Use of concurrent pupil dilation assessment to inform
interpretation and analysis of fMRI data

Greg J. Siegle[1], Stuart R. Steinhauer[1,2], V. Andrew Stenger[1], Roma Konecky[1], Cameron S. Carter[1]
[1]University of Pittsburgh, [2]VA Pittsburgh Healthcare System

Corresponding author: Greg Siegle, Ph.D., Western Psychiatric Institute and Clinic, 3811 O'Hara St., Pittsburgh, PA 15213, ph: 412-624-3172, fax: 412-624-0223, e-mail: gsiegle+@pitt.edu

Abbreviated running title: Pupil dilation and fMRI

Abstract
   Potential contributions of concurrently acquired pupil dilation data to fMRI experiments were examined. Sixteen healthy participants completed a working memory task (digit sorting) during measurement of pupil dilation outside the fMRI environment and during concurrent 3T fMRI assessment. Pupil dilation increased parametrically with task difficulty inside and outside the scanner, on a similar time-course suggesting that task demand was similar in both environments. The time-course of pupil dilation during digit sorting was similar to the time course of the fMRI signal in the middle frontal gyrus, suggesting that middle-frontal gyrus activity indexed the engagement working memory processes. Incorporating individual differences in pupil dilation improved the sensitivity and specificity of general linear modeling analyses of activity in the middle frontal gyrus, above and beyond standard analytic techniques. Results suggest concurrent pupil dilation during fMRI assessment can help to 1) specify whether task demand is the same inside and outside the fMRI environment, 2) resolve the extent to which fMRI signals reflect different aspects of event-related designs, and 3) explain variation in fMRI data due to individual differences in information processing.

Introduction
   Functional magnetic resonance imaging (fMRI) is frequently used to understand brain correlates of cognitive processing. Interpreting fMRI data is often difficult because it is unclear whether non-ecological features of the fMRI environment affect task-related cognition. It is also difficult to generate specific predictions about the expected time-course of brain activity in response to individual stimuli, with which to constrain fMRI analysis. This study examines the extent to which concurrent collection of pupil dilation and fMRI data can aid in resolving these challenges.
   Many studies have demonstrated pupil dilation to be a reliable correlate of cognitive processing; the pupil dilates more under conditions of higher attentional allocation, memory use, and interpretation of more difficult material (see Beatty 1982a, Steinhauer & Hakerem, 1992, for reviews). Pupil dilation persists if the demand is sustained (e.g., Beatty, 1982b). As individuals are asked to remember larger numbers of digits, for example, pupil dilation increases proportionally (e.g., Kahneman & Beatty, 1966; Granholm et al., 1996). Of particular relevance, the pupil is innervated by brain areas associated with both cognitive and emotional processing (e.g., Szabadi & Bradshaw, 1996) suggesting that it may represent a general summative index of brain activity associated with performing cognitive and emotional tasks. Technology for concurrent acquisition of pupil dilation and fMRI data is becoming increasingly available. Pupil size is a standard variable collected by most popular eye-trackers used in fMRI environments, e.g., ASL[TM], SMI[TM], ISCAN[TM]. Pupil dilation may thus be an informative and easily acquired peripheral indicator of cognitive activity that could help to guide analysis and interpretation of fMRI data.
   In this paper, we will address three primary questions that pupil dilation can help to answer:
   1) Are cognitive processes similar inside and outside the scanner? The fMRI environment is intensely non-ecological – it is a dark room in which participants lie down and complete tasks while lying perfectly still, often in a cramped position. The extent to which cognitive processes observed inside the scanner are the same outside the scanner are difficult to estimate. By acquiring pupil dilation in response to tasks outside and inside the fMRI environment, the extent to which cognitive processes change in the fMRI environment can be estimated.
   2) What aspects of cognitive processing on information processing tasks are represented by the BOLD signal? Analyzing fMRI data presumes that participants are actively attending to a presented task. Often condition-related differences in behavioral data are used to infer participant effort. Yet, behavioral data is often a coarse measure of

cognitive processing. For example, a participant's reaction time is, by definition, measured one time in a trial, which precludes detailed analysis of variation in information processing before or after the reaction time. Moreover, not all tasks lend themselves readily to behavioral measures. In these cases, an ancillary physiological measure may allow inference of task-related cognitive processing.

For example, we have used pupil dilation measured outside the scanner to assist in interpreting the results of a previous fMRI study in which covert responses were generated (Brown et al, 1999; participants silently responded to a Stroop color-naming task and no verbal or button-press responses were acquired). Since behavioral measures were unavailable from which to infer participant engagement in the task, we collected pupil dilation data outside the scanner during the same task with overt (verbal) and covert (silent) responses. By demonstrating that pupil dilation varied consistently with task condition, and was similar during overt and covert responding conditions, participant engagement in the covert task was supported.

The information gained from measuring pupil dilation can then be parlayed to examining relationships between the time course of cognitive processing and the BOLD signal. One such application could be to aid in understanding the extent to which observed activity in a region is associated with different aspects of a cognitive task. By examining the extent to which variation in pupil dilation, which has higher temporal resolution than fMRI, during different temporal segments of a task, explains observed variation in fMRI data, it may be possible to enhance our understanding of the extent to which brain activity in a region is associated with specific cognitive operations on a finer time scale than fMRI can measure.

An extension of this approach involves using pupil-dilation to constrain assumptions about the time-course of brain activity for fMRI analysis. For example, general-linear-modeling (GLM), a dominant method of fMRI analysis, involves examining the extent to which fMRI activity correlates with a prediction of the fMRI signal. The predicted signal is commonly derived by convolving a prediction of cognitive activity with a hemodynamic filter. This process presumes a knowledge of the time-course of cognitive activity in a task, often estimated as a pulse at the onset of a trial; if cognitive activity is more temporally extended, e.g., on the order of a few seconds, it may be important to provide for such a sustained process in the computation of a regressor. Towards that end, some groups have used a sequence of pulses extending from the beginning of a trial through the execution of a behavioral response to improve power over a single pulse regressor (Christoff et al, 2001). Similarly, pupil dilation can be used to estimate the time-course of cognitive load occurring both before and after a participant's reaction time. In this way, pupil dilation waveforms could inform the construction of regressors or selection of scans for analysis.

