Current Biology, Volume *23*

Supplemental Information

Temporal Jitter of the BOLD Signal Reveals a Reliable Initial Dip and Improved Spatial Resolution

Masataka Watanabe, Andreas Bartels, Jakob H. Macke, Yusuke Murayama, and Nikos K. Logothetis

Supplemental Inventory

Figure S1. The temporal jitter of the positive BOLD has a biological origin, related to Figure 1
Figure S2. Jitter-uncovered initial dip in motor areas (supplemental experiment 1), related to Figure 2
Figure S3. Voxel population analysis of the jitter-uncovered initial dip with complementary wedge stimuli (supplemental experiment 2), related to Figure 3
Supplemental Experimental Procedures
Supplemental References

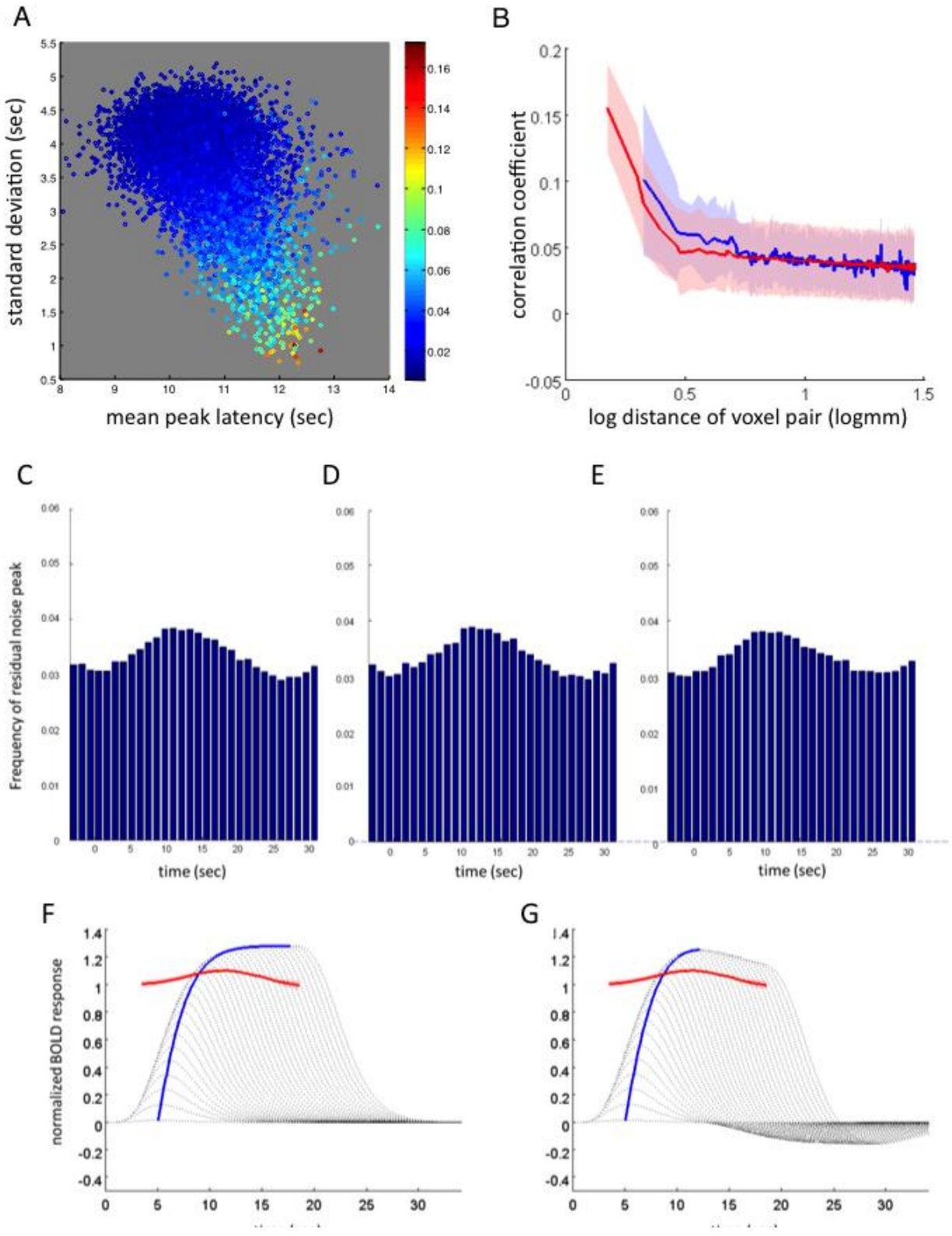
A
standard deviation (sec)
mean peak latency (sec)
B
correlation coefficient
log distance of voxel pair (logmm)
C
D
E
Frequency of residual noise peak
time (sec)
time (sec)
time (sec)
F
G
normalized BOLD response

Figure S1.

