NeuroImage 78 (2013) 249–260

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

# Pattern classification precedes region-average hemodynamic response in early visual cortex

Peter J. Kohler [a,*,1], Sergey V. Fogelson [a,1], Eric A. Reavis [a], Ming Meng [a], J. Swaroop Guntupalli [a], Michael Hanke [b], Yaroslav O. Halchenko [a], Andrew C. Connolly [a], James V. Haxby [a], Peter U. Tse [a]

[a] Department of Psychological and Brain Sciences, Dartmouth College, NH 03755, USA
[b] Institut für Psychologie II, Otto-von-Guericke-Universität Magdeburg, D-39106 Magdeburg, Germany

ARTICLE INFO

*Article history:*
Accepted 4 April 2013
Available online 13 April 2013

*Keywords:*
Visual cortex
fMRI
Rapid temporal sampling
Multi-voxel pattern analysis
Hemodynamic response
Temporal dynamics
Early classification

ABSTRACT

How quickly can information about the neural response to a visual stimulus be detected in the hemodynamic response measured using fMRI? Multi-voxel pattern analysis (MVPA) uses pattern classification to detect subtle stimulus-specific information from patterns of responses among voxels, including information that cannot be detected in the average response across a given brain region. Here we use MVPA in combination with rapid temporal sampling of the fMRI signal to investigate the temporal evolution of classification accuracy and its relationship to the average regional hemodynamic response. In primary visual cortex (V1) stimulus information can be detected in the pattern of voxel responses more than a second before the average hemodynamic response of V1 deviates from baseline, and classification accuracy peaks before the peak of the average hemodynamic response. Both of these effects are restricted to early visual cortex, with higher level areas showing no difference or, in some cases, the opposite temporal relationship. These results have methodological implications for fMRI studies using MVPA because they demonstrate that information can be decoded from hemodynamic activity more quickly than previously assumed.



## Introduction

Although the neural activity in visual cortex associated with a visual task such as category recognition begins within a few hundred milliseconds after stimulus presentation (Rust and Dicarlo, 2010), associated change in blood flow measured with functional magnetic resonance imaging (fMRI), known as the hemodynamic response (HR), begins seconds later, and evolves over the course of several seconds, originating in a constrained spatial region, and spreading outward from that point while rising in amplitude (Shmuel et al., 2007). The complex spatiotemporal dynamics of the HR make it difficult to predict when the maximum amount of information about neural activity should be recoverable from the signal. Moreover, the time at which maximal information about brain activity is recoverable may not be the same in different types of analyses or different brain regions. Because knowledge of when maximal information is recoverable from fMRI data is of great utility for optimizing experimental designs and analyses, we undertook a systematic investigation of the timecourse of information availability in the HR.

There are two types of analyses commonly performed on fMRI data: univariate and multivariate. In univariate analyses, a general linear model (GLM) is applied to each voxel individually. By contrast, multivariate pattern analyses (MVPA) of fMRI data take into account relationships in the activity of multiple voxels. Several recent reports have begun to investigate the timecourse of MVPA classification accuracy during a range of cognitive tasks, showing that although the MVPA timecourse roughly tracks the region-average HR timecourse, classification accuracy can have temporal dynamics that differ from the region-average HR (Bode and Haynes, 2009; Greenberg et al., 2010). In fact, under certain conditions, accurate classification can persist even after the region-average HR has returned to baseline (Harrison and Tong, 2009).

We hypothesized that the reverse might also be the case, namely that HR patterns between voxels in a region would be able to support MVPA classification *prior* to a significant rise in the region-average HR. We will call this the "early onset" hypothesis. This hypothesis could be true if the HRs of individual voxels deviated reliably from baseline early in the timecourse, without being uniform enough to cause the region-wide average to deviate from baseline. Similarly, we hypothesized that peak classification would not always occur at the same time as the peak of the region-average HR. We will refer to this as the "early peak" hypothesis. In univariate analyses, peak region-average HR will by definition yield the largest difference between conditions and hence the largest effect size, but this is not the case for MVPA, where HR patterns could potentially contain more information about experimental conditions at timepoints before (or after) the peak region-average HR. This could be the case,

* Corresponding author at: HB 6207 Moore Hall, Dartmouth College, Hanover, New Hampshire 03755. Fax: +1 603 646 1419.
*E-mail address:* peter.j.kohler.gr@dartmouth.edu (P.J. Kohler).
[1] P. J. Kohler and S. V. Fogelson contributed equally to this research.

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.04.019

for example, if the patterns were made less informative by the spatial spreading of blood through the capillary bed that may occur as the HR approaches its peak.

We tested early onset and early peak hypotheses in a number of functionally defined regions in visual cortex. In order to characterize the timecourse of the region-average HR and classification accuracy as precisely as possible, we collected fMRI data over the occipital and temporal lobes at a high temporal resolution (one acquisition = 739 ms), using a slow event-related design, while participants viewed pictures of faces and houses. We analyzed stimulus category classification at each acquisition and compared classification accuracy to the region-average HR within each predefined ROI. We find that both the onset (first above-chance classification) and peak (most statistically significant classification) of the MVPA analysis precede the onset (first significant deviation from baseline) and peak (most statistically significant increase from baseline) of the region-average HR in V1, but not in other visual areas.

## Methods

### *Participants*

11 participants (mean age = 26.5, five female) were recruited from Dartmouth College. All participants had normal or corrected-to-normal vision and, prior to participating, gave written, informed consent under a protocol approved by the Dartmouth Committee for the Protection of Human Subjects.

### *Experimental design*

During each experimental run, participants viewed images of human faces and houses (10 of each) from the stimulus set used by Haxby and colleagues (Haxby et al., 2001). Of the 10 faces, five were males. All pictures were presented foveally within a 7.6° (degrees of visual angle) by 8.3° rectangular window. On average, house stimuli were ~6.0° wide by ~3.8° tall, and face stimuli were ~4.5° wide by ~7.6° tall.

We used a slow event-related design, in which subjects were presented with a single stimulus for one acquisition (739 ms), which was always followed by an inter-stimulus interval spanning 14 acquisitions (10.3 s). Subjects were instructed to perform a one-back task during which they pressed a button whenever the currently presented stimulus was identical to the previously presented stimulus. This one-back task was employed to ensure that subjects remained alert and awake. All 20 stimuli were shown in a random order once per run, with an additional three being selected randomly for "same" trials, where the stimulus was identical to the preceding stimulus. "Same" trials were excluded from the analysis. With a total of 23 presentations, each run had 360 acquisitions. 10 runs were collected per subject. Stimulus timing and presentation were controlled using MATLAB (The Mathworks; Natick, MA) in combination with Psychophysics Toolbox Version 3 Software (Brainard, 1997; Pelli, 1997). Stimuli were projected onto a screen behind the scanner bore which subjects viewed through a mirror mounted on the head coil.

### *Data acquisition*

Images were acquired with a Philips 3 T Achieva Intera scanner with an eight-channel head coil at the Dartmouth Brain Imaging Center. In order to maximize the temporal resolution of the acquired volumes, we used a rapid scanning protocol, PRESTO-SENSE (Golay et al., 2000), two-shot acquisition scheme with an EPI factor of 17, which allowed us to acquire a partial brain volume every 739 ms (17 axially-oriented slices, 3.0 × 3.0 mm in-plane voxel resolution, 3 mm slice thickness, no gap, interleaved slice acquisition, FOV = 240 × 240 × 51, TR = 21 ms per slice, TE = 14 ms (effective TE = 35 ms), flip angle = 10°, acquired matrix size = 80 × 63, reconstructed matrix size (zero filled) = 80 × 80, P reduction (RL) sense factor of 2, S reduction (FH) sense factor of 1). It should be noted that the acquired matrix size is not what one might expect mathematically. This is because the PHILIPS software attempts to protect the user from SENSE artifacts by enlarging the FOV before SENSE is applied. Without SENSE, the FOV is 240. However, at SENSE R = 2, the value is 129 and not the expected 120 (240/2). Once the image is reconstructed internally, it is cropped to the desired FOV of 240 (instead of 2 * 129 = 258) and the resulting images have a matrix dimension of 63 voxels.

