



Published in final edited form as:

Neuroimage. 2016 June ; 133: 529–548. doi:10.1016/j.neuroimage.2016.02.071.

A method for real-time visual stimulus selection in the study of cortical object perception

Daniel D. Leeds^{a,b,*} and Michael J. Tarr^{c,b}

^aFordham University, Computer and Information Science Department, Bronx, New York, USA

^bCarnegie Mellon University, Center for the Neural Basis of Cognition, Pittsburgh, Pennsylvania, USA

^cCarnegie Mellon University, Psychology Department, Pittsburgh, Pennsylvania, USA

Abstract

The properties utilized by visual object perception in the mid- and high-level ventral visual pathway are poorly understood. To better establish and explore possible models of these properties, we adopt a data-driven approach in which we repeatedly interrogate neural units using functional Magnetic Resonance Imaging (fMRI) to establish each unit's image selectivity. This approach to imaging necessitates a search through a broad space of stimulus properties using a limited number of samples. To more quickly identify the complex visual features underlying human cortical object perception, we implemented a new functional magnetic resonance imaging protocol in which visual stimuli are selected in real-time based on BOLD responses to recently shown images. Two variations of this protocol were developed, one relying on natural object stimuli and a second based on synthetic object stimuli, both embedded in feature spaces based on the complex visual properties of the objects. During fMRI scanning, we continuously controlled stimulus selection in the context of a real-time search through these image spaces in order to maximize neural responses across predetermined 1 cm^3 brain regions. Elsewhere we have reported the patterns of cortical selectivity revealed by this approach (Leeds 2014). In contrast, here our objective is to present more detailed methods and explore the technical and biological factors influencing the behavior of our real-time stimulus search. We observe that: 1) Searches converged more reliably when exploring a more precisely parameterized space of synthetic objects; 2) Real-time estimation of cortical responses to stimuli are reasonably consistent; 3) Search behavior was acceptably robust to delays in stimulus displays and subject motion effects. Overall, our results indicate that real-time fMRI methods may provide a valuable platform for continuing study of localized neural selectivity, both for visual object representation and beyond.

*Corresponding author. ; Email: dleeds@fordham.edu (Daniel D. Leeds), ; Email: michaeltarr@cmu.edu (Michael J. Tarr)

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Real-time stimulus selection; functional magnetic resonance imaging; object recognition; real-time signal processing; computational modeling

1. Introduction

How do humans visually recognize objects? Broadly speaking, it is held that the primate ventral occipito-temporal pathway of the human brain implements a feedforward architecture in which the features of representation progressively increase in complexity as information moves up the hierarchy (Felleman and Essen, 1991; Riesenhuber and Poggio, 1999). In almost all such models, the top layers of the hierarchy are construed as high-level *object representations* that correspond to and allow the assignment of category-level or semantic labels. Critically, there is also the presupposition that while early levels along the pathway encode information about edge locations and orientations (Hubel and Wiesel, 1968) and information about textures (Freeman et al., 2013), one or more levels, between what we think of as early vision and high-level vision, encode *intermediate* visual features. Such features, while less complex than entire objects, nonetheless capture important — and possibly compositional — object-level visual properties (Ullman et al., 2002). Remarkably, for all of the interest in biological vision, the nature of these presumed intermediate features remains frustratingly elusive. To help address this knowledge gap, we introduce new methods that leverage human fMRI to explore the intermediate properties encoded in regions of human visual cortex.

Any study investigating the visual properties employed in cortical object perception faces multiple challenges. First, the number of candidate properties present in real-world objects is large. Second, these properties are carried by millions to billions of potential stimulus images. Third, feature and image space can be parameterized by an uncountable number of potential models. Fourth, the time available in a given human fMRI experiment is limited. Scanning time for an individual subject is limited to several hours across several days. Fifth, during a given scan session, the slow evolution of the blood-flow dependent fMRI signal necessarily limits the frequency of single stimulus display trials to one every 8 to 10 seconds; more frequent displays produce an overlay of hemodynamic responses difficult to recover without carefully tuned pre-processing or careful dissociation of temporally adjacent stimuli. Moreover, even with these considerations, the neural data recovered will be noisier and less amenable to use on a trial-by-trial basis. As such, assuming a minimum of 8 seconds to display each trial, at most several hundred stimuli can be displayed to a subject per an hour.

Here we suggest that dynamic stimulus selection, that is, choosing new images to present based on a subject's neural responses to recently shown images, enables a more effective investigation of visual feature coding. Our methods build on the dynamic selection of stimuli in studies of object vision in primate neurophysiology. For example, Tanaka (2003) explored the minimal visual stimulus sufficient to drive a given cortical neuron at a level equivalent to the complete object. He found that individual neurons in area TE were selective for a wide

variety of simple patterns and that these patterns bore some resemblance to image features embedded within the objects initially used to elicit a response. Tanaka hypothesized that this pattern-specific selectivity has a columnar structure that maps out a high-dimensional feature space for representing visual objects. In more recent neurophysiological work, Yamane et al. (2008) and Hung et al. (2012) used a search procedure somewhat different from Tanaka and a highly-constrained, parameterized stimulus space to identify the contour selectivity of individual neurons in primate IT. They found that most contour-selective neurons in IT encoded a subset of the parameter space. Moreover, each 2D contour within this space mapped to specific 3D surface properties meaning that collections of these contour-selective units would be sufficient to capture the 3D appearance of an object or part.

At the same time, there has been recent interest in real-time human neuroimaging. For example, Shibata et al. (2011) used neurofeedback from visual areas V1 and V2 to control the size of a circular stimulus displayed to subjects and Ward et al. (2011) explored real-time mapping of the early visual field using Kalman filtering. Most recently, Sato et al. (2013) have developed a toolbox (“FRIEND”) that implements neural feedback applications in fMRI, applying classification and connectivity analyses to study the encoding of emotion. These studies support the idea of incorporating real-time analysis and feedback into neuroimaging work to expanding fields, such as the study of object perception.

Here we explore new methods for the real-time analysis of fMRI data and the dynamic selection of stimuli. More specifically, our procedure selects new images to display based on the neural responses to previously-presented images as measured in pre-selected brain regions. Our overall objective is to maximize localized neural activity and to identify the associated complex featural selectivity within image spaces that are organized on the basis of insights from earlier studies in object perception (Leeds et al., 2013; Williams and Simons, 2000). We employ two sets of objects and their corresponding spaces — real-world objects organized based on similarities computed by the SIFT computer vision method (Lowe, 2004) and synthetic “Fribble” objects (Williams and Simons, 2000) organized based on morphs in the shapes of their component appendages (see Fig. 5 below).

In previously published results, we reported the nature of the cortical selectivities uncovered by this novel approach (Leeds et al., 2014). Here we study the technical and biological factors influencing the performance of our real-time stimulus search, as well as the behavior of our search across subjects and stimulus sets. In particular, using synthetic stimuli, we found that searches exhibited some convergence onto a small number of preferred visual features and consistency across repeated searches for a given brain region within an individual subject. In contrast, using real-world object stimuli, we found only weak convergence and consistency, possibly as a result of the visual diversity of the real-world stimuli included in this image space. More generally, we observe that our methods are robust to undesired actions from subjects (e.g., head motions) and program flaws (e.g., stimulus selection delays), suggesting that our methods offer an important first-step in developing effective methods for real-time human neuroimaging.

2. Material and methods

2.1. Stimulus selection method

Our study is unique in that it relies on the dynamic selection of stimuli in a parameterized stimulus space, choosing new images to display based on the BOLD responses to previous images within a given pre-selected brain region. More specifically, we automatically choose the next stimulus to be shown by considering a space of visual properties and probing locations in this space (corresponding to stimuli with particular visual properties) in order to efficiently identify those locations that are likely — based on prior neural responses to other stimuli in this space — to elicit maximal activity from the brain region under study. As discussed in Secs. 2.8.3 and 2.9.3, we employed two somewhat different representational spaces, one based on SIFT features derived from real-world images, and one based on synthetic “Fribbles” objects (see Fig. 5). SIFT was used for the first group of ten subjects, while Fribbles were used for the second group of ten subjects. For both groups, each stimulus i that could be displayed is assigned a point in space p_i based on its visual properties. The measured response of a given brain region to this stimulus r_i is understood as:

$$r_i = f(p_i) + \eta \quad (1)$$

That is, a function f of the stimulus’ visual properties as encoded by its location in the representational space plus a noise term η , drawn from a zero-centered Gaussian distribution. The process of displaying an image, recording the ensuing cortical activity via fMRI, and isolating the response of the brain region of interest using the preprocessing program we model as performing an evaluation under noise of the function describing the region’s response. For simplicity’s sake, we perform stimulus selection assuming our chosen brain region has a selectivity function f that reaches a maximum at a certain point in the representational space and falls off with increasing Euclidean distance from this point. Our assumption is consistent with prior work in primate neurophysiology, such as Tanaka (2003), Hung et al. (2012), and Yamane et al. (2008), in which stimuli were progressively adapted to maximize response of a single neural unit to converge on the single (complex) visual selectivity presumed to be associated with the unit. We also note that our assumption is consistent with recent work in human fMRI that finds that selectivity for object categories is organized in a smooth gradient across cortex whereby the amount of neural “real estate” apportioned to shared features across visually-similar categories is minimized Huth et al. (2012). Under these assumptions, we use a modified version of the simplex simulated annealing Matlab code available from Donckels (2012), implementing the algorithm from Cardoso et al. (1996). This method seeks to identify new points (corresponding to stimuli) that evoke the highest responses from the selected cortical region. An idealized example of what a search run might look like based on this algorithm is shown in Fig. 1b. The results of our study indicate our assumption of a single peak in cortical response is not always accurate. Nonetheless, the simplex simulated annealing method achieves convergence for several real-time stimulus searches.

For each of four distinct stimulus classes — mammals, human-forms, cars, and containers for real-world objects and four classes distinguished by core body shape and appendage

orientation for Fribble objects (described further in Sec. 2.3 and in Leeds et al. (2014)) — we performed searches in each of two scan sessions. To probe the consistency of our search results across different initial simplex settings, we began the search within each session at a distinct point in the relevant stimulus representational space. In the first session, the starting position was set to the origin for a given stimulus class, as specific stimulus exemplars were distributed in each space relatively evenly around the origin. In the second scan session, the starting position was manually selected to be in a location opposite from the regions in which stimuli were visited most frequently and which produced the highest magnitude responses in the previous session. Additionally, if a given stimulus dimension was not explored during the first session, a random offset from the origin along that axis was selected for the beginning of the second session.

The starting locations for the simplexes for each display run beyond the first run in the session, that is, the i^{th} run, is set to be the simplex point that evoked the largest response from the associated cortical region in the $(i-1)^{\text{th}}$ run. At the start of the i^{th} display/search run, each simplex is initialized with the starting point $x_{i,1}$, as defined above, and D further points, $x_{i,d+1} = x_{i,1} + U_d v_d$ where D is the dimensionality of the space, U_d is a scalar value drawn from a uniform distribution between -1 and 1 , and v_d is a vector with d^{th} element 1 and all other elements 0 . In other words, each initial simplex for each run consists of the initial point and, for each dimension of the space, an additional point randomly perturbed from the initial point only along that dimension. The redefinition of each simplex at the start of each new run constitutes a partial search reset to more fully explore all corners of the feature space, yet maintaining some hysteresis from the location from the previous run that produced the most activity.

Further details of the simplex simulated annealing method are provided by Leeds (2013) and Cardoso et al. (1996).

2.2. Inter-program communication

Three programs run throughout each real-time search to permit dynamic selection and display of new stimuli most effectively probing the visual selectivity of a chosen cortical region. The programs — focusing on fMRI preprocessing, visual property search, and stimulus display tasks, respectively — are written and executed separately to more easily permit implementation and application of alternate approaches to each task. Furthermore, the display program runs on a separate machine from the other two processes, shown in Fig. 2, to ensure sufficient computational resources are dedicated to each task, particularly as analysis and display computations must occur simultaneously throughout each scan. The first machine, called the “analysis machine,” is an Optiplex 960 running Red Hat on an Intel Core 2 Duo processor at 3 GHz with 4 GB memory; the second machine, called the “Display machine,” is an Apple MacBook Pro (2008) running OS X on an Intel Core 2 Duo processor at 2.5 GHz with 4 GB memory.

Due to the division of tasks into three separate programs, each task relies on information determined by a different program and/or processor, as indicated in Fig. 2. The methods used to communicate the information necessary for preprocessing, search, and stimulus display are as follows:

- **Preprocessing program input** The scanner console machine receives brain volumes from the fMRI scanner and sends these volumes to the analysis machine disk in real-time. The preprocessing program checks the disk every 0.2 seconds to determine whether all the volumes for the newest block of search results—the full 10 s cortical responses to recently-shown stimuli — are available for analysis. The preprocessing program uses the data to compute one number to represent the response of the corresponding pre-selected brain region to its respective stimulus. The program proceeds to write the response into a file labeled `responseN` and then creates a second empty file named `semaphoreN`, where $N \in \{1, 2, 3, 4\}$ in each file is the number of the search being processed (see Sec. 2.3). The files are written into a pre-determined directory that is monitored by the search program, so the search program can find information saved by the preprocessing program. The creation of the `semaphoreN` file signals to the search program that the response of the brain region studied in the N^{th} search has been written fully to disk. This approach prevents the search program from reading an incomplete or outdated `responseN` file and acting on incorrect information.
- **Search program input** The search program rotates among four simultaneous searches for the visual feature selectivities of four different brain regions, that is, searching for the stimulus images containing features producing the highest possible activity in pre-selected cortical regions. At any given time during a real-time scan, the search program either is computing the next stimulus to display for a search whose most recent cortical response has recently been computed, or is waiting for the responses of the next block of searches to be computed. While waiting, the search program checks the pre-determined directory every 0.2 seconds for the presence of the semaphore file of the current search, created by the preprocessing program. Once the search program finds this file, the program deletes the semaphore file and loads the relevant brain region's response from the response file. The search program proceeds to compute the next stimulus to display, intended to evoke a high response from the brain region and sends the stimulus label to the display program running on the display machine.
- **Display program input** Two different methods were used for the transmission of stimulus labels between the search and display programs.
 - **Update Method 1** For our initial group of subjects ($N = 5$) — all presented with the real-world object stimuli — the search program sent each label to the display program by saving it in a file, `rtMsgOutN`, in a directory of the analysis computer mounted by the display computer. Immediately prior to showing the stimulus for the current search $N \in \{1, 2, 3, 4\}$ — the display program looked for the corresponding file in the mounted directory (rotating between four searches, as did the preprocessing and search programs).
 - **Update Method 2** For our remaining subjects — presented with either real-world or Fribble object stimuli — labels were

passed over an open socket from the Matlab (MATLAB, 2012) instance running the search program to the Matlab instance running the display program. In the socket communication, the search program paired each label with the number identifier N of the search for which it was computed. Immediately prior to showing the stimulus for any given current search, the display program read all available search stimulus updates from the socket until it found and processed the update for the current search and then showed the current stimulus to display for the current search.

Ordinarily, both techniques allowed the display program to present the correct new stimulus for each new trial, based on the computations of the search program. However, when preprocessing and search computations did not complete before the time the new stimulus was needed for display, the two communication techniques between the search and display programs had differing behaviors. As discussed in Sec. 3.1, we find the second method is preferable in that socket communication enables direct and immediate communication between the search and display programs once the search program has selected new stimuli to display. In contrast, the first method's writing of files to a mounted directory relies on periodic updates to shared files across the network, which is performed by operating system functions that may be delayed in execution beyond the control of our Matlab programs. Thus, the display program sometimes acts on outdated information in the local copy of its shared file before file updates have been completed.

It is also worth noting the first (file-update) method provides an occasional benefit over the second (socket-communication) method. At certain iterations, the search program will refrain from exploring a new simplex point for a given stimulus class. In this case, using the second method, the search program will not send a stimulus update over the socket and the display program will pause several seconds while awaiting an update through the socket. Using the first method, the display program will present a stimulus at the proper time interval regardless, using the stimulus saved in the shared file at the previous iteration. In practice, this beneficial behavior of update method 1 is outweighed by method 1's relatively slower communication of new stimulus choices. Furthermore, the second method can be repaired in future studies by implementing a simple alteration of our code in which the search program selects a blank screen or a default object stimulus each time a simplex computation is skipped.

2.3. Interleaving searches

We explored the selectivity to specific visual images for four distinct preselected brain regions within ventral cortex. Brain regions were selected based on criteria discussed in Secs. 2.8.5 and 2.9.5 using data collected from an earlier scanning session for each subject. For each brain region, a distinct search was performed using stimuli drawn from a single class of visual objects. For each brain region, a unique search was performed using a distinct class of visual object stimuli. For real-world object stimuli, the four stimulus classes were mammals, human-forms, cars, and containers, as shown in Fig. 4a. For Fribble object

stimuli, the four classes were distinguished by core body shape and color as well as by orientation of appendages, as shown in Fig. 5a.

To use scanning time most efficiently, four searches were performed that interrogated the four different pre-selected brain regions during each scan. Following the stimulus onset, a 10–14 s interval¹ is required to gather the 10 s cortical response to the stimulus and an additional ~10 s is required to process the response and to select the next stimulus for display. While the next stimulus for a given search is being selected, the display program rotates to another search, maximizing the use of limited scan time to study multiple brain regions. The display and analysis programs also rotate in sequence among the four searches — that is, Search 1 → Search 2 → Search 3 → Search 4 → Search 1 ... As discussed above, different classes of real-world and Frabble objects were employed for each of the four searches. More generally, alternation among visually-distinct classes is an advantage to our approach in that it decreases the risk of cortical adaptation present if multiple similar stimuli are viewed in direct succession. Note that the specific nature of each visual class is not critical to our methods. While we studied cars and mammals, we anticipate a search would work equally well for any two relatively unrelated categories, for example, buildings and fish.

The preprocessing program evaluates cortical responses in blocks of two searches at a time — that is, the program waits to collect data from the current stimulus displays for Search 1 and Search 2, analyzes the block of data, waits to collect data from the current stimulus displays for Search 3 and Search 4, analyzes this block of data, and then repeats the sequence. This grouping of stimulus responses increases overall analysis speed. Several steps of preprocessing require the execution of AFNI (Pittman, 2011) command-line functions. Computation time is expended to initialize and terminate each function each time it is called, independent of the time required for data analysis. By applying each function to data from two searches together, the “non-analysis” time across function calls is minimized.