Figure S1. The Temporal Jitter of the Positive BOLD Has a Biological Origin, Related to Figure 1

(A) Negative correlation between mean BOLD peak latencies and standard deviation of BOLD peak latencies across trials, plotted for all visually activated voxels (T-scores higher than 4 using SPM5 GLM analysis) from all subjects. The negative correlation is statistically significant for all subjects. The marker color denotes percent signal change of the trial-averaged BOLD response peak.

(B) Trial by trial correlation of positive component peak latency for contra-lateral (blue) and ipsi-lateral (red) voxel pairs. The correlation decreases steeply within a distance below 1 cm and reaches a plateau of significantly positive values for larger distances. Hatched regions denote 95% confidence interval.

(C-E) Temporal distributions of residual response peak for three representative subjects.

(F and G) Comparison between simulated and observed BOLD on the relation between peak heights and peak latency. Blue line denotes simulated bold (F:'boynton', G:'two-gamma') whereas red line and hashed region denote observed mean and confidence interval for observed BOLD.

To test whether the temporal jitter has a biological origin, we first examined dependencies between a given voxels' mean peak latency, latency variance, and mean peak amplitude, as all would be commonly affected by the voxels' overlap with capillaries versus downstream venules. Second, we tested for similarities between voxel properties as a function of their spatial distance to examine influences of local processing versus global effects. Third, we rule out a potential alternative account for the temporal jitter, namely that it arises due to the superposition of auto-correlated (or non-auto-correlated) noise to a voxel specific temporally fixed BOLD time-course, with analysis of residual noise in the BOLD response. Finally, the possibility that the trial-to-trial temporal jitter is simply a byproduct of trial-to-trial variability in the peak magnitude of the evoked BOLD response is ruled out by comparing peak-height vs. peak-latency relationships in observed and simulated BOLD signals. The following sections refer to the figure panels (A-G) of supplemental figure S1.

(A) As a first attempt to investigate the origin of the detected temporal jitter, we plotted the standard deviations (SDs) as a function of mean peak latency and positive peak amplitude (result of a representative subject shown in Figure S1A). A marker corresponds to a visually evoked voxel and its color denotes the mean amplitude of the positive peak across trials. We find a significant negative correlation between the SD of the peak latency and the mean latency ($P < 0.001$ for all subjects), and also between SD of the peak latency and positive peak amplitude ($P < 0.005$ for all subjects). That is to say, voxels with larger overall peak latency had lower variability in their latencies and higher peak amplitudes. The most plausible explanation for this is the differential overlap of distinct voxels with draining veins. Voxels overlapping with draining vessels are known to have larger signal change in gradient-echo fMRI. Since they are downstream from other voxels in terms of vascular hierarchy, their overall peak latency is expected to be long, and since they reflect the average response from many upstream voxels, also a reduced variability in latencies is expected according to the rule of large numbers.

(B) The considerable variability of temporal delays across trials raises the question to which extent this variability originated from local, voxel-specific mechanisms or from more

global, spatially extensive or brain-wide sources. To examine this, we took pairs of voxels with varying cortical distances and correlated their peak latencies across all trials. Voxel pairs were taken either from the same hemisphere or from opposite hemispheres. Figure S1B shows that correlations were generally low, indicating that voxel-specific mechanisms dominated the temporal jitter. However, there was also a clear influence of spatial distance as both the contra-lateral and ipsi-lateral pairs decreased their correlations with increasing distance, reaching a plateau around 10mm where correlations remained constant at a significant positive value independent of the distance. These results imply the existence of relatively strong local and of weaker global mechanisms that influence temporal fluctuation, which is consistent with influences of local neural processing on temporal jitter, and with the horizontally connected architecture of cortical vasculature.