This partial brain volume covered most of the occipital lobes, including all of retinotopic cortex, and parts of the temporal lobes, including lateral occipital cortex (LOC), the occipital face area (OFA), the fusiform face area (FFA), and the parahippocampal place area (PPA). 2 dummy scan acquisitions (~1.5 s) followed by 14 partial volume acquisitions (~10 s) of blank stimuli were collected at the beginning of every scan. High-resolution structural T1-weighted MPRAGE full-brain scans were also acquired for each subject (160 sagittal slices, 0.94 × 0.94 mm in-plane voxel resolution, 1 mm slice thickness, acquired matrix size = 256 × 232, reconstructed matrix size = 256 × 256, FOV = 240 × 240 × 160, TR/TE = 9.9/4.6 ms, flip angle = 8°). This high-resolution scan and scans collected with the same parameters during other sessions were used to construct flattened cortical meshes for retinotopic mapping. A T1-weighted coplanar anatomical image with the same slice orientation as the PRESTO data was also collected. This image was used to aid co-registration of the high-resolution anatomical scan to the functional data.

### *Localization of regions of interest*

We performed retinotopic mapping using 22.5° rotating monochromatic checkerboard bowties. Intact and scrambled objects were used for localizing LOC (Kourtzi and Kanwisher, 2001). In order to localize areas OFA, FFA and PPA, we contrasted short (2 s) dynamic videos of non-translationally moving objects and scenes with analogous videos of faces (Fox et al., 2009). All localization data were acquired during separate scanning sessions using standard fMRI EPI acquisition sequences. Functional mapping was conducted using procedures described previously (Caplovitz and Tse, 2010; Sereno et al., 1995; Slotnick and Yantis, 2003).

Polar angle representation in visual cortex was measured using two symmetrical wedges of a black and white polar checkerboard grating flickering at 8 Hz (Sereno et al., 1995; Slotnick and Yantis, 2003). Each wedge subtended 22.5° of 360° and occupied a given location for one TR (2000 ms) before moving to the adjacent location in a clockwise or counter-clockwise manner (direction alternated across runs). Cortical representation of eccentricity was measured using expanding concentric checkerboard-patterned rings flickering at 8 Hz that each spanned ~1° of visual angle in ring width. For every TR, a given ring was replaced by its outward neighbor, except that the outermost ring was replaced by the innermost ring at the end of the cycle. This process was repeated until the end of the run. For each participant, five runs of each direction were collected for the wedge stimulus and three runs were collected for the concentric rings. Retinotopic areas (V1d, V1v V2d, V2v, V3d, V3v, V4v and V3A/B) were defined as masks on the basis of standard criteria (Sereno et al., 1995), assuming a contralateral quadrant representation for V1d, V1v, V2d, V2v, V3d and V3v, and a contralateral hemifield representation for V4v/VO1 and V3A/B (Tootell et al., 1997). Visual areas V1, V2, V3 and V3A/B were created by merging the dorsally and ventrally defined regions of each respective area. V4v and the hemifield representation just anterior to it, called VO1 (Brewer et al., 2005), were combined because the border between these regions was not distinct in most subjects.

Retinotopic mapping, cortical reconstruction and volumetric segmentation, as well as cortical inflation and flattening, were performed using the FreeSurfer image analysis suite (Dale et al., 1999; Fischl et al., 1999). Once anatomical maps of the occipital lobes were flattened using

FreeSurfer, these surface models were imported into SUMA (Saad et al., 2004), where ROIs for areas V1–V4v were drawn by hand based on functional retinotopic activation maps created using AFNI software (Cox, 1996). Any overlapping voxels between adjacent ROIs were removed to guarantee a conservative specification of each ROI.

We created individual masks for LOC, OFA, FFA, and PPA for each subject via GLM contrasts (intact > scrambled objects for LOC, dynamic face videos > dynamic object videos for OFA and FFA, and dynamic object videos > dynamic face videos for PPA), using AFNI's 3dDeconvolve program (Cox, 1996). We then determined the smallest cluster size for each unilateral area among all of our subjects. In order to create bilateral masks of each area with approximately equal size, we increased *t*-value thresholds for the hemisphere where the ROI was larger, so that the ROI for each subject had an approximately equal size on both sides of the brain, and then combined the two regions to create a single ROI. Subjects who did not show bilateral activation for a given area were excluded from later analyses in that ROI.

*Data preprocessing*

Each functional run was motion-corrected to the first functional volume of the first run (that is, the volume that was acquired closest to the anatomical image) for each subject, using AFNI's 3dvolreg program (Cox, 1996). High-resolution anatomical images were coregistered to the first functional volume using AFNI's align_epi_anat.py program (Saad et al., 2009). We then co-registered an anatomical image aligned to our previously collected functional ROIs, to the anatomical image aligned to the experiment, and applied this same transformation to the ROIs, so that they were aligned to the experimental data. Functional runs were also high-pass Fourier filtered using 3dFourier with a cutoff frequency of 0.01 Hz, linearly detrended and de-meaned, with all motion regressors removed, and normalized to the voxel by voxel unit norm within a run (making the sum-of-squares for each voxel time series equal to 1, using AFNI's 3dDetrend program (Cox, 1996).

*Unparameterized analysis of region-average HR and multi-voxel pattern analysis*

All analyses were performed separately for each subject within each of the nine pre-defined, subject-specific ROIs (V1, V2, V3, V3A/B, V4v, LOC, OFA, FFA, and PPA). In order to determine the region-average HR across time, we found the average unit-normalized values within each ROI that were associated with each category across all runs (see Fig. 1 and Supplementary Fig. S1). This was done at each timepoint beginning with stimulus onset. To approximate the baseline for each category-associated region-average HR within an ROI, we found the average of the normalized values associated with each category at the first timepoint across runs, and subtracted this number from the signal change values for both categories at every timepoint. We then performed *t*-tests between the percent signal change for a given category and zero at every timepoint, using the run-wise averages, and identified the first timepoint that gave a significant (uncorrected $p < 0.05$) two-tailed *t*-value for either category. This was identified as the first timepoint at which one of the category-dependent region-average HRs was different from baseline for that subject within the given ROI; this was, in other words, the timepoint of the onset of the region-average HR.

For the multi-voxel pattern analysis, we trained two different classifiers, linear support vector machines (SVM) and sparse multinomial logistic regression (SMLR), as implemented in PyMVPA (Hanke et al., 2009) on the preprocessed and normalized pattern of activity within each ROI labeled with the stimulus categories for nine of the functional runs, and then proceeded to test on the patterns of activity from the tenth functional run. This was done with all possible combinations of nine runs of training data and one run of testing data (leave-one-run out cross-validation). This allowed us to obtain classification accuracies for each timepoint using each classifier. We then performed nonparametric Monte Carlo simulations by reshuffling face and house labels for the voxel response patterns, and training and testing 1000 times on these reshuffled datasets for each timepoint. We thus obtained a chance level distribution for each timepoint and identified the first timepoint during which classification accuracy was greater than the top 5% of simulated accuracies. This was identified as the first timepoint at which classification was significantly above chance. We then subtracted the time of earliest above-chance classification using each classifier from the time of onset of the region-average HR. This yielded a single value per ROI for each subject that quantified the latency difference between first above-chance classification and the onset of the average HR.

We performed a similar analysis to compare the time of region-average HR peak to the time of peak classification for each subject. The peak of the region-average HR was identified as the timepoint with maximum positive *t*-value associated with either category. We identified the time of peak classification as the timepoint for which the Monte Carlo simulation returned the lowest *p*-values. We then subtracted the time of peak classification using each classifier from the time of region-average HR peak. This yielded a single value per ROI for each subject that quantified the latency difference between peak classification and peak region-average HR. We also investigated the influence of smoothing on our onset and peak effects, by repeating the MVPA analysis, using both classifiers, on data smoothed with a kernel 2–10 times the voxel size (6–30 mm).

*Parameterized analysis of region-average HR and multi-voxel pattern analysis*

In order to further characterize the timecourse of classification accuracy and compare this to the timecourse of the region-average HR, we parameterized the region-average HRs associated with each condition and the unsmoothed SVM and SMLR classification timecourse for each subject within each of our ROIs. We used a double gamma function for parameterization because such functions reasonably describe the shape of the hemodynamic response of visual cortex with easily interpretable parameters (Glover, 1999). It is an open question what kind of function best approximates the MVPA timecourse. For the sake of consistency, however, we used the same double gamma-based parameterizations for the MVPA accuracy timecourses and the condition-associated region-average HR timecourses across all our ROIs. In order to properly parameterize the individual timecourses, we anchored them by zeroing the first timepoint of each timecourse and then normalized all three timecourses (classification, face HR, house HR) within a given ROI by subtracting the minimum value from the timecourse and then dividing it by the sum of values in that timecourse. Anchoring facilitates function fitting under the assumption that any deviation from zero at stimulus onset would be due to noise. We also limited our parameterization to the first eleven timepoints of the timecourse, under the assumption that any deviations from zero near the end of the trial might arise from internal factors, such as preparation for the upcoming one-back task (as shown by Harrison and Tong, 2009), rather than from perception of the stimulus. Following zeroing and normalization we used MATLAB's fit function with the nonlinear least squares method of fit and located the onset of each timecourse as the first timepoint where each fitted double gamma curve reached a normalized threshold value. The average values of face HR, house HR and MVPA (SVM and SMLR) accuracy at the time of onset as determined in the unparameterized analysis were used as threshold values for each individual function. In some cases the fitted double gamma curve failed to reach this threshold, at which point that dataset was excluded from further onset analysis for that ROI. The timepoint at which the fitted function reached its maximum was considered the peak. The duration of each function was defined as the full duration at half the maximum value of each function (FDHM).

