Novemeber 22, 1999

The Role of Context in Visual Perception: An fMRI Study

Thomas E. Nichols, William F. Eddy, and James L. McClelland

We use Functional Magnetic Resonance Imaging (fMRI) to test two competing theories of visual perception. Expressed as connectionist models, the two theories make different predictions about the role context plays in visual perception. McClelland and Rumelhard's interactive activation model (1981) asserts that context effects the actual formation of the percept, while the "bottom up" model asserts that context is integrated with visual stimuli at higher cognitive levels. Using pairs of ambiguous stimuli (quickly flashed words) preceded by context (sentences), we image the primary visual cortex looking for context-dependent activity. We use a combination of classical and Bayesian statistical tools to model the signal. Four normal controls were studied, two subjects yielded usable data; in just one subject there was context-dependent activity, but not in the manner hypothesized. This result and the small amount of usable data precludes making any solid conclusions about the two theories of visual perception.

1. INTRODUCTION

Efforts to understand visual perception have produced two competing theories. Each of these theories can be expressed as a connectionist model (see Figure 1) and they are each distinguished by the way information, or activation, flows through the connectionist network. One model specifies that information only flows upward; for example, the activity related to identification of letters is only influenced by external stimuli. This is the "bottom up" model. The other model allows for information to move in both directions; for example, the activity related to identification of letters can be influenced by word-related activity as well as external stimuli. This is the "interactive activation" or bidirectional model. While there has been a growing amount of experimental evidence for the bidirectional model none of the experiments have provided results that cannot be explained under the bottom up model. The central goal of this work is to use fMRI to look for evidence for "top down" effects in visual perception, that is, evidence for the bidirectional model that is inconsistent with the bottom up model.

Thomas E. Nichols is Doctoral candidate and William F. Eddy is Professor, Department of Statistics & Center for the Neural Basis of Cognition and James L. McClelland is Professor, Department of Psychology & Center for the Neural Basis of Cognition, Carnegie Mellon Unviersity, Pittsburgh, PA 15213. The authors would like to thank Keith Thulborn for guidance on fMRI fundamentals, Jim Voyvodic for help with sound delivery and paradigm presentation, and Denise Davis for help with subject preparation and MR data acquisition. This research was supported by a pilot grant from the Functional Imaging Research Program, University of Pittsburgh Medical Center. This is the written component of TE Nichols's "Advanced Data Analysis" project, submitted as partial fulfillment of requirements for a Ph.D. at CMU Statistics.

[Figure 1 about here.]

These two models are distinguished phenomenologically in the perception of ambiguous stimuli. Consider the experiment illustrated in Figure 2, where a subject fixates on a cross-hair and after a pause a pair of words are flashed parafoveally for a very brief duration. If the words are far from the center (e.g. 10°) and flashed quickly (e.g. 10 ms), the subject will be unable to identify either of the words, yet, if the flash is preceded by a sentence that supplies a context for *one* of the words, the subject will be able to identify the word that matches the context but will be unable to identify the other word (McClelland and O'Regan 1981). This demonstrates that context can facilitate perception at a low level. The bidirectional model can easily account for this effect, as top down influences can effect word area activity, which can in turn affect letter area activity, etc; the bottom up model, where only external stimuli can affect perception, can only explain this effect by the integration of contextual information outside of the model. Hence we use fMRI to look for context-dependent changes in the visual cortex, the portion of the brain that performs the visual processing these models attempt to explain.

[Figure 2 about here.]

We focus on two questions that can be answered with fMRI: Is there measurable change in brain activity due to flashed words? Second, if there is such a change, is it modulated by context? fMRI produces images of the BOLD effect, a signal change shown to be related to neuronal activity. We could use fMRI to look for context-dependent changes in neuronal activity in any of the brain areas corresponding to the levels of the visual perception models, but we choose to examine the primary visual area, which would correspond to the input of the bottom layer of the two connectionist models. A finding of context-dependent changes here would be very strong evidence for the bidirectional model. A lack of finding would not disprove this model, as the context effects could be present in the upper layers of the model.

2. METHODS

The experiment had three components, two supporting components and the visual perception experiment. The No Context Calibration adjusted the stimuli presentation based on the sensitivity of each subject's visual system; V1 Localization was done to find what part of the subject's visual cortex was sensitive to the stimuli; and the Context Experiment is the focus of the study.

Stimuli

For all of three components a rear projection video display system projected a video image to a screen approximately 30 cm from the subject's face. At all times a black cross-hair on a white background was present in the center of the screen.

For the No Context Calibration component, pairs of words were presented for 13.9 ms at varying distances from the center to determine the smallest distance at which the subject would fail to reliably identify the two words; we call the visual angle from center to center of each word the *flash angle*. A series of 25 word pairs were flashed, one pair every two seconds; each word in a pair was matched for length and pairs were randomly blocked such that every 5 pairs had one pair of words of each word length: 3, 4, 5, 6, and 7 letters. The fixed-width Courier font was used, with an on-screen x-height of about .68°; 3 and 7 letter words were 2.9° and 6.7° wide, respectively. The subject was instructed to fixate on the cross-hair and for each flash to speak the words he or she could perceive. At the end of each block the flash angle was adjusted on the basis the subject accuracy: if a subject correctly identified 1 word on more than two trials or if they identified both words on any of the 5 trials, then the flash angle was increased; if no word was identified on any of the 5 trials then the flash angle was decreased. Hence, through the subject's own responses, the flash angle was adjusted so that with no context the subject could just barely identify the words. This flash angle value was modestly increased (10 pixels, about .68°) to decrease the chance of identifying a word with no context.

For V1 LOCALIZATION, the words were replaced with two 7 Hz flashing checkerboards, 2.0 by 6.8 degrees, centered on the locations of the word pairs, as set by the calibration just described. The checkerboards were alternated with cross-hair alone during a 6 minute acquisition: the pattern was 30 seconds of flashing checkerboard, 30 seconds of cross-hair, repeated 6 times.

The CONTEXT EXPERIMENT utilized the same type of visual stimuli as the NO CONTEXT CALIBRATION, except context sentences were presented aurally before each flash of words; each sentence was constructed such that the last word could be imputed from the rest of the sentence. 100 sentences provided context for 50 pairs of words; for each word pair one sentence provided context for the left word, another sentence provided context for the right word. Each word pair was presented twice and each of the words in the pair matched the context sentence for one the two trials. For a given trial, the word which matches the context is called the *target word*. From a list of 116 sentences we eliminated the 16 sentences that provided the weakest context, as measured by a survey of 10 subjects; the 16 worst sentences were used for practice. In the survey each sentence was missing its last word and subjects were asked to fill in the first word that came to their mind or the word that seem most likely to finish the sentence. The 100 stimuli were divided into two blocks of 50 such that each word pair was presented only once in each block; order of the stimuli was randomized within block.

For the first subject each trial lasted 15 seconds, which was determined to be too short (see results), so all remaining subjects had 20 second trials. Figure 3 shows the structure of one trial. Sentences were all shorter than 5 seconds and were delayed such that they ended exactly 5 seconds after the trial began. The word pair was flashed 1 second after the sentence presentation finished. Since we were interested in the brain's response to a single trial we wait 14 seconds until the hemodynamic response returned to equilibrium (i.e. the BOLD response returned to zero). This is known as Event Related fMRI; it is in contrast to using a continuous stimuli for long durations, as in the V1 LOCALIZATION paradigm.

[Figure 3 about here.]

Subject Preparation

Four normal subjects were recruited from the University of Pittsburgh and Carnegie Mellon University graduate student population. Subjects were screened for claustrophobia and any unsafe ferro-magnetic objects embedded in their body. Informed consent was obtained from each subject under procedures approved by the Institutional Review Board of the University of Pittsburgh Medical Center.

Outside of the scanner subjects were instructed how to perform the CONTEXT EXPERIMENT and practiced it for 8 to 16 trials. They were instructed to press the left button if they could identify the left word, the right button if they could identify the right word and both buttons if they could identify both. We defined "identify" as that they "could see the word and were able to determine the identity of the word." We added that they do not need to be absolutely certain but reasonably certain. The instructions were repeated again in the magnet just before the experiment. Once in the magnet, we adjusted the flash angle of the stimuli for each subject as described above (NO CONTEXT CALIBRATION).

MR Parameters

Subjects were scanned in a 3 Tesla GE scanner (General Electric, Milwaukee WI) equipped with Echo Planar Imaging (Advanced Nuclear Magnetic Resonance, MA). For each subject at least 5 image series were acquired: A 13 slice sagittal scout (TR 500 ms, TE 8 ms, FA 90°, 0.9375 mm pixels in 256×256 images, 5 skip 1 mm slices), a 7 slice oblique axial-coronal "scout" functional acquisition (TR 1500ms, TE 25 ms, FA 90°, 3.125 mm pixels in 128×64 images, 3 skip 1mm slices, 240 NEX per slice), two 4 plane oblique axial-coronal primary functional acquisitions (TR 1000ms, TE 25 ms, FA 90°, 3.125 mm pixels in 128×64 images, 3 skip 1 mm slices, 1000 NEX per slice), a 4 plane oblique axial-coronal "two shot catch and hold" high resolution anatomical (TR 10,000 ms, TE 114 ms, FA 90°, 0.78125 mm pixels in 512×128 images, 3mm skip 1mm slices), and a 124 slice coronal spoiled gradient (SPGR) anatomical (TR 25 ms, TE 4 ms, FA 40°, 0.9375 mm pixels 256×256 images, 1.5 mm contiguous slices). All of the oblique acquisitions were at the same angle and the 4 slice anatomicals had the same prescriptions as the 4 slice functionals.