3) *Is pupil dilation a sensitive and specific predictor of regional brain activity?*

Even if pupil dilation predicts brain activity on a cognitive task, the additional effort required to employ this methodology is only justified if pupil dilation provides information above and beyond standard analysis paths that account for condition. That is, the method should ideally be more sensitive and specific to variation in activity in brain regions of theoretical interest than traditional methods. There are a number of reasons to believe that accounting for pupil dilation could provide increased signal detection. First, as stated in the previous section, the time course of cognitive activity across subjects may be more accurately estimated using a collateral physiological responses. Second, by incorporating condition-related differences in pupil dilation for each subject into analyses, individual differences in the expected time-course of cognitive processing can be entered into the GLM. Third, by using the entire pupil dilation waveform throughout an experiment as a regressor, trial-to-trial effects and differences in processing throughout a task can be used to predict brain activity throughout a task. Because, each of these innovations introduces potential prediction error due to the potential instability and unreliability of the pupil dilation signal, it is useful to see whether incorporating information about individual differences and trial-to-trial effects improves or decreases sensitivity and specificity.

In the following experiment, we used concurrent acquisition of pupil dilation and fMRI data to examine these considerations. A slow event-related design was used to capitalize on the pupil's ability to aid in understanding the time-course of cognitive events. Healthy participants completed a working memory task that involved sorting digits outside and inside the scanner. Digit sorting was chosen because it accesses a variety of working memory processes (e.g., Baddely's (1986) central executive as well as maintenence components), and was therefore likely to produce activity parametric with task difficulty in brain areas associated with working memory. In particular the dorso-lateral prefrontal cortex (DLPFC) activity has been detected in many other fMRI studies of working memory (e.g., Braver et al, 1997; Callicott et al, 1999; Cohen et al, 1997; Rypma & D'Esposito, 1999; Rypma et al, 1999). Other desirable properties of the task are that it takes 2-7 seconds to complete cognitive operations involved in a trial making it appropriate for event-related analyses involving sustained processing, and it is sufficiently easy that participants rarely make errors. Healthy individuals have been shown to perform well and to display remarkable consistency in using a single strategy of systematically identifying successive lowest digits on a similar task (Piaget, 1965). This strategy suggests that the time involved in cognitive processing may vary with the number of digits to be sorted. The task was particularly appropriate for analysis using pupil dilation because

the cognitive operations of interest (sorting the digits) happen long before a behavioral response is made; augmenting our understanding of behavioral responses with additional physiological could thus provide particular insight into the dynamics of cognitive processing on the task.

Materials and Methods

Participants.

Participants included 16 healthy individuals (4 Male, 13 Caucasian, ages 21-48, M(SD)$_{age}$=24(7.0)). Participants stated they had no significant eye problems; a subset who were assessed with a hand-held eye chart had normal corrected vision (at least 20/30) with both eyes open; the remainder reported no difficulty reading text on a computer screen much smaller than the text used in the experiment. Participants had no Axis I psychopathology as assessed by the SCID (First et al, 1997) interview. Participants reported no psychoactive drug abuse within the six months prior to testing. Fifteen participants reported no alcohol abuse within the past six months on the SCID; one participant's data was ambiguous in this regard, and the participant could not be recontacted. Participants reported no history of head trauma, and were not impaired on a cognitive screen (Blair & Spreen, 1989).

Procedure.

Three appointments were scheduled with participants within a three-week period. At their first appointment participants were told about the experiment, signed IRB-approved consent forms, completed an initial clinical interview and a brief vision test. At the second appointment participants completed the digit-sorting task followed by a battery of other cognitive and emotional processing tasks during which pupil dilation and facial EMG data were collected. Time of day was not controlled for. Participants returned for a third appointment in which the tasks were repeated during concurrent pupil dilation and fMRI assessment.

Apparatus – Pupil dilation only.

Testing occurred in a moderately lit room (.56 foot-candles or 6.03 lux illuminance) in which the experimenter was not present. Participants sat approximately 65.5 cm from stimuli. Stimuli were digits approximately 1.59 cm high, subtending 1.4 degrees of visual angle. Reaction times were recorded using a game pad capable of reading reaction times with millisecond resolution.

Pupil dilation was recorded at the VA Pittsburgh using methods previously described and tested (e.g., Steinhauer, Condray, & Kasparek, 2000). In brief, data were collected using an ISCAN RK406 pupillometer. The pupillometer consisted of a video camera and infrared light source that were pointed at a participant's eye, and a device that tracked the size of the pupil using these tools. Pupil size was recorded at 62Hz (every 16ms) and passed in analog form from the pupillometer to a computer that digitally stored the acquired data along with signals marking the beginning of trials, the end of fixation, stimulus onset time, and reaction time. The pupillometer's resolution for a typical participant was better than 0.025 mm pupil diameter. Data collection was managed using EEGSYS software (Hartwell, 1995).

Apparatus – Concurrent pupil dilation and fMRI assessment

As participants lay in the moderately lit scanner, stimuli were projected on a back-projection screen approximately 47cm from a mirror placed approximately 13cm above their eye (varied slightly by head size). Stimuli were digits approximately 2mm high in the mirror, yielding a visual angle of .88 degrees. Thus, the stimulus was effectively perceived as 1.5cm tall at a distance of 60cm. Reaction times were recorded using a Psychology Software Tools™ glove capable of reading reaction times with millisecond resolution.

Thirty-four 3.2mm slices were acquired parallel to the AC-PC line using a reverse spiral pulse sequence (T2*-weighted images depicting BOLD contrast; TR=1500ms, TE=5ms, FOV=24cm, flip=60 on a 3T GE scanner), yielding 12 whole-brain images per 18 second trial. Pupil dilation was acquired using an ASL model 504 eye-tracker. This device consists of a video camera and infra-red light source positioned outside the magnet's bore. The pupil was automatically tracked through a mirror anchored to the head-coil. Pupil size was recorded at 60Hz (every 16.7ms) along with signals marking the beginning of trials, the end of fixation, stimulus onset time, and reaction time. Because of the camera's considerable distance from the participant's eye (over 10 feet) pupil resolution was poorer than for dilation measured outside the scanner; because of intersubject variability in head-size it was not possible to get absolute measures of diameter; rather a relative metric was used. Data collection was managed using ASL's e5win software.

Task.

A digit-ordering task similar to that used by MacDonald et al (2001) was employed. Participants viewed a fixation mask (row of X's with vertical prongs over the center) for one second followed by simultaneous presentation of three to five different digits, which remained on screen for two seconds after which they were covered by a mask (row of X's) for five seconds. A target digit appeared for the following ten seconds. Participants were instructed to read the digits from left to right, sort them in memory, and remember the middle digit – if there were an even number of digits, the higher of the middle digits was to be remembered. Directions suggested they should not use special strategies such as sorting the digits by moving their eyes on the screen or sorting only some digits, because we were examining the process of sorting items in

memory. When the target came on, participants were instructed to push a button for "Yes" if the target was the middle digit, and "No" if it was not. To account for differential response latencies, the order of these buttons was counterbalanced across participants. Stimuli were displayed in black, on a white background. These characteristics minimized dilation responses to changes in illumination associated with stimulus onset and offset.