(C-E) An additional analysis was conducted to test for an alternative explanation of the observed temporal jitter of the BOLD peak latency; superposition of auto-correlated noise [S1] or uncorrelated noise on a voxel specific temporally fixed BOLD time-course. In the presence of additive noise, there would still be observed jitter in the detected peaks, but it would be explainable by the conventional way of looking at fMRI data, namely models which are either explicitly or implicitly based on linear systems with additive noise. In order to test for this alternative explanation, we extracted the trial-to-trial noise components and analyzed its temporal properties.

We first calculated the trial-averaged BOLD response independently for all visually evoked voxels and subtracted it from individual trial responses to obtain time-courses of residual noise. Next, as in the procedure of peak latency detection, we applied a Gaussian filter (tau=6s) to temporally smooth the residual noise and detected the latency of the positive peak. If the alternative explanation, additive noise on a voxel specific fixed time-course is valid, the temporal distribution of the detected peaks should be uniform, since, by definition, the timing of residual noise should be independent of the stimulus evoked BOLD response. On the other hand, if the BOLD peak latency possesses a genuine temporal jitter, apart from the effect of additive auto-correlated noise, the detected peak latency should have a distinct distribution, namely, a peak at the timing of the original BOLD response peak latency.

Temporal distributions of this residual response peak are given in figure S1 C-E for three representative subjects. Here the results from all visually evoked voxels are concatenated into one histogram. The positive peak around the timing of the original BOLD response peak indicates that the alternative explanation can be disclaimed. For all 7 subjects, likewise, there was a clear positive peak at the timing of the original BOLD peak. To test for statistical significance independently for all subjects, we calculated the variance of the residual noise peak latency distribution independently for all voxels. Then we conducted a t-test on the obtained values of variance (number of voxels) to test if they are smaller than the variance values (number of voxels) obtained from an uniform distribution with identical sample size (number of trials). For all 7 subjects, the actual variance values proved to be significantly smaller ($P < 1E-8$) then that derived from a uniform distribution.

(F and G) To investigate the effect of neural response variability on the temporal jitter of the BOLD positive peak, we further analyzed data from our first main experiment and compared it with two types of simulated BOLD responses. Dependencies of simulated BOLD peak latency on the magnitude of the neural response were calculated by convolving

a standard hemodynamic response function (Figure S1 F:'boynton', G:'two-gamma') with box-car functions of variable duration, under the assumption that a fixed duration sensory stimulus may result in neural activity with variable temporal profiles (e.g. altered balance between phasic and tonic activity), and in the extreme case, with variable duration. Note that when the duration of the box-car functions are kept constant and only the height is altered, peak latency of simulated BOLD remains constant. The simulated BOLD peak latency and the corresponding peak heights are plotted as blue lines (Figure S1 F,G).

Next, we analyzed the above relation in experimentally observed BOLD responses. Firstly, for all visually evoked voxels, single trial responses were pooled dependent on peak latency. Next, voxel-averaged peak heights and corresponding peak latencies for each subject were obtained by normalization of evoked BOLD amplitude. Finally, subject mean and its confidence interval of peak height were calculated and overlaid in figure S1 F,G as red lines and hashed regions, respectively. Although there exists some non-monotonic dependency of peak amplitude on peak latency, it is qualitatively and quantitatively very different from that of the simulated BOLD. In case of the simulated BOLD, regardless of the hemodynamic response function in use, the peak height only monotonically increases with peak latency, and more importantly, the changes are much greater compared to the experimentally obtained BOLD response. In case of the more biologically plausible hemodynamic response function (two-gamma), the width of peak latency variability is limited and does not reach that of the experimentally obtained BOLD response. Therefore we conclude that the effect of evoked neural response variability on the trial-to-trial temporal jitter of the observed BOLD is minimal, if any.