2.4. Stimulus display

All stimuli were presented using MATLAB (2012) and the Psychophysics Toolbox (Brainard, 1997, Pelli, 1997) running on an Apple MacBook Pro (2008) running OS X. Images were displayed on a BOLDscreen (Cambridge Research, Inc.) 24 inch MR compatible LCD display located at the head end of the scanner bore. Subjects viewed the screen through a mirror attached to the head coil with object stimuli subtending a visual angle of approximately $8.3 \text{ deg} \times 8.3 \text{ deg}$. During the real-time search scans, each stimulus was displayed for 1 s followed by a centered fixation cross that remained displayed until the end of each 8 s trial, at which point the next trial began. The 8 s trial duration is chosen to be as short as possible while providing sufficient time for the real-time methods to compute and determine the next stimuli to display based on the previous cortical responses. Further experimental design details are provided in Secs. 2.8.4 and 2.9.4.

¹The 4 s beyond the duration of the cortical response accounts for communication delay between the fMRI scanner and the machine running the preprocessing and search programs.

2.5. fMRI Procedures

Subjects were scanned using a 3 T Siemens Verio MRI scanner with a 32-channel head coil. Functional images were acquired with a $T2^*$ -weighted echoplanar imaging (EPI) pulse sequence (31 oblique axial slices, in-plane resolution $2\text{mm} \times 2\text{mm}$, 3mm slice thickness, no gap, sequential descending acquisition, repetition time $\text{TR} = 2000\text{ms}$, echo time $\text{TE} = 29\text{ms}$, flip angle = 72° , GRAPPA = 2, matrix size = 96×96 , field of view $\text{FOV} = 192\text{ mm}$). An MP-RAGE sequence ($1\text{mm} \times 1\text{mm} \times 1\text{mm}$, 176 sagittal slices, $\text{TR} = 1870$, $\text{TI} = 1100$, $\text{FA} = 8^\circ$, GRAPPA = 2) was used for anatomical imaging.

2.6. Experimental design

For each subject, our study was divided into an initial reference scanning session and two real-time scanning sessions (Fig. 3a). In the reference session we gathered cortical responses to four classes of object stimuli to identify cortical regions selective for each separate stimulus class. As discussed in Secs. 2.8 and 2.9, two different stimulus sets, comprised of four visually-similar object classes, were used to explore visual feature selectivity: real-world objects and synthetic Fribble objects; each subject viewed stimuli from only one set. In the real-time scan sessions we used our real-time imaging methods to search for stimuli producing the highest possible responses from each of the four cortical brain regions identified during the reference scan, dynamically choosing new stimuli based on the response of each region to recently shown stimulus images.

Runs in the **reference scan session** used a slow event-related design. Each stimulus was displayed in the center of the screen for 2 s followed by a blank 53% gray screen shown for a time period randomly selected to be between 500 and 3000 ms, followed by a centered fixation cross that remained displayed until the end of each 10 s trial, at which point the next trial began. As such, the SOA between consecutive stimulus displays was fixed at 10 s. Subjects were instructed to press a button when the fixation cross appeared. The fixation onset detection task was used to engage subject attention throughout the experiment. No other task was required of subjects, as such, the scan assessed object perception under passive viewing conditions. Further details about the reference scan are provided in Leeds et al. (2014).

Across two 1.5-hour **real-time scan sessions**, we explored the selectivity to specific visual images for four distinct brain regions within ventral cortex, each assigned to a distinct search. Stimuli were presented for each search in 8.5-minute “search” runs (4 to 8 runs were used per subject depending on other factors). Each stimulus was selected by the real-time search program based on responses of a pre-selected region of interest (ROI) to stimuli previously shown from the same object category. Task details are provided in Secs. 2.8.4 and 2.9.4.

Each real-time session began with a 318-second functional scan performed with a viewing task to engage subject attention. For a given subject, the first functional volume scanned for this task was used to align the ROI masks (defined in Secs. 2.8.5 and 2.9.5) selected in the reference session to that subject’s brain position in the current session. Further details of this initial scan are provided in Leeds et al. (2014).

2.7. Preprocessing

During real-time scan sessions, at the beginning of each run functional volumes were motion corrected using AFNI. Polynomial trends of orders one through three were removed. The data then were normalized for each voxel by subtracting the average and dividing by the standard deviation, obtained from the currently analyzed response and from the previous reference scan session, respectively, to approximate zero-mean and unit variance (Just et al., 2010). The standard deviation was determined from ~1 hour of recorded signal from the reference scan session to gain a more reliable estimate of signal variability in each voxel. Due to variations in baseline signal magnitude across and within scans, each voxel's mean signal value required updating based on activity in each block (the time covering the responses for two consecutive trials). To allow multivariate analysis to exploit information present at high spatial frequencies, no spatial smoothing was performed (Swisher et al., 2010).

Matlab was used to perform further processing on the fMRI time courses for the voxels in the cortical region of interest for the associated search. For each stimulus presentation, the measured response of each voxel consisted of five data samples starting 2 s/1 TR after onset. Each five-sample response was consolidated into a weighted sum by computing the dot product of the response and the average hemodynamic response function (HRF) for the associated region. The HRF was determined based on data from the reference scan session. The pattern of voxel responses across the region was consolidated further into a single scalar response value by computing a similar weighted sum. Like the HRF, the voxel weights were determined from reference scan data. The weights corresponded to the most common multi-voxel pattern observed in the region during the earlier scan; that is, the first principal component of the set of multi-voxel patterns. This projection of recorded real-time responses onto the first principal component treats the activity across the region of interest as a single locally-distributed code, emphasizing voxels whose contributions to this code are most significant and de-emphasizing those voxels with typically weak contributions to the average pattern.

During the alignment run of each real-time session, AFNI was used to compute an alignment transformation between the initial functional volume of the localizer and the first functional volume recorded during the reference scan session. The transformation computed between the first real-time volume and the first reference volume was applied in reverse to each voxel in the four ROIs determined from the reference scan.

More standard preprocessing methods were used for the reference scan; as such, preprocessing steps for the reference scan can be found in Leeds et al. (2014).

2.8. Real-world objects embedded in SIFT space

We pursued two methods to search for visual feature selectivity. In our first method, we focused on the perception of real-world objects with visual features represented by the scale invariant feature transform (SIFT, Lowe (2004)).

2.8.1. Subjects—Ten subjects (four female, age range 19 to 31) from the Carnegie Mellon University community participated, provided written informed consent, and were monetarily compensated for their participation. All procedures were approved by the Institutional Review Board of Carnegie Mellon University.

2.8.2. Stimuli—Stimulus images were drawn from a picture set comprised of 400 distinct color object photos displayed on 53% gray backgrounds (Fig. 4a). The photographic images were taken from the Hemera Photo Objects dataset (Hemera Technologies, 2000–2003). The number of distinct exemplars in each object class varied from 68 to 150 object images. Note that our use of real-world images of objects rather than the hand-drawn or computer-generated stimuli employed in past studies of intermediate-level visual coding (e.g., Cadieu et al. (2007) and Yamane et al. (2008)) was intended to better capture a broad set of naturally-occurring visual features.

2.8.3. Defining SIFT space—Our real-world stimuli were organized into a Euclidean space that was constructed to reflect a scale invariant feature transform (SIFT) representation of object images (Lowe, 2004). Leeds et al. (2013) found that a SIFT-based representation of visual objects was the best match among several machine vision models in accounting for the neural encoding of objects in mid-level visual areas along the ventral visual pathway. The past success of SIFT as a model for mid-level visual representation in the brain (Leeds et al., 2013) lends the model to study of visual properties of interest for diverse visual classes, from the cars and mammals examined in our current study to faces, tools, dwelling-places and beyond. The SIFT measure groups stimuli according to a distance matrix for object pairs (Leeds et al., 2013). In our present work, we defined a Euclidean space based on the distance matrix using Matlab's implementation of metric multidimensional scaling (MDS, Seber (1984)). MDS finds a space in which the original pairwise distances between data points — that is, SIFT distances between stimuli — are maximally preserved for any given n dimensions. This focus on maintaining the SIFT-defined visual similarity groupings among stimuli — using MDS — was motivated by the observations of Kriegeskorte et al. (2008) and Edelman and Shahbazi (2012), both of whom argued for the value of studying representational similarities to understand cortical vision.

The specific Euclidean space used in our study was derived from a SIFT-based distance matrix for 1600 Hemera photo objects, containing the 500 stimuli available for display across the real-time searches, as well as 1100 additional stimuli included to further capture visual diversity across the appearances of real-world objects (nb. ideally, the object space would be better covered by many more than 1600 objects, however, we necessarily had to restrict the total number of objects in order to limit the computation time required to generate large distance matrices). Details on the computation of the distance matrix are provided by Leeds et al. (2014). MDS was then used to generate a Euclidean space into which all stimulus images were projected. The real-time searches for each object class operated within the same MDS space. This method produced an MDS space containing over 600 dimensions. Unfortunately, as the number of dimensions in a search space increases, the sparsity of data in the space will increase exponentially. As such, any conclusions regarding the underlying selectivity function will become increasingly more uncertain absent further

search constraints. To address this challenge, we constrained our real-time searches to use only the four most-representative dimensions from MDS space.²

2.8.4. Experimental design—Search runs in the real-time scan sessions employed a one-back location task to engage subject attention throughout the experiment. Each stimulus was displayed centered on one of nine locations on the screen for 1 s followed by a centered fixation cross that remained until the end of each 8 s trial, at which point the next trial began. Subjects were instructed to press a button when the image shown in this subsequent trial was centered on the same location as the image shown in the previous trial. The specific nine locations were defined by centering the stimulus at +2.5, 0, or –2.5 degrees horizontally and/or vertically displaced from the screen center. From one trial to the next, the stimulus center shifted with a 30% probability. Differences in the display location across stimuli were kept small to enhance subject attention in a difficult task and to minimize the effects of location shift on visual responses in the brain regions of interest.

2.8.5. Selection of regions of interest (ROIs)—Reference scan data was used to select ROIs for further study in real-time scan sessions. Single-voxel and voxel-searchlight analyses, described by Leeds et al. (2014), were used to find class-selective and SIFT-representational regions in the ventral stream. For each class, a 125 voxel cube-shaped ROI was selected. The use of relatively small — one cubic centimeter — cortical regions make it more likely that our methods will reveal information regarding local neural selectivities for complex visual properties. This assumption is based on analyses that were successfully pursued on similar spatial scales in Leeds et al. (2013), using 123-voxel searchlights.

2.9. Fribble objects embedded in Fribble space

Our second approach to searching for visual feature selectivity focused on the perception of synthetic novel objects — Fribbles — in which visual features were parameterized as interchangeable 3D components (Williams and Simons, 2000).

2.9.1. Subjects—Ten subjects (six female, age range 21 to 43) from the Carnegie Mellon University community participated, provided written informed consent, and were monetarily compensated for their participation. All procedures were approved by the Institutional Review Board of Carnegie Mellon University.

2.9.2. Stimuli—Stimulus images were generated based on a library of synthetic objects known as Fribbles (Williams and Simons, 2000; Tarr, 2013), and were displayed on 54% gray backgrounds as in Sec. 2.8.2. Fribbles are creature-like objects composed of colored, textured, geometric volumes. They are divided into classes, each defined by a specific body form and a set of four locations for attached parts. In the library, each appendage has three potential shapes, for example, a circle, star, or square head for the first class in Fig. 5a, with potentially variable corresponding textures. In contrast to the more natural, but less

²Given n dimensions, it is preferable to explore at least 2^n stimuli, more preferably 3^n , constituting two to three locations along each dimensional axis. 16 stimulus displays are performed for each of the four object classes in each of five to seven runs during each real-time session. Thus, there are a total of $16 \times 5 = 80$ to $16 \times 7 = 112$ stimulus displays per object class. Employing the more-conservative 3^n stimuli requirement, $\log_3 80 \approx 4$ dimensions are most appropriate for the number of available stimulus displays.

parameterized real-world objects, Fribble stimuli provide good control for the varying properties shown to subjects.

2.9.3. Defining Fribble space—We organized our Fribble stimuli into Euclidean spaces. In the space for a given Fribble class, movement along an axis corresponded to morphing the shape of an associated appendage. For example, for the purple-bodied Fribble class, the axes were assigned to: 1) the tan head; 2) the green tail tip; and 3) the brown legs, with the legs grouped and morphed together as a single appendage type. Valid locations on each axis spanned from -1 to 1 representing two end-point shapes for the associated appendage (e.g., a circle head or a star head). Appendage appearance at intermediate locations was computed through the morphing program Norrkross MorphX (Wennerberg, 2009) based on the two end-point shapes. Example morphs can be seen in the Fribble space visualization in Fig. 5b.

For each Fribble class, stimuli were generated for each of 7 locations — the end-points -1 and 1 as well as coordinates -0.66 , -0.33 , 0 , 0.33 , and 0.66 — on each of 3 axes, that is, $7^3 = 343$ locations. A separate space was searched for each class of Fribble objects.

Note that, in contrast to our approach to building a space for real-world objects (which might apply to any set of images), the methods used to build an object space for Fribbles necessarily rely on using objects that are, or can be, parametrized across all members of the class. As such, the real-world object method, while perhaps not ideal in all respects, is likely to be applicable to a much wider range of experimental designs.

2.9.4. Experimental design—Search runs in the real-time scan sessions employed a dimness detection task to engage subject attention throughout the experiment. Each stimulus was displayed in the center of the screen for 1 s followed by a centered fixation cross that remained displayed until the end of each 8 s trial, at which point the next trial began. On any trial there was a 10% chance the stimulus would be displayed as a darker version of itself — namely, the stimulus' red, green, and blue color values each would be decreased by 50 (max intensity 256). Subjects were instructed to press a button when the image appeared to be “dim or dark.” For the Fribble stimuli, the dimness detection task was used to address a specific concern with the one-back location task: that it required subjects to hold two objects in memory simultaneously, thereby possibly adding noise to any measure of the neural responses associated with single objects. Indeed, this issue may have limited the strength of real-world object search results. As such, somewhat cleaner results were expected using the dimness detection task.

2.9.5. Selection of Fribble class regions of interest—We employed the representational dissimilarity matrix-searchlight procedure discussed in Leeds et al. (2013) to identify those cortical areas whose encoding of visual information was well characterized by each Fribble space. ROIs were selected manually from these areas for study during the real-time scan sessions in which we searched for complex featural selectivities within the associated Fribble space.

2.10. Metrics for search performance

We expected each search in visual feature space to show the following two properties:

1. **Convergence** onto one, or a few, location(s) in the associated visual space producing greatest cortical response, corresponding to local neural selectivity.
2. **Consistency** in stimuli found to be preferred by the ROI, despite differing search starting points in visual feature space across the two scanning sessions.

Metrics were defined for both Convergence and Consistency and applied to all search results to assess the behavior of our real-time stimulus selection method.

Due to the variability of cortical responses and the noise in fMRI recordings, analyses were focused on stimuli that were visited three or more times. The average response magnitude for stimuli visited multiple times is more reliable with respect to conclusions of underlying ROI selectivity. Furthermore, repeat visits may indicate the implicit importance of a stimulus to the response of a given ROI, in that, as determined by the search, increased interrogation of a local region in feature space indicates that the algorithm “expects” higher responses in that region.

2.10.1. Convergence—For a given class, convergence was computed based on the feature space locations of the visited stimuli \mathcal{S} , and particularly the locations of stimuli visited three or more times, $\mathbf{S}_{\text{thresh}}$. The points in $\mathbf{S}_{\text{thresh}}$ were clustered into groups spanning no more than d distance in the associated space based on average linkage, where $d = 0.8$ for Frubble spaces and $d = 0.26$ for SIFT space.³ The result of clustering was the vector $\mathbf{clusters}_{\mathbf{S}_{\text{thresh}}}$, where each element contained the numeric cluster assignment (from 1 to N) of each point in $\mathbf{S}_{\text{thresh}}$. The distribution of cluster labels in $\mathbf{clusters}_{\mathbf{S}_{\text{thresh}}}$ was represented as $\mathbf{p}_{\text{clust}}$, where the n^{th} entry $p_{\text{clus}}(n)$ is the fraction of $\mathbf{clusters}_{\mathbf{S}_{\text{thresh}}}$ entries with the cluster assignment n .

Conceptually, convergence is assessed as follows based on the distribution of points, that is, stimuli visited at least three times:

- If all points are close together, that is, in the same cluster, the search is considered to have converged.
- If most points are in the same cluster and there are a “small number” of outliers in other clusters, the search is considered to have converged sufficiently.
- If points are spread widely across the space, each with its own cluster, there is no convergence.

Set as an equation, the convergence metric is:

$$\text{metric}(\mathbf{S}) = \|\mathbf{p}_{\text{clust}}\|_2 - .1\|\mathbf{p}_{\text{clust}}\|_0 \quad (2)$$

where $\|\mathbf{p}_{\text{clust}}\|_2 = \sqrt{p_{\text{clust}}(1)^2 + \dots + p_{\text{clust}}(N)^2}$ and $\|\mathbf{p}_{\text{clust}}\|_0$ is the number of non-zero entries of $\mathbf{p}_{\text{clust}}$. The metric awards higher values when $\mathbf{p}_{\text{clust}}$ element entries are high (most points are in a small number of clusters) and the number of non-zero entries is small (there are few clusters in total). Eqn. 2 pursues a strategy related to that of the elastic net, in which

³The distance thresholds were chosen based on empirical observations of clusterings across regions and subjects in each space.

\mathcal{L} and \mathcal{L} norms are added to award a vector that contains a small number of non-zero entries, all of which have small values (Zou and Hastie, 2005).

2.10.2. Consistency—For each subject and each stimulus class, search consistency was determined by starting the real-time search at a different location in feature space at the beginning of each of the two search scan sessions. If the second search returns to the locations frequently visited by the first search, despite starting distant from those locations, the search method shows consistency across initial conditions.

The metric for determining consistency of results across search sessions was a slight modification of the convergence metric. The locations of the stimuli visited three or more times in the first and second searches were stored in $\mathbf{S}_{\text{thresh}}^1$ and $\mathbf{S}_{\text{thresh}}^2$, respectively. The two groups were concatenated into $\mathbf{S}_{\text{thresh}}^{\text{both}}$ taking note which entries came from the first and second searches. Clustering was performed as above and labels were assigned into the variable $\mathbf{clusters}_{\mathbf{S}_{\text{thresh}}^{\text{both}}}$. The distribution of cluster labels was represented as probabilities $\mathbf{P}_{\text{clustBoth}}$.