Data Selection and Cleaning

Aggregation of Reaction Times. Harmonic means of reaction times were used to reliably index the central tendency of an individual's reaction times within a condition (as recommended by Ratcliff, 1993). To eliminate spurious skew due to outliers while preserving rank-ordering of data, outliers more than 1.5 times the interquartile range from the median harmonic mean on any variable were scaled to the closest obtained value below this cutoff plus the difference between this value and the next closest value as in Siegle et al (2001). This technique was adopted rather than other techniques (e.g., trimmed means) to preserve as much valid data as possible, while not decreasing statistical power due to inclusion of outliers.

Preparation of fMRI data for analysis. Statistical analyses were conducted in the Neuroimaging Software (NIS) data stream using software developed locally. Data were prepared using methods described by Carter et al (2000). Following motion correction using the 6 parameter AIR algorithm (Woods, Cherry, & Mazzoitta 1992), linear trends in fMRI data, calculated over blocks of 40 trials (5.5 minutes) were removed to eliminate effects of slow drift in the fMRI signal that were not related to trial characteristics. fMRI data were then temporally smoothed via convolution with a five-point middle-peaked filter, cross-registered to a standard reference brain using the 12 parameter AIR algorithm, and spatially smoothed using a 6 mm FWHM filter.

Preparation of pupil-dilation data for analysis. Data were cleaned using methodology previously described by Granholm et al (e.g., 1996). Artifacts, including blinks, were identified as large changes in pupil dilation occurring too rapidly to signify actual dilation or contraction. Linear interpolations replaced artifacts throughout the data set. Data were smoothed using a five point unweighted average filter applied twice. Linear trends in pupil dilation calculated over blocks of 20 trials were then removed from pupil dilation data to eliminate effects of slow drift in pupil diameter. Pupil diameter, measured as the average dilation over the one second preceding the onset of the stimulus, was subtracted from pupil diameter after stimulus onset to produce pupil dilation indices. Because absolute dilation could not be calculated for data collected in the fMRI scanner, an index representing the proportion of maximal dilation was used as a proxy for diameter.

Selection of stimuli for analysis. Data were only analyzed for correct trials on which the reaction time was between 50 ms and 5000 msec. Data outside this range was considered to have been generated by processes unrelated to the task. Pupil dilation data was analyzed only for trials composed of <50% interpolated area for data acquired outside the scanner <70% interpolated area for data acquired in the scanner, in which dilation remained within 5 standard deviations of its baseline. Different cutoffs were used because pupil dilation data acquired in the scanner had lower resolution and considerably more noise than data acquired outside the scanner.

Creation of regressors for fMRI analysis. All results employed random-effects hierarchical general linear modeling (GLM). Regressors were constructed by convolving Friston et al's (1994) canonical hemodynamic response with waveforms representing expected cognitive activity. Regressors that accounted for pupil-dilation included a baseline regressor, constructed by convolving a hemodynamic response (sampled at 60 Hz) with a waveform in which the grand mean pupil dilation waveform during stimulus presentation was used at each trial, along with a separate regressor for the pupil dilation during the target portion; condition-related regressors for the 4 and 5 digit conditions were constructed from the grand pupil dilation waveform during stimulus presentation (tables 2 and 3, R4). Together with regressors from the first step these regressors could thus capture differences between the 3, 4, and 5 digit conditions. Individual differences were accounted for by adding analogous regressors constructed from each subject's mean pupil dilation waveform for each condition (R5). Finally trial-by-trial variation was accounted for by including regressors formed from each subject's pupil dilation data at each trial (R5 in table 2 and R6 in table 3). All of these regressors were downsampled to the rate at which scans were acquired using a polyphase conversion. The difference in R^2 between regressors entered on the first and second steps (henceforth, ΔR^2) was used as a measure of the variance in fMRI signal uniquely associated with trial-type.

Sensitivity and specificity analyses compared regressors that accounted for pupil dilation to regressors that accounted only for variables associated with aspects of the task and condition, constructed analogously. To capture trial-related activity, a regressor was constructed from a waveform in which there was a pulse at the beginning of each trial (tables 2 & 3, R1). A second regressor coded the response to the target with a pulse on each trial. In hierarchical analyses, 2 regressors were added that dummy-coded condition; one had a pulse at the beginning of 4-digit trials and the other had a pulse at the beginning of 5 digit trials.

We contrasted this approach with application of regressors informed by the direction and time-course of results from the pupil dilation assessment but which did not actually use the pupil dilation data. As a first constraint, rather than convolving the canonical hemodynamic response with a single pulse, a consecutive series of 3 pulses, or a "boxcar"

regressor, was used to capture activity lasting for the approximate time in which significant differences in pupil dilation were observed in the task. This analysis was examined both assuming a linear increase in the bold signal with trial type (R2; with condition coded as a single regressor), and without assumptions about the relative weights of conditions (R3; condition dummy coded in two regressors).

Results

Results are organized around the three questions described in the introduction.

Behavioral data

Behavioral data are presented in Table 1. No reliable condition-related differences in behavioral data (reaction times, d') were detected. Performance was excellent for all participants across testing sessions.

Table 1 about here
----

1) Is the time-course of cognitive processing similar in and out of the scanner?

To examine the extent to which cognitive processing differed in and out of the fMRI assessment, we compared pupil-dilation at each assessment. Condition-related differences in mean pupil dilation are shown in Figure 1. As shown in the figure, the pattern of results is nearly identical on each administration, suggesting that cognitive processes can be considered comparable in and out of the scanner. Decreased resolution lead to the more jagged appearance of waveforms acquired in the scanner. Both outside and inside the scanner, pupil dilation increased parametrically with the number of presented digits in both amplitude and duration during the period in which digit-sorting was directed to occur (highlighted areas in Figure 1 represent statistically significant differences between the waveforms[1]). As expected, condition-related pupil dilation waveforms did not reveal condition-related differences during the period in which participants matched the middle digit with a target. Specifically, pupil dilation data acquired inside the scanner suggested that participants' cognitive load usually peaked 4-5 seconds after the trial's onset; condition-related differences were reliable from 2.3-7.5 seconds after the trial's onset, $F(2,14)=11.94$, $p<.005$. Similarly, when pupil dilation was acquired outside the scanner cognitive load also peaked 4-5 seconds after the trial's onset. Condition-related differences were reliable from 2.0-9.8 seconds after the trial's onset, $F(2,14)=29.2$, $p<.005$[2]; this general pattern was present for all participants. The low error rates suggested that participants were engaged in the task. Pupil dilation data further revealed that participants were engaged throughout trials. Since initial pupil dilation associated with sorting digits had subsided by the time participants were required to make a behavioral response, we can infer that participants had enough time to sort the digits, and that the process of sorting was generally separated in time from the process of matching the middle digit to something in memory.