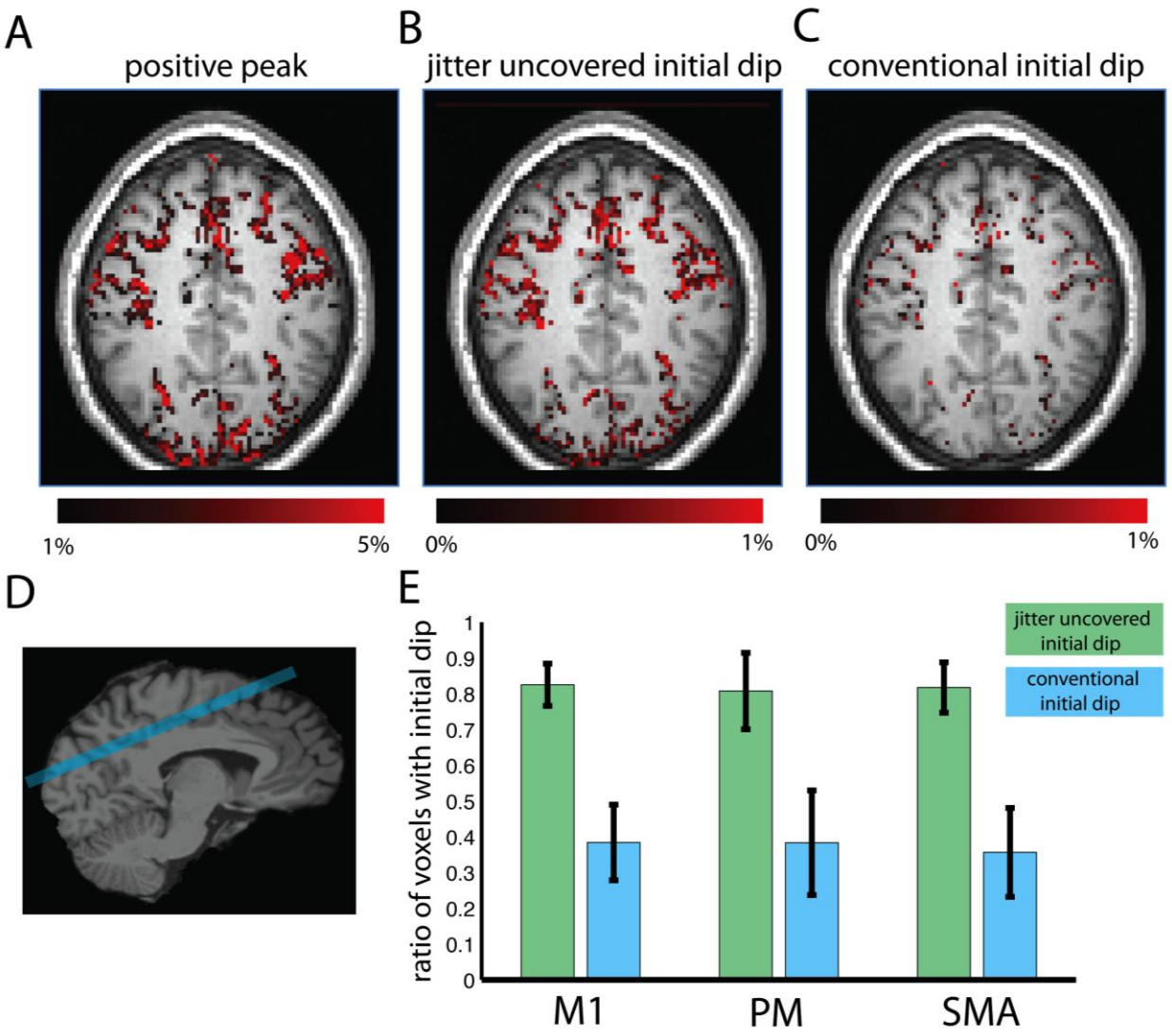

Figure S2. Jitter-Uncovered Initial Dip in Motor Areas, Related to Figure 2

Comparison of the BOLD positive peak, the jitter-uncovered initial dip and the conventional dip with a visuomotor task.

(A-C) Activation obtained by the BOLD positive peak (A), jitter-uncovered initial dip (B) and the conventional initial dip (C).

(D) Scan slice adjusted so as to include the primary motor, premotor and the supplemental motor area.

(E) Ratio of voxels showing the jitter-uncovered initial dip and the conventional initial dip, relative to significantly activated voxels as determined using a SPM GLM analysis based on the standard positive BOLD response ($p<0.05$, FWE corrected). Error bars denote 95% confidence interval.

We conducted an experiment (supplemental experiment 1) to further confirm that the jitter-uncovered initial dip is robust compared to the conventional initial dip, focusing on areas outside the visual cortex. The experimental design was identical to that of the first

main experiment using a radial checker stimulus except that the subjects were asked to perform a complex finger tapping task during visual presentation which lasted 4 seconds (see Supplemental Experimental Procedure). Figures S2A, S2B and S2C show activation maps from a slice including primary motor, secondary motor and the Supplemental motor area. Three types of activity indices, the positive BOLD signal, the jitter-uncovered initial dip and the conventional initial dip, were calculated as defined in the previous sections.