Since we were interested in a modulation of a subtle effect, that is, a response to a 14 ms visual stimulus, we selected MR parameters to optimize signal to noise under the constraint of high temporal resolution. Hence we limited the primary acquisition to 4 slices with moderate spatial resolution. The anatomical area of interest was defined functionally as the cortex in V1 sensitive to the part of the visual field occupied by the two words. Since there is considerable variability in the functional anatomy of V1, we used a "scout" functional acquisition of 7 slices parallel to and covering the calcarine fissure. The V1 LOCALIZATION checkerboard stimulus was used for a 6 minute acquisition and the data was analyzed in "near real-time," just minutes after the acquisition was completed (CIGAL software, Voyvodic 1996). The four contiguous slices with the greatest activation along the calcarine fissure were identified and were the slices used for the subsequent 4 slice functional acquisitions used for the CONTEXT EXPERIMENT.

Data Analysis

Data analysis proceeded in two stages, image construction and statistical inference. Image construction refers not just to the process of converting the raw frequency-domain MRI data into images, but the removal of systematic variation from the data that would otherwise degrade signal-to-noise and bias the results. In the statistical inference stage the constructed image data is used to estimate the experimental effects and their significance.

Image construction was performed with FIASCO software (Functional Imaging Analysis Software, Computational Olio, Eddy et al. 1996). FIASCO estimates and removes variation due to analog-to-digital converter miscalibration, mean image intensity variation, echo planar ghost artifacts, subject movement, outliers and linear drift. Except for the outlier removal and linear detrending, each step is performed in the frequency-domain. The results of each adjustment produces diagnostics plots which were examined for each subject, for each functional acquisition. Any problems were identified and noted with respect to the interpretation of the final results. Of particular concern was subject movement: At best, movement is a large source of structured variability in the data, at worst, movement artifacts can be mistaken as signal.

Because of these concerns we rejected data with substantial movement: If an acquisition had an image with average voxel movement greater than 1/2 of a voxel, that acquisition was discarded. Voxel movement is the Euclidean distance a voxel moves relative to the reference image. For each 2D image, the voxel movement is calculated at every voxel and the average voxel movement is just the mean over all image voxels. The reference image was the $n/2$-th image, where n images per slice are acquired.

The motivation for this conservative approach is the inability to correct 3D motion: only in-plane shifts and rotation are correctable with FIASCO[1]. If a subject moves more than 1/2 a voxel it is unlikely they moved strictly in the plane of the acquisition, especially since the most common motion is a pitching motion and the coronal-axial images we acquired are not coplanar with this motion. Movement was only adjusted within acquisition, meaning that we must assume negligible head motion between the different experiments. This is a very strong assumption and failures of this assumption will impede cross-acquisition acquisition comparisons, a key component of the analysis discussed below.

Statistical inference was performed with both FIASCO and BRAIN (Bayesian Response Analysis and Inference for Neuroimaging, Genovese and Sweeney 1999); both packages were developed in the Carnegie Mellon University Statistics Department. Inference in FIASCO consists of two sample t tests of groups of images. BRAIN also uses the data from FIASCO image construction but models each fMRI time course as a sum of a drift function and an experimental response function; the drift is constrained to be smooth in time and the experimental response function is based on the current understanding of the BOLD signal. (See Figure 4). The experimental response is modeled with five parameters, one for response amplitude, four for shape; the four shape parameters describe the delay and duration of the rise and fall of the BOLD signal relative to the presentation of experimental stimuli; while none of these shape parameters are of specific research interest here, they allow considerable flexibility and result in a well-fitting model. BRAIN produces maps of standardized response

[1] FIASCO 3D motion correction is currently under development.

amplitude, the amplitude of the experimental response divided by its standard deviation, analogous to t statistics. As BRAIN uses Bayesian statistics, prior distributions are specified for each parameter. Prior probability of an experimental response was set to 0.001, representing a belief of one voxel being active per 1000, roughly corresponding to one voxel per plane. Otherwise all default priors were used, except for the CONTEXT EXPERIMENT, where the shape parameters were modified to reflect the very brief stimuli.

[Figure 4 about here.]

The V1 LOCALIZATION data was analyzed with BRAIN using the standardized response amplitude images. For a voxel to be considered part of the tissue sensitive to the parafoveal stimulation we required both a large statistic value *and* that several heuristic criterion be met. A "large" standardized response was considered to be greater than a 0.05 α threshold Bonferroni corrected for searching over the whole volume of brain imaged ($\approx$ 3500 voxels). The heuristic criterion guaranteed that the identified voxels conformed to the well-known facts about the organization of the visual cortex. First, we only considered activated voxels in contiguous groups of two or more voxels in the mid-anterior calcarine sulcus, the area of the V1 corresponding to the lateral visual field where the checkerboards are presented. Further, only one pair of bilateral loci were considered; since V1 is organized retinotopically, the contiguous checkerboard should map to a contiguous region of tissue, so only one group of voxels in V1 in each hemisphere should correspond to the checkerboard stimuli; also each hemisphere's locus should exhibit symmetry with the opposing hemisphere, corresponding to the symmetry of the retnotopic map, though perfect symmetry is not expected.

The final result of the V1 LOCALIZATION was a voxel mask identifying regions in V1 stimulated by the bilateral checkerboards.

The CONTEXT EXPERIMENT data was analyzed with standard t tests using FIASCO[2]. Three conditions were defined as follows: The first 8 scans were labeled as LISTEN, the next 7 scans as FLASH and the last 5 scans as REST (See Figure 5). These labelings were based on a literature review of the BOLD response to a very short stimuli (Savoy et al. 1995; Vazquez and Noll 1998; Buckner et al. 1996; Dale and Buckner 1997; Rosen et al. 1998): Investigators have found that the response is delayed by about 2 seconds and does not return to baseline until as much as 7 seconds after stimulation. To test for a main effect of the word flash, maps of pairwise t statistics of FLASH versus LISTEN were created, one set for each block of 50 trials. The effect of context was assessed by replacing the FLASH label with FLASH-CONTEXT-L and FLASH-CONTEXT-R labels, such that trials with the target word on the left and right were labeled differently. Then t maps comparing FLASH-CONTEXT-L versus FLASH-CONTEXT-R were created. Only voxels identified by the results of the V1 LOCALIZATION were studied, with the following consideration to account for between acquisition subject movement.

If there was head movement between acquisitions then the direction of the movement was estimated with informal methods and voxels adjacent to the region of interest in that direc-

[2] BRAIN was not used because the current version did not model the overshoot, the tendency of the BOLD response go negative from baseline after cessation of the stimulus. The CONTEXT EXPERIMENT data showed strong signs of overshoot and the BRAIN model fit poorly relative to FIASCO, where the overshoot could be accounted for as REST. This shortcoming is currently being addressed.

tion were considered as well. The informal method consisted of looking for signs of motion on the difference of acquisition mean images; edge artifacts or structured inhomogeneities in the difference images are signs of motion and can be used to infer the direction of the movement. Since the signal immediately outside the brain is essentially zero, the magnitude of the edge artifact can be used to estimate the magnitude of the motion; for example, if the edge artifact has a magnitude that is %50 of the average brain intensity, then the brain moved approximate 1/2 of a pixel.

[Figure 5 about here.]

With the results of V1 LOCALIZATION and the t maps of the main and context-dependent effects of the word flash, we were able to address the questions of interest.

Is there measurable change in brain activity due to flashed words? Focusing solely on the voxels identified in the V1 LOCALIZATION, we looked for voxels with large t values in FLASH versus LISTEN t maps from *both* blocks. Large t were considered values greater than an 0.05 threshold Bonferroni corrected for the number of voxels searched.

If there is a response to flashed words, is the response modulated by context? Again focusing only on the voxels identified in the V1 LOCALIZATION, we identified voxels with large t values in the FLASH-CONTEXT-L versus FLASH-CONTEXT-R maps from both blocks, using the same significance threshold as before. We hypothesize that activity in left V1 should be greater when the target word is on the right, since the left visual cortex corresponding to the right visual field; *vice versa* for activity in right V1.

3. RESULTS

All axial and coronal images are shown in the radiological convention, were the left side of the brain is presented on the right side of the image (as if one was standing at the feet of a subject).

[Figure 6 about here.]

Subject 1

The data for the first subject are different from the other subjects in three important ways. First, the length of each trial was only 15 seconds, not 20 seconds. After examining this subject's results it was clear that the duration after the word flash, 9 seconds, was not sufficiently long enough to allow the BOLD response to return to baseline equilibrium; hence each trial was extended by 5 seconds for the subsequent subjects. Second, the prescription was mistakenly made perpendicular to calcarine fissure instead of parallel. Third, acquisition problems prevented collection of the first block of the CONTEXT EXPERIMENT, hence only 50 of the 100 trials were imaged.

Image Construction

The first subject had limited head movement, with average movement no worse than 1.0 mm (0.33 voxels). For the Context Experiment there were hints of sinusoidal Y-shift movements with 15 second periodicity, corresponding to the experimental paradigm. Since the amplitude of these oscillations was about 80 microns (0.025 voxels) these movements were less a cause for concern than a demonstration of the high precision of the movement estimation. No other problems were detected in the image construction.

V1 Localization

The optimum flash angle for this subject was determined to be about 14°. The V1 Localization produced very clean activation profiles, with just two contiguous asymmetrical bilateral regions of activation (Figure 7). From the quick analysis at the time of scanning, we selected the bottom four slices, but the completed analysis shows signs of activation on two of the rejected slices.

[Figure 7 about here.]

A problem with these results is the absence of any activation along the midline. This is a result of orienting the slices perpendicular instead of parallel to the calcarine fissure: The 4 slice, 16mm slab of the calcarine fissure that we imaged did not catch the retinotopic mapping of the flashing checkerboard. The prescription was probably posterior to the V1 activation site.

Context Experiment

Since we had no voxels in V1 this subject offers no information on our hypothesis, but it should be noted that there was a response to flash in some of the lateral areas in the V1 Localization results. These areas most likely correspond to MT (V5), the area of the visual cortex sensitive to motion.

Subject 2

The second and all remaining subjects had the standard protocol described in the methods.

Image Construction

This subject exhibited very little head movement, with average movement no greater than 0.75 mm (0.24 voxels). There were no signs of stimulus-related movement in any of the acquisitions for this subject. In some slices there were signs of periodic patterns in the mean image intensity, with periodicity matching the epoch length (60 seconds for V1 Localization, 20 seconds for Context Experiment). Since the effect was subtle and not present in all subjects we took no remedial action.