We excluded data in which the root-mean-square-error was bigger than 0.05, indicating a suboptimal fit. On average, across all subjects

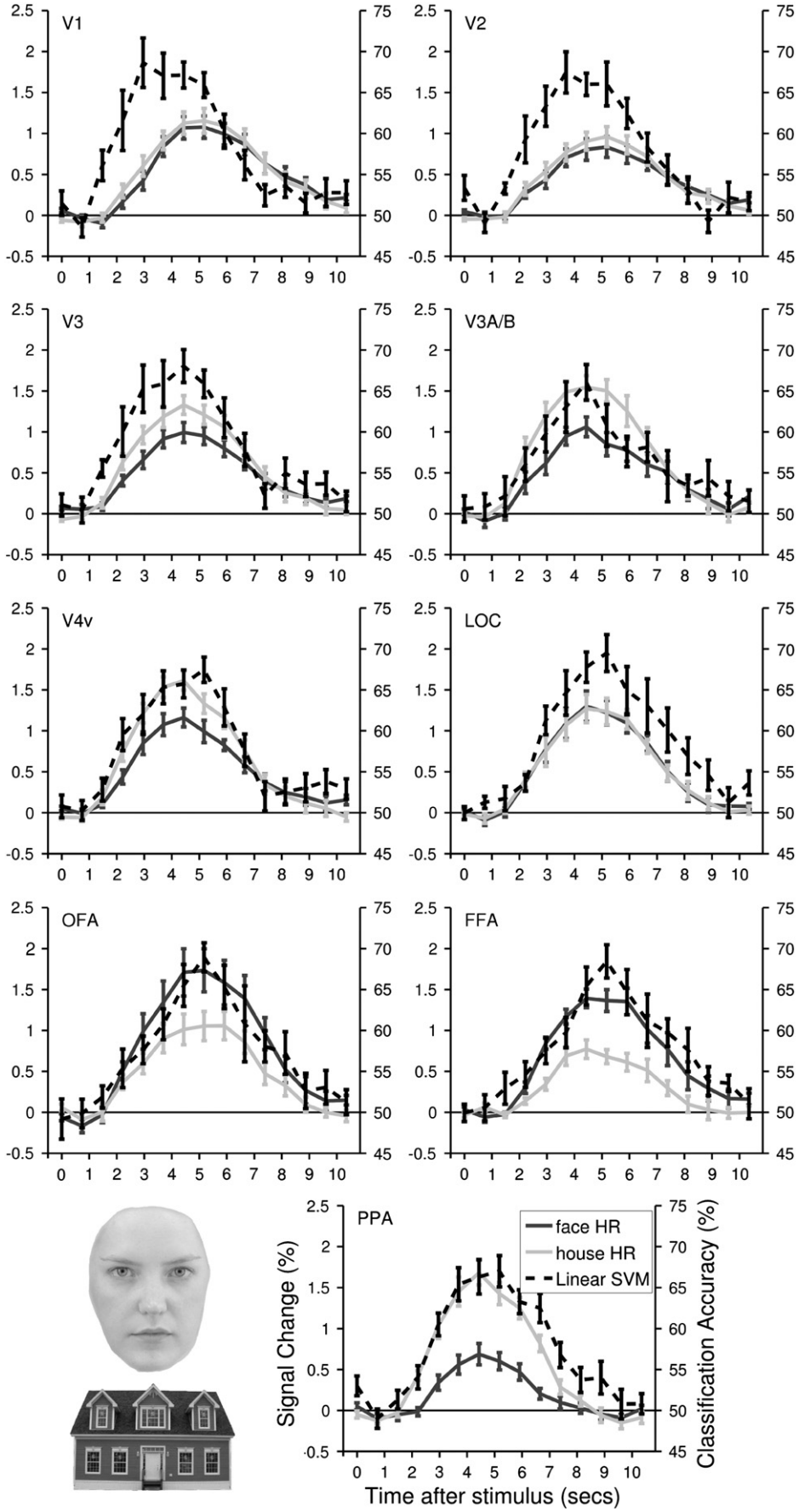


**Fig. 1.** Average timecourses across subjects (n = 11) for the nine ROIs considered. Error bars indicate the standard error of the mean across subjects. Example face and house stimuli are shown on the lower left and lower right, respectively. Individual subject timecourses within area V1 are plotted in Supplementary Fig. S1.

and for each analysis, fewer than two datasets per area were excluded because of poor fits or failure to reach onset threshold. No more than three datasets were excluded for any given dataset-analysis combination. Once all fitted estimates (onset, peak and duration) were found, we calculated difference scores by subtracting the classification onset time from the earliest region-average HR onset time, and the time of classification peak from the time of region-average HR peak, as described above. Statistical comparisons of duration were performed directly on the duration parameter estimates.

*The contribution of individual voxels to early classification*

We used two different methods to determine the contribution of each voxel within an ROI to early classification with each of the two classifiers. The first method, feature selection, used an *F*-score computed for each voxel by performing a one-way 1 × 2 ANOVA for both conditions different from baseline, based on the HR at the onset of above-chance classification. With only two conditions, the one-way ANOVA is equivalent to a paired *t*-test, with *F*-scores equal to $t^2$. We computed the *F*-scores based on the training data alone, and averaged across the cross-validation folds. This ensured that the test was orthogonal to our classification of interest (faces vs. houses), while allowing us to rank the voxels according to the robustness of their response. Because the onset of above-chance classification was different for each subject, the *F*-scores were computed based on face and house HRs at different TRs for each subject. Feature selection also allowed us to determine the minimum number of voxels required for classification within our ROIs, by selecting the subsets of voxels that had the most robust HR at above-chance classification onset, again based only on the training data, and attempt classification using just those subsets applied to each subject's hold-out dataset. As a second method of determining the contribution of each voxel to classification, we performed a voxel sensitivity analyses, by extracting the actual weights assigned to each voxel by the two classifiers during early classification; in this case, a larger weight for a given voxel meant that its response contributed to classification more strongly.

## Results

*Unparameterized analysis*

The average timecourse of the HR for faces and houses, as well as the average classification accuracy at each timepoint across all subjects, is plotted in Fig. 1. Our timepoint-by-timepoint analysis of region-average HR associated with each category and classification accuracy yielded two onset difference scores for each subject. These difference scores, one for each method of classification used (SVM and SMLR), indicated the onset time difference between the earliest above-baseline region-average HR and classification within each ROI. In order to compare the onsets of region-average HR and classification we performed a 2-way ANOVA on the difference scores with ROI as the first factor and classification type as the second factor. This revealed a main effect of ROI ($F(8,166) = 5.13$, $p < 0.05$), but not of classification type ($F(1,166) = .0168$, $p = 0.90$), and no significant interaction ($F(8,166) = 1.78$, $p = 0.08$). Planned, two-tailed comparisons revealed that both classifiers tested could reliably decode stimulus category before the onset of the region-average HR associated with either category in V1 (SVM: $t(10) = 6.71$, $p < 0.005$; SMLR: $t(10) = 4.28$, $p < 0.005$), and that this was also possible using the SMLR classifier in V2 ($t(10) = 3.71$, $p < 0.05$). Throughout the text, unless otherwise noted, all *p*-values are FDR-corrected for multiple comparisons. For several ROIs, we found that the onset of above-chance classification occurred *later* than the onset of the region-average HR, although this was not consistent across the two classifiers (V3 using SMLR: $t(10) = 7.46$, $p < 0.005$; LOC using SVM: $t(10) = 3.13$, $p < .05$; PPA using SMLR: $t(10) = 2.55$, $p < .05$). For all other areas, both comparisons fell short of significance (uncorrected $p > 0.05$). Fig. 2A shows the average onset difference scores for both classifiers in every ROI tested.