Jitter-uncovered initial dips can be seen in the majority of voxels that elicit positive BOLD (Figure S2B). In contrast, voxels showing conventional initial dips are drastically sparse (Figure S2C). This observation is confirmed by subject averaged ratios of voxels showing the two types of initial dips among significantly activated voxels (SPM GLM analysis using box-car regressors: $P < 0.05$ corrected for multiple comparisons based on Family-Wise Error (FWE)). The differences in ratio between two types of activity are statistically significant for all three areas, the primary motor, the premotor and the Supplemental motor area ($P < 0.001$ t-test). These results show that the proposed method of calculating the jitter-uncovered initial dip is several times more robust compared to the identification of the conventional dip and that the method generalizes to regions beyond the visual cortex. Together, the present results reveal a striking property of the positive BOLD signal, that both accounts for the elusive nature of the conventional initial dip and provides a robust new way to increase spatial resolution based on the jitter-uncovered initial dip.

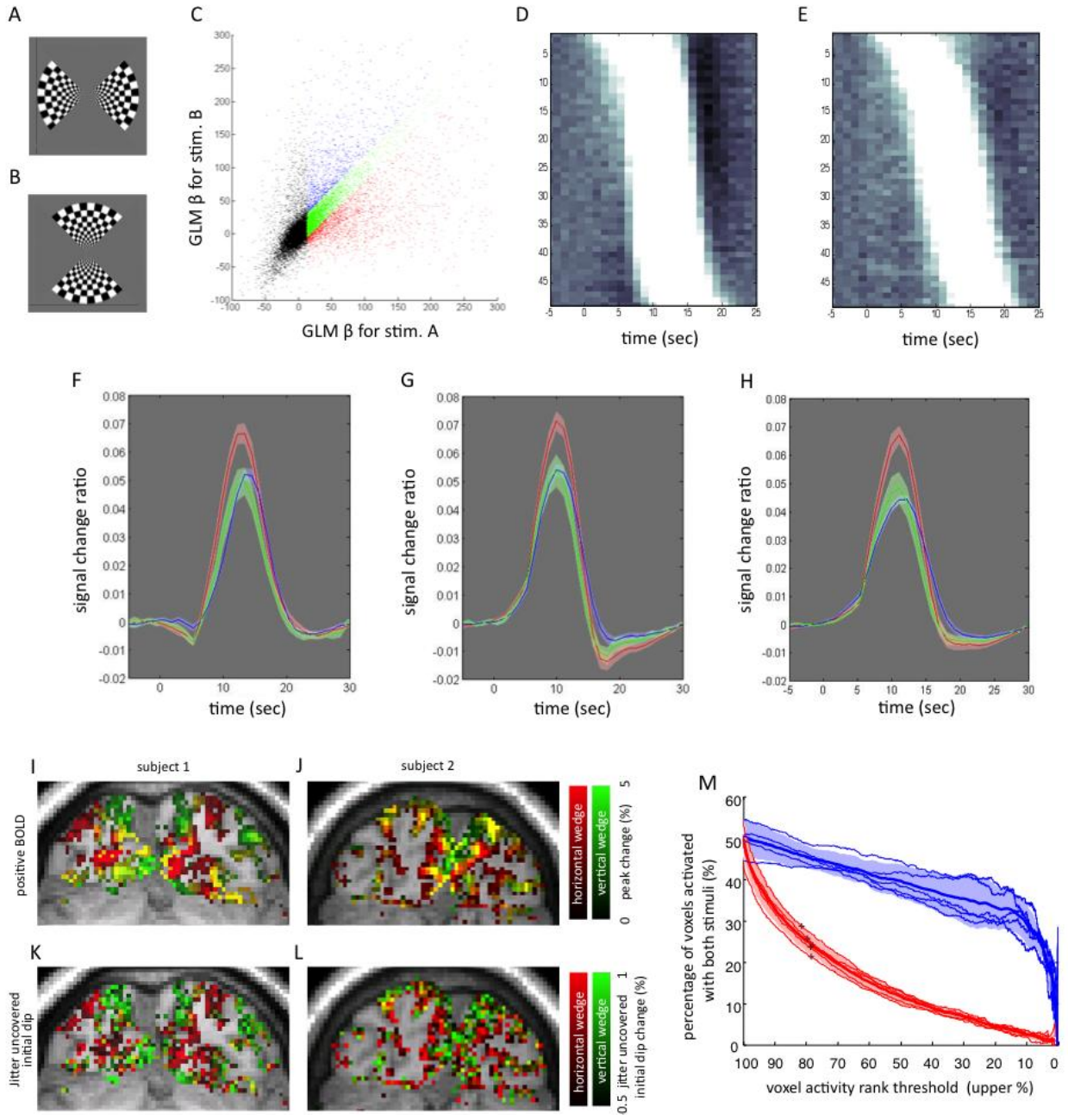

Figure S3. Voxel Population Analysis of the Jitter-Uncovered Initial Dip with Complementary Wedge Stimuli, Related to Figure 3

Stimulus configuration and results of the voxel population analysis.