V1 Localization

The optimum flash angle for this subject was determined to be about 20°. The V1 LOCALIZATION produced several loci of activation, many in bilateral pairs (See Figure 8). We focused on the most medial bilateral activations, those that were centered about and 12 mm from the midline, 19 mm anterior from the posterior edge of the brain and covering two slices (Slices 3 and 4 on Figure 8). From the results of the near real-time analysis, we discarded the top two and very bottom slices.

[Figure 8 about here.]

The four slices selected for further study were thresholded at 4.29, corresponding to a Bonferroni 0.05 one sided t_{211} test for 3661 voxels. With this threshold the right activation consisted of 8 voxels, the left 11 voxels.

Context Experiment

Difference images comparing the mean images of the V1 LOCALIZATION and the two CONTEXT EXPERIMENT acquisitions showed signs of between acquisition movement, while there did not appear to be significant movement between the two blocks of the CONTEXT EXPERIMENT (See Figure 9). The edge artifact has magnitude 500 to 1000, 1/4 to 1/2 of the typical brain intensity, hence the subject moved at most a 1/2 a voxel; the direction of motion is anterior relative to the V1 LOCALIZATION acquisition. Hence we focused not only on the 19 voxels identified above but also considered 8 adjacent voxels in the anterior direction. The 0.05 Bonferroni threshold for searching over 27 t_{997} statistics is 2.91. Based on this criterion we found a response to flashed words in the region of interest: 7 and 8 voxels were significant in the right region, for block 1 and 2 respectively, and 11 and 12 voxels were significant in the left region.

[Figure 9 about here.]

Using the same threshold, the FLASH-CONTEXT-R versus FLASH-CONTEXT-L t map had small number of voxels identifed as significant in the left and right regions. 1 voxel was significant in the right region (both blocks) and 3 and 4 voxels were significant in the left region, for each block. All significant voxels were larger for FLASH-CONTEXT-L than for FLASH-CONTEXT-R. While this is evidence of context related effects, it is inconsistent with our hypothesis, which is that right V1 voxels should show increased activity under the FLASH-CONTEXT-L condition and left V1 voxels should show increased activity under the FLASH-CONTEXT-R condition.

Figure 10 shows the results for the most anterior slice containing the region of interest. The top row shows the high resolution anatomical image (acquired after all functional acquisitions) and V1 LOCALIZATION standardized response amplitude map, thresholded for searching over all brain voxels. The region of interest was defined based on the V1 LOCALIZATION image; the dotted lines indicate voxels added to account for the cross-acquisition movement. The middle row shows the results for block 1 of the CONTEXT EXPERIMENT the bottom row shows block 2 results; The left images are t maps comparing FLASH to LISTEN, the right images are t maps comparing FLASH-CONTEXT-L to FLASH-CONTEXT-R; all t maps are thresholded *only* for searching over the 25 voxel region of interest.

[Figure 10 about here.]

Subject 3

Image Construction

The third subject was less satisfactory in terms of head movement; the first block of the CONTEXT EXPERIMENT had average movement as large as 1.6 mm (0.5 voxels) and clear indications of out of plane motion in the form of large edge artifacts in the variance images. For the other two functional acquisitions average movement was no worse than 1 mm (0.32 voxels), within our tolerance limits. Hence we discarded the first CONTEXT EXPERIMENT acquisition data and analyzed the remaining data.

V1 Localization

The optimum flash angle for this subject was determined to be about 20°. An asymetrical pair of activations were identified as the V1 loci of interest; both loci appear to span three slices, though the right locus starts two slices anterior relative to the left locus (Figure 11). From the near real time results we discarded the top two and bottom slices (losing the most anterior part of the right V1 activation).

[Figure 11 about here.]

For the four slices studied in the CONTEXT EXPERIMENT, the Bonferonni corrected 0.05 one sided t_{211} threshold for 4652 brain voxels is 4.35. With this threshold the right locus consisted of 13 voxels, the left locus 18 voxels.

Context Experiment

[Figure 12 about here.]

As previously mentioned, the first block of the CONTEXT EXPERIMENT was discarded due to severe motion. Not surprisingly, motion was detected between the V1 LOCALIZATION and second CONTEXT EXPERIMENT block by examination of the difference of mean images (Figure 12). From these difference images we can estimate that for the CONTEXT EXPERIMENT acquisition the the brain is shifted anterior and right (left in the figure), as that is the direction that the edge artifact points. Since the mean brain intensity is around 2000 and the edge artifact has magnitude of 500 to 1000, the movement is on the order of 1/4 to 1/2 of a pixel. Hence we need to expand our V1 LOCALIZATION regions of interest anterior and right. "Bleeding" the regions of interest in this direction added 38 voxels (13 to left, 25 to right); then the appropriate threshold for the FLASH minus LISTEN is the Bonferionni corrected 0.05 one sided t_{997} threshold for 69 voxels is 3.14.

Application of this criteron identified 7 activated voxels in the right region of interest and 5 activated voxels in the left region. The maximal t in the right region was 5.84, in the left region 8.26; the other significant voxels ranged from 3.22 to 4.49. The time course data for the two voxels with the largest t values (one from each side) are shown in Figure 13. There were no voxels within the regions of interest with significant decreases.

[Figure 13 about here.]

The FLASH-CONTEXT-L versus FLASH-CONTEXT-R comparison yeilded no significant voxels in the regions of interest. Searching over all brain voxels in all 4 slices (and raising the threshold appropriately to adjust for the additional tests) only 3 voxels were identified with any significant change between the FLASH-CONTEXT-L and FLASH-CONTEXT-R conditions, none of which were located near the V1 regions.

Subject 4

The last subject yeilded virtually no data due to head motion and subject distress.

Image Construction

Subject 3 had the greatest problems with movement. All acquisitions had average movements as large as 2.44 mm (0.78 voxels). The V1 LOCALIZATION acquisition had large in-plane motions that, while over our 1/2 voxel movement threshold, seemed to be corrected satisfactorially by FIASCO registration. Unfortunately the two CONTEXT EXPERIMENT acquisitions were completely corrupted by large out of plane motions; each had images with average movements over 1 pixel. Physiological data collected revealed bizzare patterns in respiration, with the subject holding his breath for over 20 seconds, multiple times.

While the CONTEXT EXPERIMENT data was completely unusable we analyzed the V1 LOCALIZATION since the motion, while great, seemed to be corrected.

V1 Localization

The optimum flash angle for this subject was about 22°. While the standardized response amplitude maps showed clear signs of motion artifact, there was also a pair of asymmetrical bilateral loci that corresponded to the expected V1 activation (Figure 14).

[Figure 14 about here.]

Context Experiment

No usable data was obtained from the either block of the CONTEXT EXPERIMENT.

4. DISCUSSION

We have examined four normal controls for evidence of context-dependent effects in the primary visual cortex. Only two subjects yielded any data on the CONTEXT EXPERIMENT; of these two, only one showed any context dependent effects, and that one subject showed increased bold response bilaterally for the FLASH-CONTEXT-L condition relative to FLASH-CONTEXT-R. The interaction was hypothesized to be hemisphere-dependent, that is greater activity in left hemisphere for the FLASH-CONTEXT-R condition, greater activity in right hemisphere for the FLASH-CONTEXT-L condition. Hence we have a found no support for the research hypothesis.

There are three plausible hypotheses for what we have observed. The first is that the imaging method is not sensitive enough for the subtle effect of interest. The second is that context is working in the visual system, as predicted by the interactive activation model, but not in V1. The third is that the interactive activation model is incorrect and context is working outside of the visual system. With a sample size of 2 subjects we do not feel there is sufficient evidence strongly support any one of these hypotheses over another, and hence the results are solidly inconclusive. This introduces the first and obvious aspect of future work needed: Additional subjects. At the time of writing three additional subjects have been scanned but their data not yet analyzed.

There are several other aspects of future work that break down into experimental design and data analytical issues. The former involves a major shift in focus away from V1 and into secondary visual areas which are more likely to show context effects; this will require reconfiguration of the experimental paradigm and the collection of new data. The latter involves investigating alternative approaches that will provide more sensitive models and schemes to address the problem of localization across acquisitions; we focus on the future direction of data analysis methods.

The main shortcoming in the statistical modeling is the inconsistent use of t tests and the drift plus functional response model of BRAIN. Subsequent work will consistently use BRAIN to model both the V1 LOCALIZATION *and* CONTEXT EXPERIMENT data. The modeling of the drift in BRAIN may add substantial sensitivity over simple t tests. There are several effects in the data that were not modeled: Epoch and word length are two effects that could improve model fit. Fitting an epoch effect makes the experimental effects be relative to a local-epoch baseline instead of an experiment-wide baseline. Fitting a word-length effect would account for the fact that longer words will retnotopically map to greater extents of of V1. Another effect, or pairing, in the experiment was the double presentation of word stimuli: each word pair was presented twice, once when the context was for the left word, once when it was for the right word. We were not able to exploit this pairing as we split the word stimuli across the two blocks of the CONTEXT EXPERIMENT, and, as seen in the results, cross acquisition comparisons are troublesome.

Which leads to next challenge for future work: Cross-acquisition analysis. A fundamental part of this work is the use of functionally defined regions of interest; the regions are derived from one acquisition and then applied to other acquisitions. While this a very powerful approach in that it focuses the inference to a handful of relevant voxels, we must assume negligible movement between acquisitions; this is a onerous assumption and one that is clearly not true when significant motion is present *within* acquisitions. Future work will include estimation of cross acquisition movement with FIASCO and another method, Automated Image Registration (AIR, Woods et al. 1992).

Lastly we comment that the potential for fMRI to image human brain function with high temporal and spatial resolution is limited by its sensitivity to artifacts due to motion and physiological effects. All of the subjects in the study were cooperative graduate students who were repeatedly informed of the need to remain absolutely still. Despite this, two of the four subjects had data corrupted by head motion. For this reason additional head restraint is also an area for future work.