To measure consistency, the final metric in Eqn. 2 was applied only to entries of $\mathbf{P}_{\text{clustBoth}}$ for which elements of $\mathbf{S}_{\text{thresh}}^1$ and $\mathbf{S}_{\text{thresh}}^2$ were present:

$$\text{metric}(\mathbf{S}^{\text{both}}) = \|\mathbf{P}_{\text{clustBoth}}(i \in \mathbf{B})\|_2 - .1 \|\mathbf{P}_{\text{clustBoth}}(i \in \mathbf{B})\|_0 \quad (3)$$

where \mathbf{B} is the set of indices i such that cluster i contains at least one point from $\mathbf{S}_{\text{thresh}}^1$ and from $\mathbf{S}_{\text{thresh}}^2$. The metric awards the highest values for convergence if there is one single cluster across search sessions. A spread of points across the whole search space visited consistently between sessions would return a lower value. Complete inconsistency would leave no $\mathbf{P}_{\text{clustBoth}}$ entries to be added, returning the minimum value of 0.

2.10.3. Testing against chance—A variant of the permutation test is used to assess the metric results. The null hypothesis is that the convergence or consistency measure computed for a given search or pair of searches, based on clustering of the k stimuli visited three or more times during the search(es), would be equally likely to be found if the measure were based on clustering of a **random** set of k stimuli; this random set is chosen from the stimuli visited **one** or more times during the same search(es). The group of stimuli visited one or more times is considered a conservative estimate of all stimuli that could have been emphasized by the search algorithm through frequent visits. In the permutation test, the designation “displayed three or more times” is randomly reassigned among the larger set of stimuli displayed one or more times to determine if a random set of stimuli would be considered similarly convergent or consistent as the set of stimuli frequently visited in my study. More specifically, indices are assigned to all points visited in Search 1 and Search 2, \mathcal{S}^1 and \mathcal{S}^2 , respectively, the indices and recorded number of visits are randomly permuted, and $\text{metric}(\mathcal{S}^1)$, $\text{metric}(\mathcal{S}^2)$ and $\text{metric}(\mathcal{S}^{\text{both}})$ are computed based on the locations randomly assigned to each frequently-visited point. For each subject and each search, this process is repeated 500 times, the mean and standard deviation are computed, and the Z score for the

original search result metrics are calculated. Based on visual inspection, searches with z 1.8 are considered to mark notably non-random convergence or consistency.

2.10.4. Temporal evolution—We studied the movement of the search through visual space for each stimulus class search and each session by comparing the distribution of stimulus locations visited during the first and second half of the session. We characterized these distributions by their mean and variance.

To assess the changing breadth of visual space examined across a search session, we divide the stimulus-points into those visited in the first half of the session and those visited in the second half:

$$\Delta \text{var} = \sum_j (\sigma^2(X_2^j)) - \sum_j (\sigma^2(X_1^j)) \quad (4)$$

where $\text{sigma}^2(\cdot)$ is the variance function and X_i^j is the set of coordinates on the j^{th} axis for the i^{th} half of the session. var pools variance across dimensions by summing. More fine covariance structure is ignored as the measure is intended to test overall contraction across all dimensions rather than changes in the general shape of the distribution.

To assess the changing regions within visual space examined across a search session, we again compared points visited in the first half of the session with those in the second half of the session:

$$\text{dist} = \sqrt{\sum_j \frac{(X_1^j - X_2^j)^2}{s_j^2}} \quad (5)$$

where X_i^j are as defined for Eqn. 4 and $s_j^2 = \frac{[\sigma_j^2]_1 + [\sigma_j^2]_2}{2}$ is the mean variance along the j^{th} dimension of the point locations visited in the two halves of the search session. dist measures the distance between the mean location of points visited in the first and second halves of the search session, normalized by the standard deviation of the distributions along each dimension — similar to the Mahalanobis distance (Mahalanobis, 1936) using a diagonal covariance matrix. A shift of 0.5 on a dimension with variance 0.1 will produce a larger metric value than a shift of 0.5 on a dimension with variance 1.0.

3. Results

Our methods are designed to more rapidly identify complex visual properties used in the neural representation of objects within the human ventral pathway. Because these search methods are somewhat novel, we also assessed and confirmed their expected performance. Specifically, we first studied the timing of the stimulus displays, as executed by the display program, as well as the stability of real-time computation of ROI responses to stimuli, as executed by the preprocessing program. We then proceeded to examine the locations in visual space visited by each real-time search.

3.1. Display program behavior

Within our real-time approach to studying the visual cortex, **the display program's central task is to display each intended stimulus** as chosen by the search program **at its intended time** (i.e., at the beginning of its associated 8 s trial, described in Sec. 2.4). Unfortunately, in the course of each real-time session, challenges periodically arose to the prompt display of the next stimulus to explore in each real-time search. The computations required to determine ROI response to a recent stimulus and to determine the next stimulus to display did not always (and were not guaranteed to) complete before the time required by the display program to show the next search selection. When the new stimulus choice was not made sufficiently quickly, the stimulus displayed to the subject could be shown seconds delayed from its intended onset time or could incorrectly reflect the choice made from the previous iteration of the search, depending on the stimulus update method used by the display program.

Of the two stimulus update methods used by the display program, as explained in Sec. 2.2, **Update Method 1** was more sensitive to this potential problem. As such, only for the first five subjects viewing real-world objects did the display program receive the search program's next stimulus choice by reading a file in a directory shared between the machines respectively running the display program and the search program. For the remainder of the subjects, five viewing real-world objects and ten viewing Fribble objects, we employed **Update Method 2** in which the display program received the search program's next stimulus choice through a dedicated socket connection. This method improved on the notable delays in updates to the shared files observed for Method 1 (Table 1).

The relative benefits of Update Method 2 over Update Method 1 are studied in the context of the hardware and software configurations of our Analysis and Display machines. As such, our findings in this section provide important technical insights in engineering that we hope will ultimately be useful in advancing our understanding of human high-level vision.

3.1.1. Real-world objects search—The number of displays that appeared late or showed the wrong stimulus for subjects viewing real-world objects is presented in Table 1 for each subject, object class, and scan session. Stimulus presentations were considered delayed if they were shown 0.5 s or more past the intended display time. Below, we first discuss display errors for stimulus update method 1, then we discuss display errors for stimulus update method 2.

Update Method 1 When updates for display stimuli were performed through inspection of shared files, for S6, S7, S8, S9, and S10, showing of incorrect stimuli dominated the display errors. S6, S7, and S8 were shown incorrect stimuli for 15 to 42% of trials for search 1 and search 3, corresponding to the mammal and car classes. Among these three subjects, incorrect displays for Search 2 and Search 4 only were observed in Session 2 for S6. S9 was shown no incorrect stimuli; S10 was shown incorrect stimuli on ~10% of trials for all searches in Session 1 and no incorrect stimuli in Session 2. Despite the frequency of incorrect stimuli displayed in searches for Object Classes 1 and 3, it is important to note that even in the worst case, correct stimuli were displayed on over half of the trials.

Note that even when stimuli are chosen 1 s prior to display time, updates through the shared files read over the mounted folder may require as much as 3 s to complete, resulting in the display program reading and acting on old stimulus choices. These sources of typically 1 to 5 s delays past display time in conjunction with the block processing method result in the strong discrepancy in incorrect display frequency of Search 1 and Search 3—whose updates sometimes did not arrive to the display computer by the required time — compared with that of Search 2 and Search 4 — whose updates usually arrived at least 3 s before they are needed.

When updates for display stimuli were performed through inspection of shared files, display errors also included a limited number of delayed displays. S9 and S10 had delayed stimulus displays for 1 to 6% of trials, with delay on at least one trial for every session and for each of the four searches. In most cases, there were more delays for Search 1 than for any of the other searches. These delays likely resulted from the directory update performed by the display program prior to reading the file containing the stimulus choice for the current search. The update operation usually executes in a fraction of a second, but occasionally runs noticeably longer. Chances of a longer-duration update are greater when the operation has not been performed recently, such as at the start of a real-time search run following a ~2 minute break between runs. As Search 1 starts every run, it may be slightly more likely to experience display delays.

Update Method 2 When updates for display stimuli were performed through a socket, for S1, S2, S3, S4, and S5, display delays dominated the errors in display program performance. Most subjects had delayed stimulus displays for 1 to 9% of trials, with delay on at least one trial for every session and for each of the four searches. However, the second session for S1 showed no delayed displays, nor did Search 4 for the second session for S2. The number of delays for Search 1 was greater than or (occasionally) equal to the number of delays for any of the other searches, except for Session 1 for S3 for which Search 3 had the most delays. Across the five subjects, Search 3 had the second, or sometimes first, highest number of delayed displays. The discrepancy in display error frequency between the first searches of each processing “block” as described above, that is, Search 1 and Search 3, and the second searches of each processing block, Search 2 and Search 4, are significantly less pronounced than they were for the frequency of incorrect stimuli for S6, S7, S8, S9, and S10, though the pattern remains weakly observable. For S1, S2, S3, S4, and S5, display delays can result from delays in completing processing of cortical responses for the block of two recently viewed stimuli — causing a greater number of delays for Search 1 and Search 3. As described in Sec. 2.3, the next stimuli to display are computed and provided to the display in blocks of two — Search 1 and Search 2 selections are provided together and Search 3 and Search 4 selections are provided together.

A limited number of incorrect stimulus displays also occurred when updating display stimuli through a socket. S3 and S4 were shown incorrect stimuli on 1 to 3% of trials for one or two searches in each scan session. The source of these errors was not determined, though they may have resulted from skipped evaluations in the simplex search. These errors did not occur using socket updates for searches of Frizzle object stimuli reported below.

Far fewer display errors occurred when updates for display stimuli were performed over a socket than when they were performed through inspection of a shared file. Indeed, the socket update approach was introduced to improve communication speed between the search program and the display program and, thereby, to decrease display errors. Reflecting on the increased performance caused by use of sockets, we employed only socket communication for the Fribble objects searches.

3.1.2. Fribble objects search—The number of displays that appeared late for subjects viewing Fribble objects — which always relied on Update Method 2 — is shown in Table 2 for each subject, object class, and scan session. Stimulus presentations were considered delayed if they were shown 0.5 s or more past the intended display time. There were no displays showing the wrong stimuli, because the display program waited for updates to each stimulus over an open socket with the search program before proceeding with the next display.

All subjects had delayed stimulus displays in each scan session in one or more of the four searches. Across subjects, a total of ~70% of searches showed delayed displays, with errors occurring in 1 to 10% of trials. The number of delays for Search 1 was greater than the number of delays for any of the other searches; across subjects, Search 1 had roughly three times as many errors as any of the other classes. As with the real-world objects, these delays in displaying Fribble stimuli were produced by delays in the completion of fMRI signal preprocessing and by skipped simplex search evaluations.

The first block processed for each run requires slightly extra time for processing than does any other block, because the first block contains six extra volumes, corresponding to the cortical activity prior to the start of the first display trial. Often, this extra processing time causes a delay for the first update of Search 1. This slow start to preprocessing also contributes to the larger number of delayed displays for Search 1 observed in subjects viewing real-world objects, shown in Table 1, though the effects are much more pronounced for Fribble subjects than for real-world object subjects.

Overall, display program performance was quite good for subjects viewing Fribble stimuli. Correct stimuli were displayed on at least 90% of trials, and usually more, for each subject, session, and search.

3.2. Preprocessing program behavior

The preprocessing program's central task was to act in real-time to compute the responses of pre-selected ROIs to recently shown stimuli. To rapidly convert raw fMRI signal to ROI response values, standard preprocessing methods were used to remove scanner and motion effects from blocks of fMRI data, followed by methods for extracting and summarizing over selected voxel activities. In more typical, that is, non-real-time, analyses, a larger array of preprocessing methods would be employed over data from the full session to more thoroughly remove signal effects irrelevant to analysis. However, a somewhat more conservative approach to preprocessing was used here to enable reasonable performance for real-time analysis, real-time stimulus selection, and real-time search of stimulus spaces.

Of note, this truncated preprocessing may lead to inaccuracies in measures of brain region responses, misinforming future search choices. To investigate this potential concern, we compared the correlation between computed ROI responses computed using preprocessing employed during the real-time sessions (Sec. 2.7) and the computed responses using “offline” preprocessing considering all runs in a scan session together, and following the drift and motion correction as well as normalization methods of Leeds et al. (2014). We considered the effects of correcting for subject motion in the scanner using real-time preprocessing over a limited set of volumes compared to offline preprocessing across BOLD data from the full session.

Our preprocessing program aligned fMRI volumes in each time block to the first volume of the current 8.5-minute run, rather than to the first volume recorded in the scanning session. To extract brain region responses for each displayed stimulus, voxel selection is performed based on ROI masks aligned to the brain using the first volume recorded in the scan session (Sec. 2.7), under the assumption voxel positions will stay relatively fixed across the session. Significant motion across the scan session could potentially place voxels of interest outside the initially-aligned ROI mask as the session proceeds, or cause voxels to be misaligned from their intended weights used in computing the overall ROI stimulus response (Sec. 2.7). In our analysis of preprocessing program performance, we track subject motion in each scan session and note its effects on the consistency between responses computed in real-time and offline.

While there were some inconsistencies between responses computed by the offline and real-time methods, particularly under conditions of greater subject motion, we observe that real-time computations are generally reliable across subjects and sessions. This reliability is particularly strong for subjects viewing Frubble objects rather than real-world objects, for reasons detailed below.

3.2.1. Real-world objects search—Consistency between ROI responses computed in real-time and responses computed offline for subjects viewing real-world objects are shown in Table 3 for each subject, object class, and scan session. Consistency was measured as the correlation between responses computed by the two methods for each display of each trial.

Correlation values were modestly strong and positive. Approximately 50% of searches produced correlations of 0.3 or above, and 20% produced correlations of 0.5 or above. Notably, 5 of the 17 searches producing negative correlations showed values below -0.3 , pointing to a marked negative trend between the two methods. Consistent misalignment of positive and negative voxel weights when combining voxel activity to form a single regional response to a stimulus may consistently invert the sign of the computed real-time response. Effects of this inversion on search behavior are considered in Sec. 3.3.

Correlation values can vary dramatically within a given subject and session across ROIs. Real-Time Session 1 for Subject S4 and real-time Session 2 for Subjects S6 and S7 show correlations that are high and low, positive and negative across stimulus class searches. Searches for Stimulus class 1 and 3 show high correlations across subjects. At first consideration, this within-session variability is quite surprising, as all regions presumably

are affected by the same subject movement and scanner drift. However, brain regions differ in the form of the multi-voxel patterns that constitute their response. Patterns that are more broad in spatial resolution, with voxels responding similarly to their neighbors, are less affected in their appearance if subject movement shifts the ROI ~2 mm from its expected location. High-resolution patterns, in which neighboring voxels exhibit opposite-magnitude responses to a stimulus, are harder to analyze correctly when shifted. Significant angular motion also could produce differing magnitudes of voxel displacement for ROIs closer and farther from the center of brain rotation.

We considered head motion as an important potential source of inconsistency between computed responses. In particular, we expected increased motion would cause increased inconsistency between real-time and offline computations. The expected pattern is weak but apparent when viewing correlation values sorted by subject motion, shown in Fig. 6. Sessions with the least motion are in the top rows and sessions with the most motion are in the bottom rows; colors correspond to correlation values and are sorted from lowest to highest in each row for ease of visualization. Studying the search with the lowest correlation — the left-most column — per subject and session reveals sessions containing two to three searches with low correlation values, corresponding to green and cyan colors, are predominantly seen when there is greater subject motion. However, all sessions contain searches with high correlations, and the search with the most motion, S3₂ in the bottom row, contains three high-correlation searches.

3.2.2. Fribble objects search—Consistency between ROI responses computed in real-time and responses computed offline for subjects viewing Fribble objects are shown in Table 4 for each subject, object class, and session. Consistency was measured as the correlation between responses computed by the two methods for each display of each trial.

Correlation values were low but generally positive, and higher than those observed in the real-time objects searches. 75% of searches produced correlations of 0.2 or above, and more than 50% produced correlations above 0.45. 7 of the 17 searches producing negative correlations showed values equal to or below -0.4, pointing to a marked negative trend between the two methods. The potential mechanism for a consistent inversion in the sign, for example, +3 becomes -3, of the computed ROI responses is discussed above for subjects viewing real-world objects.

Notably, correlation values can vary dramatically within a given subject and session across ROIs. Real-Time Session 1 for Subject S20 shows correlations that are high and low, positive and negative across stimulus class searches. Nonetheless, within-session variation is notably less pronounced for subjects viewing Fribble objects compared to subjects viewing real-world objects. Importantly, 12 of the 20 sessions, with each session corresponding to a row in the Figure, contain three or four searches with consistently high real-time/offline result correlations. Searches for Stimulus Classes 1 and 2 show high correlations across subjects. Searches from Stimulus Classes 3 and 4 show high magnitude correlations across subjects, alternating between positive and negative correlations.

Table 4 shows the maximum motion for subjects viewing Fribble objects, pooled across translational and rotational dimensions, between the start of the scan session and the start of each scan run. Motion for subjects viewing Fribble objects is generally reduced from that of subjects viewing real-world objects shown in Table 3. For 11 of 20 Fribble sessions, maximum motion falls under 2 millimeters/degrees in a given direction, while the motion along other directions is usually less than 1 millimeter/degree. In contrast, 5 of 20 real-world object sessions achieve this limit to their motion. Thus, by the end of each Fribble-viewing session, true ROI locations usually stay within a voxel-width's distance of their expected locations.

This decreased motion may be due to the differing tasks performed for the two object types. For real-world objects, subjects were asked to perform a one-back location task in which they were to judge the relative location of consecutively-displayed objects (Sec. 2.8.4). In contrast, for Fribble objects, subjects were asked to perform a dimness-detection task in which they were to judge whether the object, always displayed in the same central location, was dimmed (Sec. 2.9.4). We suggest that slight movement of real-world objects around the screen may have encouraged slight head motion during stimulus viewings.

Comparing between real-world object and Fribble object viewing groups, there appears to be a relation between subject motion and consistency for real-time and offline computations. Fribble subjects, who moved less as a whole, showed a higher number of searches with high correlation values, as well as more pronounced negative correlation values for several searches. To consider motion effects within the Fribble sessions, we study correlation values sorted by subject motion, shown in Fig. 7. In this Figure, there is no clear smooth transition from high (red) to low (green/cyan) correlations with increasing motion (moving from higher to lower rows). However, the two sessions with unusually high motion, S12₁ and S15₂, contain searches with consistently lower real-time/offline result correlations as shown in the bottom two rows. Even these two sessions contain at least one search with a correlation value above 0.3.

3.3. Real-Time search performance

3.3.1. Visualized feature spaces—To search for those visual properties selectively driving different cortical regions within the ventral pathway we constructed two types of visual feature spaces. Each of these spaces — Euclidean in nature — represented an array of complex visual properties through the spatial grouping of image stimuli that were considered similar according to the defining visual metric, as in Secs. 2.8.3 and 2.9.3.