Our timepoint-by-timepoint analysis of peak differences also yielded two peak difference scores per region (one per classifier), indicating the time difference between peak region-average HR and peak classification. We performed an analogous 2-way ANOVA, which revealed a main effect of ROI ($F(8,166) = 9.70$, $p < 0.005$), with no effect of classification type ($F(1,166) = 2.26$, $p = 0.17$) and no interaction ($F(8,166) = 0.79$, $p = 0.61$). Planned comparisons revealed that peak classification preceded the peak of both condition-associated region-average HRs in V1 (SVM: $t(10) = 3.72$, $p < 0.05$; SMLR: $t(10) = 2.8$, $p < 0.05$). Peak classification occurred later than the peak region-average HR in V3A/B (SMLR: $t(9) = 3.72$, $p < 0.05$), V4v (SVM: $t(10) = 4.89$, $p < 0.005$; SMLR: $t(10) = 3.18$, $p < 0.05$), and PPA (SVM: $t(10) = 2.95$, $p < 0.05$; SMLR: $t(10) = 3.5$, $p < 0.05$). In two additional areas, this late peak effect was significant at the uncorrected $p < 0.05$ level (LOC using SMLR: $t(10) = 2.78$; FFA using SVM: $t(7) = 2.39$). The peak difference scores are shown in Fig. 2B.

The timecourses of classification in V1 using each classifier, at every level of smoothing, as well as with no smoothing are shown in Fig. 3A. We calculated the onset and peak effects for each classifier, at each degree of smoothing, and performed two-way repeated measures ANOVAs, for peak and onset, with degree of smoothing as the first

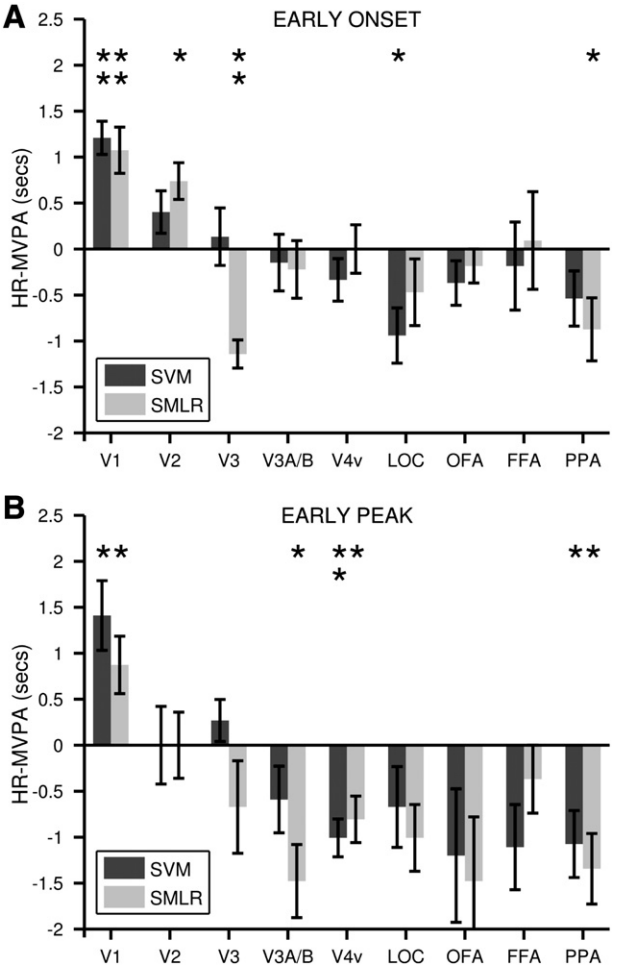


**Fig. 2.** Unparameterized results. Average classification onset and peak differences for classification using SVM (dark gray bars) and SMLR (light gray bars). The Y-axis indicates the time difference between region-average HR onset/peak and onset/peak of classification. Error bars indicate the standard error of the mean across subjects and comparisons. The parameterized results are plotted analogously in Supplementary Fig. S2.

level (0–30 mm) and classifier as the second. We found a main effect of smoothing on onset ($F(9,90) = 3.26, p < 0.005$), with no other main effects or interactions. We also found a main effect of smoothing on peak ($F(9,90) = 4.58, p < 0.005$), with no other main effects or interactions. However, when doing *t*-tests to test the significance of the onset and peak effects at each level of smoothing we find that both effects persist with smoothing up to 18 mm (for SVM) and 21 mm (for SMLR, see Fig. 3B). The unsmoothed data yielded the largest onset effect for both classifiers, whereas the peak effect was largest with a modest amount of smoothing (6 mm for SVM and 9 mm for SMLR). To confirm that classification is in fact relying on patterns of voxel HRs, even at high levels of smoothing, we attempted the same analysis using the region-wide HR (that is, using a single "voxel" value per timepoint), and performed classification at each timepoint (the timecourse can be seen in Fig. 3A). As expected, this classification remained at chance throughout the timecourse for both classifiers.

Finally, we compared the two classifiers, and found that they were quite similar in performance. The onset times are not significantly different ($t(10) = 0.056, p = 0.59$), while the peak time for SVM occurred slightly earlier than SMLR on average, albeit not quite significantly so ($t(10) = 2.19, p = 0.054$). For both classifiers the maximum overall classification accuracy in V1 occurred with 6 mm smoothing

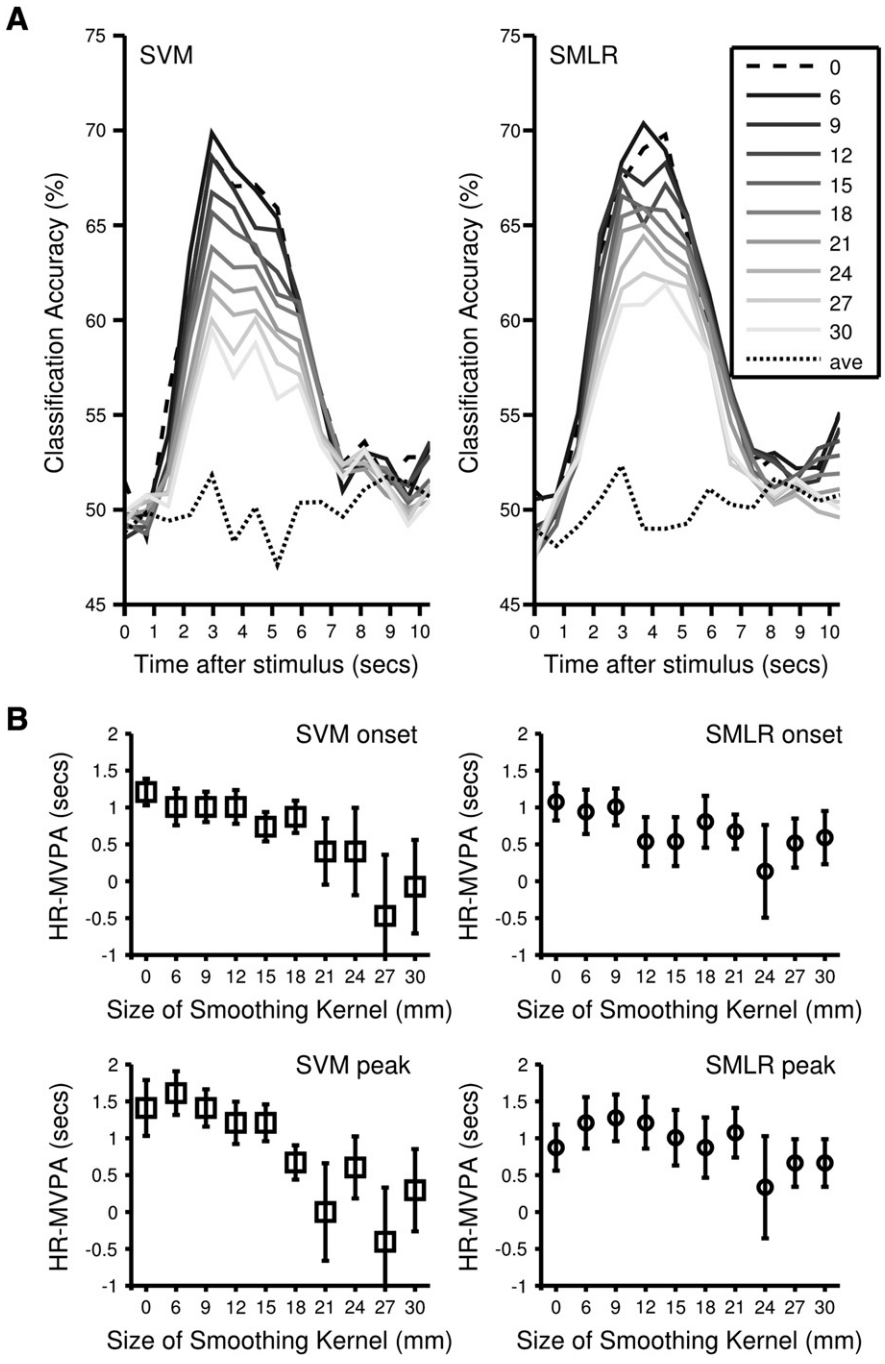


**Fig. 3.** The effects of smoothing on classification in V1. 3A shows the classification timecourses for each classifier, at every level of smoothing, as well as with no smoothing ("0") and when classifying based on the region-wide averages ("ave"). 3B shows the effect of smoothing on the onset and peak effects for each classifier. The x-axis indicates level of smoothing (with "0" again meaning no smoothing), while the y-axis shows the size of the effect in seconds.

and the two values were practically indistinguishable (SVM: 69.9%; SMLR: 70.0%).