5. CONCLUSION

The interactive activation model predicts the presence of context dependent activity in the early visual areas of the brain. We hypothesized that activity in the primary visual cortex would be modulated by context; in particular, that the fMRI BOLD response in V1 due to flashed word pairs would be modulated by context. Two out of four subjects yielded usable data; both showed a response to flashed word pairs, but only one showed context-dependent effects and not in the hypothesized manner. Hence we have found no evidence to support the stated hypothesis, and have discussed alternative hypotheses, but the general paucity of data prevents any specific conclusion. We are collecting more data pursing several avenues for analyzing the existing data with increased sensitivity.

REFERENCES

Buckner, R., Bandettini, P., O'Craven, K., Savoy, R., Petersen, S., Raichle, M., and Rosen, B. (1996), "Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging," *Proceedings of the National Academy of Science*, 93, 14878–14883.

Dale, A. and Buckner, R. (1997), "Selective Averaging of Rapidly Presented Individual Trials Using fMRI," *Human Brain Mapping*, 5, 329–340.

Eddy, W., Fitzgerald, M., Genovese, C., Mockus, A., and Noll, D. (1996), *Proceedings in Computational Statistics*, volume 12, pp. 39–49, Heidelberg: Physica-Verlag.

Genovese, C. and Sweeney, J. (1999), *Case Studies in Bayesian Statistics*, volume 4, chapter Functional connectivity in the cortical circuits subserving eye movements, New York: Springer-Verlag.

McClelland, J. and O'Regan, J. (1981), "Expectations Increase the Benefit Derived From Parafoveal Visual Information in Reading Words Aloud," *Journal of Experimental Psychology*, 7(3), 634–644.

McClelland, J. and Rumelhard, D. (1981), "An Interactive Actvation Model of Context Effects in Letter Perception: Part 1. An Account of Basic Findings," *Psychological Review*, 88(5), 375–407.

Rosen, B., Buckner, R., and Dale, A. (1998), "Event Related fMRI: Past, present and future," *Proceedings of the National Academy of Science*, 95(3), 773–780.

Savoy, R., Bandettini, P., O'Craven, K., Kwong, K., Davis, T., Baker, J., Weisskoff, R., and Rosen, B. (1995), "Pushing the Temporal Resolution of fMRI: Studies of Very Brief Visual Stimuli, Onset Variability and Asynchrony, and Stimulus-Correlated Changes in Noise," in *Proceedings of the International Society for Magnetic Resonance in Medicine*, pp. 450.

Vazquez, A. and Noll, D. (1998), "A Study of the Time-invariant Properties of the BOLD Response," in *Proceedings of the International Society for Magnetic Resonance in Medicine*, pp. 1424.

Voyvodic, J. (1996), "Real-time fMRI paradigm control software for integrating stimulus presentation, behavioral and physiological monitoring, and statistical analysis," *Proceedings of the International Society for Magnetic Resonance in Medicine*, 1835, Abstract.

Woods, R., Cherry, S., and Mazziotta, J. (1992), "Rapid automated algorithm for aligning and reslicing PET images," *Journal of Computer Assisted Tomography*, 16, 620–633.

Id: TechRpt.tex,v 1.8 1999/11/23 04:49:35 nicholst Exp

List of Figures

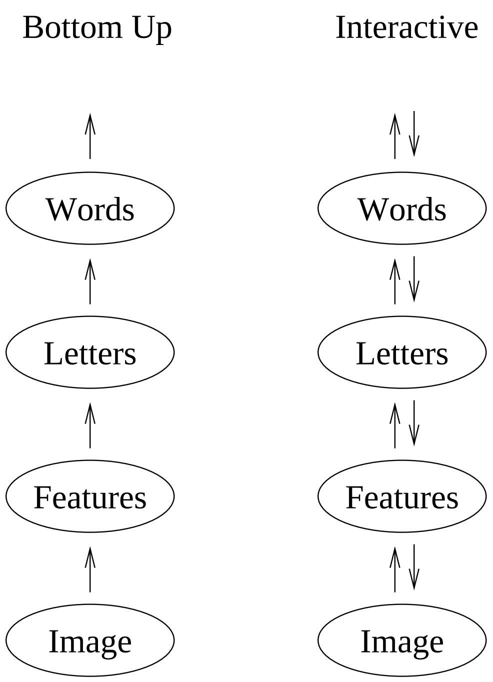

Figure 1. Illustrations of the two connectionist models for visual perception, the “Bottom Up” or unidirectional model and the “Interactive” or bidirectional model. In the bottom up model information only flows upward, that is, the activity at the top layer is solely a function of the external stimuli. In the bidirectional model information can flow either direction, for example, the activity in the word layer can effect the activity in the letter layer, which can in turn effect activity in the feature layer, etc.

• Fixate:		+	
• Flash two words briefly (10 ms):	space	+	radio
⇒ Cannot identify either word, words flashed too quickly.			
• Display Context Sentence (5 s):	I like my coffee with cream and		
• Fixate:		+	
• Flash two words briefly (10 ms):	sugar	+	march
⇒ *Can* identify left word, context aided perception.			

Figure 2. Visual Perception Context Experiment. Shown are the structure of two types of trials, first a no-context trial, then a context trial. For the no-context trial the subject fixates on the cross-hair; after a pause a pair of words are flashed to either side; the words are flashed so quickly and so far apart that the subject cannot read either word, but if flashed slightly more slowly or slightly more centrally the subject could begin to see them. For the context trial, before the subject fixates on the cross-hair a sentence is presented which provides context for one of the words to be flashed; using the same presentation parameters as the other trial (i.e. same timing and distance from center) the subject will be able, on average, to identify the word that matches the context but not the other word. This demonstrates that context can affect the actual formation of a percept.

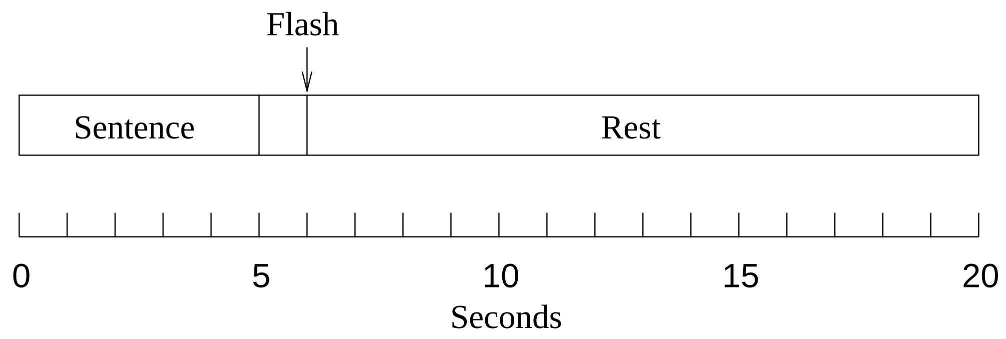

Figure 3. Timing of Context Experiment. Sentences aurally presented and delayed such that they end at the 5 second mark. After a 1 second pause the pair of words is flashed parafoveally. The subject rests for 14 seconds until the next trial.

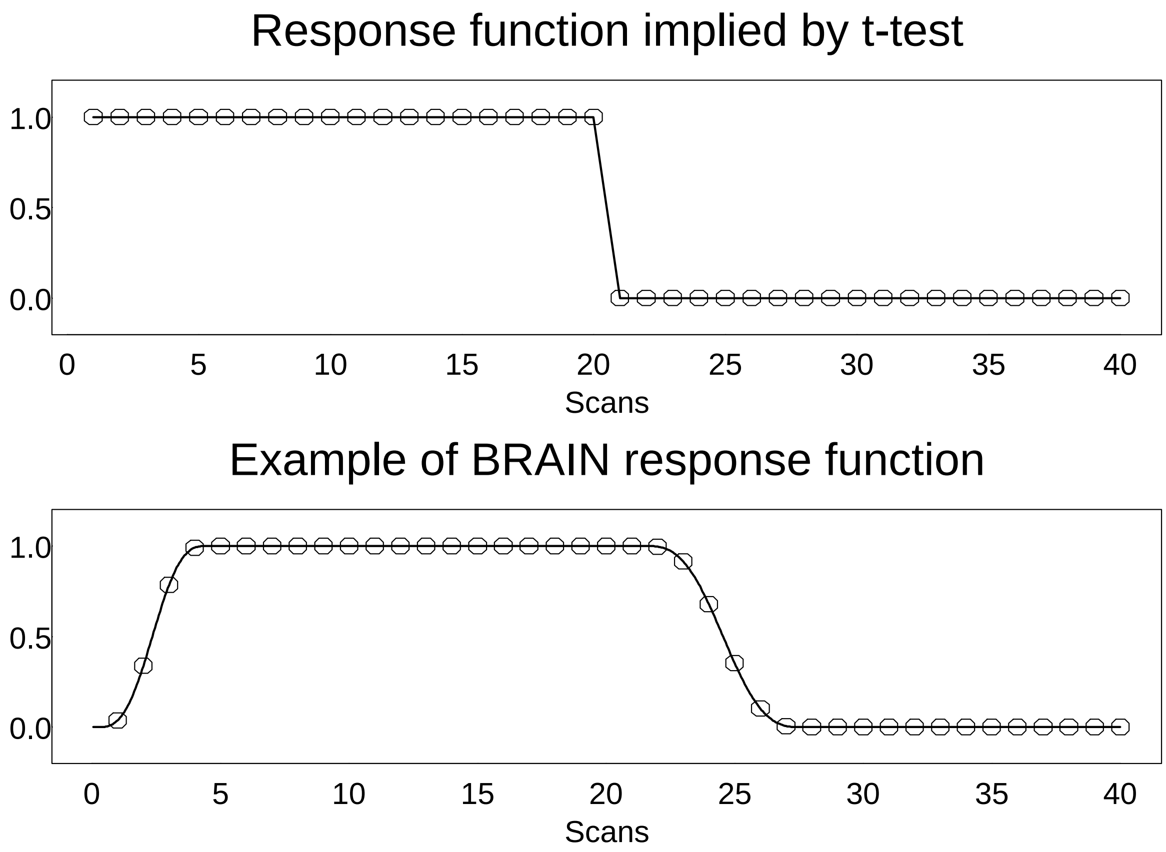

Figure 4. Comparison of basic and BRAIN hemodynamic response functions for a stimuli lasting from scans 1 through 20. A standard t test implies a model of constant activity during scans 1 through 20, and a independent level during scans 21 through 40; the difference between the two levels is the parameter of interest. The BRAIN response function models a smooth response parameterized with an amplitude parameter and 4 shape parameters; the shape parameters describe a variable delay until rise begins and reaches plateau, relative to the onset of the stimuli at scan 1, and a variable delay until the fall begins and returns to baseline, relative to the cessation of the stimuli at scan 20.