(A and B) Complementary visual stimuli A and B that were shown (flickering at 4Hz) during 4 sec each in supplemental experiment 2, separated by 24-28 sec blank.

(C) Illustration of the distinct voxel population pools that differed in their responses to stimuli A or B based on a GLM analysis. Each dot represents the beta values for stimulus A and B for a given voxel. Red: core and spread voxels for stimulus A and B respectively. Blue:

core and spread voxels for stimulus B and A respectively. Green: neutral voxels not used for analysis.

(D and E) Single-trial responses of the "core" (D) and "spread" (E) response pool, sorted according to positive peak times.

(F-H) Average responses of late 50% percentile (F), early 50% percentile (G), and all (H) voxel responses, plotted for the core response pool (red line), spread response pool (blue line), and for core response pool voxels whose peak-amplitudes did not exceed those of the spread pool (green line). Only core response pool voxels with latency-sorted trials showed significant initial dip or post-stimulus-undershoot. Hatched regions denote 95% confidence interval.

(I-L) Maps of visually responsive voxels from two representative subjects. (I,J) activation maps based on the analysis of the conventional BOLD positive peak (K,L) maps based on the jitter-uncovered initial dip analysis. Green and red indicate percent signal change related to the two wedge stimuli of (A) and (B). Yellow indicates overlapping activity for both stimuli.

(M) Percentage of voxels with overlapping response for the two complementary stimuli as a function of the voxel activity rank threshold. Analysis based on positive BOLD signal led to higher fractions of overlapping voxels (blue) compared to analysis based on the initial dip (red). Thin lines show individual subject results, thick lines group averages. Hatched regions denote the 95% confidence interval. Crosses indicate the zero point of the jitter-uncovered initial dip.

To provide further evidence that the jitter-uncovered initial dip leads to higher spatial resolution, we conducted an additional experiment (supplemental experiment 2) that allowed for a voxel population analysis. This experiment went beyond main experiment 2 in that we used fMRI slice orientations that were perpendicular to the calcarine sulcus and we used two wedge stimuli. Together, this allowed us to observe an increased number of stimulus activation borders including ones beyond the primary visual cortex.

We used two spatially complementary checker wedge stimuli with common edges (Figure S3 A, B). The basic idea was - as in main experiment 2 - to test whether analysis of voxel responses based on the jitter-uncovered dip would reveal less spatial overlap between responses to the two wedge stimuli than traditional analysis based on the positive BOLD signal. At the neural level, we expected each checker wedge to induce spatially largely segregated neural activity, even though there would also be limited overlap due to the finite size of receptive fields and distortions in retinotopy at finer spatial scales [S2]. However, the overlap of neural activity would be much smaller than that of the positive BOLD response that is on the order of few millimeters [S3]. Intrinsic optical imaging in non-human primates suggests that the spatial spread of tissue and veins/venules oxygenation (which corresponds to the BOLD positive peak) is several times larger than that of the initial dip (which should be slightly larger than the increased local neural activity) [S4]. The experimental setup and analyses were identical to those of the first main experiment.

First, we identified voxels that were significantly activated by both stimuli (i.e. overlapping responses, $p<0.05$, FWE corrected). From these, we defined two sub-populations: 1) *horizontal wedge core voxels* that exhibited significantly larger ($P < 0.05$, FWE corrected) BOLD activity for the horizontal wedge and 2) *vertical wedge core voxels*

that exhibited significantly larger ($P < 0.05$, FWE corrected) activity for the vertical wedge. The motivation for this lies in a hypothetical border between activation sites corresponding to the two stimuli (Figure S3C). Actual estimation and use of the stimulation border were performed in main experiment 2. Figure S3D illustrates the two voxel populations as a scatter plot in beta value space for the horizontal and the vertical wedges (standard SPM5 GLM analysis). Using the set of voxels of both sub-populations, we align each voxel's responses in two different ways, leading to two pools of voxel time-series; 1) *core response pool* which are the horizontal wedge core voxel responses aligned to the onsets of the horizontal wedge stimuli and vertical wedge core voxel responses aligned to the onsets of the vertical wedge stimuli; 2) *spread response pool* which are responses of the horizontal wedge core voxels aligned to the onsets of the vertical wedge stimuli and vertical wedge core voxel responses aligned to the onsets of the horizontal wedge stimuli respectively.