*Parameterized analysis*

In an additional set of analyses, we characterized and compared the HR and classification accuracy data in a manner analogous to that found in the "Unparameterized analysis" section, after both had been fit with a double gamma function (Glover, 1999). This parameterized approach largely confirmed the results of the unparameterized analysis. A 2-way ANOVA on the parameterized onset difference scores found a main effect of ROI ($F(8,140) = 17.5$, $p < 0.005$), with no main effect of classifier ($F(1,140) = 0.70$, $p = 0.42$), and no interaction ($F(8,140) = 0.34$, $p = 0.95$). Planned, two-tailed comparisons indicated significantly earlier onset of classification in V1 (SMLR: $t(10) = 3.19$, $p < 0.05$). In V2 this early effect was marginally significant at the uncorrected $p < 0.05$ level (SVM: $t(9) = 2.23$, $p = 0.053$; SMLR: $t(9) = 1.97$, $p = 0.08$). Additionally, we found that the onset of classification occurred later than the onset of the region-average HR in LOC (SVM: $t(7) = 3.16$, $p < 0.05$; SMLR: $t(7) = 3.02$, $p < 0.05$). For all other areas, both comparisons fell short of significance in either direction ($p > 0.05$). A 2-way ANOVA on the parameterized peak difference scores revealed a main effect of ROI ($F(8,160) = 17.1$, $p < 0.005$), no main effect of classifier ($F(1,160) = 0.02$, $p = 0.90$) and no interaction ($F(8,160) = 0.31$, $p = 0.96$). Planned two-tailed comparisons revealed that peak classification occurred earlier than the region-average HR peak for both classifiers in V1 (SVM: $t(10) = 3.66$, $p < 0.01$; SMLR: $t(10) = 3.41$, $p < 0.01$). This effect was also significant at the uncorrected $p < 0.05$ level in V2 (SVM: $t(10) = 2.4$; SMLR: $t(9) = 2.29$). Our unparameterized finding that peak classification occurred later than the region-average HR peak in V3A/B, V4v, and PPA, however, was not found using the parameterized analysis, as no other areas reached significance.

The parameterized analysis also allowed us to calculate the full duration at half maximum (FDHM) of each curve. This allowed us to compare the duration of the face- and house-associated region-average HRs with the duration of above-chance classification. We performed a one-way ANOVA with a planned contrast (1 1 −1 −1) for each ROI, testing whether the durations of above-chance classification for both classifiers were either longer or shorter than the duration of both region-average HRs. In V1, the ANOVA was not statistically significant ($F(3,40) = 1.62$, $p = 0.20$), but the planned contrast indicated that the timecourse of above-chance classification was systematically shorter than the region-average HR ($t(40) = 2.13$, $p < 0.05$). In PPA, the ANOVA was significant ($F(3,40) = 3.59$, $p < 0.05$) and the planned contrast revealed that the timecourse of above-chance classification was systematically longer than that of the region-average HRs ($t(40) = -3.16$, $p < 0.005$). None of the ANOVAs or planned comparisons reached significance in any other area. The durations for all areas are plotted in Supplementary Fig. S3.

*Controlling for very late classification*

We performed a control analysis to directly address the potential confound that our early classification could in fact have been very late classification; that is, early classification could in principle rely on residual information about stimulus type remaining in the region-average HR from the previous stimulus presentation, if stimuli were improperly counterbalanced. In order to rule out this possibility we attempted to classify stimulus category prior to stimulus onset. If early classification did in fact rely on signal 'left over' from the previous stimulus, we should have been able to classify stimulus categories prior to the actual presentation of the stimulus. We found that in all areas and all subjects, classification accuracy was never significantly above chance prior to stimulus onset (uncorrected $p > 0.05$ in all areas), indicating that early classification exploited HR patterns that arose from the current stimulus and not from the previous stimulus.

*Controlling for ROI size*

Since the V1 ROIs were considerably larger than any of our other ROIs (average V1 size: 304 voxels), it could be that size alone distinguished V1 from the other ROIs and allowed early classification. To investigate this possibility, we merged the LOC, OFA, FFA and PPA masks to create a new combined mask (CM) that was comparable in size to V1 (in fact, it was substantially larger than V1 in most subjects, average size: 517 voxels). We performed the same analysis, comparing onset and peak of classification to onset and peak HR, within CM. If V1 supports early classification due to a larger number of voxels, then early classification should be possible within the large CM ROI as well. This was not the case. For both classifiers, above-chance classification actually occurred *later* than average HR deviation from baseline in CM (SVM: $t(10) = 2.95$, $p < 0.05$; SMLR: $t(10) = 3.46$, $p < 0.01$), and classification and HR peaked at the same time (SVM: $t(10) = 1.90$, $p > 0.05$; SMLR: $t(10) = 1.32$, $p > 0.05$). We conclude that neither the onset nor the peak effects in V1 can be explained by a difference in size between V1 and other areas.

*Controlling for temporal variability*

In order to further characterize the behavior of voxels in V1 and CM, we compared the temporal variability in early and late visual areas by calculating the HR onset and peak for each voxel in V1 and CM using the same procedure for defining HR onset and peak as we used for the entire mask. We then calculated the subject-wise variance in both onset and peak time across all voxels in CM and V1, and subtracted the two to obtain a variance difference score for onset and peak. We then did permutation testing for both onset and peak, randomly assigning signs (negative or positive) to each subject's score 5000 times, and calculating the mean difference score for each permutation. This allowed us to calculate a *p*-value for our measured mean variance difference score by comparing it with the distribution of random scores created by the permutation analysis. We found that for the onsets, CM had marginally significantly more variance than V1 (mean difference: 0.402 s, $p = 0.0778$), whereas there was no significant difference in the peak variance (mean difference: 0.0859 s, $p = 0.846$).

*The contribution of individual voxels to early classification*

We used two different methods to determine the contribution of each voxel within an ROI to early classification using SVM and SMLR: feature selection and a sensitivity analysis of the weights assigned to each voxel by the classifiers. Both methods were implemented in PyMVPA (Hanke et al., 2009) and since both rank the voxels according to their ability to contribute to classification, we would expect them to be largely in correspondence. For SVM, this is what we found: *F*-scores found by feature selection and SVM voxel weights were highly correlated (average Spearman rank correlation coefficient = 0.605; the correlations were highly significant at $p < 0.001$ in all individual subjects), which can be seen in Fig. 4 where *F*-scores and SVM voxel weights in V1 are plotted together. We limited the analysis to the top 100 voxels so that voxels contributing little to classification (with low *F*-scores and voxel weights) would not artificially inflate correlations. For SMLR, however, we found that SMLR weights were not very correlated with *F*-scores (average Spearman rank correlation coefficient = 0.108). Only 4 out of 11 subjects had significant correlations ($p < 0.05$). Because of this, the following analysis based on feature selection was done using only the SVM classifier.