Figure 1 about here
----

2) What aspects of cognitive processing on information processing tasks are represented by the BOLD signal?

This question can be answered in two ways. First, the finer time-course available through pupil dilation assessment can be used to understand what aspects of cognitive processing are captured by temporal variation in a single brain region of interest. Second, as pupil dilation waveforms can be constructed for each subject, individual differences can be accounted for in understanding variability in activity throughout the brain in response to information processing tasks.

Using pupil dilation to estimate the time-course of cognitive processing in a single brain region of interest. To demonstrate the utility of pupil dilation in interpreting fMRI data, we created predictions for fMRI activity by convolving the grand mean pupil dilation in each condition, acquired in the scanner, with a canonical hemodynamic response (Friston et al, 1994), downsampled the resulting waveforms to the rate at which scans were acquired, and compared them to activity in the left middle frontal gyrus, extracted using AFNI's (Cox, 1996) Talairach tool (left panel, Figure 2). As shown in this panel, activity in the left middle frontal gyrus was parametric with the number of digits, and differences were significant using repeated-measures ANOVAs on scans from 6-12 seconds, $F(2,14)=5.0$, $p<.02$ for the region. The middle panel of Figure 2 demonstrates predicted fMRI signal when the portion of the pupil dilation waveform associated only with the digit-sorting portion of the scan is used to construct a regressor, $r_{\text{mean signal for each valence}} = .89$, versus the case in which pupil dilation over the entire trial was employed, $r_{\text{mean signal for each valence}} = .73$. These correlations represent the

---

[1] Statistically significant differences between waveforms were computed using Guthrie and Buchwald's (1991) procedure, in which significance is determined based on contiguous chains of time-points that reliably differ. Based on the temporal autocorrelation of pupil dilation waveforms, regions of points that differed at $p=.1$ for 1 second were used to determine regions in which differences were statistically significant. Described regions were further restricted to those in which the overall effect in the region was significant at $p<.05$.

[2] When measured as proportion-of-maximum dilation. When calculations were performed in raw millimeter units the graphs looked nearly identical. Significant condition-related differences were present from 3.0-10.6 seconds, $F(2,14)=27.0$, $p<.005$.

correlation of a waveform constructed of the consecutive means of pupil dilation regressors in each condition with a waveform constructed from the consecutive means of the fMRI signal in each condition. Steiger's (1980) test of dependent correlations confirms that convolution with only the digit-sorting part of the trial better represents the observed signal in the left middle-frontal-gyrus, $t(33)=3.5$, $p<.001$. In effect, the top middle panel suggested that if a brain region responded primarily to the digit sorting portion of a trial it's measured fMRI signal would return to baseline before the next trial; the bottom shows that if it responded to the whole trial, the signal would not return to baseline. Thus, analysis suggested that the digit-sorting portion of the trial yielded a better fit to middle frontal gyrus, potentially implicating this area most in the portion of the task associated with working memory.

Figure 2 about here

Effects of using pupil dilation to tune fMRI whole-brain analyses.

   To examine the extent to which pupil dilation could be used as a predictor of activity throughout the brain, we examined results from a whole-brain voxel-by-voxel regression in which regressors that accounted for variation in the grand mean as well as individual differences in mean pupil dilation per condition were included. Regions were considered significant in which $\Delta R^2>.05$, and accounted for a statistically significant portion of variance $p<.001$, with a 56 voxel contiguity threshold (computed using Cox's (1996) method for controlling type I error at $p=.05$ based on the spatial autocorrelation present in the ΔR^2 map). As shown in Figure 3, three regions of interest (ROI's) were identified comprising 1) a left prefrontal/inferior frontal region including left Middle Frontal Gyrus (centroid of Talairach coordinates: -30,21,26, 134 voxels) and Inferior Frontal Gyrus (–47, 18,7; 42 voxels), $r_{prediction,average\ waveform}=.97$, 2) a right parietal region including Precuneus (25,-67,48; 49 voxels), Superior Parietal Lobule (41,-53,55; 81 voxels), Inferior Parietal Lobule (54,-33,44; 96 voxels), Supermarginal Gyrus (2 centroids: 54,-37,44; 41,-47,38; 13 voxels), Postcentral Gyrus (54,-30,45; 18 voxels), $r_{prediction,average\ waveform}= .96$, and 3) a left parietal region including Precuneus (2 centroids: -21,-57,52; -17,-70,44; 102 voxels) and Superior Parietal Lobule (-24,-67,52; 31 voxels), $r_{prediction,average\ waveform}=.96$. Each demonstrated the expected condition-related variation. Of note, no regions of interest were detected using boxcar regressors at these thresholds, though boxcar regressors did detect middle frontal gyrus and parietal activity using less stringent ΔR^2 thresholds. Thus, using the pupil as a regressor demonstrated sensitivity above and beyond boxcar analyses as well as specificity for activity in plausible regions.

Figure 3 about here
----

3) *Is pupil dilation a sensitive and specific predictor of regional brain activity?*

   The previous analyses suggested that a pupil dilation regressor that accounted for individual differences sensitively and specifically detected brain activity in expected regions of interest. To justify using this procedure it is also useful to know whether using pupil dilation as a regressor provides a more sensitive and specific index than a task-related regressor, and whether it is important to account for individual or trial-by-trial related differences in the analysis. Thus, we formally examined the differential sensitivity and specificity of regressors that did and did not account for pupil dilation.

   Sensitivity of pupil dilation to middle frontal gyrus activity. Sensitivity was demonstrated by examining the incremental variance accounted for by regressors based on pupil dilation over standard analyses in GLM's on fMRI data from the talairach extracted left middle frontal gyrus region (Table 2). Using a pre-defined ROI is important in that it did not rely on a particular style of empirical analysis for selection. As shown in Table 2, explained variance in both the raw data and mean data for each condition increased when pupil-dilation related regressors were included. Boxcar regressors informed by the overall shape of pupil dilation data explained more variance than a single pulse regressor, and the major increase in explained variance came when terms reflecting individual and trial-by-trial differences were included (R5). Of note, a regression in which individual subject variability was modeled using pupil dilation, but in which trial-to-trial related differences were not included explained the most variance in the average waveform ($r_{mean\ signal\ for\ each\ valence}=.99$). This regression explained more variance in the average waveform than did the regression in which trial-to-trial variability was included because trials with outlying pupil dilation did not need to be removed, yielding a more representative estimate of the average MRI signal.