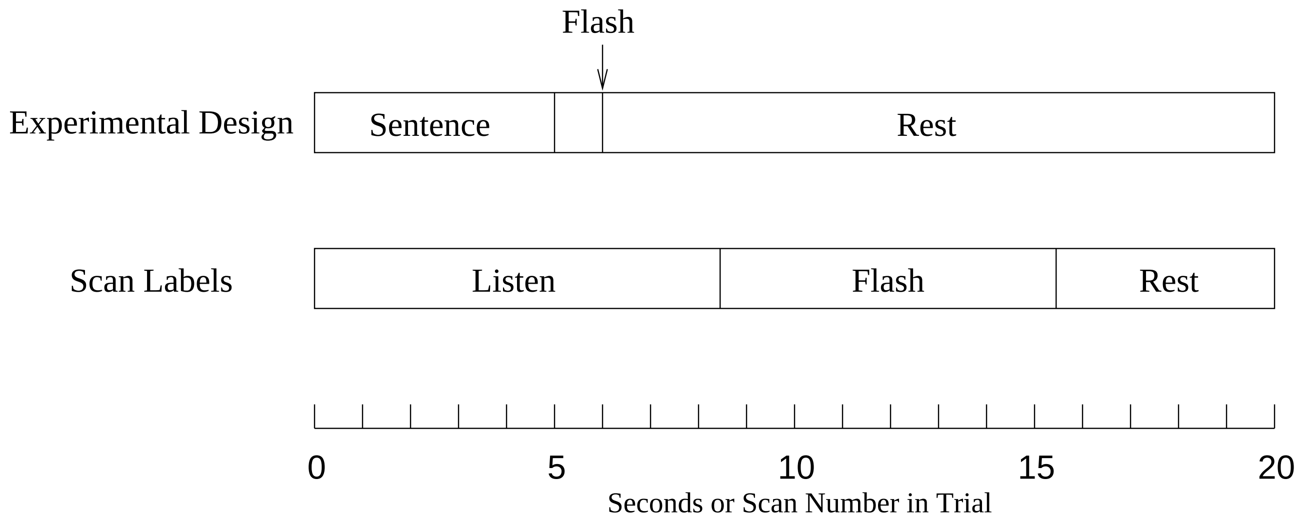

Figure 5. Comparison of experimental design and assignment of scan labels for t tests. The BOLD response is delayed and dispersed relative to the experimental stimuli, so the scan labels must reflect this. The BOLD response has a well-known 2 second delay until a substantial rise is seen, hence the LISTEN condition encompasses the two scans after the flash. Single trial literature on very brief visual stimuli suggest that the signal returns to baseline after approximately 7 seconds, hence the FLASH condition is 7 seconds long. The remaining 5 scans were assigned to the REST condition.

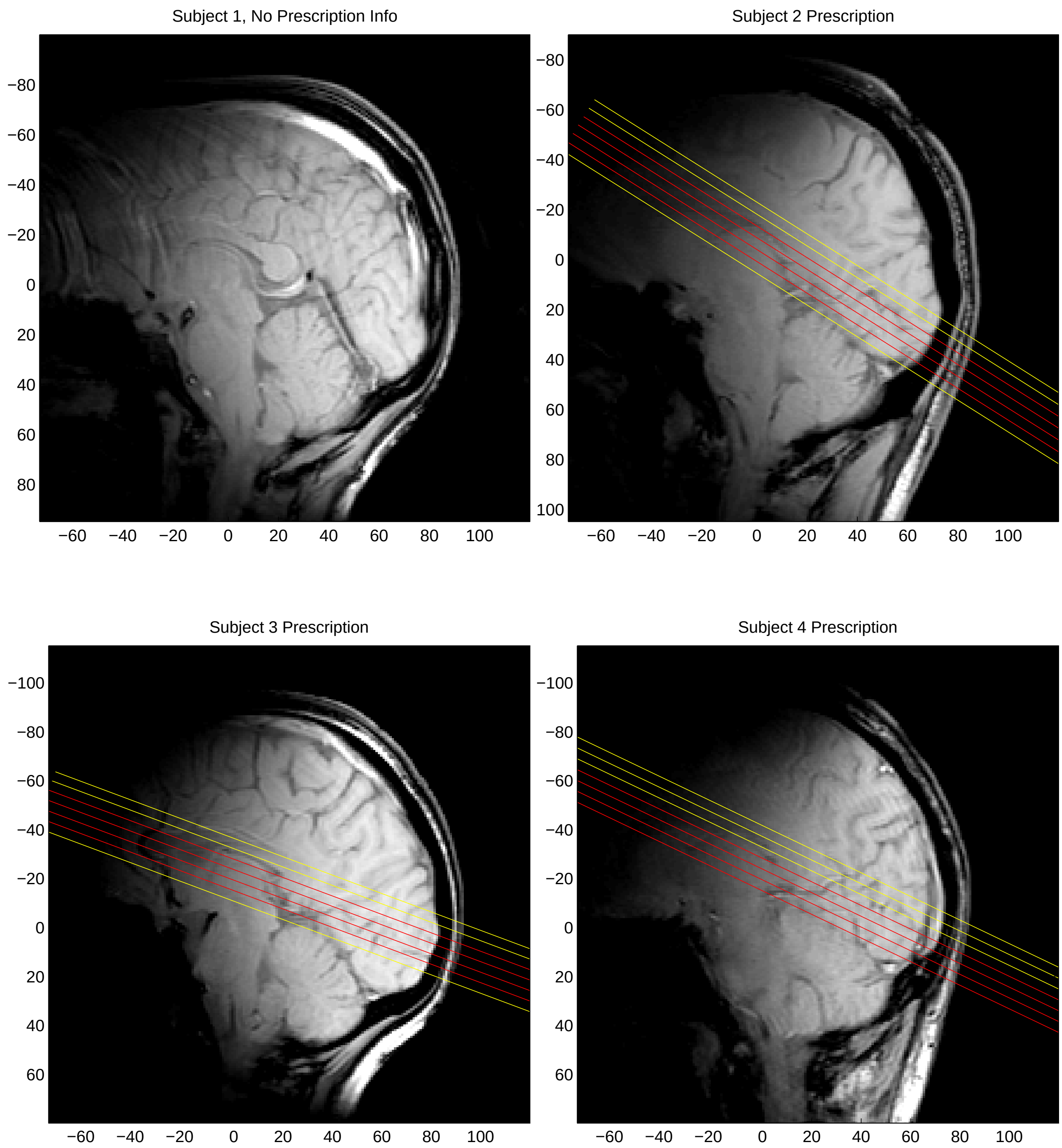

Figure 6. Prescriptions of functional slices for all subjects except first, where precise localization information was lost. Prescriptions were chosen to cover and be be parallel with calcarine fissure. Red lines indicate slices that were chosen for study in the CONTEXT EXPERIMENT.