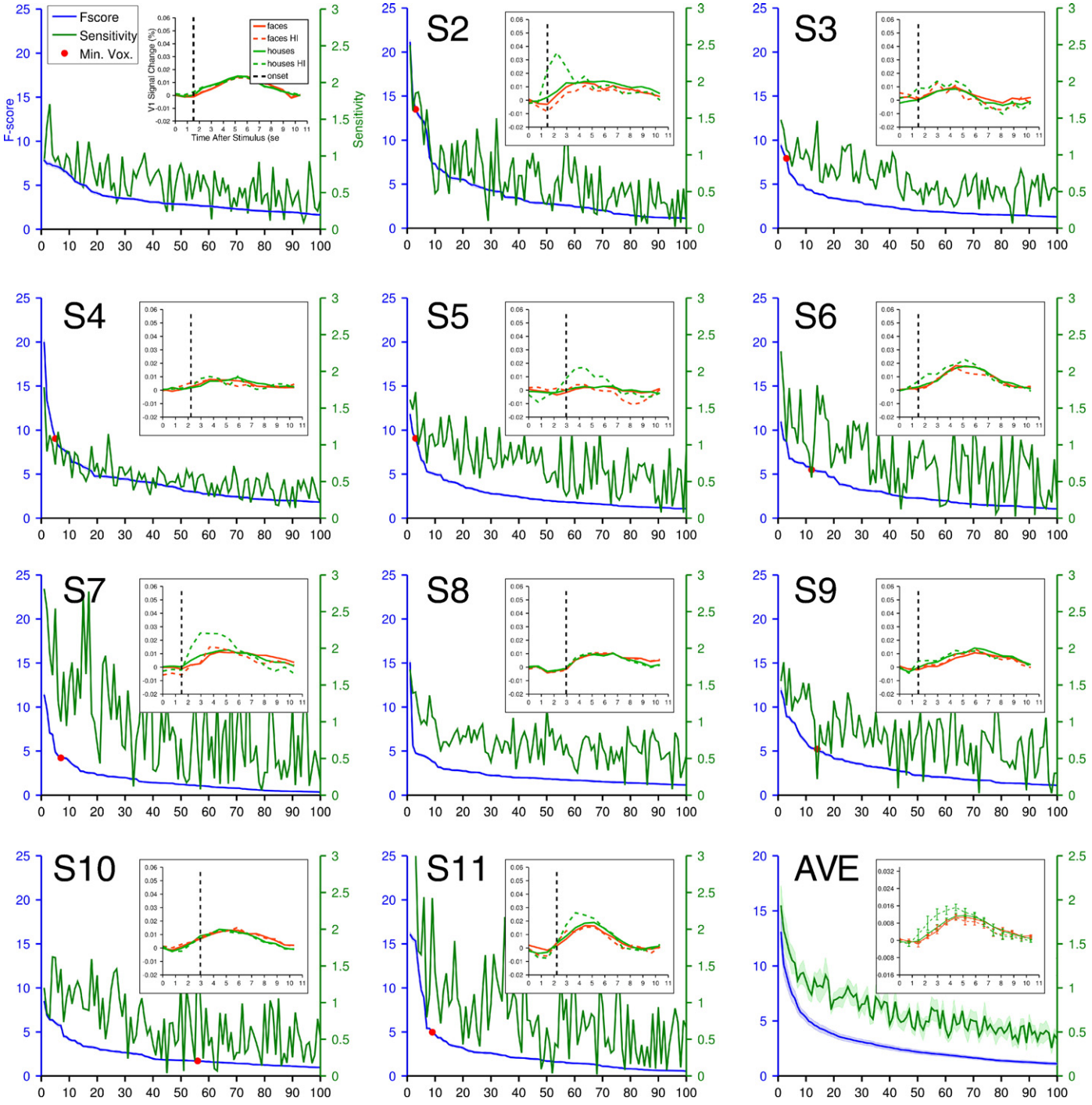


**Fig. 4.** *F*-scores and SVM weights in V1 plotted for each individual subject (S1–S11) and for the average across all subjects (bottom right graph). Data from subject 1 is plotted on the graph in the upper left corner, which is unlabeled in order to make room for the figure legend. The top 100 voxels are shown, ranked according to *F*-scores, computed based on the face and house HR at above-chance classification onset. Blue lines indicate the *F*-score for each voxel, and green lines indicate the SVM voxel weights. The red dots indicate the minimum number of voxels required for above chance classification. For three subjects, this dot is not present because >100 voxels were required for classification. The inset plot indicates the average HR timecourse for faces (in red) and houses (in green) for all voxels (solid lines) and for the minimum highly informative (HI) voxels required for classification (dashed lines). The black dashed line on each individual subject plot indicates the onset of above-chance classification. In the interest of clarity, axis labels and legends are shown only on the graph in the upper left corner.

Feature selection also allowed us to determine the minimum number of voxels required for classification within our ROIs. We selected the subsets of voxels that had the most robust HR at above-chance classification onset, based only on the training data (1%, 5%, 10%, 20%, 30%, 40%, 50% with the highest *F*-scores) and attempted classification using just those subsets applied to each subject's hold-out dataset (the testing dataset). This was done using every possible combination of leave-one-run out training and testing dataset splits. For 8 out of 11 subjects the number of *F*-score selected voxels necessary for SVM classification at onset was very low (on average 7 voxels, ~2% of the overall V1 mask), and for 3 others it was quite high (on average 112 voxels ~ 37% of the overall V1 mask).

We performed the same analysis within CM (consisting of LOC, OFA, FFA and PPA), at the time of earliest classification, and determined the

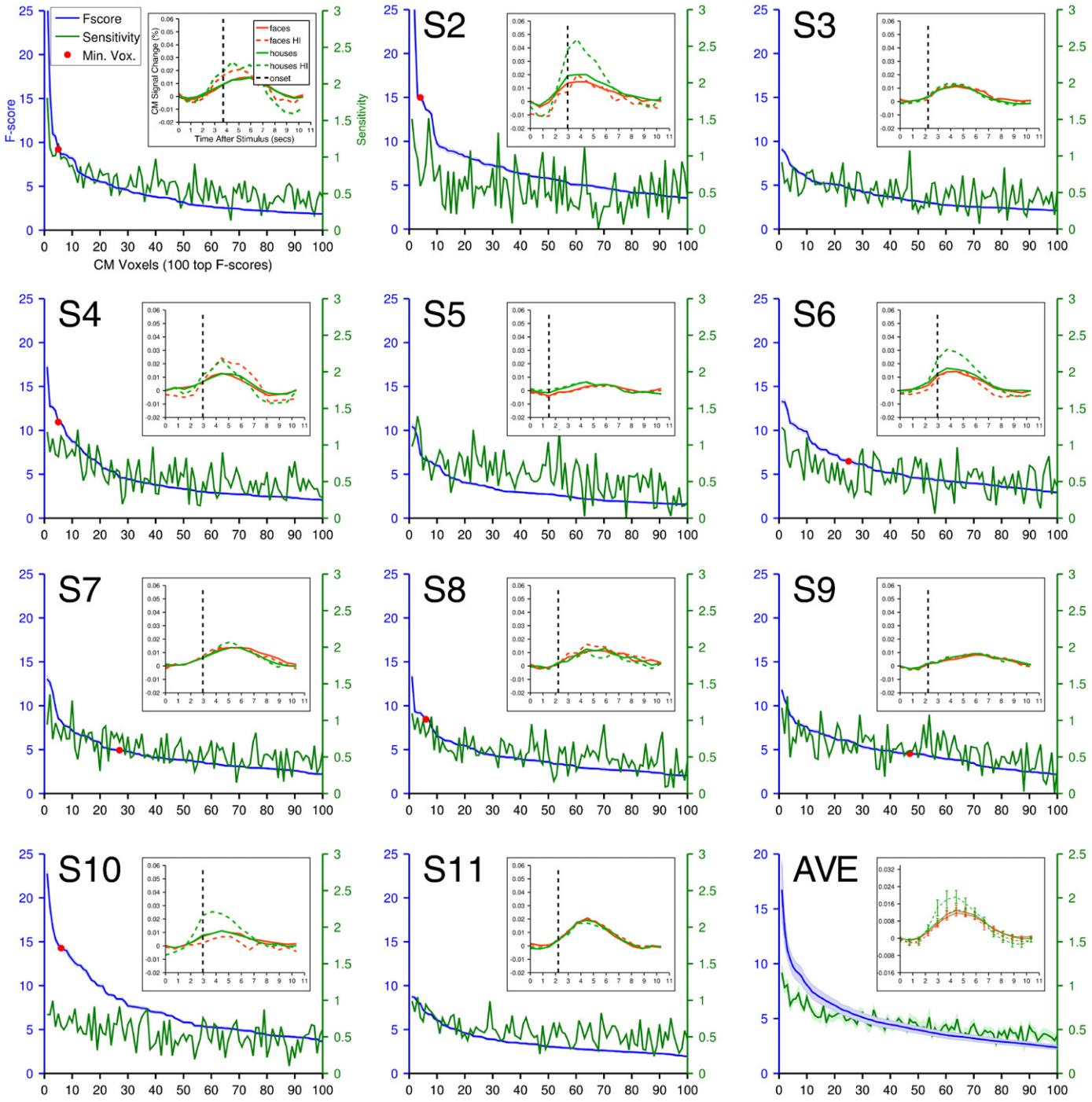


**Fig. 5.** *F*-scores and SVM voxel weights in the combined mask (LOC, OFA, FFA and PPA) plotted for each individual subject and for the average across all subjects. This figure follows the same conventions as Fig. 4.

