figs. 3 and 4) must be used with caution when used to optimize task paradigm timing (Bandettini *et al.*, 1998).

Even though the linear convolution model does not accurately predict the response to a long task from the impulse response, individual episodes of motor activity can be separated using Wiener deconvolution provided a temporal spacing ≥4 s (i.e., a minimum somewhat less than the FWHM) is maintained. Thus, as long as individual occurrences of a task can be measured in order to provide the impulse response, the amplitude of the response to concatenated tasks of the same type can be approximately recovered.

The variable-rate experiments show that quantitation of task responses may be obtained by deconvolution and that without deconvolution the measured response can be substantially distorted. The distortion takes the form both of reduced separation of discrete tasks and of amplitude distortion from the overlapping response components. However, no attempt was made to draw global inferences about response versus finger-tapping rate as only two subjects were examined in this series. Indeed, some of the disparities noted in Fig. 8 may be due to lack of reproducibility in performing the multiple-episode tasks compared to the single-burst task. Nevertheless, the activation levels tend to be correlated with finger-tapping rate, as observed by Cohen and Rao (Rao *et al.*, 1996).

In general, it is well known that deconvolution is a noisy process. In the limiting case of an inversal filter where $d(t) = h^{-1}(t)$, zeroes in the spectrum of $H(\omega)$ can cause the deconvolved response to exhibit poles, i.e., to ring at the frequency of such zeroes. Thus the noise