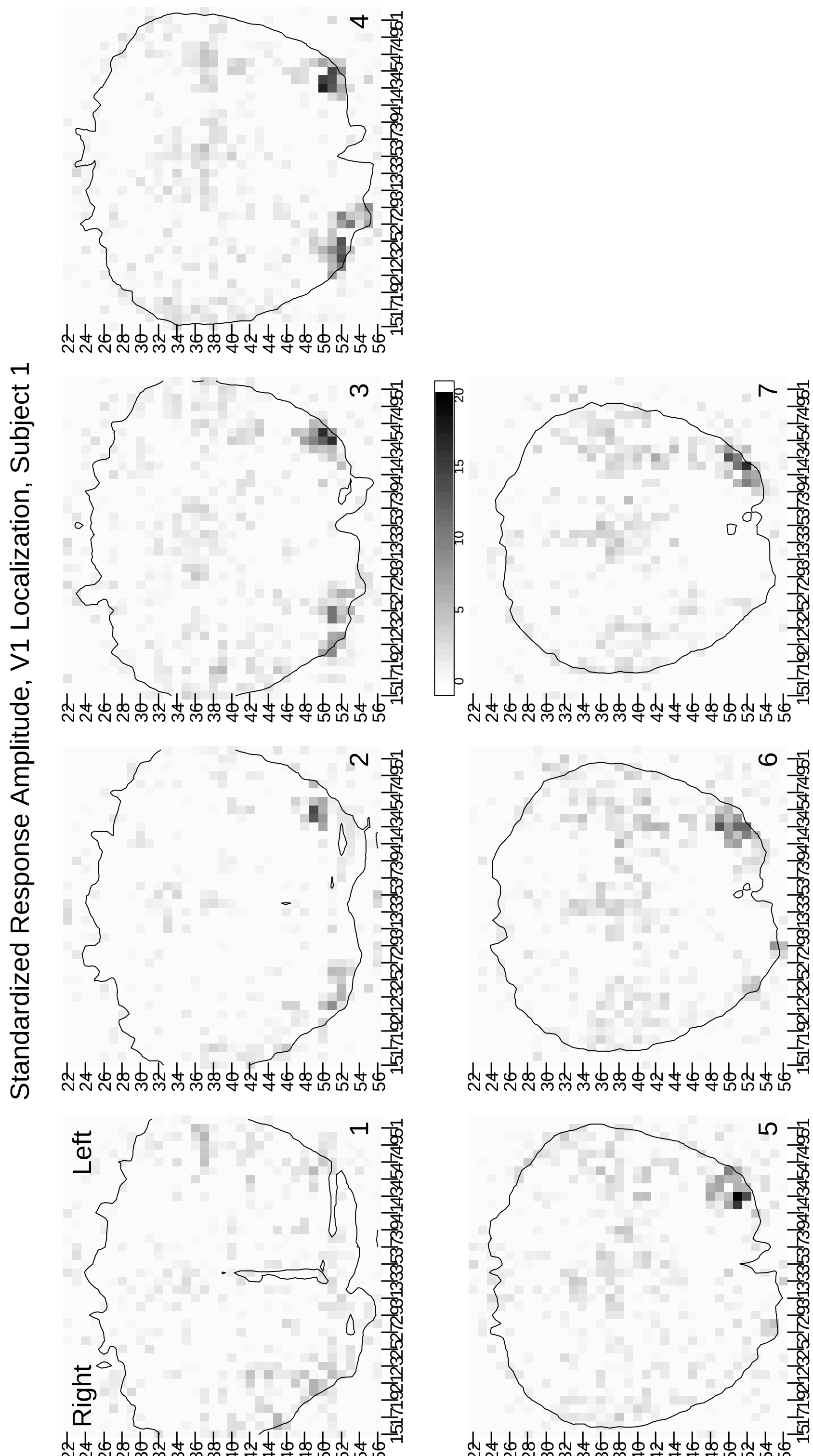

Figure 7. Unthresholded map of standardized response amplitude for the V1 LOCALIZATION experiment, subject 1. First slice is most inferior and anterior; prescription is oblique coronal-axial. Black line is contour of brain tissue producing significant signal. Dark regions indicate areas where amplitude of response was substantially greater than the standard error of the response. Note the lack of medial activation.

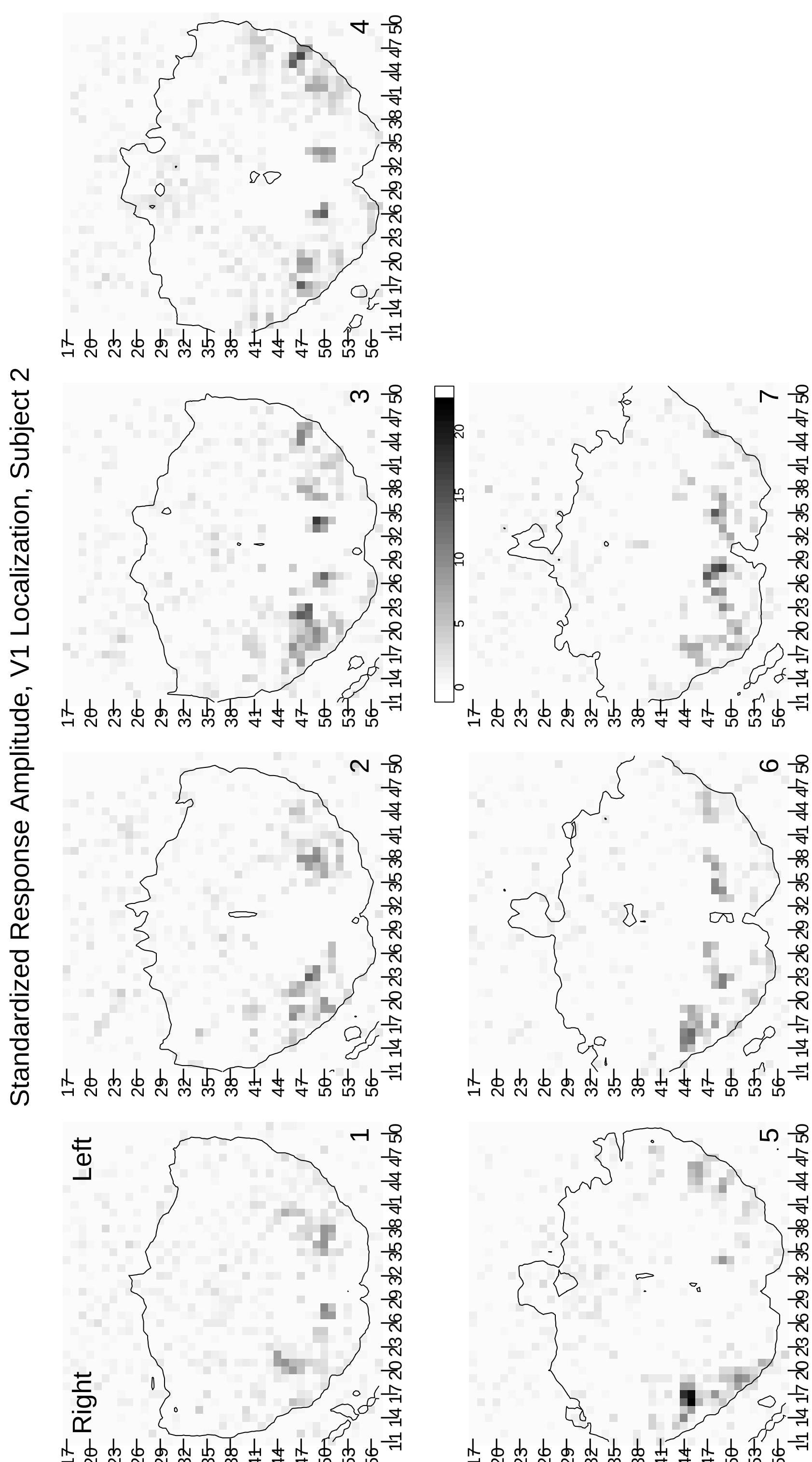

Figure 8. Unthresholded map of standardized response amplitude for the V1 Localization experiment, subject 2. First slice is most superior. Black line is contour of brain tissue producing significant signal. Dark regions have greatest amplitude of response relative to the standard error. Medial bilateral activations on slices 3 and 4 were identified as V1.

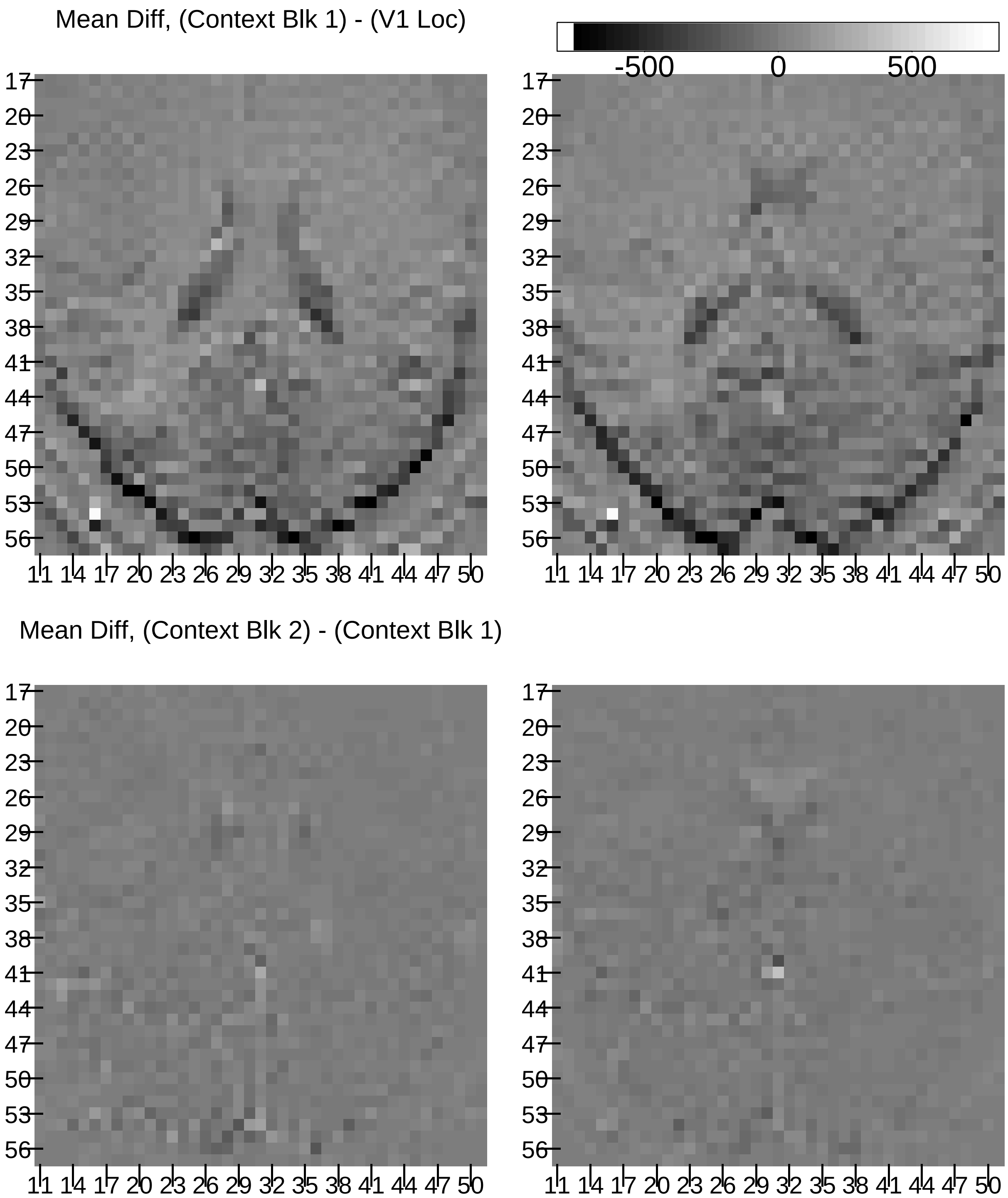

Figure 9. Interacquisition difference images. The top slices show the difference between the mean of the V1 LOCALIZATION acquisition and the first block of the CONTEXT EXPERIMENT; a forward shift is evident. The bottom slices show, on the same color scale, the difference between the means of the two acquisitions of the CONTEXT EXPERIMENT; virtually no motion is evident.