Critically, each space contained a low number of dimensions — four dimensions for SIFT and three dimensions for each Fribble class — to allow the searches for visual selectivity to converge in the limited number of simple steps that can be evaluated in real-time over the course of a scanning session. These low-dimensional spaces also permit visualization of search activity over each scan session and visualization of general ROI response intensities across the continuum of visual properties represented by a given space. We display this information through colored scatter plots. For example, representing each stimulus as a point in feature space, Fig. 8 shows the locations in SIFT-based space selected or visited by the search for human-form images evoking high activity in the pre-selected SIFT/“human-form”

region of Subject S3, and shows the localized neural response to each of the displayed stimuli. The four dimensions of SIFT-based space are projected onto its first two and second two dimensions in Figs. 8a and b, respectively. Stimuli visited during the first and second realtime sessions are shown as circles and diamonds, respectively, centered at each stimulus' corresponding coordinates in the space. (Black dots correspond to the locations of all stimuli in the human-form class that were available for selection by the search program.) The magnitude of the average ROI response to a given visited stimulus is reflected in the color of its corresponding shape. For stimuli visited three or more times, colors span blue–dark blue–dark red–red for low through high average responses. Size of circles and diamonds indicates time in search when stimulus was visited, with small shapes for early visits and larger shapes for later visits.

Inspection of Fig. 8 reveals two patterns visible in the stimulus searches: (1) There are multiple distinct selectivities — multiple locations of search concentration — for single ROIs; and (2) There are marked changes in the cortical responses arising from slight deviations in visual properties (i.e., slight changes in location in each visual space). We report and discuss the implications of this observed local cortical selectivity in Leeds et al. (2014). In contrast, here we incorporate these observations into our formulations of metrics for optimizing search performance, as introduced in Sec. 2.10.

3.3.2. Real-world objects search—Convergence of real-time searches, that is, the focus of searches on one or a small number of locations across a session, is shown for real-world object searches in Table 5 for each subject, object class, and session.

Above-threshold convergence, $z \geq 1.8$, occurred for only 9 of 80 searches performed across all sessions and object classes. Interestingly, 8 of the 9 converged searches were performed for Stimulus Classes 2 (human-forms) and 4 (containers), with 4 performed for each class. The greater success of searches for these stimulus classes seems to coincide with reduced errors in stimulus display throughout the search, as shown in Table 1. However, it is worth noting there are a large number of display errors for Search 4 of S6₂, as well as for Search 1 of S8₁, despite their high convergence. Motion and preprocessing factors underlying the rare above-threshold convergence results are not apparent. Less than 50% of convergent searches showed high (above 0.2) correlations between real-time and offline calculations of ROI responses to stimuli (Table 3). The scan session with greatest head motion, S1₁, shows a high Z value for Stimulus Class 4.

Below-threshold convergence Z values ranged widely. Several searches showed values $Z < -1.3$, seeming to indicate that a random set of stimuli was markedly more convergent than the stimuli actually visited frequently. To some extent, this phenomenon may point to an unexpected feature of our significance test, as defined in Sec. 2.10. Convergence measures the clustering of stimuli visited by the search three or more times, while stimuli visited one or two times are ignored. For our permutation test, we randomly reassigned each frequently-visited label to one of the stimuli visited any number of times by the search. This approach was intended to judge the convergence of frequently visited stimuli in light of the distribution of stimuli that were visited but not considered sufficiently close to the ROI selectivity center to be re-visited. However, if several stimuli are nearby in space and close

to the location producing highest cortical response, their neighborhood may be visited many times but each stimulus visited only visited once or twice. This non-frequently visited clustering may be indicated by extreme negative Z values. At the same time, it is worth noting that convergence Z values did not fall below -2 , while the majority of above-threshold values were greater than 2 .

Consistency of real-time searches, that is, the focus of a search on the same location or locations in visual space when initialized at two different points in the space in two different scan sessions, is shown for real-world object searches in Table 6 for each subject and object class.

Above-threshold consistency, $z > 1.8$, occurred for only 2 of 40 searches performed across all subjects and object classes. The searches were performed for Stimulus Class 2 (human-forms). From the two above-threshold results, no clear pattern for successful consistency could be deduced. Each consistent search coincided with a low number of errors in stimulus display (Table 1), but many searches with fewer display errors do not exhibit high consistency. Motion and preprocessing factors underlying the rare above-threshold consistency results are not apparent. Neither of the two subjects, S1 and S5, showed above-threshold convergence for Class 2 searches. The lack of consistency for searches with above threshold convergence — particularly for Search 2 for S7, which converged in both session but shows a consistency score of $Z = 0.74$ — indicates the potential presence of multiple regions in SIFT-based space producing high responses from a given ROI. Consideration of further sources of difficulty for search performance of real-world objects are discussed in Sec. 4.3.

Below-threshold consistency Z values ranged widely. For example, 6 searches showed values $Z < -1.3$, seeming to indicate a random set of stimuli selected for each of two sessions would be markedly more consistent than the stimuli actually visited frequently by the search. Reasons for extreme low Z scores are discussed above.

The change in the distribution of locations visited by real-time searches, as reflected by change in the distribution's mean (dist) and variance (var), is shown for real-world object searches in Table 7 for each subject, object class, and session.

Change in the variance of locations explored from the first half to the second half of each session was quite small across all searches. var generally falls between -0.02 and 0.02 , while the variance of locations explored in each half of a session fall between 0.02 and 0.07 . Visited points are just as likely to be more dispersed (positive var values) as they are to be more concentrated (negative var values) as the search progresses.

The lack of convergence over time as indicated by the var measure in part may reflect the reinitialization of the simplex at the start of each new run within the scan session, as described in Sec. 2.1. Existence of multiple locations in search space evoking high cortical responses also may account for lack of convergence over time. In contrast to var , the time-independent convergence measure defined in Eqn. 2 can reach high Z values while converging on multiple locations in space, provided the number of locations is small.

Changes in the center of the distribution of locations explored from the first half to the second half of each session is notable for several searches, with $dist = 2$ for 9 out of 80 searches and $dist = 1.5$ for 24 out of 80 searches. The 9 high shifts in distribution focus occurred with roughly equal frequency for searches of Stimulus Classes 2, 3, and 4. Most high shifts in focus (7 of 9) occur in the second session. In the second session, the starting locations were selected to be distant from the center of focus from the first session, as discussed in Sec. 2.10.2; in the first session, the starting locations were set to be the origin, around which stimuli are distributed in a roughly Gaussian manner. While this observation indicates a step towards cross-session consistency for several searches, the corresponding Z scores for the consistency metric defined in Eqn. 3 are predominantly negative.

3.3.3. Fribble objects search—Convergence of real-time searches, as defined in Sec. 2.10, is shown for Fribble objects in Table 8 for each subject, object class, and session.

Above-threshold convergence occurred for 20 of 80 searches performed across all sessions and object classes. Converged searches were performed for all stimulus classes, though more frequently for Classes 1, 3, and 4 as compared to Class 2. Higher frequency of delayed displays for Search 1 compared to the frequency of delays for other searches (Table 2) did not appear to adversely affect performance of Search 1 as it had for subjects viewing real-world objects. In part, this may be attributable to the smaller number of display errors for Fribble object searches overall, especially compared to the number of incorrect real-world stimuli displayed for Search 1 and Search 3 reported in Table 1. Motion and preprocessing factors underlying above-threshold convergence results are not apparent. Only 55% of convergent searches showed high (above 0.2) correlations between real-time and offline calculations of ROI responses to stimuli (Table 4). Several real-time sessions contained multiple searches with above-threshold convergence; three of four searches converged in session 1 for S19. However, decreased subject head motion was not an apparent underlying factor in successful search convergence.

Above-threshold convergence Z scores generally were higher for Fribble object searches than they were for real-world object searches; 50% of above-threshold Fribble object searches showed $Z > 2.5$, compared to 33% of above-threshold real-world object searches. This greater frequency and magnitude of successful search convergence for Fribble objects may reflect the lesser motion of the subjects in these sessions or, potentially related, the seemingly more reliable results of fMRI signal processing during these sessions, reported in Sec. 3.2. The structure of the Fribble search spaces also may pose advantages over the SIFT-based image space, as discussed in Sec. 4.3.

Below-threshold convergence Z values still were assigned to 60 of the 80 searches, and ranged somewhat widely. However, unlike in real-world objects searches, negative Z values were much more infrequent and were relatively small in magnitude, that is, $Z > -1.3$. Furthermore, many sub-threshold searches exhibited degrees of convergence, for example, 22 searches have $1.0 < Z < 1.8$, compared to 6 searches fitting this criterion for real-world objects sessions.

Consistency of real-time searches, as defined in Sec. 2.10, is shown for Fribble objects in Table 9 for each subject and object class.

Above-threshold consistency occurred for 7 of 40 searches performed across all subjects and object classes. The searches were performed for all stimulus classes, though somewhat more frequently for Classes 1 and 4. There was no clear pattern based on frequency of stimulus display errors, nor based on motion and preprocessing factors. Nonetheless, 5 out of 7 search pairs showed a correlation greater than or equal to 0.3 between real-time and offline calculations of ROI responses to stimuli. Several real-time sessions contained multiple searches with above-threshold consistency; three of four searches were consistent for S17. However, decreased subject head motion was not an apparent underlying factor in successful search consistency within Fribble-viewing subjects. Almost all searches showing above-threshold consistency also showed convergence in one scan session, and in both sessions for S19 Search 4. It is possible that the lack of convergence we observed for both scan sessions may reflect the fact that convergence may require more observations and time than was possible in our fMRI study. This may be particularly true for our second session in which we started the search at locations distant from the potentially high-activity regions of visual space, thereby leading the search to probe many suboptimal locations.

Below-threshold consistency Z values ranged widely. Several searches show values $Z < -1.3$.

The change in the distribution of locations visited by real-time searches, as reflected by change in the distribution's mean ($dist$) and variance (var), is shown for Fribble object searches in Table 10 for each subject, object class, and session.

Equivalent to observations made for search behavior using real-world objects, the change in the variance of locations explored from the first half to the second half of each session was small across all searches. var generally falls between -0.1 and 0.1 . Visited points were observed to be just as likely to be dispersed (positive var values) as to be concentrated (negative var values) as the search progressed. Potential contributions to this lack of convergence over time as indicated by the var measure are discussed above in the context of search performance for real-world objects, where we observed a similar lack of decrease in variance of stimuli explored over time.

Also similar to real-world objects searches, changes in the center of the distribution of locations explored from the first half to the second half of each session is notable for several searches, with $dist = 2$ for 12 out of 80 searches and $dist = 1.5$ for 23 out of 80 searches. The 12 high shifts in distribution focus occurred with roughly equal frequency for searches of all stimulus classes. Unlike in real-world objects searches, high shifts in focus occurred with equal frequency across the first and second sessions.

Starting from the origin in the first session, each search initially probed stimuli whose component shapes were morphed intermediate appearances between two better-established shapes at the extremes: -1 and 1 coordinates on each axis. Intuitively, a region involved in object perception would be expected to be selective for salient visual features, as opposed to the less well-defined shapes generated by morphing within the Fribble space (Fig. 5). As

such, large shifts from the origin in Session 1 are unsurprising. In contrast, the definition of the real-world object feature space through SIFT and multi-dimensional scaling placed groups of salient visual features throughout the space, not just at its extremes, making large shifts from the origin less likely in the first session.

As in real-world objects search, in the second Fribble session, the starting locations were selected to be distant from the center of focus from the first session (Sec. 2.10.2) — thus, a significant shift in search focus would be required to identify the same stimuli producing high activity in the pre-selected cortical regions. While these second session observations indicate a step towards cross-session consistency for several searches, the corresponding Z scores for the consistency metric defined in Eqn. 3 are predominantly below threshold, though for all but S11, $Z = 1.4$.

Overall, all measures of Fribble object search behavior indicate more stability and more consistency in identified visual selectivities when compared with search of real-world objects. In particular, Fribble searches benefit from more tightly-defined feature spaces describing stimuli with known dimensions of variation, similar to the artificial stimuli used in past neurophysiological studies of neural visual selectivity (Hung et al., 2012). However, there remains significant room for improvement in achieving convergence, both over space and while operating across time. Nonetheless, even the current success rate of a relatively simple search method—simplex simulated annealing—to investigate a rather complex problem in visual encoding constitutes a promising basis for further development of real-time fMRI methods.

4. Discussion

We develop and assess a novel method to explore the complex visual feature selectivities of localized regions in the human ventral visual pathway. Our work introduces a set of algorithms and their implementations for dynamically selecting stimuli to display during each scan session based on cortical responses to stimuli displayed seconds earlier in the same session. This dynamic, or real-time, stimulus selection was implemented to most effectively search defined visual spaces in limited scan time. That is, we developed methods to rapidly identify those images of objects that produced the highest responses from localized brain regions. This application of real-time search to identify maximally preferred stimuli for regions of the ventral pathway has not been pursued previously in human neuroimaging to our knowledge, and algorithmically-driven real-time neural data analysis is quite new across all neuroscientific studies of vision (Yamane et al., 2008; Hung et al., 2012). Despite the large number of technical and biological challenges this approach faces, our real-time BOLD response processing and stimulus display system showed acceptable errors across twenty subjects. Furthermore, stimulus selection through searches for preferred stimuli exhibited significant convergence and consistency in many subjects. Adjustments in system settings applied throughout our study steadily improved the operation of our methods, thereby suggesting valuable ways forward in the continued development of methods for the real-time analysis of neural data.

4.1. Socket communication to reduce stimulus display errors

Ordinarily during real-time scanning sessions, the search program selected each new stimulus image at least 1 to 2 seconds before it was needed by the display program. The shared-file method for communication between the search and display programs proved problematic when stimuli were selected less than 4 seconds before they were to be displayed. While the new stimulus was correctly recorded in the file, the file updates sometimes required over 5 seconds to be visible on the display machine. In contrast, socket communications for selected stimuli were available to the display machine immediately after their computation by the search program. **Once sockets were used for inter-computer communication, display errors dropped dramatically, providing the first major technical insight in our study.**

Delays in stimulus selection past the required stimulus display time are relatively limited, but do occur on occasion. These late displays reflect variable speed of real-time fMRI signal processing, which can be slowed by irregular subject motion and scanner magnet behavior during cortical response to previous stimuli. Other programs running on the “analysis machine” — in parallel with the preprocessing and search programs — also can unexpectedly take up processor resources, slowing down real-time analysis. While we initiate no extra programs on the analysis machine during real-time sessions, we also do not reconfigure the machine to suspend potentially unnecessary background processes.

Additional delays can occur when the search program occasionally skips simplex computations for a given class of objects at a given step. Thus, stimulus displays on some occasions occurred at 20 second delays, followed in succession by the displays of the other stimuli whose trials had passed during the time waiting. This problem did not occur often, but requires further code development for future versions of the real-time search study.

Delays in stimulus selection and delays in updates to shared files both are truly technical issues, rather than fundamental scientific challenges. These delays may be decreased by improvements to the hardware and software used for real-time analyses. Our study employs Red Hat on an Intel Core 2 Duo processor running at 3 GHz with 4 GB memory. Increasing the number of cores on our processor, increasing machine clock speed, and adjusting the operating system policies on directory update timings likely will decrease display errors using either method of inter-program communication used in this study. Nevertheless, our observations provide engineering insight for those using systems similar to our own. For more advanced hardware, our insights can still guide real-time investigators wishing to use smaller inter-stimulus intervals, such as 4s, where processing speed may still remain an issue.

4.2. Moderate stability of truncated real-time BOLD signal processing

To compute ROI BOLD responses with sufficient speed for dynamic stimulus selection, several typical BOLD response preprocessing steps are truncated or removed in our system. These steps are designed to better account for factors such as scanner drift and subject head motion. We find truncated real-time preprocessing achieves somewhat similar ROI response estimates to full offline preprocessing — correlation of responses between real-time and

offline computations was 0.3 on average across all searches, and 0.5 on average across the two stimulus classes with the most stable search results for real-world and Fribble stimuli. These results appear fairly robust to sub-voxel/single-voxel motion, that is, up to 2–3mm translation (Tables 3 and 4). For larger motions, evidence suggests truncated preprocessing may break down; however, there is insufficient data to judge this condition with confidence.

It is important to note that correlation in real-time and offline preprocessing computations still remain below 0.9, and correlations below 0.1 do exist for several searches. These results potentially can motivate future development in real-time scanning and search methods. At the same time, 0.3 and 0.5 correlations are relatively promising results. While our method for correlation expresses the consistency between real-time and offline preprocessing results on a trial-by-trial basis, the consistency of computed cortical responses considered for study of ROI selectivity likely is higher. As discussed in Sec. 3.3.1, ROI responses across the associated visual space are examined only for stimuli shown three or more times. Responses for each of these stimuli are averaged across displays to reduce variability from noise. This noise removal may mimic offline preprocessing effects, increasing the correlation between the two methods' results.

4.3. Contributing factors to search convergence and consistency

Over 30 preferred-stimulus searches exhibited significant convergence and consistency across our study. However, over 100 searches did not pass significance in either measure. Our results suggest potential room for improvement in the technical and conceptual underpinnings of each of the programs in our system. Revised settings employed over the course of our study were accompanied by marked improvement in search performance, suggesting promising future directions in the development of real-time stimulus selection methods. The convergence and consistency that we observe throughout the study instills confidence in the robust nature of our search method to succeed in the face of sub-optimal assumptions and operation parameters.

Display errors show effect on performance—Insufficiently fast computation and network-communication times prevented the display program from showing subjects the correct stimuli at the proper time. The stimulus selection method assumes the correct stimulus is shown for each trial and selects each new stimulus based on past ROI responses regardless of the validity of the visual stimuli actually reaching the subject. Incorrect displays misinform the simplex search about stimulus responses and can lead to sub-optimal exploration and acceptance of future simplex points. Looking at real-world and Fribble object results together, all but one of the significantly convergent and consistent searches used socket communication for the display program, significantly limiting the number of display errors. The assumption of noisy stimulus response measurements embedded in the simplex simulated annealing approach may contribute to the partial robustness of real-time search to a relatively small number of display errors.

Preprocessing errors show less clear effect on performance—Shortcomings in motion correction during preprocessing also may mislead real-time stimulus selection. Similar to the risks of undetected display errors, incorrect response calculations could lead

to sub-optimal exploration and acceptance of future simplex points. However, counter to this theoretical concern, we observe more than half of significant convergence results correspond to sessions with low correlations in offline-vs-real-time response computations. Limiting convergence and consistency measures to stimuli visited three or more times may permit averaging activity over multiple trials to overcome errors in individual measurements. Alternatively, for sessions with highly negative correlations, particularly noticeable in Fribble spaces, searches may effectively be searching for stimuli evoking particularly low responses; this strategy may successfully identify maxima in stimulus space as well because of the observed phenomenon of local inhibition (Leeds et al., 2014).