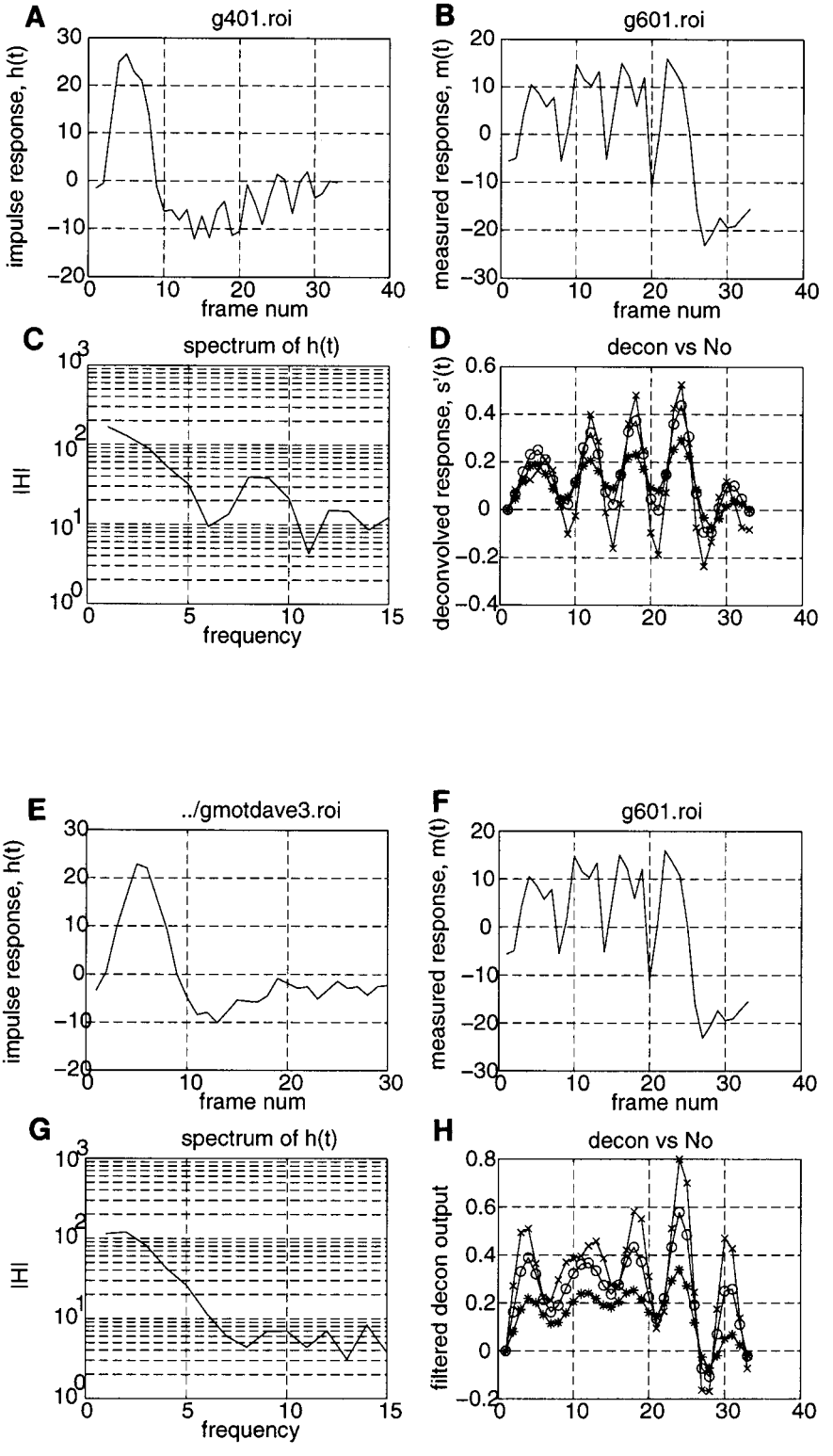

FIG. 9. Effect of noise parameter on deconvolution, using $m(t)$ data from Fig. 7G. (A) $h(t)$ from Fig. 7F, 3 Hz; (B) $m(t)$; (C) spectrum of $h(t)$; (D) deconvolution results for $N_0 = 10$ (x), 20 (o), 40 (*). (E) $h(t)$ from averaged response, Fig. 3A; (F–H) as in (B–D).

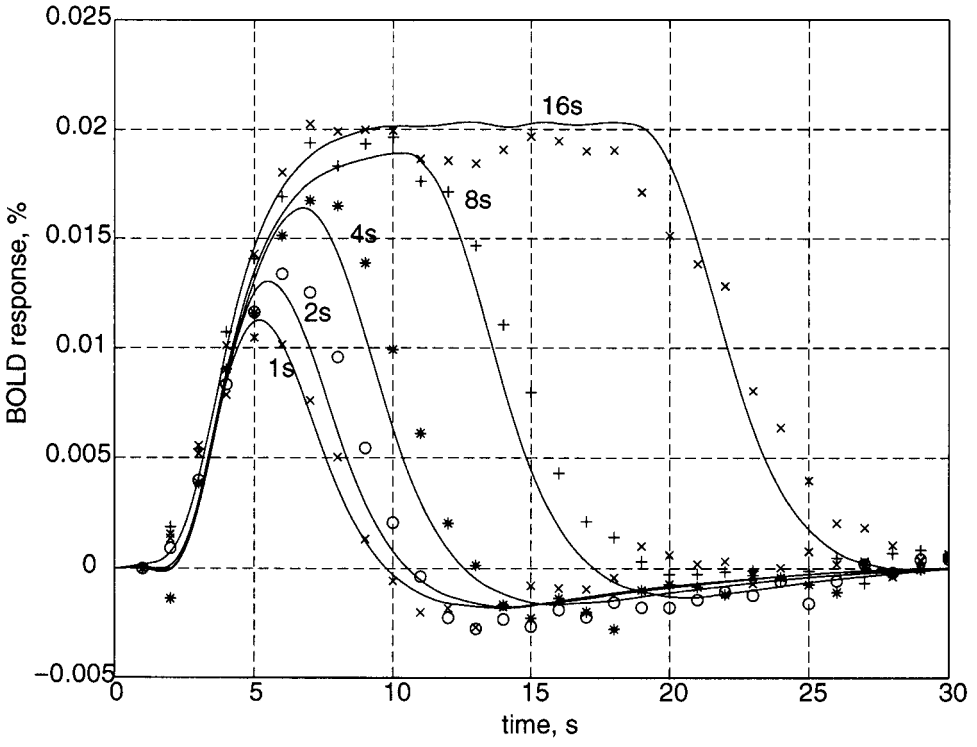

FIG. 10. Comparison of calculated response using balloon model (solid lines) and measured averaged 1-s response data from Fig. 3A. Calculations use modified Buxton's model.

character of the deconvolved response can be markedly altered. The degree of noise enhancement by deconvolution depends on the SNR of the measured data and the impulse response characteristics.

Thus the deconvolution parameters must be chosen carefully. In this study $N(\omega)$ was assumed to be white noise, and its amplitude N_0 was obtained from $H(\omega)$. If too low a value of N_0 is used, substantial ringing occurs in $s'(t)$, while too high a value results in poor recovery of the activation time series (Fig. 9A). The success of deconvolution methods depends on having high SNR in the data used to obtain $h(t)$ and $m(t)$. Even if the random noise is low, however, the impulse response must be matched to the measurement conditions or distortion results, as shown in Fig. 9B, in which the average $h(t)$ did not provide a satisfactory deconvolution filter. Thus it is important to measure $h(t)$ on the subject to avoid systematic "noise," preferably during the same scanning session.

In summary, the BOLD fMRI response to short bursts of sensorimotor and auditory activity is delayed (about 5.5 s) and blurred by the finite response time of the hemodynamic system. The relatively broad impulse response function (FWHM = 5.0 s for sensorimotor) reduces the temporal resolution of dynamic fMRI experiments, while the long negative undershoot can create strong distortions in measured time series (e.g., Fig. 6). These characteristics are reasonably well predicted by a modified version of Buxton's balloon model. Deconvolution using a Wiener filter has been found to be effective in reducing the blurring and distortion of the time series measured during event-related activation experiments. The degree to which deblurring can be achieved, and thus the ultimate temporal resolution, depends on the signal-to-noise ratio and the validity of the impulse response function. The intersubject variability in impulse response necessitates measuring it for every subject in the cortical region of interest. Despite the nonlinearity of sensorimotor and auditory cortex observed in this study and others, significant enhancement of the apparent temporal resolution can be obtained using deconvolution. The fact that robust signal is obtained even with brief episodes of activation therefore suggests that this method may have a role in quantitative analysis of event-related fMRI data.