Table 2 about here

   To illustrate the increased sensitivity of including pupil dilation, accounting for individual differences, as a regressor, the right column of Figure 1 shows the predicted average condition-related waveform for a few of the regressors. As shown in the figure, the regression in which a regressor is included for only scan 1does not produce a sustained enough

response to model the data adequately. Using a boxcar predicted a more sustained response. Using pupil dilation predicted an even more sustained response at 5 digits, which better fits the observed data.

Specificity: Unique empirical detection of middle frontal gyrus ROIs. We analyzed specificity in a manner similar to that suggested by Genovese et al (2002), in which a strict threshold was adopted to control for false, or non-predicted detected activations, and surviving activity in areas of theoretical interest were examined. This method controlled for the potential confound that including more regressors must account for more variance, by controlling for variance accounted for in areas of non-interest. Specifically, the following strategy was adopted on whole-brain voxel-by-voxel analyses. A ΔR^2 threshold for considering voxels as varying with condition was increased until fewer than 20 voxels were detected as active in the entire brain excluding a few regions of plausible activation during working memory tasks suggested by similar studies in the literature (middle frontal gyrus, inferior frontal gyrus, precuneus, superior parietal lobule, inferior parietal lobule).[3] Voxels considered active were restricted to those having at least 6 similarly active neighbors and to have significance of p=.001 or better in a t-test of the multiple R against zero. This analysis selected for specificity to areas of theoretical interest.

Table 3 contrasts various measures of effect size for each style of analysis. These include the number of significant detected voxels in a-priori specified plausible regions of functional interest, the number of voxels detected specifically in middle frontal gyrus, and standardized indices of model-fit. As shown in Table 3, incorporating information about pupil dilation into prediction equations increased the number of detected voxels in all plausible regions, particularly the middle frontal gyrus, as well as employed thresholds, relative to other areas of non-interest.

The analysis in which activity was modeled by a convolved pulse (i.e., without consideration of pupil dilation data) revealed no activity in predicted regions when there were 20 voxels active in other regions, suggesting that this analysis had little specificity. Similarly, when the time-course of the fMRI data was accounted for by a boxcar regressor, a few voxels were activated in the expected regions, but none of them were in middle frontal gyrus. In contrast, when the mean pupil dilation regressor was used, more voxels were identified in the expected areas and a large left middle-frontal-gyrus ROI was detected. Incorporating information about individual differences allowed a higher threshold to be employed, indicating even more specificity, and yielding the largest number of detected activations in left middle-frontal gyrus. The highest thresholds were available when regressors including trial-by-trial variation were employed, though these high thresholds decreased the number of detected voxels in areas of theoretical interest relative to regressors with fewer degrees of freedom; in addition, the detected middle frontal gyrus ROI was on the right rather than the left. Once pupil dilation was added, also adding aspects of condition dramatically decreased specificity. All employed regressions that accounted for aspects of pupil dilation accounted for nearly all of the variation in the average fMRI waveform.[4]

Table 3 about here

Discussion

We examined the extent to which pupil dilation was informative for analysis and interpretation of fMRI data. Pupil dilation was used to inform key aspects of fMRI data interpretation. Specifically 1) the time course of cognitive processing, estimated from pupil dilation was demonstrated to be similar inside and outside the scanner on a cognitive task and 2) this time course was used to aid in interpretation and power for fMRI analysis, and 3) pupil dilation was a more sensitive and specific predictor of brain activity than task-condition. Thus we suggest that concurrent acquisition of pupil dilation data during cognitive tasks may aid in understanding and analyzing fMRI data.

Analyses that incorporated information about individual differences in pupil dilation appeared particularly sensitive and specific to relevant brain activity. Incorporating information about trial-to-trial variation in pupil dilation allowed more variance to be explained in the raw data in areas of theoretical interest. Yet, preserving high specificity to areas of theoretical interest required raising the threshold for signal to a level that considerably decreased sensitivity to activity in relevant areas.

In particular, pupil dilation data helped to suggest that middle frontal gyrus activity primarily reflected the part of a working memory task involving sorting digits rather than the part of the task in which the sorted digit is matched to a

---

[3] Standard false-detection-rate FDR procedures (e.g., Genovese et al, 2002) set a threshold based upon a theoretical ratio of false positive activations to the total number of detected activations. Calculations used to estimate the FDR are based on estimates of signal-to-noise throughout the brain. This analysis used a similar logic. Since we had explicit expectations about areas in which activity was expected to occur we could explicitly control the number of false positive activations by controlling the number of activations outside these areas.

[4] Of note, the largest and strongest ROIs detected with all regressions was actually in the precuneus; examination of this region also revealed parametric variation with difficulty. Non-PFC ROIs are not analyzed explicitly because doing so would not have added explanatory power to the paper's primary message regarding the role of pupil dilation in detecting prefrontal activity.

target. Furthermore, it suggested that prefrontal and parietal areas varied in activity with difficulty on the task. These functional observations are consistent with existing theories of prefrontal cortex's involvement in executive and working memory processes (e.g., Jonides, 1993; Fiez et al, 1996; Braver et al, 1997; Rypma & D'Esposito, 1999) as well as observed parietal cortex activity in studies involving a phonological loop (e.g., Na et al, 2000; Chein et al, 2001).

   The examined applications are not necessarily the only ways in which pupil dilation can be useful to the neuroimager. Pupil dilation can also help to detect behavioral anomalies such as participants closing their eyes during stimulus presentation. Pupil dilation may also help to interpret null fMRI findings. Interpretation of null results in fMRI data is particularly difficult, because it is unclear whether the lack of results is due to poorly specified regressors or the absence of expected cognitive load (e.g., because participants were not sufficiently engaged in the task). If pupil dilation varies with task conditions, it is thus expected that brain activity will, as well. Failure to demonstrate changes in brain activity is thus more likely to be due to analytical rather than behavioral factors. Alternately if condition-related differences are found in neither pupil dilation nor fMRI data, a true lack of condition-related differences in brain activity may be supported.

   Pupil dilation may also allow estimation of activity in brain areas difficult to capture with fMRI. Small, deep brain areas, such as the locus coeruleus cannot be easily measured with fMRI. To the extent that pupil dilation captures activity in such areas it may be used as an index of brain activity that complements acquired fMRI-data (Jonathan Cohen, personal communication). For example, fairly direct projections of the locus-coeruleus to the pupil have been observed in animals, and locus coeruleus activity has been suggested for human pupil dilation (Szabadi & Bradshaw, 1996). Because the locus coeruleus is so small, it's activity is difficult to detect using fMRI. Pupil dilation may thus be used to augment fMRI data.