In the next step, BOLD peak latencies were determined for individual responses and sorted by their values. Figures S3D and S3E show the sorted responses of a representative subject for the core and spread response pools, respectively, each row indicating an average time course of 124 BOLD time courses. In the core response pool an initial dip is clearly visible in the late peak latency responses and a sharp post-stimulus undershoot is equally clearly visible in the short peak latency responses (Figure S3D). In contrast to this, the spread response pool does not reveal either of the two characteristics (Figure S3E). Figures S3F and S3G show the subject averaged time courses of the late and early 50 percentile peak latency responses respectively. The jitter-uncovered initial dip proves to be statistically significant ($p < 0.01$) compared to the pre-stimulus baseline only in the case of the core response pool (Figure S3F red line) but not for the spread response pool (blue line). To control for the difference in positive peak amplitude between the core and spread response pools, we further constructed a response pool based on the core response pool but excluding voxels with GLM Beta values higher than 60 (green line), thus matching the amplitudes of the spread response pool. Also here the average response of the late 50 percentile peak latency had a significant initial dip ($p < 0.01$) (Figure S3F). Equally, the characteristic sharp post-stimulus undershoot was statistically significant ($p <0.01$) relative to the baseline BOLD level (chosen here at 22 seconds after stimulus onset where it returned to baseline) only for the core response pool (with and without amplitude correction: red and green lines) but not for the spread response pool (blue line)(Figure S3G). However, when all voxel-trials were averaged (without sorting according to core or spread pools, nor to latencies), neither the initial dip nor the sharp post-undershoot could be observed (Figure S3H). Moreover, the amplitudes of the conventional post-stimulus undershoot (i.e. the minimum value after the positive peak) was not statistically different between the spread response pool and the core response pool (with and without amplitude correction). These results demonstrate that the initial dip as well as the sharp response undershoot can only be observed in voxels that are neurally involved in stimulus processing, and not in fringe voxels that may have a significant but non-specific response to a stimulus, in the sense that the BOLD response of the latter is likely due to vascular spread beyond sites of neural involvement. Our results thus support the notion that the initial dip (but also the sharp response undershoot) is spatially tightly coupled to the site of neural metabolism, while the positive BOLD response is comparably spatially more spread-out.

Finally, we performed two analyses of the same visually responsive voxels in order to test whether using jitter-uncovered initial dips would yield higher spatial accuracy compared to using the conventional positive response. Voxels that showed significant visual activity (T-score > 4) for either or both of the stimuli were used for analyses. As previously mentioned, the magnitude of the jitter-uncovered initial dip was defined as the percent signal difference between 3 seconds prior and 3 seconds after stimulus onset using the 50% of trials with longer positive peak latencies (Fig 2B). The positive signal change was defined as the peak signal amplitude of trial averaged BOLD time courses. Color-coded activity maps for both analyses are shown for two representative subjects in figures S3I-L. Figures S3I and S3J show the conventional positive BOLD signal change map, while figures S3K and S3L show the jitter-uncovered initial dip map, for the same set of voxels as specified above. The color scales in red and green correspond to activity evoked from the horizontal (Figure S3A) and vertical wedge (Figure S3B) respectively. Additive color codes are used so that voxels that were activated by both stimuli are shown in yellow. It is evident that the jitter-uncovered initial dip map shows considerably less voxels that are shared by the two stimuli in comparison to the positive percent signal change map. To verify that this was not the result of higher statistical power of the positive signal change map, we quantified the fraction of overlapping voxels across a range of statistical thresholds for both methods. We calculated the ratio of voxels that were shared by the two stimuli relative to all visually activated voxels, for a range of systematically altered activity rank thresholds of voxels to be included in the analysis (see Experimental Procedures). The jitter-uncovered initial dip proved to have a considerably lower ratio of commonly activated voxels for all subjects ($P < 0.001$ for ranges 95% to 3% of voxels used for analysis, see figure S3M). These results indicate that the spread of activity is considerably smaller when assessed using the jitter-uncovered initial dip. In main experiment 2, we estimated the activity border of two stimuli in order to compare the point spread functions of activity measured in terms of the jitter-uncovered initial dip or in terms of positive BOLD signal, respectively.