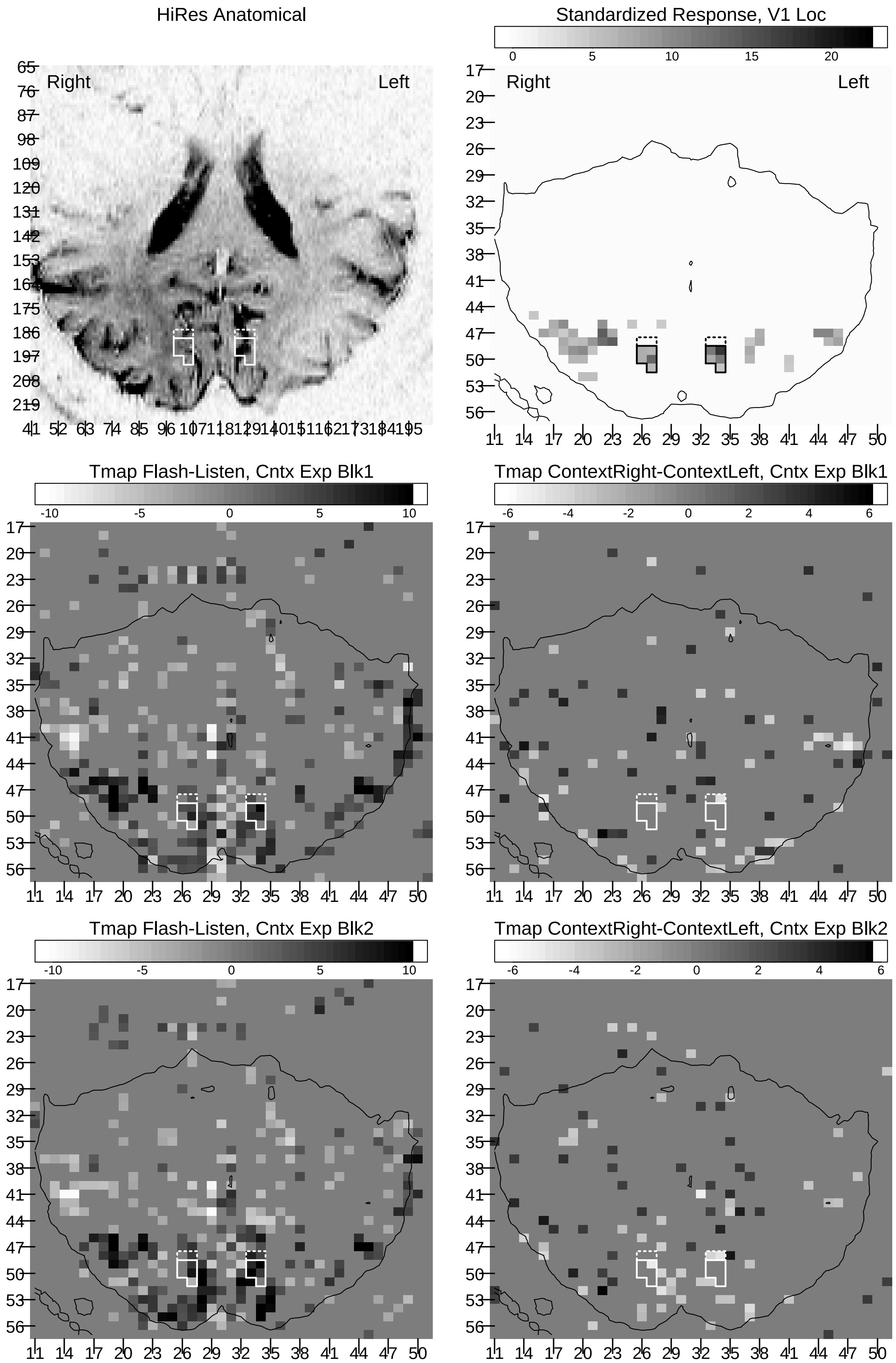

Figure 10. Results for subject 2, anterior-most slice. All images show region of interest defined by V1 Localization; dotted lines indicate voxels added to account for movement. Top Left: High resolution anatomical image. Top Right: Standardized response amplitude for V1 Localization, thresholded for searching over all brain voxels. Middle & Bottom Left: t map of Flash-Listen for block 1 (middle) and block 2 (bottom) of Context Experiment, thresholded for searching the 20 voxel region of interest. Dark voxels correspond to Flash greater than Listen. Note the consistency of the activated voxels (within the region) between the two blocks. Middle & Bottom Right: t map of Flash-Context-L - Flash-Context-R for blocks 1 and 2 of the Context Experiment, threshold for the 25 voxel region. All activated voxels, for both sides, are in favor of the Flash-Context-L > Flash-Context-R alternative; this is inconsistent with the experimental hypothesis which predicts such changes only in the right hemisphere.

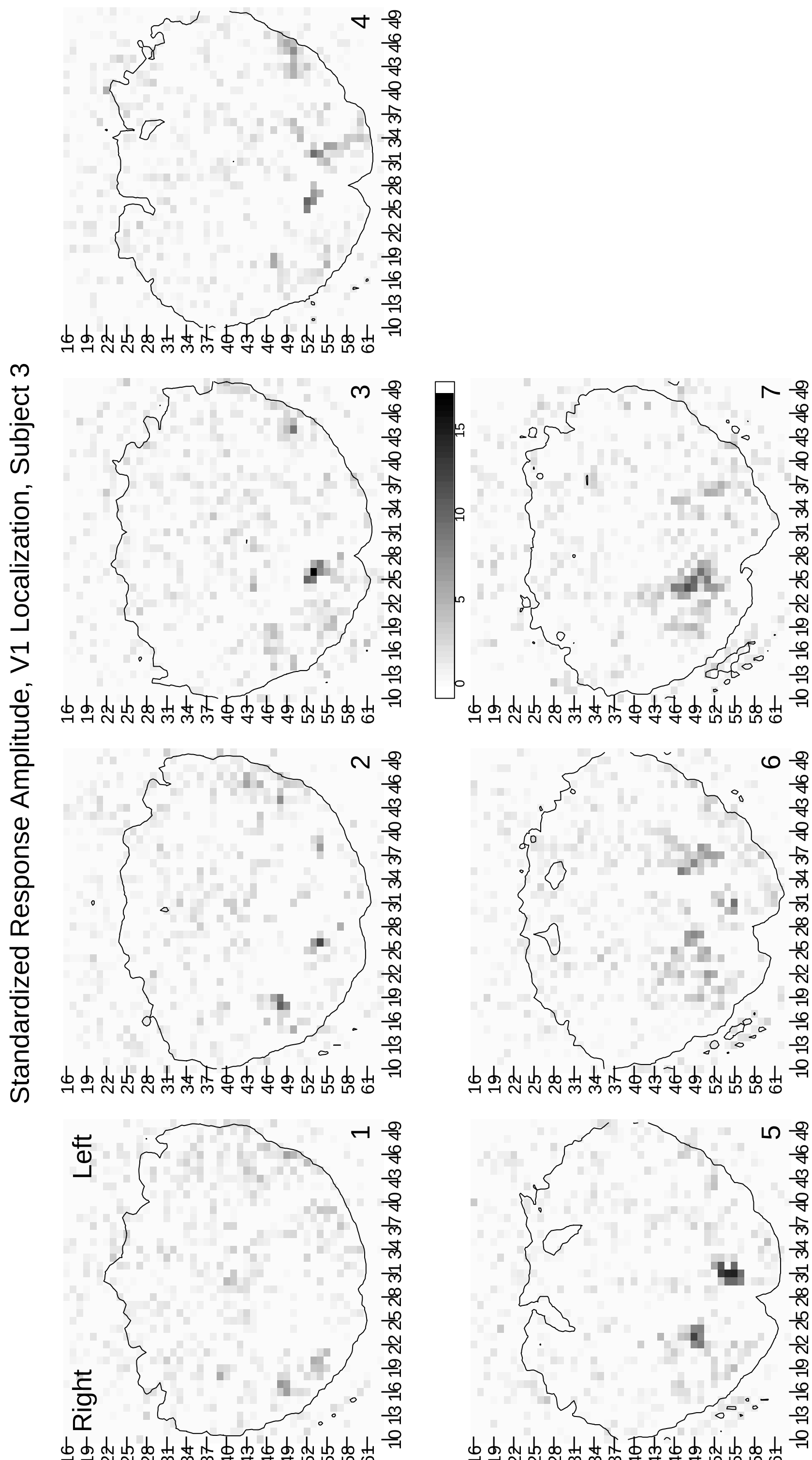

Figure 11. Unthresholded map of standardized response amplitude for the V1 LOCALIZATION experiment, subject 3. First slice is most superior. Black line is contour of brain tissue producing significant signal. Asymmetrical medial bilaterial activations on slice 4 were identified as V1.