   Alternately, concurrent acquisition of pupil dilation and fMRI data may be turned around to help to explain physiological mechanisms associated with pupil dilation, which is traditionally believed to be a summative measure of brain activity. While methods for relating activity throughout the brain to a single physiological signal are in their infancy the technology is promising. For example, in the current data, from Figure 3, adding the parietal ROIs to the DLPFC roi in a regression on the pupil dilation data increased explained variance by a mean of 6% and by as much as 16% in some subjects suggesting that different brain regions may explain independent variance in pupil dilation. In contrast, their interaction accounted for a mean of 2% more variance than the main effects, and never more than 4% of the variance in any subject supporting the idea that the pupil acts primarily as a summative measure of activity in these regions.

   A number of limitations in the current study can be addressed by future research. Methodological limitations involve the small employed sample, the use of a single task, and the fact that assessment dates were not counterbalanced (i.e., everyone was assessed outside the scanner first, and then inside the scanner). The fixed assessment order could have contributed to the marginally (though not significantly) smaller pupil dilation in the scanner than outside. Technical limitations involved our inability to measure absolute light levels in the scanner (we did not have an MR-compatible light-meter) and absolute units on pupil dilation in the scanner. Thus issues of differential brightness and the resolution, in mm, of pupil responses in the scanner could not be addressed. An analytical limitation is that the obtained full-model R^2's with the raw data were relatively low even when trial-by-trial variation in pupil dilation was accounted for, suggesting that a great deal of observed variance was not explained by the employed regressors. Use of a canonical rather than idiosyncratically empirically generated hemodynamic response could have contributed to this observation. Not being able to obtain an absolute measure of pupil dilation in the scanner is a technical limitation sure to be overcome in the near future; though use of a relative metric did not appear to affect results. Of note, two common limitations of the fMRI design were addressed by acquiring pupil dilation data, including having no behavioral data that reliably discriminated between conditions and having no deviation from alternation between trial parts, which would allow event-related deconvolution.

   It is also useful to consider the potential of pupil dilation, as a measure to be administered *before* fMRI data are acquired. For example, our lab often pilots tasks using pupil dilation as a proxy for brain activity, and tunes task parameters using pupil dilation to maximize expected condition-related differences in brain activity. Similarly, pupil dilation can be used to predict group differences. For example, we have used pupil dilation to establish expected differences in the time course of brain activity in response to emotional stimuli between depressed and control participants (Siegle et al, in press), and only after they were established, collected fMRI data to confirm predictions (Siegle et al, 2002). As such, this paper represents but a first step in using pupil dilation, collected during fMRI to better understand brain mechanisms associated with cognition.

Acknowledgements

Correspondence concerning this article should be addressed to Greg J. Siegle, Ph.D., Western Psychiatric Institute and Clinic, 3811 O'Hara St. Pittsburgh, PA 15213-2593. email: gsiegle@pitt.edu

References

Baddeley, A. 1986. *Working Memory,* Oxford Science Publications, Oxford.

Beatty, J. 1982a. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin* **91**: 276-292.

Beatty, J. 1982b. Phasic not tonic pupillary responses vary with auditory vigilance performance. *Psychophysiology* **19**:167-172.

Blair, J. R., & Spreen, O. 1989. Predicting premorbid IQ: A revision of the National Adult Reading Test. *The Clinical Neuropsychologist* **3**, 129–136.

Braver TS. Cohen JD. Nystrom LE. Jonides J. Smith EE. Noll DC. 1997. A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage* **5***:49-62.*

Brown, G., Kinderman, S., Siegle, G. J., Granholm, E., Wong, E. C., Buxton, R. B. 1999. Brain activation and pupil response during covert performance of the Stroop color word task. *Journal of the International Neuropsychological Society* **5**, 308-319.

Callicott, J.H., Mattay, V.S., Bertolino A., Finn, K., Coppola, R., Frank, J.A., Goldberg, T.E. & Weinberger, D.R. 1999. Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. *Cerebral Cortex* **9**: 20-6.

Carter C.S., Macdonald A.M., Botvinick M., Ross L.L., Stenger V.A., Noll D., Cohen J.D. 2000. Parsing executive processes: strategic vs. evaluative functions of the anterior cingulate cortex. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 1944-8.

Chein, J. M.. Fiez, J. A.. 2001. Dissociation of Verbal Working Memory System Components Using a Delayed Serial Recall Task. *Cerebral Cortex.* **11**:1003-1014.

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J.K., Holyoak, K. J., Gabrieli, J.D.E. 2001. Rostrolateral Prefrontal Cortex Involvement in Relational Integration during Reasoning. *Neuroimage* **14**: 1136-1149.

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

Cox, R.W. 1996. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers in Biomedical Research* **29**: 162-173.

Fiez, J.A., Raife, E.A., Balota, D.A., Scharz, J.P., Raichle, M.E. & Petersen, S.E. 1996. A positron emission tomography study of the short-term maintenance of verbal information. *Journal of Neuroscience* **16:** 808-822.

First, M.B., Spitzer, R.L., Gibbon, M. and Williams, J.B.W. 1997. *Structured Clinical Interview for DSM-IV Axis I Disorders-Clinician Version (SCID-CV)*, American Psychiatric Press, Washington, DC.

Friston KJ, Jezzard P, Turner R 1994. *Human Brain Mapping* **1**:153-171.

Genovese C. R., Lazar, N.A., and Nichols, T. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* **15**: 870–878.

Granholm, E., Asarnow, R.F., Sarkin, A.J., & Dykes, K.L. 1996. Pupillary responses index cognitive resource limitations. *Psychophysiology* **33***:457-461.

Guthrie, D. & Buchwald, J. 1991. Significance testing of difference potentials, *Psychophysiology* **28***, 240-244.

Hartwell, J.W. 1995. *EEGSYS Version 5.7 User's Guide.* Baltimore, MD: Friends Medical Science Research Center, Inc.

Jonides, J., Smith, E.E., Koeppe, R.A., Awh, E., Minoshima, S., Mintun, M.A. (1993). Spatial working memory in humans as revealed by PET. *Nature* **363**: 623-625.

Kahneman D, Beatty J. 1966. Pupil diameter and load on memory. *Science* **154**: 1583-1585.

MacDonald, M.C., Almor, A., Henderson, V.W., Kempler, D. & Andersen, E.S. 2001. Assessing Working Memory and Language Comprehension in Alzheimer's Disease. *Brain and Language* **78**: 17–42.

Piaget, J. 1965. *The child's conception of number.* Norton, New York.

Ratcliff R 1993: Methods for dealing with reaction time outliers. *Psychological Bulletin* **114**:510-532.