*F*-scores and SVM weights in a manner identical to that described above. These values are plotted across all subjects in Fig. 5. As in V1, the two were highly correlated (top 100 *F*-score voxels, average Spearman rank correlation coefficient = 0.531; the correlations were highly significant at $p < 0.005$ in all individual subjects). As in V1, we found that SMLR weights were on average less correlated with onset *F*-scores than SVM weights (average Spearman rank correlation coefficient = 0.211). However, the majority of subjects (7 out of 11) did have significant correlations ($p < 0.05$).

Furthermore, the feature selection analysis reveals that, as was the case with V1, very few voxels can drive SVM classification in 8 out of 11 subjects (on average 16 voxels ~ 3% of CM), with 3 subjects requiring more voxels (on average 177 voxels, ~34% of CM). The three subjects who required more voxels for accurate classification in CM were not the same three who required more voxels for accurate classification in V1.

We also compared the onsets of highly informative voxels, as identified using our feature selection method, to the onset of the entire ROI. In V1, highly informative voxels on average leave baseline at 1.68 s after stimulus presentation, which is 1.54 s before the average

across all voxels in the region. Highly informative voxels in the combined mask, however, leave baseline 1.94 s after stimulus presentation, which is 0.168 s before the onset of the combined mask region average. For both ROIs, the earliest that highly informative voxels leave baseline is 1 TR (0.739 s) after onset.

In V1, at the timepoint of earliest classification, highly informative voxels have a marginally stronger response to house stimuli compared to the overall region-average ($t(10) = 2.21$, $p = 0.051$), whereas their face responses are indistinguishable from the overall region-average ($t(10) = 0.15$, $p = 0.88$). The same pattern is present in CM, but is more pronounced (house difference: $t(10) = 2.76$, $p < 0.05$; face difference: $t(10) = 0.58$, p = 0.58), as seen in Supplementary Fig. S4. Finally, we wanted to investigate the locations of voxels that were highly informative for early classification. We calculated the percentages of highly informative voxels in the four V1 regions corresponding to each quadrant of the visual field, and performed a two-way repeated measures ANOVA with left/right as the first level and dorsal/ventral as the second level. There were no main effects and no interactions (see Supplementary Fig. S5). To characterize the locations of highly informative voxels more precisely, we also plotted V1 voxel sensitivity weights for SVM and SMLR onto flattened maps of occipital cortex created in FreeSurfer (Dale et al., 1999; Fischl et al., 1999), using AFNI's (Cox, 1996) 3dVol2Surf function. As far as we could tell by visual inspection, there were no meaningful or consistent patterns in the locations of voxels with large sensitivity weights, across subjects for either classifier (see Supplementary Fig. S6). This suggests that highly informative voxels are not restricted to specific cortical locations in V1. In CM, there were more highly informative voxels in LOC (average percentage of selected voxels ~ 38%) and PPA (~43%) with a much smaller percentage of voxels in FFA and OFA (~17% and ~2%, respectively).

## Discussion

We find that above-chance MVPA classification in V1 is possible before the average HR across the region leaves baseline, and that classification peaks earlier than the region-average HR, while having a shorter overall duration. This consistent early classification is unique to early visual cortex, and is not an artifact of the large size of V1 compared with other cortical regions, as evidenced by the fact that an ROI of comparable size to V1, (CM, consisting of LOC, OFA, FFA and PPA) does not show early classification. Early onset classification is driven by a small number of interspersed voxels that are approximately evenly distributed among the regions of V1 corresponding to the four quadrants of the visual field, and do not appear to be restricted to specific regions of each quadrant (see Supplementary Fig. S6).

Early onset classification occurs for both SVM and SMLR. The onset times for the two classifiers are not significantly different ($t(10) = 0.056$, $p = 0.59$), while the peak time for SVM occurred slightly earlier than SMLR on average, albeit not quite significantly so ($t(10) = 2.19$, $p = 0.054$). For both classifiers the maximum mean overall classification accuracy in V1 occurred with 6 mm smoothing and was practically indistinguishable between the two (SVM: 69.9%; SMLR: 70.0%). Unlike SVM, SMLR voxel weights are not very well correlated with *F*-scores, with only 4 out of 11 subjects showing a significant correlation. This discrepancy may explain why SVM and SMLR classification had different onsets for 8 out of 11 subjects.

Early onset classification cannot be explained by greater temporal variance in V1, compared to areas later in the visual processing stream, as V1 and CM are comparable in terms of temporal variance, with CM actually having slightly more onset variance than V1. V1 on average has a later onset than the CM average (V1: 3.22 s, CM: 1.75 s), but the highly informative V1 voxels leave baseline as early as in later areas, or slightly earlier, thereby allowing classification before the onset of the region-average HR.

The initial negative dip that is sometimes seen in HRs in visual cortex (Menon et al., 1995), could potentially drive early classification. Although we performed two-tailed *t*-tests on the average signal which would have identified negative as well as positive deviations from baseline, it is possible that a subset of the voxels within V1 that exhibit the initial dip are driving the effect, without leading to a significant negative difference in average signal. If this were the case, the highly informative voxels identified using our feature selection method should show an initial negative dip in HR. Another possibility is that voxels with evenly balanced positive and negative responses to the two categories are driving classification, which would create an average response close to the baseline. If this were the case, the average response of the informative voxels would be at baseline. In fact, we find neither to be true; highly informative voxels are quite variable in their responses to both houses and faces, but on average they have slightly larger responses to houses than faces, compared with the region-wide average.

Our early classification results cannot be explained by the superior sensitivity of MVPA compared to univariate analyses (Haynes, 2009). The fact that we find an early peak in classification accuracy, as well as early onset, and that both findings are restricted to early visual cortex, cannot easily be explained by superior sensitivity. If the results were due to superior sensitivity of MVPA, classification should also persist for longer than the region-average HR duration. The longer duration of classification in PPA may in fact be indicative of such an effect. In primary visual cortex, we find the opposite: classification duration is shorter than region-average HR duration (see Supplementary Fig. S3) — the information extracted by MVPA before region-average HR onset is no longer available as HRs return to baseline.

A potentially important difference between early and later visual areas is that early visual areas are not specialized for processing faces or houses, whereas areas like OFA, FFA and PPA are (Epstein and Kanwisher, 1998; Kanwisher et al., 1997). Classification in V1 is likely to rely on local image-level differences between the two categories (such as image extent, spatial frequency, or the presence of high contrast straight edges in the house stimuli), rather than differences between the categories themselves. This could mean that a few voxels highly sensitive to local features might drive classification, in which case classification accuracy would presumably follow the HR of these highly informative voxels, rather than the ROI-wide average. In later areas such as FFA and PPA, on the other hand, the majority of voxels have a categorical response to faces and houses as a result of the way those areas are defined, which would mean that many voxels are likely to contribute to classification. Hence, classification would most likely follow the average hemodynamic response. We do in fact find that for the majority of subjects a few highly informative voxels can drive classification in V1, but our CM analysis indicates that the same appears to be the case for category-sensitive areas later in the visual processing stream. As was the case for V1, the average responses of the highly informative voxels in CM differ from the ROI-wide averages in terms of their increased response to houses, but not to faces. These results suggest that house stimuli evoke more reproducible, coherent voxel response patterns, which can benefit early classification in both V1 and CM. These findings make us conclude that both early and late visual areas have small numbers of voxels that are highly informative and can drive classification. Rather than a difference in the number of voxels driving classification, the relevant difference between V1 and later areas seems to be that in V1, the highly-weighted voxels become informative prior to average HR onset, whereas in later areas this happens only at or after average HR onset.

These findings relate to previous results in several interesting ways. Harrison and Tong (2009) found that orientation of a grating held in memory could be reliably classified after region-average HRs had returned to baseline in visual areas including V1 and V2. In our results, one object-sensitive cortical region, LOC, shows above-chance classification consistently later than region-average HR onset. However, our results in V1 show the converse of Harrison and Tong's finding, namely that classification in early visual cortex is possible even before

the region-average HR deviates from baseline. Regarding peak classification, previous results have indicated that peak information availability should *follow* the region-average HR peak (Fuhrmann Alpert et al., 2007). Our findings show that for V1 this is not the case; peak information availability in MVPA occurs prior to the peak of the region-average HR. Interestingly, Greenberg et al. (2010) found that peak classification performance during an attentional switching task occurred at approximately the same time in parietal regions as the region-average HR peak evoked by attentional shifts, whereas the overall shape of the classification curve was markedly different from the region-average HR timecourse, being much less symmetrical and having a faster rise time and a slower decay to baseline. Given that their temporal resolution was coarser than that used here, it is possible that more rapid temporal sampling would reveal peak effect dissociations similar to those we report here.