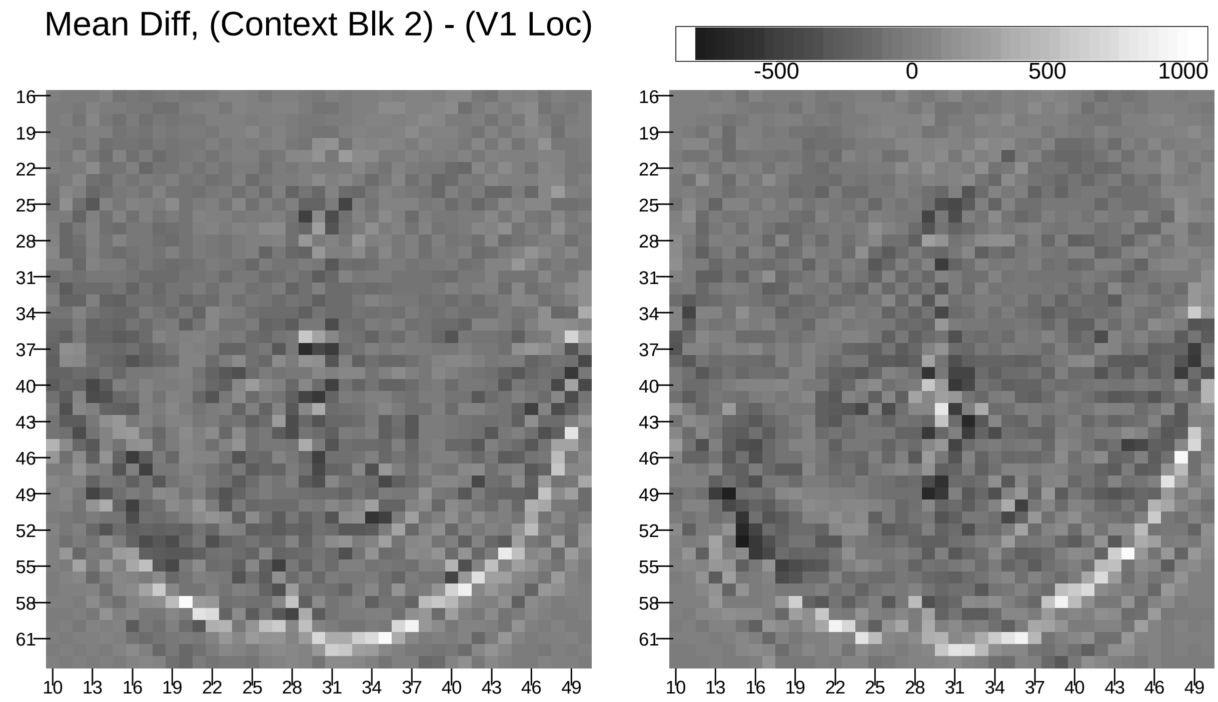

Figure 12. Two most anterior slices of difference image of CONTEXT EXPERIMENT (block 2) versus V1 LOCALIZATION for subject 3. Edge artifact indicates motion; positive intensity of artifact indicates the brain in the CONTEXT EXPERIMENT block 2 acquisition has moved posteriorly.

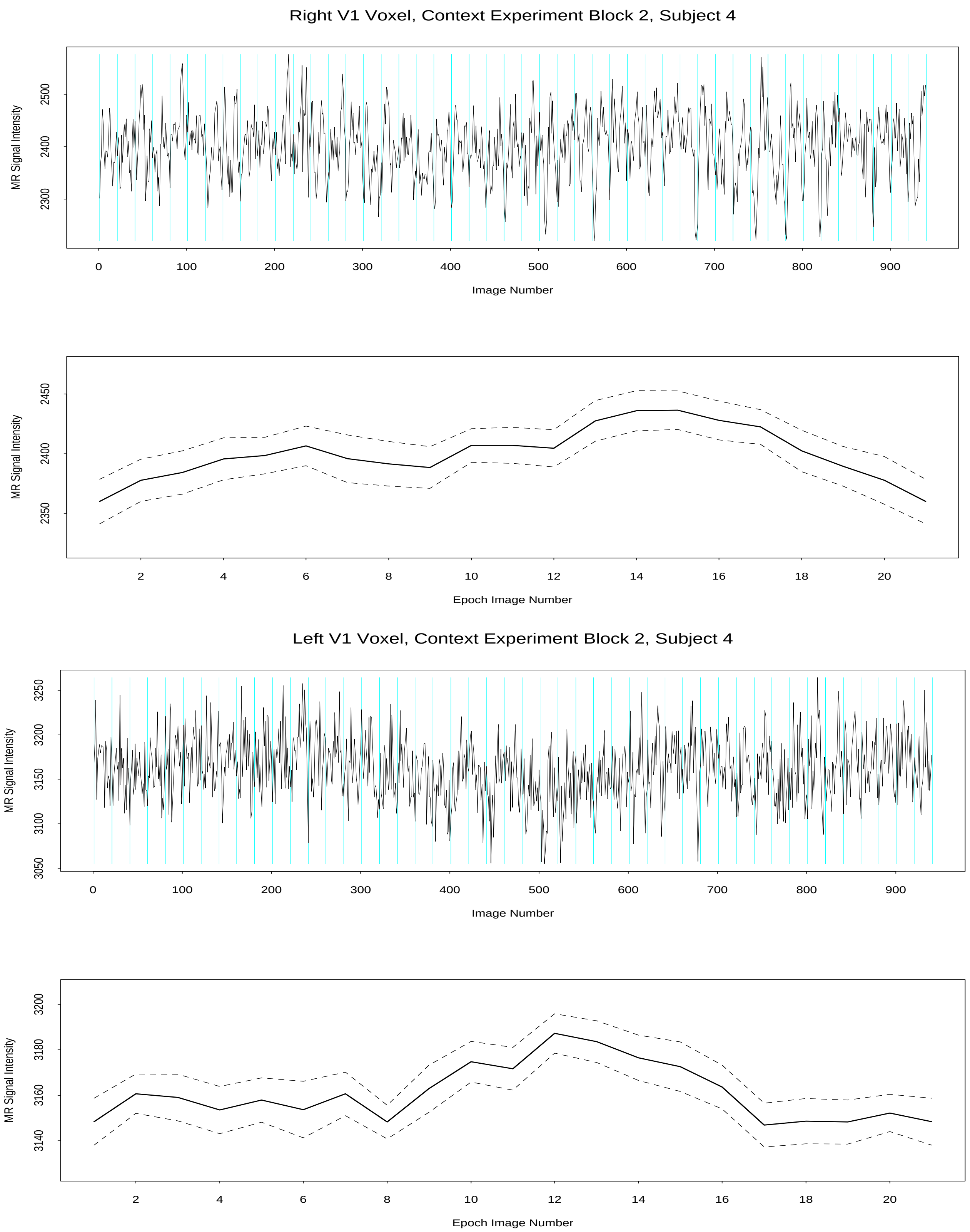

Figure 13. Raw and averaged data for two voxels identified as significant in the bilateral V1 areas of interest, Subject 3. Raw data plots shows every data point collected at the voxel of interest; the averaged data plots show the mean activity averaged over 20 second epochs; the dotted lines indicate univariate 95% confidence intervals for each time point in the epoch.

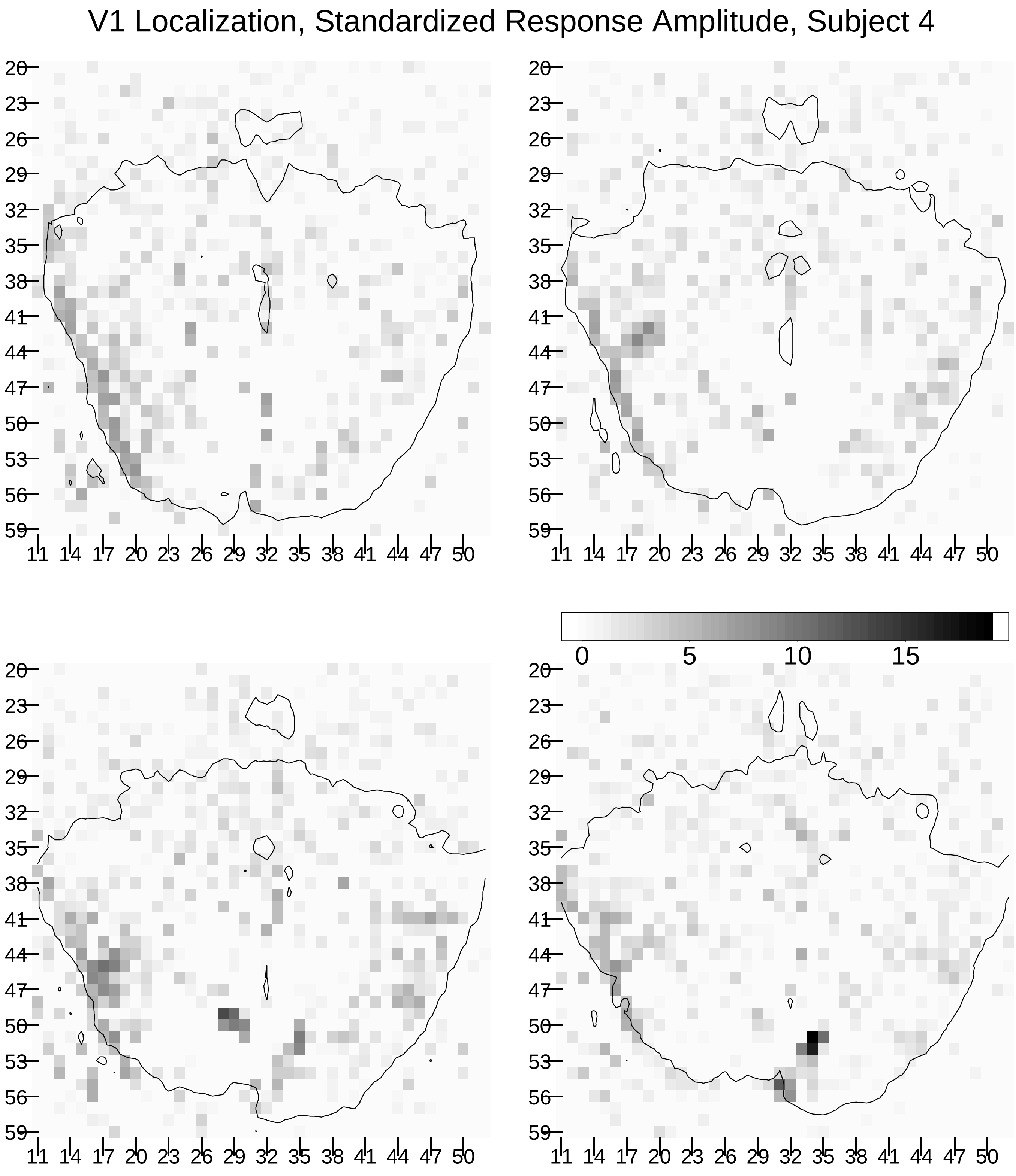

Figure 14. Unthresholded map of standardized response amplitude for subject 4, slices 2 through 5 of the V1 Localization acquisition. First two slices show the movement artifact; second two show bilateral medial V1 activation.