Na, D.G., Ryu, J.W., Byun, H.S., Choi, D.S., Lee, E.J., Chung, W.I., Cho, J.M., Han, B.K., 2000, Functional MR Imaging of Working Memory in the Human Brain, *Korean Journal of Radiology,* **1**:19-24

Rypma, B. & D'Esposito, M. 1999. The roles of prefrontal brain regions in components of working memory: Effects of memory load and individual differences. *Proceedings of the National Academy of Sciences* **96***:* 6558-6563.

Rypma,B. Prabhakaran,V., Desmond, J.E., Glover, G.H. & Gabrieli, J.D.E., 1999. Load-Dependent Roles of Frontal Brain Regions in the Maintenance of Working Memory *NeuroImage* **9**: 216–226.

Siegle, G. J., Granholm, E., Ingram, R. E., & Matt, G. E. 2001. Pupillary response and reaction time measures of sustained processing of negative information in depression. *Biological Psychiatry* **49***:* 624-636.

Siegle, G. J., Steinhauer, S.R., Carter, C. S., Thase, M.E. in press. Do the seconds turn into hours? Relationships between sustained pupil dilation in response to emotional information and self-reported rumination. *Cognitive Therapy and Research.*

Siegle, G. J., Steinhauer, S.R., Thase, M.E., Stenger, V.A., Carter, C. S., 2002. Can't Shake that Feeling: fMRI Assessment of Sustained Amygdala Activity in Response to Emotional Information in Depressed Individuals. *Biological Psychiatry* **51***: 693-707.*

Steiger, J.H. 1980. Tests for comparing elements of a correlation matrix. *Psychological Bulletin* **87***:* 245-251.

Steinhauer S.R. & Hakerem G. 1992: The pupillary response in cognitive psychophysiology and schizophrenia. In *Psychophysiology and Experimental Psychopathology: A Tribute to Samuel Sutton*. (D. Friedman & G. E. Bruder, Eds), pp. 182-204. New York Academy of Sciences, New York.

Steinhauer, S.R., Condray, R., Kasparek A. 2000. Cognitive modulation of midbrain function: task-induced reduction of the pupillary light reflex. *International Journal of Psychophysiology* **39**: 21-30.

Szabadi, E. & Bradshaw, C. M. 1996. Autonomic pharmacology of α_2-adrenoceptors. *Journal of Psychopharmacology* **10, S.3**: 6-18.

Woods, R, Cherry, S, & Mazzoitta, J 1992: Rapid automated algorithm for aligning and reslicing PET images. *Journal of Computer Assisted Tomography,* **16**: 620-633.

Table 1
Behavioral data for digit sorting task at each assessment. Reported statistics are for multivariate repeated measures ANOVAs

	Reaction times			Signal Detection % Correct			Signal Detection d'		
	Digits	M	SD	Digits	M	SD	Digits	M	SD
Pupil day 1 (n=16)	3	1002.91	245.45	3	0.98	0.05	3	5.59	0.90
	4	997.10	246.94	4	0.94	0.12	4	4.85	1.79
	5	1045.51	320.69	5	0.91	0.15	5	4.25	2.35
	$F_{(2,14)}=0.61$, p=0.55			$F_{(2,14)}=1.83$, p=0.20			$F_{(2,14)}=2.27$, p=0.14		
FMRI (n=16)	3	882.73	334.43	3	0.92	0.10	3	4.65	1.54
	4	863.24	332.22	4	0.94	0.09	4	4.91	1.62
	5	859.45	314.60	5	0.91	0.15	5	4.51	2.94
	$F_{(2,14)}=0.70$, p=0.51			$F_{(2,14)}=0.64$, p=0.54			$F_{(2,9)}=0.56$, p=0.58		

Table 2:

Sensitivity analysis: Effect magnitude for GLM's in Talairach extracted left middle frontal gyrus on the digit sorting task.

	# of regressors (Restricted / Full model)	Full model R^2	ΔR^2*	$r_{mean\ signal\ for}$ each valence
R1: convolved pulse at scan 1	2/4	.04	.02	.75
R2: convolved boxcar assuming linear effect	2/4	.05	.02	.96
R3: convolved pulse at s1 + boxcar	3/7	.06	.04	.97
R4: convolved mean pupil waveform	2/4	.06	.03	.87
R5: R4+ individual and trial-by-trial differences in pupil dilation	4/9	.15	.11	.94
R3+R5	7/14	.17	.13	.96

*ΔR^2 represents the change for inclusion of condition-related regressors from a model including only analogous every-trial regressors.

Table 3:

Specificity analysis: Effect magnitude in empirically detected regions for GLM's at ΔR^2 (change for condition-related regressors from analogous every-trial regressors) threshold for which <20 voxels were active in areas of noninterest, and for which p<.001, contiguity >5 voxels.

	# regressors Restricted / Full model	ΔR^2 threshold for <10 voxels detected in areas of non-interest	# active voxels in areas of theoretical interest	# active voxels in middle frontal gyrus	ΔR^2 for middle frontal gyrus ROI	$r_{mean\ signal\ for}$ each valence, middle frontal gyrus ROI
R1: convolved pulse at scan 1	2/4	.04	0	0	N/A	N/A
R2: convolved boxcar assuming linear effect	2/4	.03	7	0	N/A	N/A
R3: convolved pulse at s1 and boxcar	3/7	.05	36	0	N/A	N/A
R4: convolved mean pupil waveform	2/4	.04	69	24	.06	.93
R5: R4 + individual differences in pupil dilation	4/8	.06	104	40	.07	.98
R6: R5+ trial-by-trial variation in pupil dilation	4/9	.13	15	6	.11	.93
R3+R5	7/13	.09	2	0	N/A	N/A
R3+R6	7/14	.15	20	7	.14	.97

Note: All GLM ΔR^2 effects represent variance accounted for by a regressor that accounts for condition-related differences above and beyond a regressor representing a hemodynamic response for each trial.