Regular resetting of simplex points may limit convergence—The persistent variability of visited stimulus locations across each search session reflects the simplex re-initialization strategy described in Sec. 2.1. At the beginning of each run, a new simplex was defined centered at the termination point of the search for the previous run. The four additional simplex points were randomly selected around this center to be located +1 to -1 distance away from the center. Thus, an equal spread of points was used to investigate the first run of the session as the last run of the session. As there were 16 stimulus displays for each run to place 5 simplex points, chance for convergence was limited both within and across runs.

Complexity of visual feature space shows effect on search performance—Searches performed in Fribble spaces substantially outperformed searches performed in SIFT-based space. While there were 10% more searches with minimal display errors (5 or fewer errors per search) for Fribble object searches than for real-world object searches, there is no evidence that these display errors influenced our results. In particular, we observed 100% more convergent Fribble object searches than real-world object searches despite the higher display error rate for the former. The significant improvement afforded by Fribble-based space may be attributed to the closer relation between Fribble-space coordinates and their corresponding stimuli. Fribble space is composed of three axes indicating morphing of selected appendage properties. These axes account for the entire variability in stimuli shown. SIFT space, in contrast, sought to capture an unknown number of complex real-world visual properties using only four dimensions. The small number of dimensions was required to enable effective search over a limited number of scan trials. However, Fig. 4c shows that at least 50 dimensions would be required to explain 50% of the variance in a SIFT-based pairwise distance matrix for 1000 images.

While Fribble space is associated with markedly improved search performance, it requires markedly simplified stimuli. Future work may pursue variations on SIFT space derived from a set of real-world objects more similar in appearance but nonetheless retaining real-world visual complexities.

Viewing task may influence search performance—It also is possible that the greater frequency and magnitude of successful search convergence for Fribble objects may reflect the lesser motion of the subjects in these sessions or, potentially related, the seemingly more reliable results of fMRI signal processing during these sessions, reported in Sec. 3.2. These improvements in head-motion and BOLD signal stability, in turn, may stem from the

Fribble-object viewing task. Subjects viewing Fribbles performed a dimness detection task for stimuli presented at a fixed location on the screen, while subjects viewing real-world objects performed a one-back location task for stimuli presented at slightly varied locations on the screen. Variable stimulus locations for real-world stimuli — though location variation was very slight — may have produced further head motions that hindered ROI BOLD response calculations. The requirement to remember a past stimulus further may have weakened the signal from the stimulus being viewed presently, as it competed for representation space with the previous stimulus. It also is possible changes in display location for real-world objects elicited different location-dependent cortical responses. However, this confound seems unlikely given the relatively small degree of location change (± 2.5 degrees) and well-established invariance typically seen in mid- and high-level cortical regions within the ventral visual pathway.

Finally, in a recent experiment, we employed a real-time search using a **dimness detection task** and real-world objects for a single subject. We note that search convergence and consistency measures were consistent with those reported in our present study using the one-back location task. This admittedly somewhat anecdotal result provides some indication that the simplicity of Fribble space plays a greater role in the heightened success of real-time search than does the task.

Simplified search assumptions—The simplex method underlying our search expects a given ROI's stimulus response function to have a unique maximum in feature space. In contrast, our data often show multiple local maxima (e.g., Fig. 8). If there are three or more maxima in a region — particularly if the number of maxima is larger — it is unlikely the search will repeatedly probe a sufficient number of stimuli to associate each maximum location with a large enough cluster of stimuli to produce a high convergence value, defined in Eqn. 2. Similarly, the presence of a large number of maxima increases the likelihood that starting searches from different points in feature space will produce different sets of results, each focusing on points closest to their respective starting location, producing poor consistency measures as defined in Eqn. 3. Despite some inaccuracies in our assumptions, we are encouraged by convergence on local extrema in cortical responses to a small number of locations in visual feature space, as seen in several searches.

5. Conclusions

Promise of real-time stimulus selection

Our work employs a collection of novel methods in real-time analysis of cortical data to explore complex visual properties used in perception. This unique approach to human neuroimaging faces many technical and biological challenges — from scanner and physiological noise in fMRI recordings to uncertainty about the nature of higher level visual representations — compounded by the small number of stimuli able to be shown in the limited scanning time available in fMRI. To address this final challenge, real-time selection of stimuli based on cortical responses to recently displayed visual objects was used to optimize the use of this limited scanning time, building on similar approaches in primate neurophysiology (Tanaka, 2003, Yamane et al., 2008, Hung et al., 2012). Our present

application of simplex simulated annealing (Cardoso et al., 1996) for stimulus selection faces considerable additional challenges — from occasional faults in stimulus display to frequent simplex resets. Nonetheless, search for numerous brain regions studied converged in their search for stimuli of interest to the region. These searches, for real-time stimuli and particularly for Fribble stimuli, help us to understand the representation of objects in human ventral cortex. In particular, within the search spaces we defined, we observed evidence for local inhibition, likely reflective of local competition between neural units, and multiple sets of featural selectivities, likely reflecting the large size of the studied brain regions; we discuss both properties in further detail in Leeds et al. (2014).

In sum, examinations of the behavior of each component of our real-time system under multiple settings and for multiple subjects shows our system to be robust to undesired actions from subjects (e.g., head motions) and program flaws (e.g., stimulus selection delays). At the same time, our examination of system performance suggests methodological and experimental parameters to improve future performances in real-time search:

- Use socket communication between stimulus selection and display programs
- Use the fastest available processors to perform real-time analyses
- Use stimuli that can be described by low-dimensional visual feature spaces (and conversely...)
- Use feature spaces that require low dimensions to describe stimuli
- Continue search simplex convergence across runs
- Explore alternative stimulus selection methods that can manage local maxima
- Align functional volumes to start of each scan session

We view these first attempts at searching for effective stimuli to drive localized brain regions as measured by fMRI as a success. In parallel, similarly successful real-time stimulus selection methods have been implemented in neurophysiology to make efficient use of limited recording time. Together, these new methods point to the potential of dynamic approaches to stimulus selection in future work.

Acknowledgments

This research was funded by an NIH EUREKA Award 1R01MH084195-01, the Temporal Dynamics of Learning Center at UCSD (NSF Science of Learning Center SMA-1041755), and, in part by a grant from the Pennsylvania Department of Health's Commonwealth Universal Research Enhancement Program. Daniel D. Leeds was supported by a Faculty Research Grant through Fordham University, an NSF IGERT Fellowship through the Center for the Neural Basis of Cognition, an R.K. Mellon Fellowship through Carnegie Mellon University, and the Program in Neural Computation Training Program (NIH Grant T90 DA022762).

References

- Brainard D. The psychophysics toolbox. *Spatial Vision*. 1997; 10:443–446. [PubMed: 9176954]
- Cadiou C, Kouh M, Pasupathy A, Connor C, Riesenhuber M, Poggio T. A model of v4 shape selectivity and invariance. *Journal of Neurophysiology*. 2007; 98(3):1733–1750. [PubMed: 17596412]

- Cardoso M, Salcedo R, de Azevedo S. The simplex-simulated annealing approach to continuous non-linear optimization. *Computers and Chemical Engineering*. 1996; 20(9):1065–1080.
- Donckels B. Global optimization algorithms for matlab. 2012 URL: <http://biomath.ugent.be/~brecht/downloads.html>; Online.
- Edelman S, Shahbazi R. Renewing the respect for similarity. *Frontiers in Computational Neuroscience*. 2012; 6(45):1–19. [PubMed: 22291635]
- Felleman D, Essen DV. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*. 1991; 1:1–47. [PubMed: 1822724]
- Freeman J, Ziemba C, Jeeger D, Simoncelli E, Movshon J. A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience*. 2013; 16(7):974–981. [PubMed: 23685719]
- Hemera Technologies I. Hemera photo objects volumes i, ii, and iii. 2000–2003
- Hubel D, Wiesel T. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*. 1968; 195:215–243. [PubMed: 4966457]
- Hung C, Carlson E, Connor C. Medial axis shape coding in macaque inferotemporal cortex. *Neuron*. 2012; 74(6):1099–1113. [PubMed: 22726839]
- Huth A, Nishimoto S, Vu A, Gallant J. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*. 2012; 76(6):1210–1224. [PubMed: 23259955]
- Just M, Cherkassky V, Aryal S, Mitchell T. A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PLoSone*. 2010; 5(1)
- Kriegeskorte N, Murr M, Ruff D, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini P. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*. 2008; 60(6):1126–1141. [PubMed: 19109916]
- Leeds, D. Ph.D. thesis. Carnegie Mellon University; 2013. Searching for the visual components of object perception.
- Leeds D, Pyles J, Tarr M. Exploration of complex visual feature spaces for object perception. *Frontiers in Computational Neuroscience*. 2014; 8(106)
- Leeds D, Seibert D, Pyles J, Tarr M. Comparing visual representations across human fmri and computational vision. *Journal of Vision*. 2013; 13(13)
- Lowe D. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*. 2004; 60(2):91–110.
- Mahalanobis P. On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*. 1936; 2(1):49–55.
- MATLAB. version 8.0.0.783 (R2012b). Natick, Massachusetts: The MathWorks Inc; 2012.
- Pelli D. The videotoolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*. 1997; 10:437–442. [PubMed: 9176953]
- Pittman, B. [accessed 20-September-2011] Afni main page — afni and nifti server for nimh/nih/phs/dhhs/usa/earth. 2011. URL: <http://afni.nimh.nih.gov/afni>; Online
- Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. *Nature Neuroscience*. 1999; 2(11):1019–1025. [PubMed: 10526343]
- Sato J, Basilio R, Paiva F, Garrido G, Bramati I, Bado P, Tovar-Moll F, Zahn R, Moll J. Real-time fmri pattern decoding and neurofeedback using friend: an fsl-integrated bci toolbox. *PLoS One*. 2013; 8(12):e81658. [PubMed: 24312569]
- Seber, G. *Multivariate Observations*. Hoboken, NJ: John Wiley and Sons Inc; 1984.
- Shibata K, Watanabe T, Sasaki Y, Kawato M. Perceptual learning incepted by decoded fmri neurofeedback without stimulus presentation. *Science*. 2011; 334(6061):1413–1415. [PubMed: 22158821]
- Swisher J, Gatenby J, Gore J, Wolfe B, Moon CH, Kim SG, Tong F. Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. *Journal of Neuroscience*. 2010; 30(1): 325–330. [PubMed: 20053913]
- Tanaka K. Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. *Cerebral Cortex*. 2003; 13(1):90–99. [PubMed: 12466220]

- Tarr M. Novel object – the cnbc wiki. 2013 URL: http://wiki.cnbc.cmu.edu/Novel_Objects; Online.
- Ullman S, Vidal-Naquet M, Sali E. Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*. 2002; 5:682–687. [PubMed: 12055634]
- Ward B, Janik J, Mazaheri Y, Ma Y, DeYoe E. Adaptive kalman filtering for real-time mapping of the visual field. *Neuroimage*. 2011; 59(4):3533–3547. [PubMed: 22100663]
- Wennerberg, M. version 2.9.5. Norrkross Software; 2009.
- Williams P, Simons D. Detecting changes in novel, complex three-dimensional objects. *Visual Cognition*. 2000; 7:297–322.
- Yamane Y, Carlson E, Bowman K, Wang Z, Connor C. A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nature Neuroscience*. 2008; 11(11):1352–1360. [PubMed: 18836443]
- Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*. 2005; 67:301–320.

Highlights

- We develop real-time BOLD signal processing for efficient study of cortical vision.
- Adaptive search of stimuli converges on visual selectivities in some searches.
- Adaptive search is robust to undesired subject motion and stimulus display errors.
- Simpler visual stimuli and search spaces allow more frequent search convergence.
- Assumption of single regional selectivity seems flawed, but searches still converge.

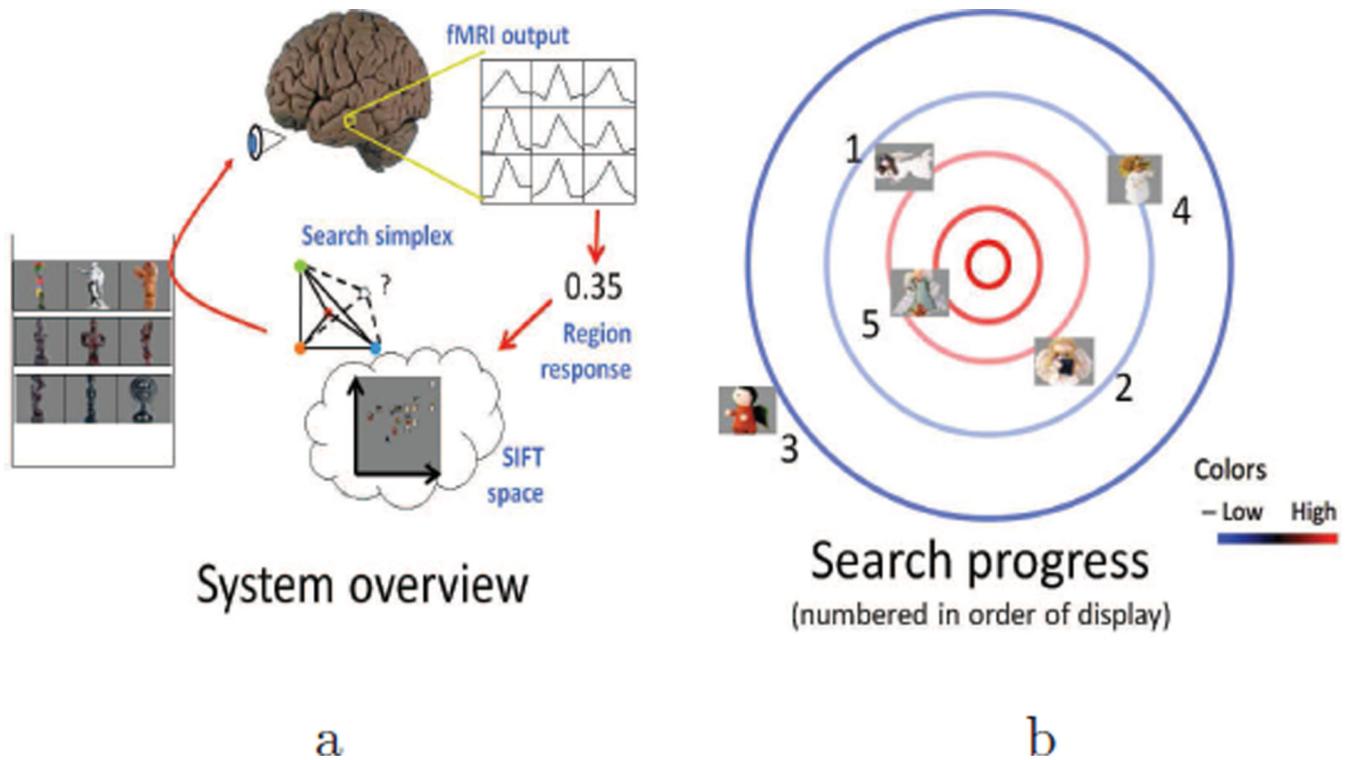


Figure 1.

(a) Schematic of loop from stimulus display to measurement and extraction of cortical region response to selection of next stimulus. (b) Example progression of desired stimulus search. Cortical response is highest towards the center of the space (red contours) and lowest towards the edges of the space (blue contours). Stimuli displayed in order listed. Cortical responses to initial stimuli, e.g., those numbered 1, 2, and 3, influence selection of further stimuli closer to maximal response region in visual space, e.g., those numbered 4 and 5. Figure adapted from Fig. 1 in Leeds et al. (2014)

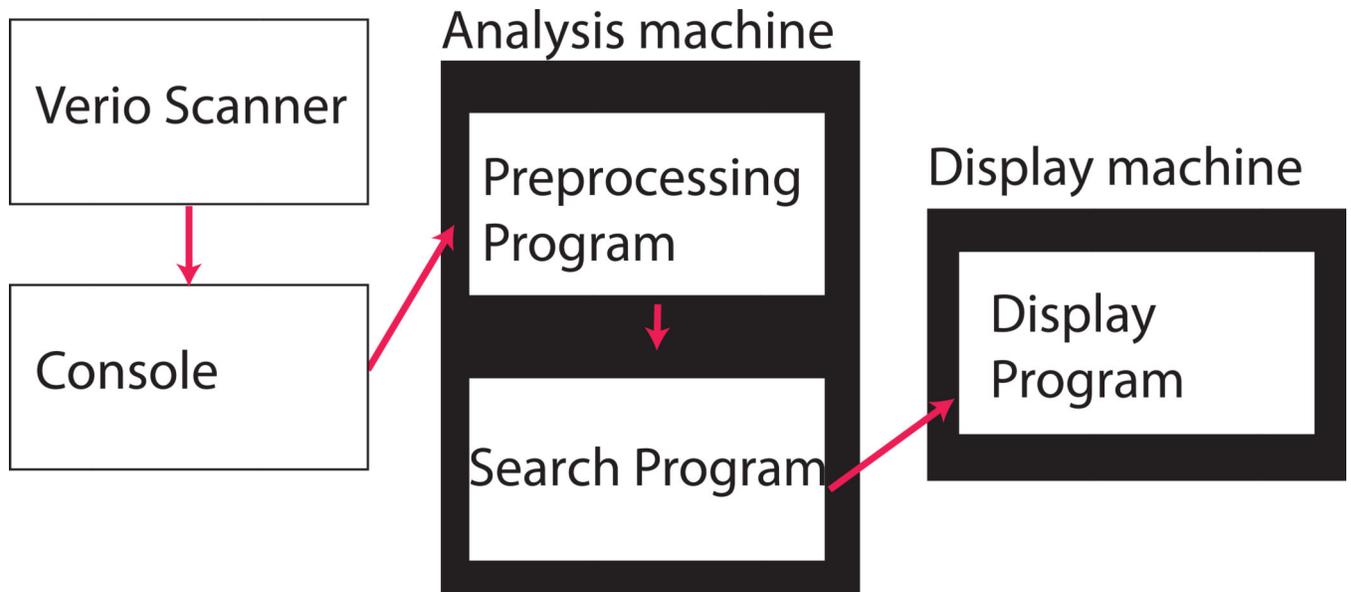


Figure 2.

Diagram of communications between the console (which collects and sends fMRI data from the scanner), the “analysis machine,” and the “display machine,” as well as communications between the analysis programs. These elements work together to analyze cortical responses to object stimuli in real-time, select new stimuli to show the subject, and display the new stimuli to the subject.