ACKNOWLEDGMENT

The author thanks A. M. Sawyer-Glover for assistance with scanning.

REFERENCES

Aguirre, G., Zarahn, E., and D'Esposito, M. 1998. The variability of human, BOLD hemodynamic responses. *NeuroImage*, **8:** 360–369.

Bandettini, P. A., and Cox, R. W. 1998. Contrast in single trial fMRI:

Interstimulus interval dependency and comparison with blocked strategies. Proceedings, ISMRM Sixth Annual Meeting, Sydney, p. 161.

Binder, J. R., Rao, S. M., Hammeke, T. A., Frost, J. A., Bandettini, P. A., and Hyde, J. S. 1994. Effects of stimulus rate on signal response during functional magnetic resonance imaging of auditory cortex. *Brain Res. Cognit. Brain Res.* **2:**31–38.

Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* **16:** 4207–4221.

Buckner, R. L., Bandettini, P. A., O'Craven, K. M., Savoy, R. L., Petersen, S. E., Raichle, M. E., and Rosen, B. R. 1996. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging [see comments]. *Proc. Natl. Acad. Sci. USA* **93:** 14878–14883.

Buxton, R., and Frank, L. 1997. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *J. Cereb. Blood Flow Metab.* **17:**64–72.

Buxton, R. B., Wong, E. C., and Frank, L. R. 1998. Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. *Magn. Reson. Med.* **39:** 855–864.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., and Smith, E. E. 1997. Temporal dynamics of brain activation during a working memory task [see comments]. *Nature* **386:** 604–608.

Cohen, M. S. 1997. Parametric analysis of fMRI data using linear systems methods. *NeuroImage* **6:** 93–103.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., and Turner, R. 1998. Event-related fMRI: Characterizing differential responses. *NeuroImage* **7:** 30–40.

Friston, K. J., Jezzard, P., and Turner, R. 1994. Analysis of functional MRI time-series. *Hum. Brain Mapp.* **1:** 153–171.

Friston, K. J., Josephs, O., Rees, G., and Turner, R. 1998. Nonlinear event-related responses in fMRI. *Magn. Reson. Med.* **39:** 41–52.

Glover, G. H., and Lai, S. 1998. Self-navigated spiral fMRI: Interleaved versus single-shot. *Magn. Reson. Med.* **39:** 361–368.

Hayes, C., and Mathias, C. 1996. Improved brain coil for fMRI and high resolution imaging. ISMRM 4th Annual Meeting Proceedings, New York, p. 1414.

Hu, X., Le, T. H., and Ugurbil, K. 1997. Evaluation of the early response in fMRI in individual subjects using short stimulus duration. *Magn. Reson. Med.* **37:**877–884.

Hykin, J., Clare, S., Bowtell, R., Humberstone, M., Coxon, R., Worthington, B., Blumhardt, L., and Morris, P. 1996. A non directed method in the analysis of functional magnetic resonance imaging (fMRI) and its application to studies of short term memory. 13th Annual Meeting of the European Society for Magnetic Resonance in Medicine, Prague, p. 180, MAG*MA Supplement to Vol. IV, No. II.

Kandel, E. R., Schwartz, J. H., and Jessell, T. M. 1991. *Principles of Neural Science*, Vol. 3. Elsevier, New York.

Krüger, G., Kleinschmidt, A., and Frahm, J. 1996. Dynamic MRI sensitized to cerebral blood oxygenation and flow during sustained activation of human visual cortex. *Magn. Reson. Med.* **35:**797–800.

Lee, A. T., and Glover, G. H. 1995. Discrimination of large venous vessels in time-course spiral blood oxygen dependent magnetic resonance functional neuroimaging. *Magn. Reson. Med.* **33:**745–754.

Papoulis, A. 1977. *Signal Analysis.* McGraw–Hill, New York.

Rao, S. M., Bandettini, P. A., Binder, J. R., Bobholz, J. A., Hammeke, T. A., Stein, E. A., and Hyde, J. S. 1996. Relationship between finger movement rate and functional magnetic resonance signal change in human primary motor cortex. *J. Cereb. Blood Flow Metab.* **16:**1250–1254.

Robson, M. D., Dorosz, J. L., and Gore, J. C. 1998. Measurements of the temporal fMRI response of the human auditory cortex to trains of tones. *NeuroImage* **7:** 185–198.

Rosen, B. R., Buckner, R. L., and Dale, A. M. 1998. Event-related functional MRI: Past, present, and future. *Proc. Natl. Acad. Sci. USA* **95:**773–780.

Vazquez, A. L., and Noll, D. C. 1998. Nonlinear aspects of the BOLD response in functional MRI. *NeuroImage* **7:** 108–118.