Figure 1.
Pupil dilation during a digit sorting task, in the same participants, outside the fMRI environment and during fMRI acquisition session (n=16). Lightly shaded areas represent significant differences between the waveforms, p<.1; darkly shaded areas represent significant differences, p<.05, determined using repeated measures ANOVAs at each time-point. Pupil dilation reliably increased with difficulty during the digit-sorting portion of the task.
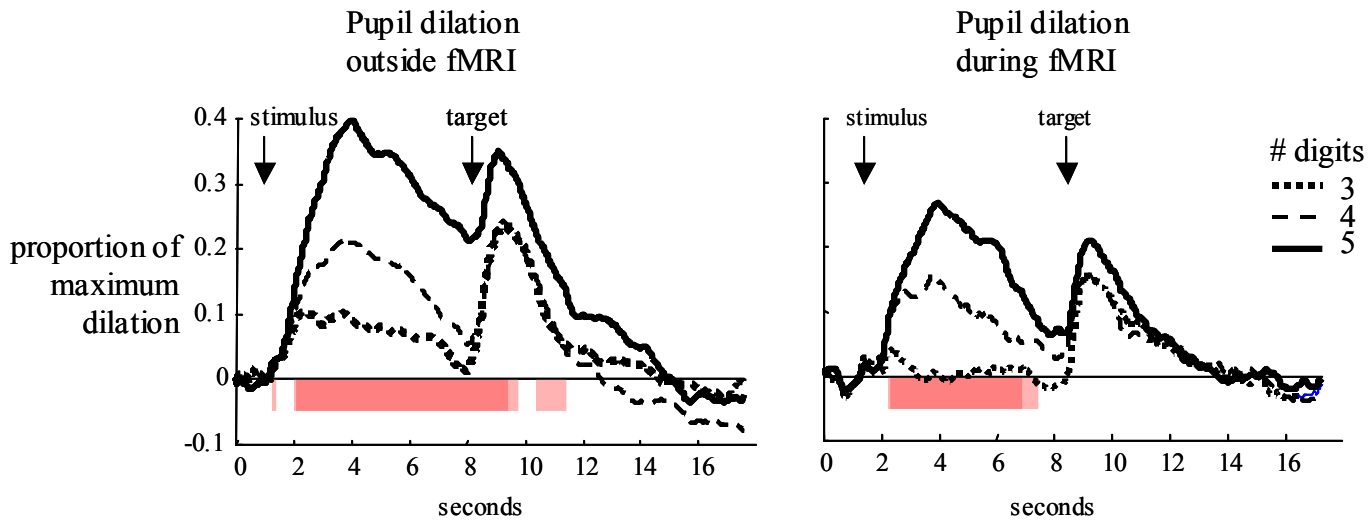
Figure 2.
Left: Mean condition-related activity in Talairach extracted middle frontal gyrus. Scans occurred every 1.5 seconds. Shaded areas represent significant differences, p<.05, determined using repeated measures ANOVAs at each scan. Middle: Predicted fMRI activity derived by convolving pupil dilation with a canonical hemodynamic response (Friston, 1994), either for just the digit sorting portion of the trial (a) or during the whole trial (b). A better fit is obtained when just the digit-sorting portion of the trial is used. Right: Predictions for brain activity based on regressors in which brain activity only occurred at scan 1, uniformly at scans 1-3, and proportional to each participant's mean pupil dilation for each condition.
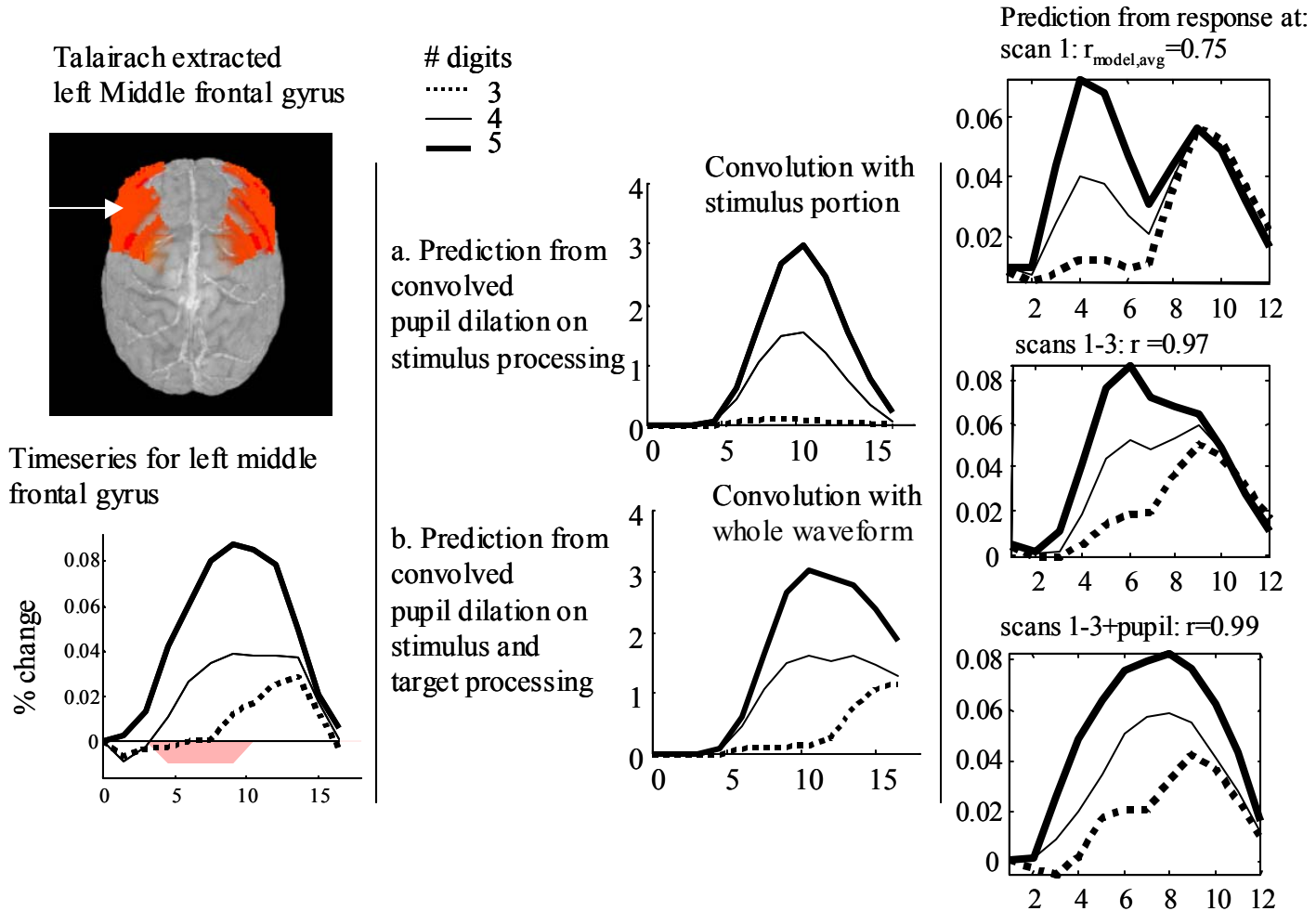
Figure 3.

Whole brain voxel-by-voxel GLM analysis of pupil dilation (regressors accounting for mean variation as well as for individual differences) on fMRI activity, $\Delta R^2 > ..05$, $p<.001$, contiguity = 56 voxels. Prefrontal and parietal activity emerged; timeseries for all areas suggested increased activity with the number of digits to be sorted. Shaded areas represent significant differences, $p<.05$, determined using repeated measures ANOVAs at each scan.