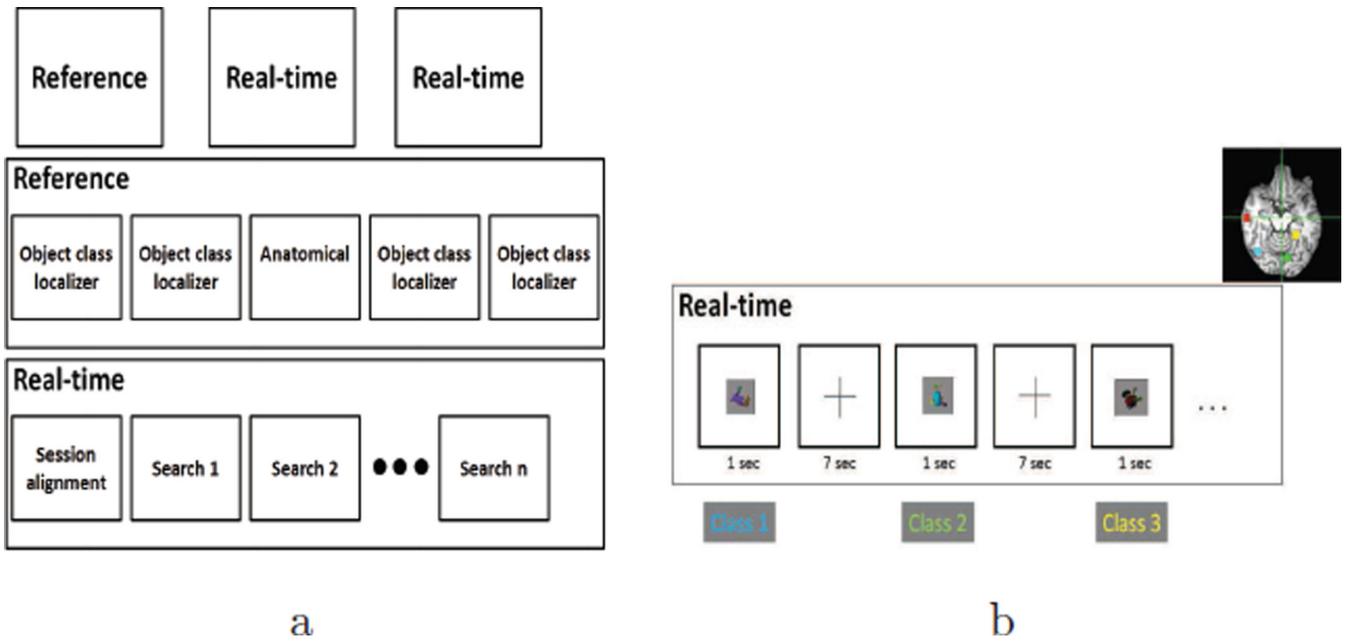
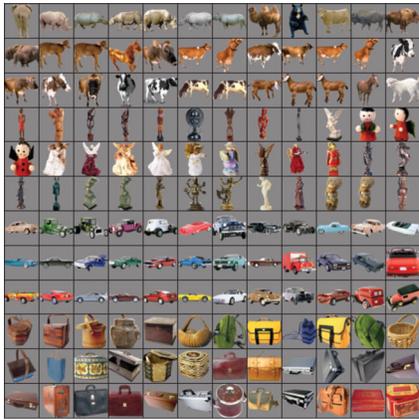


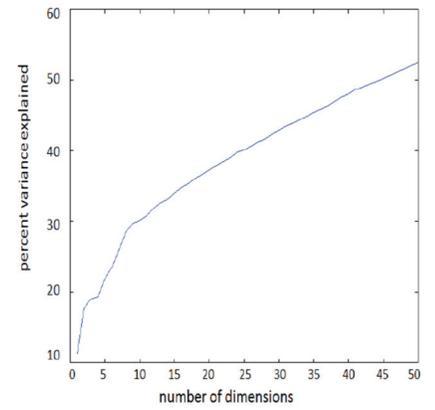
Figure 3. (a) Structure of the three scanning sessions performed for each subject. *First row* depicts the three sessions, *second row* depicts the runs for the reference session, and *third row* depicts the runs for each real-time session. (b) An example of the alternation among four stimulus class searches in a real-time search run. These four classes are comprised of mammals, human-forms, cars, and containers, and correspond to four colored brain regions shown on the upper-right of the figure. Figure adapted from Fig. 2 in Leeds et al. (2014).



a



b



c

Figure 4. Example real-world objects (a) and corresponding SIFT feature space (b). Real-world object images were selected from four object classes — mammals, human-forms, cars, and containers. Feature space shows example stimuli projected onto first two dimensions of space. (c) Percent variance explained using first n dimensions of MDS feature space for SIFT. Figure is adapted from Fig. 3 in Leeds et al. (2014).

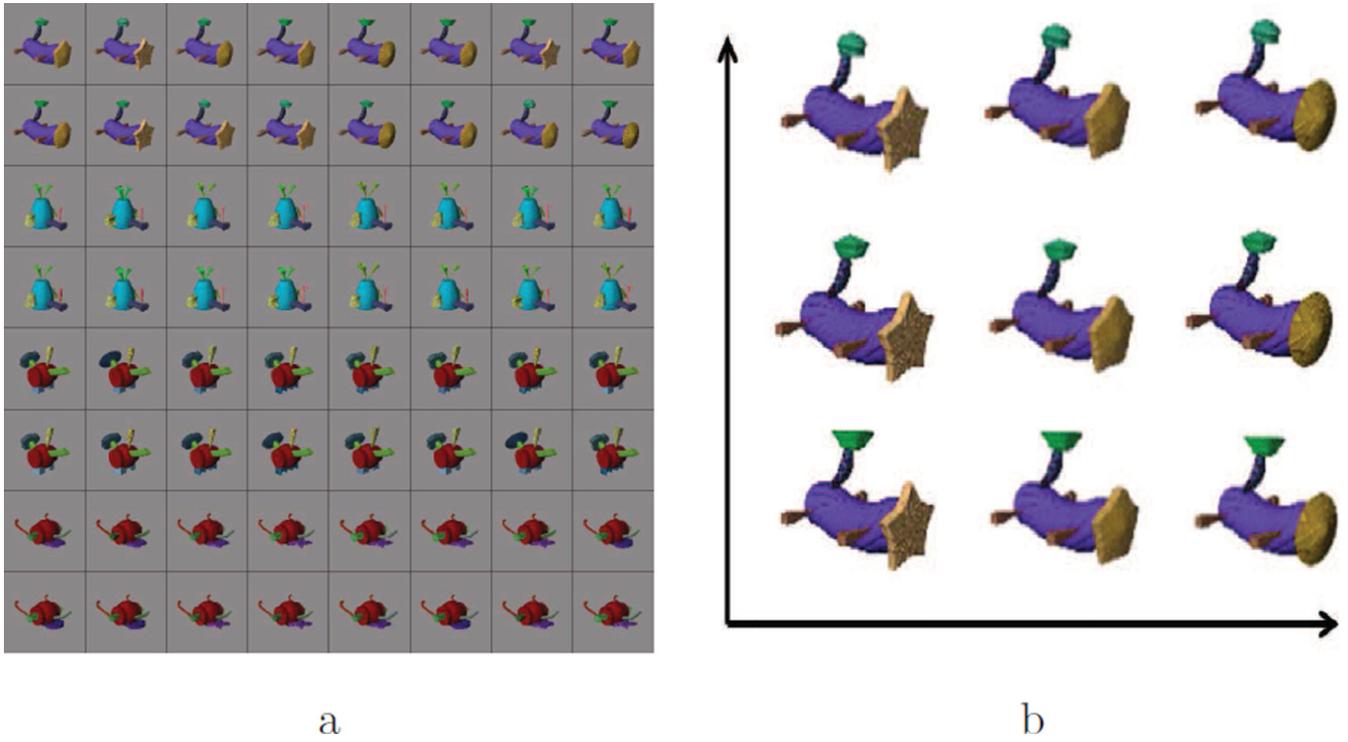


Figure 5. Example Fribble objects (a) and example corresponding Fribble feature space (b). Fribble images were selected from four synthesized classes, shown in rows 1/2, 3/4, 5/6, and 7/8, respectively. Feature space shows stimuli projected onto first two dimensions of space. Figure is adapted from Fig. 4 in Leeds et al. (2014).

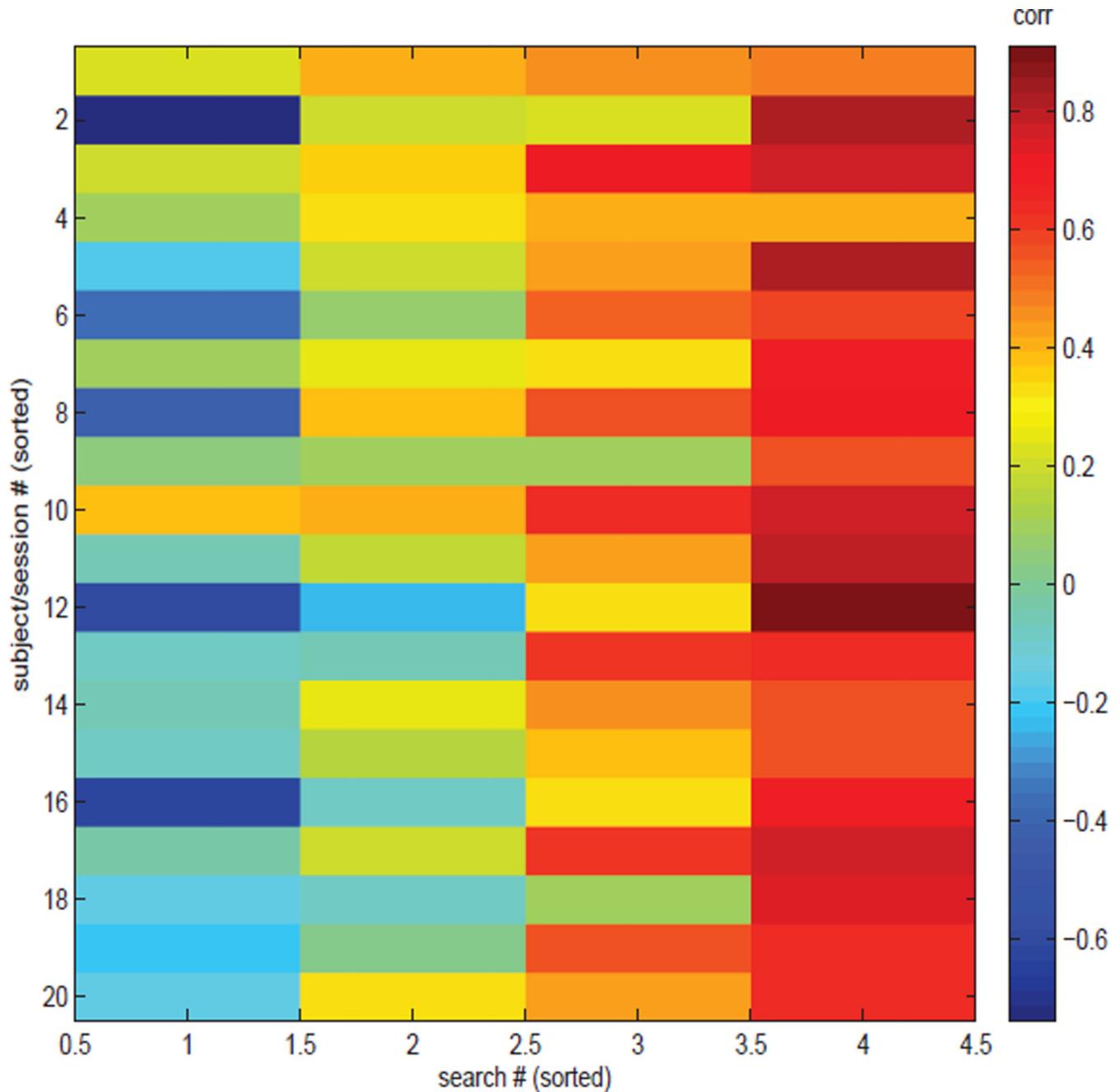


Figure 6. Motion effects on ROI computed responses for real-world objects searches, as in Table 3. Rows are sorted from lowest to highest corresponding maximum motion magnitude (values not shown), and columns within each row are sorted from lowest to highest correlation values. Correlation between computed responses for each of four class ROIs using offline preprocessing on full scan session versus real-time preprocessing on small time blocks within single runs.

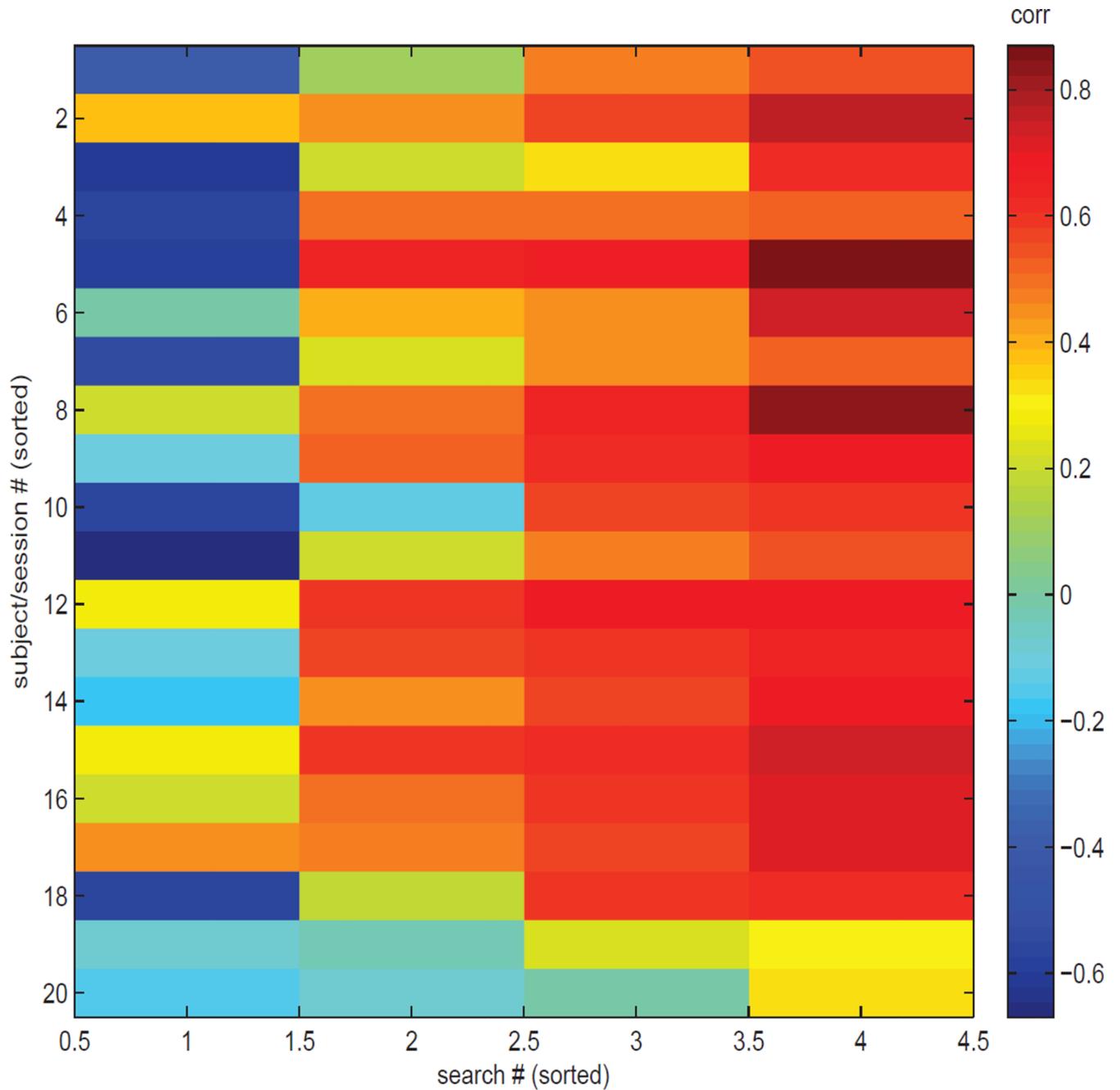


Figure 7. Motion effects on ROI computed responses, as in Table 4. Rows are sorted from lowest to highest corresponding maximum motion magnitude (values not shown), and columns within each row are sorted from lowest to highest correlation values.

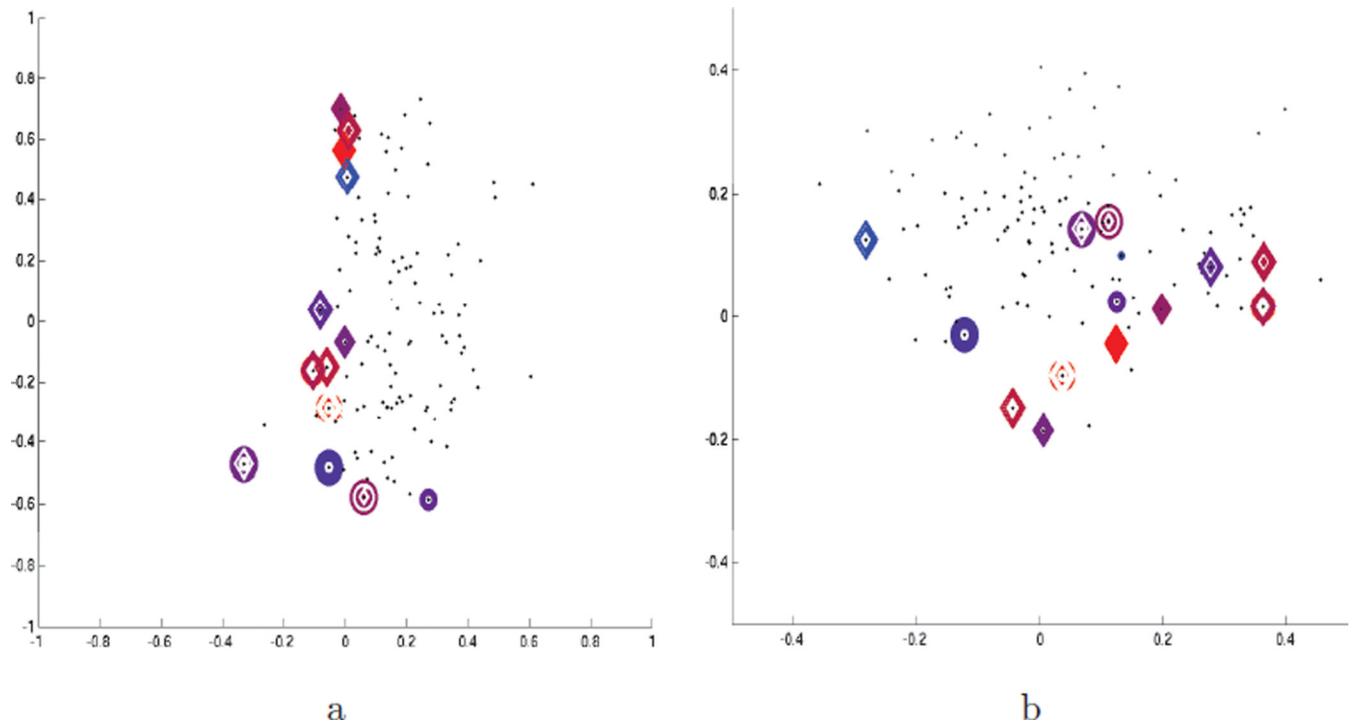


Figure 8.