Supplemental Experimental Procedures

Experiment 1

Stimuli: Stimuli were presented at a resolution of 1280 x 1024 pixels and at a screen refresh rate of 60 Hz at a viewing distance of 82 cm. Stimuli were projected onto a translucent screen at the end of the scanner bore and viewed through a tilted mirror fixed to the head coil.

A polar-transformed checkerboard (100% contrast, width= 18 deg) flickering (contrast-inversing) at 4Hz (Figure 1A) was presented during 4 seconds, followed by a random inter stimulus interval of 40 to 46 seconds during which the monitor was isoluminant grey apart from the central fixation task described below. Stimulus onsets were synchronized with fMRI image acquisitions. To ensure fixation and alertness, subjects continuously performed a central character detection task (throughout the scanning session) in which a pseudo-random sequence of white characters was presented on the fixation marker at 10 Hz. Subjects had to report occurrences of the target letter 'f' by button press. The detection rate was above 75 % for all subjects. The subjects were scanned for 10 runs of 8 minutes each resulting in 100 visual presentations.

Experiment 2

Stimuli: The exact eccentricities of the stimuli were adjusted so that their shared activation was centered on the imaged flat gray matter region (see below). To align scan slices tangential to the calcarine sulcus, a T1-weighted anatomical scan was performed before the functional scan. We collected 12 slices (slice thickness 2 mm) oriented tangential to individual subject's calcarine sulcus where the 6th and 7th slice were aligned to the upper and lower bank of the sulcus, using an interleaved sequence and the following parameters: volume repetition time (TR) 1.17 s, echo time (TE) 35 ms, 128 x 128 matrix, and voxel size 1.5 x 1.5 x 2.0 mm.

Supplemental Experiment 1: Jitter-Uncovered Initial Dip in Motor Areas

This experiment was identical to main experiment 1, with the following differences. Subjects were six healthy right-handed volunteers (three males and three females), aged 24 ± 4 years. Subjects were instructed to conduct a complex bilateral finger tapping task during presentation of the polar-transformed checkerboard (Figure 1A). The task consisted of bilateral pressing of buttons with fingers in the following order, 1) index finger 2) ring finger 3) middle finger 4) little finger. Subjects were asked to perform them as fast as possible while maintaining accuracy. The inter-stimulus interval was set randomly between 24 and 28 seconds. Subjects were asked to maintain fixation on the flow of characters at the center of the screen without a central fixation task. Scan parameters: volume repetition time (TR) 1.28 s, echo time (TE) 35 ms, 96 x 96 matrix, and voxel size 2 x 2 x 2.0 mm.

Supplemental Experiment 2: Voxel Population Analysis of the Jitter-Uncovered Initial Dip with Complementary Wedge Stimuli

This experiment was identical to main experiment 1, with the following differences. Subjects were five healthy right-handed volunteers (three males and two females), aged 25

± 4 years. Stimuli consisted of two sets of polar-transformed checkerboard wedges (100% contrast, width= 18 deg) flickering (contrast-inversing) at 4Hz (Figure S3 A, B), presented for 4 seconds, followed by a random inter stimulus interval of 24 to 28 seconds during which the monitor was isoluminant grey apart from the central fixation task described in experiment 1. fMRI data analysis was identical to that of the first main experiment.

Supplemental References

S1. Lund, T.E., Madsen, K.H., Sidaros, K., Luo, W.L., and Nichols, T.E. (2006). Non-white noise in fMRI: does modelling have an impact? Neuroimage *29*, 54-66.

S2. Das, A., and Gilbert, C.D. (1997). Distortions of visuotopic map match orientation singularities in primary visual cortex. Nature *387*, 594-598.

S3. Engel, S.A., Glover, G.H., and Wandell, B.A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb Cortex *7*, 181-192.

S4. Frostig, R.D., Lieke, E.E., Ts'o, D.Y., and Grinvald, A. (1990). Cortical functional architecture and local coupling between neuronal activity and the microcirculation revealed by in vivo high-resolution optical imaging of intrinsic signals. Proc Natl Acad Sci U S A *87*, 6082-6086.