Bode and Haynes (2009) found that peak classification of task set occurred earlier in intraparietal sulcus than in prefrontal cortex, which they interpreted as indicative of timing differences in neural coding between the two areas. Although it is tempting to propose that our results are similarly indicative of timing differences between V1 and later visual areas, the scale of the onset and peak differences we find in V1 (~1 s) pale in comparison to the millisecond scale of neural transduction between areas of visual cortex. Instead, we take our results to indicate that at least with certain types of visual stimuli, the timecourse of classification can be dissociable from the timecourse of the average HR, independent of the timing of neural events. The fact that such dissociations occur in some areas (in our case early visual cortex) and not others, could potentially have implications for the timing differences found by Bode and Haynes (2009). Future work should explore whether similar dissociations can also be found with other stimuli, and in other areas of the brain.

The effect of smoothing on the onset and peak effect is informative about the scale of the information on which early classification relies. Our analysis of the effect of smoothing in fact suggests that early classification is not picking up on fine-scale information. For both classifiers, the maximum onset effect occurs without any smoothing, whereas the peak effect is largest with modest amounts of smoothing (6 mm for SVM and 9 mm for SMLR; see Fig. 3). However, both the onset and the peak effect persist with smoothing up to 5–6 times the voxel resolution. It has been suggested that MVPA can rely on sub-voxel resolution information from regionally correlated columnar responses in early visual areas (Kamitani and Tong, 2005). If fine-scale information can indeed be extracted from patterns of voxel responses (see Chaimow et al., 2011; Freeman et al., 2011; Op de Beeck et al., 2008), it is possible that such patterns would be most informative early in the HR timecourse, and contribute less to classification later in the timecourse, as they are washed out by an increased influx of blood. Spatial smoothing, however, has been shown to reduce classifier sensitivity to information in the high spatial frequency range most likely to contain columnar information (Misaki et al., 2013). The fact that both the onset effect and the peak effect survive even large amounts of spatial smoothing makes it highly unlikely that either effect is dependent on fine-scale columnar information.

As we have described, our analyses allow us to rule out a number of potential explanations for our results, but in the end, it is unclear how early classification occurs. We do not know what makes a small subset of V1 voxels, evenly distributed across retinotopic cortex, have early responses that allow them to contribute to early classification onset. We also do not know why voxels with earlier responses appear to be more informative for classification overall, leading to early peak classification. It may be that highly informative voxels for early classification in our experiment would not be so in an experiment using different stimuli. Further work will be required to determine exactly why some voxels have early responses and whether our effects generalize beyond the specific stimuli used in our experiments.

We conclude that throughout occipital and ventral temporal cortex there can be systematic differences in the temporal dynamics of classification accuracy and the region-average HR. Most strikingly, in V1 these differences include early onset of above-chance classification and an early peak in classification accuracy. Our results show that stimulus information is not tightly coupled with the region-wide average HR in all visual areas and that in V1 stimulus information can be extracted from a small number of interspersed voxels more rapidly than previously thought. Although it remains to be seen whether these effects generalize to other areas and other stimuli, the fact that these effect *can* occur has methodological implications for selecting the time at which to perform classification analyses. These findings should be the first step in the development of HR signal data acquisition methods and analysis techniques that can extract information from the earliest components of the hemodynamic response more effectively.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2013.04.019.

## Conflict of interest

The authors declare that the research was conducted in the absence of any actual or potential conflict of interest including any financial, personal or other relationships with other people or organizations that could inappropriately influence, or be perceived to influence, their work.

## References


Bode, S., Haynes, J.-D., 2009. Decoding sequential stages of task preparation in the human brain. NeuroImage 45, 606–613.
Brainard, D., 1997. The psychophysics toolbox. Spat. Vis. 10, 433–436.
Brewer, A.A., Liu, J., Wade, A.R., Wandell, B.A., 2005. Visual field maps and stimulus selectivity in human ventral occipital cortex. Nat. Neurosci. 8, 1102–1109.
Caplovitz, G.P., Tse, P.U., 2010. Extrastriate cortical activity reflects segmentation of motion into independent sources. Neuropsychologia 48, 2699–2708.
Chaimow, D., Yacoub, E., Ugurbil, K., Shmuel, A., 2011. Modeling and analysis of mechanisms underlying fMRI-based decoding of information conveyed in cortical columns. NeuroImage 56, 627–642.
Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.
Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. NeuroImage 9, 179–194.
Epstein, R., Kanwisher, N.G., 1998. A cortical representation of the local visual environment. Nature 392, 598–601.
Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. NeuroImage 9, 195–207.
Fox, C.J., Iaria, G., Barton, J.J.S., 2009. Defining the face processing network: optimization of the functional localizer in fMRI. Hum. Brain Mapp. 30, 1637–1651.
Freeman, J., Brouwer, G.J., Heeger, D.J., Merriam, E.P., 2011. Orientation decoding depends on maps, not columns. J. Neurosci. 31, 4792–4804.
Fuhrmann Alpert, G., Sun, F.T., Handwerker, D., D'Esposito, M., Knight, R.T., 2007. Spatio-temporal information analysis of event-related BOLD responses. NeuroImage 34, 1545–1561.
Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage 9, 416–429.
Golay, X., Pruessmann, K.P., Weiger, M., Crelier, G.R., Folkers, P.J., Kollias, S.S., Boesiger, P., 2000. PRESTO-SENSE: an ultrafast whole-brain fMRI technique. Magn. Reson. Med. 43, 779–786.
Greenberg, A.S., Esterman, M., Wilson, D., Serences, J.T., Yantis, S., 2010. Control of spatial and feature-based attention in frontoparietal cortex. J. Neurosci. 30, 14330–14339.
Hanke, M., Halchenko, Y.O., Sederberg, P.B., Olivetti, E., Fründ, I., Rieger, J.W., Herrmann, C.S., Haxby, J.V., Hanson, S.J., Pollmann, S., 2009. PyMVPA: a unifying approach to the analysis of neuroscientific data. Front Neuroinform. 3, 3.
Harrison, S.A., Tong, F., 2009. Decoding reveals the contents of visual working memory in early visual areas. Nature 458, 632–635.
Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science (N.Y., N.Y.) 293, 2425–2430.
Haynes, J.-D., 2009. Decoding visual consciousness from human brain signals. Trends Cogn. Sci. 13, 194–202.
Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8, 679–685.
Kanwisher, N.G., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17, 4302–4311.
Kourtzi, Z., Kanwisher, N.G., 2001. Representation of perceived object shape by the human lateral occipital complex. Science 293, 1506–1509.

Menon, R.S., Ogawa, S., Hu, X., Strupp, J.P., Anderson, P., Uğurbil, K., 1995. BOLD based functional MRI at 4 Tesla includes a capillary bed contribution: echo-planar imaging correlates with previous optical imaging using intrinsic signals. Magn. Reson. Med. 33 (3), 453–459.

Misaki, M., Luh, W.-M., Bandettini, P.A., 2013. The effect of spatial smoothing on fMRI decoding of columnar-level organization with linear support vector machine. J. Neurosci. Methods 212, 355–361.

Op de Beeck, H.P., Torfs, K., Wagemans, J., 2008. Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. J. Neurosci. 28, 10111–10123.

Pelli, D., 1997. The VideoToolbox Software for Visual Psychophysics: Transforming Numbers into Movies. Spatial Vision.

Rust, N.C., Dicarlo, J.J., 2010. Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. J. Neurosci. 30, 12978–12995.

Saad, Z.S., Reynolds, R.C., Argall, B., Japee, S., Cox, R.W., 2004. SUMA: an interface for surface-based intra- and inter-subject analysis with AFNI. 2004 2nd IEEE International Symposium on Biomedical Imaging: Macro to Nano (IEEE Cat No. 04EX821). IEEE, pp. 1510–1513.

Saad, Z.S., Glen, D.R., Chen, G., Beauchamp, M.S., Desai, R., Cox, R.W., 2009. A new method for improving functional-to-structural MRI alignment using local Pearson correlation. NeuroImage 44, 839–848.

Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268, 889–893.

Shmuel, A., Yacoub, E., Chaimow, D., Logothetis, N., Ugurbil, K., 2007. Spatio-temporal point-spread function of fMRI signal in human gray matter at 7 Tesla. NeuroImage 35, 539–552.

Slotnick, S.D., Yantis, S., 2003. Efficient acquisition of human retinotopic maps. Hum. Brain Mapp. 18, 22–29.

Tootell, R.B., Mendola, J.D., Hadjikhani, N.K., Ledden, P.J., Liu, A.K., Reppas, J.B., Sereno, M.I., Dale, A.M., 1997. Functional analysis of V3A and related areas in human visual cortex. J. Neurosci. 17, 7060–7078.