Search results for S3, class 2 (human-forms), shown in (a) first and second SIFT space dimensions and (b) third and fourth dimensions. Location of all potential stimuli in space shown as black dots. Results from real-time scan session 1 are circles, results from real-time scan session 2 are diamonds. For stimuli “visited” (i.e., selected by the search) three or more times, colors span blue–dark blue–dark red–red for low through high responses. Size of circles and diamonds indicates time in search when stimulus was visited, with small shapes for early visits and larger shapes for later visits. Note axes for (a) are from -1 to 1 and for (b) are from -0.5 to 0.5 . Figure is adapted from Fig. 5 in Leeds et al. (2014).

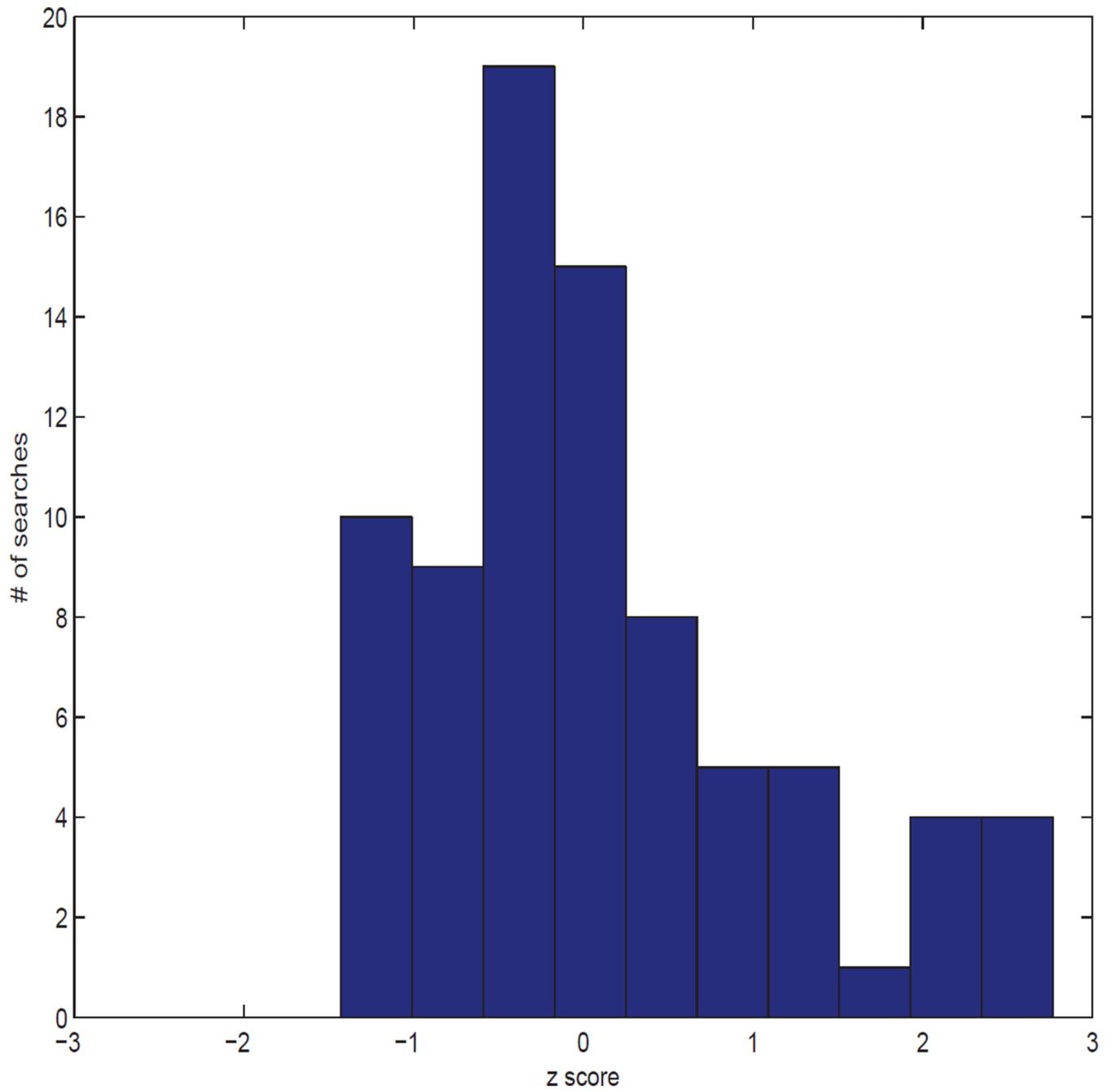


Figure 9. Histogram of convergence measures for searches in SIFT-based space.

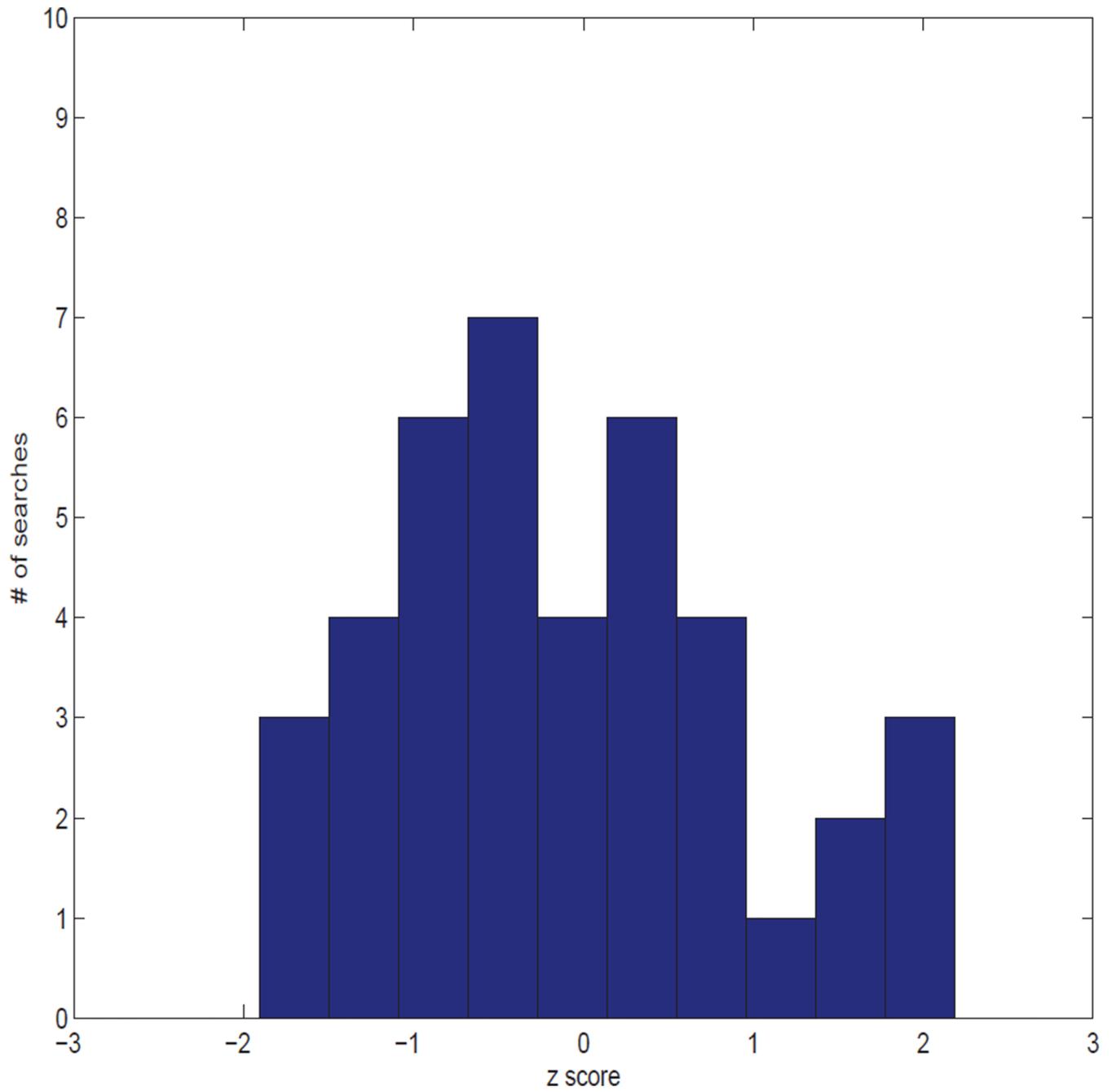


Figure 10.
Histogram of consistency measures for searches in SIFT-based space.

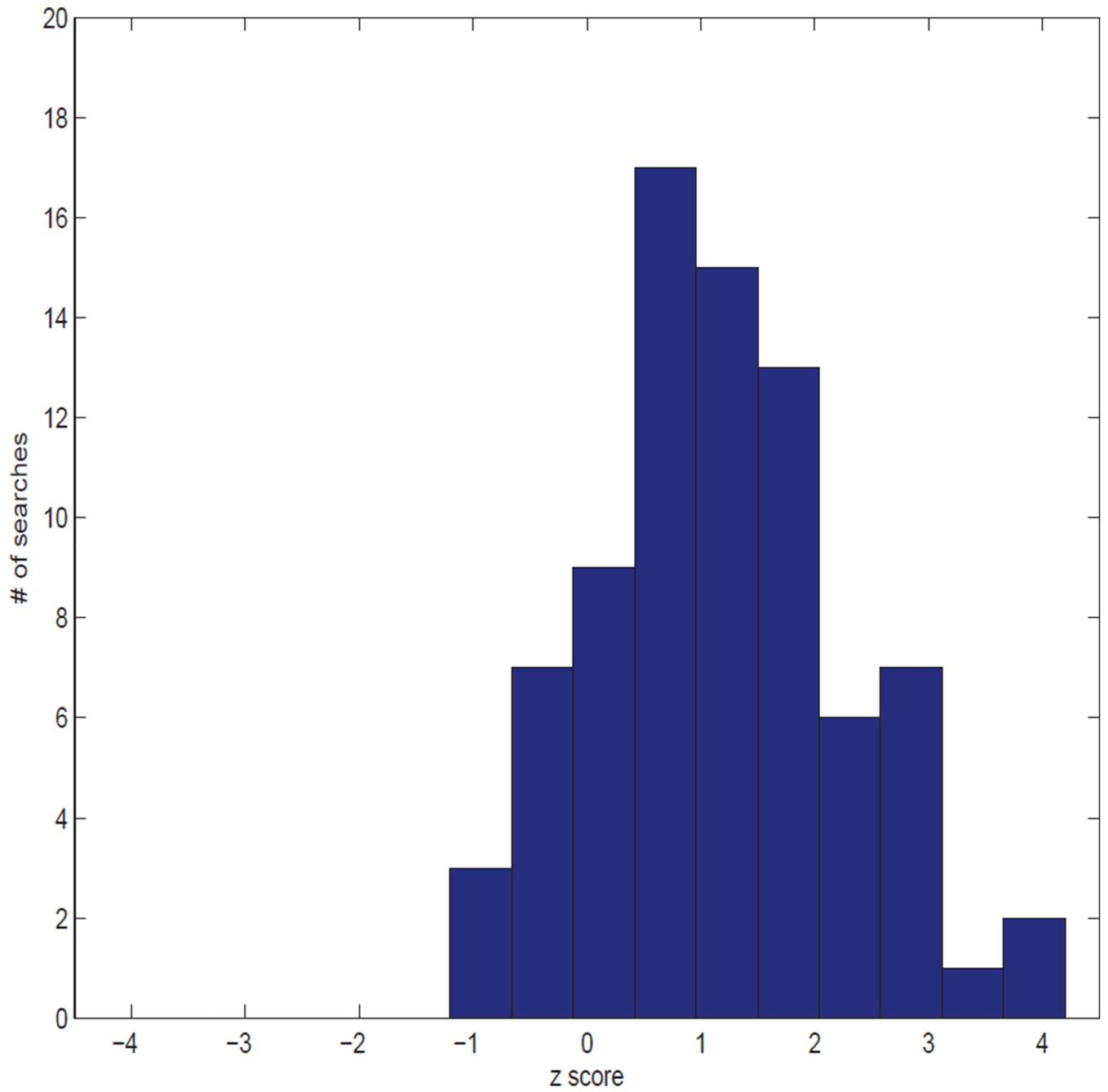


Figure 11.
Histogram of convergence measures for searches in Fribble spaces.

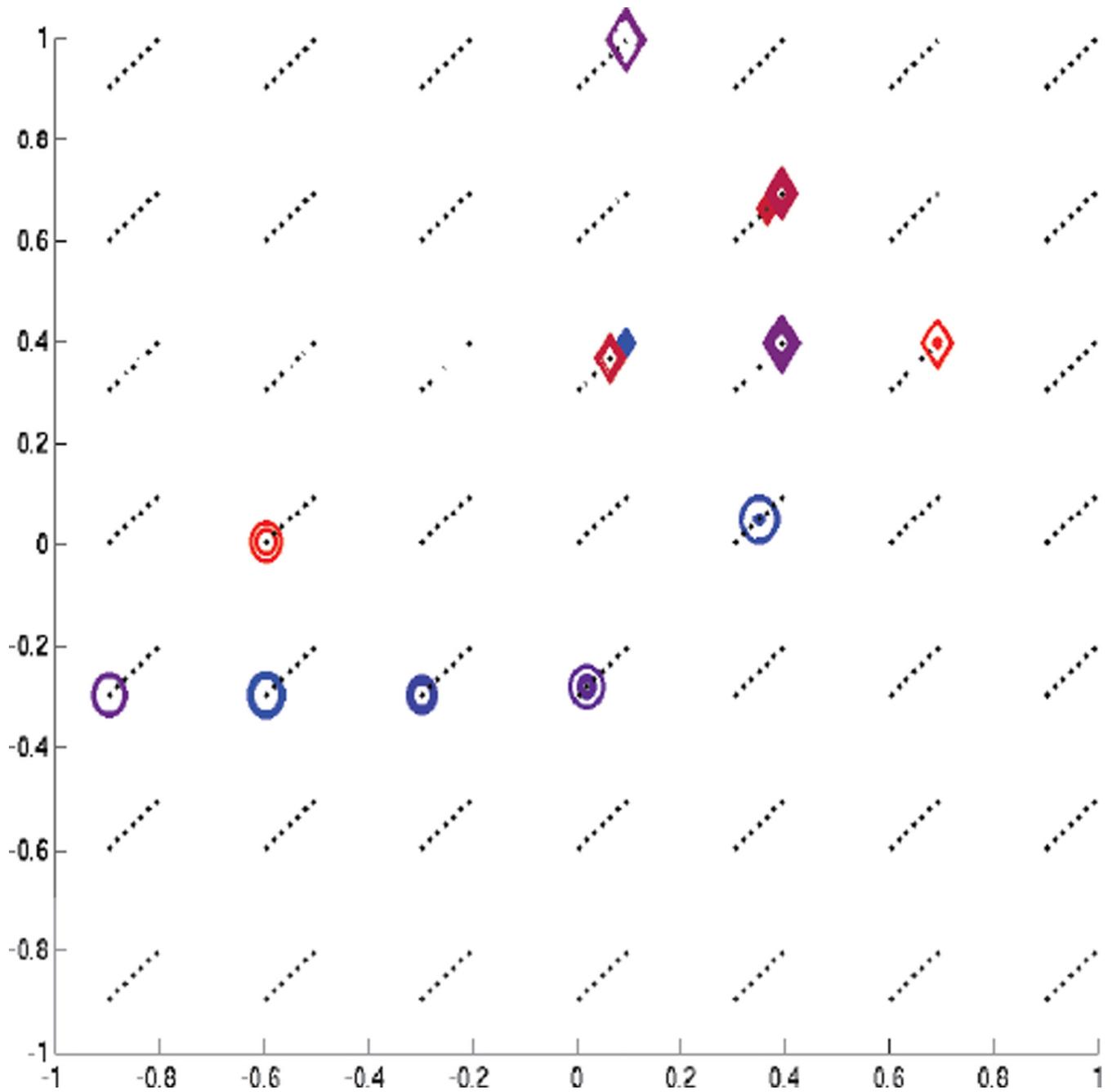


Figure 12.

Search results for S16, class 1, shown in three-dimensional Fribble space. Each dimension corresponds to morphing of a Fribble appendage. First two dimensions are represented by the horizontal and vertical axes. Because candidate stimuli are spaced at regular intervals in the space, indicated by the black dots, the third dimension coordinate is visualized as a slight diagonal offset from the location of the coordinates in the first and second dimensions. A positive third dimension coordinate results in displacement up and to the right along the corresponding diagonal line of black dots. Shapes and colors assigned as described in Fig. 8.

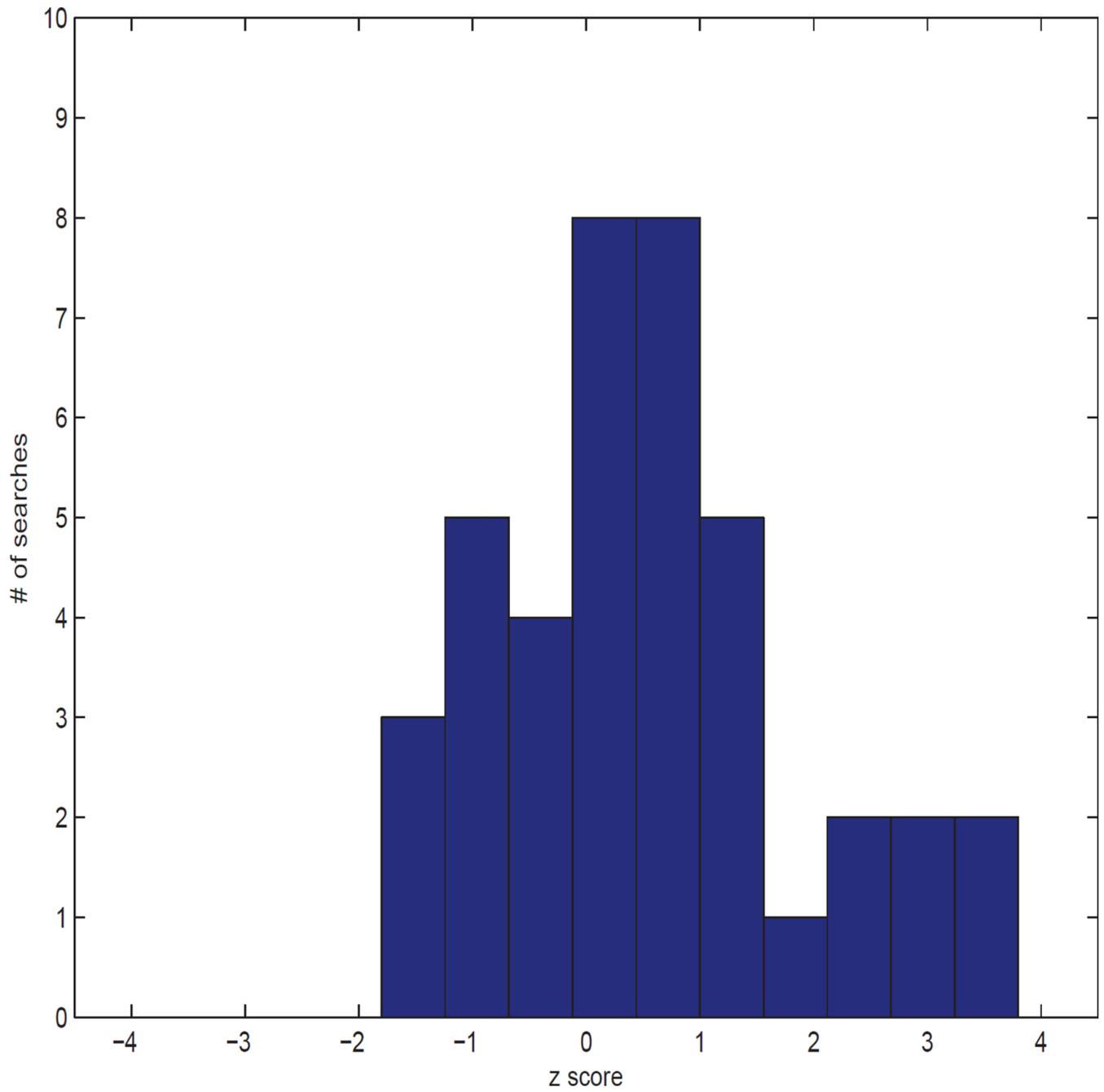


Figure 13.
Histogram of consistency measures for searches in Fribble spaces.

Number of delayed and incorrect display trials for real-world objects searches for each object class and each subject.

Table 1

Subject _{session}	late1	late2	late3	late4	wrong1	wrong2	wrong3	wrong4	# trials
S1 ₁	4	3	3	1	0	0	0	0	80
S1 ₂	0	0	0	0	0	0	0	0	96
S2 ₁	3	1	1	1	0	0	0	0	96
S2 ₂	3	1	1	0	0	0	0	0	112
S3 ₁	8	4	9	7	0	3	0	1	112
S3 ₂	3	2	2	2	3	0	3	0	112
S4 ₁	4	3	4	2	0	0	0	1	112
S4 ₂	5	4	4	1	0	0	0	1	112
S5 ₁	6	3	4	4	0	0	0	0	112
S5 ₂	7	2	3	6	0	0	0	0	112
S6 ₁	0	0	0	0	23	0	18	0	112
S6 ₂	0	0	0	0	41	25	47	23	112
S7 ₁	0	0	0	0	32	0	24	0	112
S7 ₂	0	0	0	0	21	0	30	0	112
S8 ₁	0	0	0	0	36	0	24	0	112
S8 ₂	0	0	0	0	30	0	25	0	112
S9 ₁	5	3	3	3	0	0	0	0	112
S9 ₂	5	1	1	2	0	0	0	0	112
S10 ₁	3	1	2	3	12	13	12	11	112
S10 ₂	6	2	3	3	0	0	0	0	112
total	62	30	40	35	198	41	183	37	

Delayed trials were those shown 0.5 s or more past the intended display time. Results are tallied separately for real-time sessions 1 and 2 for each subject, for example, S72 correspond to session 2 for subject S7. Results are tallied separately for each stimulus class, for example, late3 counts number of delayed display trials for stimulus class 3 and wrong4 counts number of incorrect stimuli displayed for stimulus class 4. Class numbers correspond to mammals (1), human-forms (2), cars (3), and containers (4), respectively. Number of search trials per class varied in each session, as seen in final column.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Number of delayed display trials for Fribble searches for each subject.

Subject _{session}	late1	late2	late3	late4	# trials
S11 ₁	4	0	3	3	96
S11 ₂	5	0	2	4	80
S12 ₁	4	0	1	3	80
S12 ₂	11	5	5	9	96
S13 ₁	6	2	2	1	96
S13 ₂	6	2	2	1	96
S14 ₁	3	0	0	0	96
S14 ₂	5	0	0	1	80
S15 ₁	6	2	3	1	64
S15 ₂	5	0	0	2	80
S16 ₁	5	0	0	0	80
S16 ₂	5	0	0	0	80
S17 ₁	3	0	0	0	96
S17 ₂	6	1	0	2	96
S18 ₁	6	2	2	1	80
S18 ₂	6	0	0	1	80
S19 ₁	2	1	1	1	96
S19 ₂	3	2	2	1	80
S20 ₁	3	0	0	2	96
S20 ₂	5	1	1	3	64
total	99	18	24	36	

Delayed trials were those shown 0.5 s or more past the intended display time. Results tallied separately for real-time sessions 1 and 2 for each subject, and tallied separately for each stimulus class, as in Table 1. Class numbers correspond to four distinct Fribble object classes. Number of search trials per class varied in each session, as seen in final column.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Motion effects on ROI computed responses for real-world objects searches.

Subject _{session}	max motion	corr1	corr2	corr3	corr4	average
S1 ₁	8.5	0.56	-0.22	0.63	0.03	0.25
S1 ₂	1.7	0.44	0.21	0.82	-0.19	0.32
S2 ₁	2.2	0.43	-0.06	0.79	0.17	0.33
S2 ₂	1.1	0.41	0.23	0.48	0.47	0.40
S3 ₁	2.1	0.39	0.55	0.71	-0.43	0.31
S3 ₂	9.6	0.63	0.44	0.33	-0.17	0.31
S4 ₁	2.2	0.91	-0.24	-0.59	0.34	0.11
S4 ₂	1.1	0.82	0.23	-0.74	0.20	0.13
S5 ₁	2.0	0.59	-0.37	0.54	0.08	0.21
S5 ₂	1.2	0.71	0.35	0.77	0.20	0.51
S6 ₁	2.3	0.39	0.57	0.16	-0.09	0.26
S6 ₂	2.7	0.69	0.33	-0.07	-0.62	0.08
S7 ₁	3.1	0.09	-0.15	0.74	-0.09	0.15
S7 ₂	2.2	0.64	-0.05	0.62	-0.09	0.28
S8 ₁	2.9	0.19	-0.04	0.77	0.61	0.38
S8 ₂	2.1	0.10	0.10	0.55	0.04	0.20
S9 ₁	2.0	0.70	0.34	0.24	0.10	0.35
S9 ₂	2.2	0.26	0.45	0.55	-0.06	0.30
S10 ₁	1.2	0.40	0.11	0.40	0.34	0.31
S10 ₂	2.1	0.76	0.42	0.63	0.38	0.55
average	-	0.51	0.16	0.42	0.06	0.29

Correlation between computed responses for each of four class ROIs using offline preprocessing on full scan session versus real-time preprocessing on small time blocks within single runs. Average column shows average correlation results across the four ROIs for a given subject and session. Maximum motion magnitude among the starts of all runs also included, pooled from x, y, z translations (in mm) and yaw, pitch, roll rotations (in degrees).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Motion effects on ROI computed responses.

Subject _{session}	max motion	corr1	corr2	corr3	corr4	average
S11 ₁	1.2	0.50	0.49	-0.55	0.51	0.24
S11 ₂	0.7	0.54	0.48	-0.40	0.12	0.19
S12 ₁	4.8	0.31	0.24	-0.08	-0.04	0.11
S12 ₂	1.2	0.87	0.64	0.67	-0.59	0.40
S13 ₁	2.5	0.56	0.70	0.45	-0.17	0.39
S13 ₂	1.6	0.51	0.68	0.62	-0.10	0.43
S14 ₁	2.4	0.60	0.65	-0.10	0.57	0.43
S14 ₂	1.2	0.39	0.74	-0.01	0.44	0.39
S15 ₁	1.2	0.44	0.53	-0.54	0.23	0.17
S15 ₂	7.0	0.34	-0.07	-0.01	-0.15	0.03
S16 ₁	2.7	0.60	0.72	0.50	0.20	0.51
S16 ₂	1.4	0.84	0.65	0.50	0.20	0.55
S17 ₁	0.7	0.46	0.75	0.37	0.56	0.54
S17 ₂	2.7	0.57	0.71	0.44	0.48	0.55
S18 ₁	2.7	0.59	0.62	0.19	-0.57	0.21
S18 ₂	1.9	0.47	0.54	0.20	-0.67	0.14
S19 ₁	2.0	0.60	0.70	0.69	0.29	0.57
S19 ₂	2.6	0.74	0.60	0.62	0.27	0.56
S20 ₁	1.7	0.59	0.57	-0.14	-0.57	0.14
S20 ₂	1.0	0.62	0.32	0.22	-0.60	0.14
average	-	0.56	0.56	0.19	0.01	0.33

Correlation between computed responses for each of four class ROIs using offline preprocessing on full scan session versus real-time preprocessing on small time blocks within single runs. Average column shows average correlation results across the four ROIs for a given subject and session. Maximum motion magnitude among the starts of all runs also included, pooled from x, y, z translations (in mm) and yaw, pitch, roll rotations (in degrees).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Convergence for searches of real-world objects as measured by Z score metric discussed in Sec. 2.10.

Subject _{session}	z1	z2	z3	z4
S1 ₁	-0.36	1.29	-0.34	2.14
S1 ₂	-0.82	1.26	.01	-0.68
S2 ₁	-0.09	0.15	-0.43	0.39
S2 ₂	0.01	0.38	-0.75	0.67
S3 ₁	-1.34	0.77	-0.41	-1.01
S3 ₂	-0.87	2.60	0.60	-0.32
S4 ₁	0.30	0.71	-0.35	2.27
S4 ₂	-.49	-1.04	-0.08	-0.45
S5 ₁	0.35	-0.08	-1.23	-0.95
S5 ₂	0.52	1.14	-0.32	-0.88
S6 ₁	-0.57	2.77	0.79	2.37
S6 ₂	-0.01	-1.43	-0.20	2.58
S7 ₁	-0.57	1.95	-1.01	1.00
S7 ₂	0.11	1.91	-0.54	1.30
S8 ₁	2.23	0.36	0.07	-0.37
S8 ₂	-1.26	0.14	1.23	0.83
S9 ₁	0.20	0.20	-1.38	-0.93
S9 ₂	-0.15	-0.80	0.05	-0.42
S10 ₁	-1.35	-0.34	-0.42	-1.07
S10 ₂	-0.69	-0.21	-0.18	0.14

Z scores of 1.8 and above in bold.

Table 6

Consistency between searches of real-world objects as measured by Z score metric discussed in Sec. 2.10.

Subject _{session}	z1	z2	z3	z4
S1	-1.02	1.80	-0.39	-0.59
S2	0.34	-1.40	-0.21	-1.78
S3	-1.91	-0.82	1.44	0.04
S4	-0.92	0.10	-1.35	0.44
S5	-1.12	2.19	-0.71	0.41
S6	0.20	-0.67	0.86	-0.83
S7	0.21	0.74	0.60	0.21
S8	-0.49	-0.53	1.79	1.35
S9	1.69	-0.33	0.65	-0.91
S10	-1.54	-0.59	.09	-1.36

Z scores of 1.8 and above in bold.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 7

Temporal evolution of real-world objects searches.

Subject _{session}	var1	var2	var3	var4	dist1	dist2	dist3	dist4
S1 ₁	0.01	0.01	0.02	-0.01	1.81	0.42	0.69	1.20
S1 ₂	0.01	-0.00	0.01	0.00	0.87	1.11	1.90	0.86
S2 ₁	-0.01	0.00	-0.00	-0.01	1.11	0.74	0.64	0.70
S2 ₂	0.00	-0.02	-0.00	-0.03	0.75	1.08	2.28	2.48
S3 ₁	-0.00	-0.03	0.01	0.03	1.61	1.14	1.39	2.29
S3 ₂	-0.02	0.02	0.01	-0.00	1.14	0.85	1.40	1.72
S4 ₁	0.02	0.01	0.02	0.01	1.49	2.29	0.92	1.22
S4 ₂	-0.01	0.03	0.01	-0.01	0.60	1.19	1.61	1.21
S5 ₁	-0.01	0.03	0.00	0.00	0.97	1.55	1.00	1.70
S5 ₂	-0.01	-0.03	0.00	0.02	1.22	2.55	1.06	1.45
S6 ₁	-0.01	-0.02	0.02	-0.01	1.37	1.47	1.40	0.65
S6 ₂	0.02	0.04	-0.02	-0.01	1.83	2.36	2.29	1.65
S7 ₁	-0.00	-0.01	-0.00	0.03	1.20	0.74	1.44	1.29
S7 ₂	-0.01	0.02	-0.01	0.01	1.62	0.80	1.37	1.29
S8 ₁	-0.01	-0.00	-0.00	-0.03	1.40	1.56	1.00	1.36
S8 ₂	-0.01	0.01	-0.03	-0.00	0.80	1.31	2.54	0.96
S9 ₁	0.01	0.03	-0.02	-0.01	0.45	0.48	1.12	0.60
S9 ₂	-0.01	0.01	0.00	0.01	1.45	1.57	1.65	1.23
S10 ₁	-0.02	-0.01	-0.03	0.02	1.35	1.04	1.60	1.58
S10 ₂	-0.01	0.01	0.01	0.02	1.65	1.15	2.61	1.65

var_n and dist_n, corresponding to the change in variance and mean of locations visited in the first and second half of each scan session by the search of stimulus class *n*, are as defined in Eqns. 4 and 5, respectively. Distances of 2.0 and greater in bold.

Table 8

Convergence for searches in Fribbles spaces as measured by Z score metric discussed in Sec. 2.10.

Subject _{session}	z1	z2	z3	z4
S11 ₁	-0.08	0.40	-0.13	3.90
S11 ₂	3.40	0.18	0.63	-0.38
S12 ₁	1.20	0.56	1.70	0.25
S12 ₂	0.42	1.10	0.51	1.90
S13 ₁	0.91	1.10	-0.43	0.79
S13 ₂	2.42	-1.2	1.42	2.67
S14 ₁	0.39	1.43	0.43	0.95
S14 ₂	1.45	0.60	0.52	1.40
S15 ₁	2.76	1.66	2.20	0.18
S15 ₂	1.45	1.69	-0.83	1.87
S16 ₁	1.72	2.10	1.80	0.98
S16 ₂	2.87	-1.10	-0.11	-0.22
S17 ₁	-0.47	-0.27	0.89	-0.59
S17 ₂	2.42	2.97	1.76	0.47
S18 ₁	0.54	1.57	1.82	1.72
S18 ₂	1.43	0.93	1.17	2.30
S19 ₁	2.00	0.93	3.00	4.20
S19 ₂	0.90	0.86	1.40	2.10
S20 ₁	0.77	1.07	2.86	2.84
S20 ₂	1.24	1.66	0.39	1.41

Z scores of 1.8 and above in bold.

Table 9

Consistency between searches in Fribbles spaces as measured by Z score metric discussed in Sec. 2.10.

Subject _{session}	z1	z2	z3	z4
S11	2.10	0.57	0.43	2.20
S12	-0.53	1.40	-0.03	1.40
S13	0.46	0.62	-1.20	-1.40
S14	-0.59	-0.19	-1.20	1.22
S15	-1.10	-1.10	1.43	2.96
S16	-0.29	0.85	0.39	0.54
S17	2.28	3.14	3.28	-0.99
S18	-1.70	-0.03	0.28	-1.80
S19	1.40	0.30	0.97	3.80
S20	0.63	0.15	0.46	0.05

Z scores of 1.8 and above in bold.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 10

Temporal evolution of Fribble searches.

Subject _{session}	var1	var2	var3	var4	dist1	dist2	dist3	dist4
S11 ₁	0.02	0.01	0.05	0.03	1.38	1.15	0.58	1.13
S11 ₂	0.05	0.03	-0.04	0.03	1.42	2.30	1.55	1.49
S12 ₁	0.02	0.02	-0.11	0.00	1.38	1.29	2.18	2.25
S12 ₂	-0.04	0.01	-0.01	0.07	1.03	2.49	0.87	2.28
S13 ₁	-0.02	0.12	0.03	0.01	2.39	0.89	0.99	0.50
S13 ₂	0.10	0.01	-0.03	0.03	1.44	0.60	.84	1.46
S14 ₁	-0.08	-0.01	-0.07	0.03	3.05	2.33	0.77	1.30
S14 ₂	-0.06	-0.05	-0.03	-0.06	1.68	1.19	1.35	1.95
S15 ₁	0.02	0.01	-0.08	0.08	1.39	1.47	1.89	1.31
S15 ₂	0.06	-0.02	0.01	-0.08	1.83	0.98	1.07	1.57
S16 ₁	-0.00	0.09	-0.02	-0.08	0.52	0.54	1.73	0.95
S16 ₂	0.05	0.05	-0.08	0.05	0.60	1.34	1.17	1.32
S17 ₁	0.01	0.06	0.03	0.07	1.31	1.03	1.16	1.20
S17 ₂	0.03	0.02	-0.08	-0.05	2.41	0.81	0.76	1.09
S18 ₁	0.10	-0.03	0.05	0.00	0.25	0.22	0.81	1.07
S18 ₂	0.01	-0.06	0.07	0.00	1.31	1.06	1.09	0.39
S19 ₁	-0.01	-0.03	0.01	-0.02	1.14	1.72	1.42	1.71
S19 ₂	-0.08	0.00	0.01	0.03	2.12	0.78	1.97	2.30
S20 ₁	-0.03	-0.07	-0.05	0.03	1.93	1.80	2.59	1.09
S20 ₂	-0.05	0.00	0.06	0.09	1.34	1.09	0.84	0.85

var_n and dist_n, corresponding to the change in variance and mean of locations visited in the first and second half of each scan session by the search of stimulus class *n*, are as defined in Eqns. 4 and 5, respectively. Distances of 2.0 and greater in bold.