Brain and Cognition 74 (2010) 112–122

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Temporal and spatial integration of face, object, and scene features in occipito-temporal cortex

Thomas W. James [a,c,*], Eunji Huh [a,b], Sunah Kim [c]

[a] Department of Psychological and Brain Sciences, Indiana University, United States
[b] Department of Speech and Hearing Sciences, Indiana University, United States
[c] Cognitive Science Program, Program in Neuroscience, Indiana University, United States

ARTICLE INFO

Article history:
Accepted 26 July 2010
Available online 19 August 2010

Keywords:
Face recognition
Feature integration
Fusiform face area
fMRI
Restricted viewing

ABSTRACT

In three neuroimaging experiments, face, novel object, and building stimuli were compared under conditions of restricted (aperture) viewing and normal (whole) viewing. Aperture viewing restricted the view to a single face/object feature at a time, with the subjects able to move the aperture continuously though time to reveal different features. An analysis of the proportion of time spent viewing different features showed stereotypical exploration patterns for face, object, and building stimuli, and suggested that subjects constrained their viewing to the features most relevant for recognition. Aperture viewing showed much longer response times than whole viewing, due to sequential exploration of the relevant isolated features. An analysis of BOLD activation revealed face-selective activation with both whole viewing and aperture viewing in the left and right fusiform face areas (FFA). Aperture viewing showed strong and sustained activation throughout exploration, suggesting that aperture viewing recruited similar processes as whole viewing, but for a longer time period. Face-selective recruitment of the FFA with aperture viewing suggests that the FFA is involved in the integration of isolated features for the purpose of recognition.

1. Introduction

Recognition of objects in our environment is an essential cognitive operation, and one that is performed with relative ease by humans and other primates. Decades of research into the cognitive and neural mechanisms of object recognition suggest that objects can (and potentially must) be decomposed into features for identification to occur (Biederman, 1997; Marr, 1982; Poggio & Edelman, 1990; Schyns, Goldstone, & Thibaut, 1998). These features are alternatively called parts, components, geons, or primitives, and their exact nature remains a point for debate. By most definitions, object recognition involves binding together multiple features across space and time, by recruiting spatial and temporal integration processes. For *visual* object recognition, spatial feature integration has received more intense scrutiny than temporal integration. When studied in relation to human face recognition, spatial feature integration is often called holistic, configural, or relational processing (Gauthier & Tarr, 2002; Maurer, Le Grand, & Mondloch, 2002; Moscovitch, Winocur, & Behrmann, 1997; Tanaka & Farah, 1993). Face recognition, which is a special case of object recognition, is notable due to its extreme efficiency in human observers in terms of the short time required for feature integration (Bruce & Young, 1986; Farah, Wilson, Drain, & Tanaka, 1998; Gauthier & Tarr, 2002; Maurer et al., 2002; McKone, Martini, & Nakayama, 2001; Moscovitch et al., 1997; Rhodes, 1988; Tanaka & Farah, 1993; Yovel & Duchaine, 2006). In fact, it has been suggested that holistic processing is characterized by the *simultaneous* integration of face features (Rossion, 2008).

The striking recruitment of spatial feature integration for visual face recognition is contrasted with more generic object recognition (such as discriminating a stapler from a telephone), which relies less on whole object processing, and instead relies more on what is referred to as feature-based, parts-based, sequential, componential, piecemeal, or analytic processing (Marsolek & Burgund, 1997; Tanaka & Farah, 1993; Yovel & Duchaine, 2006). In fact, in many cases, objects can be recognized based on a single diagnostic feature (Bruce & Young, 1986; Tanaka & Farah, 1993). In other cases, though, formation of a coherent object percept involves integration of multiple features in a time-consuming, sequential manner.

Faces and other objects are not only recognized visually, but can also be recognized using the sense of touch, especially when objects are actively explored haptically (Klatzky, Lederman, & Reed, 1987). Although both the visual and haptic systems are able to extract many properties of objects, to efficiently recognize them, both systems rely heavily on shape features (Kilgour & Lederman, 2002; Klatzky et al., 1987). Because the receptor and peripheral nerve

* Corresponding author. Address: 1101 E Tenth St., Bloomington, IN 47405, United States. Fax: +1 812 855 4691.
E-mail address: thwjames@indiana.edu (T.W. James).

0278-2626/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandc.2010.07.007

systems of the eyes and skin are quite different, it is no surprise that object recognition performance using haptics and vision is also different. Perhaps the most salient of these differences is the time required for haptic recognition, especially for objects where the features are spatially isolated, such as faces (Kilgour & Lederman, 2002). In this case, the visual receptor array of the eye has an advantage over the somatosensory array of the fingers, because it can sample large parts of the environment simultaneously. Haptics, in contrast, must usually sample spatially separated features sequentially (Loomis, 1981). This difference suggests that visual object recognition may involve very fast or even simultaneous spatial feature integration, whereas haptic object recognition may involve sequential feature integration. This effect would be exaggerated for a stimulus class such as faces, which requires analysis of several spatially isolated features for successful recognition.

Investigations into the neural substrates of object and face recognition have focused on the ventral occipito-temporal cortex. Although an entire network of brain regions is activated preferentially with faces (for example, see Fox, Iaria, & Barton, 2009; Haxby, Hoffman, & Gobbini, 2000; Sergent, Ohta, & Macdonald, 1992; Weiner & Grill-Spector, 2010), the most studied region in that network is the fusiform face area (FFA). The FFA is located on the posterior aspect of the fusiform gyrus (FG), is more reliably found in the right hemisphere than the left hemisphere, and is face-selective, that is, the FFA is activated more strongly with face images than with any other type of object image (for review, see Kanwisher & Yovel, 2006). An ultimate explanation for the stimulus specificity of the FFA has been hotly debated (for example, see Gauthier, 2000; McKone, Kanwisher, & Duchaine, 2006), however, recently, there appears to be some consensus that the neurocognitive mechanism instantiated in the FFA is holistic processing or spatial integration (Gauthier & Tarr, 2002; Kanwisher & Yovel, 2006; Maurer et al., 2002; Rossion & Gauthier, 2002; Schiltz & Rossion, 2006; Sergent et al., 1992; Yovel & Kanwisher, 2005).

Recently, it was shown that the FG is recruited for haptic face recognition, that is, recognition of tangible face stimuli by exploring with the digits in the absence of vision. In two studies (James, Servos, Kilgour, Huh, & Lederman, 2006; Kilgour, Kitada, Servos, James, & Lederman, 2005), it was found that the *left* FG was recruited more with haptic exploration of face stimuli than nonsense objects, and more for haptic exploration of familiar than nonfamiliar face stimuli. In contrast, most studies of visual face recognition have found a bias to stronger right FFA activation (Kanwisher, McDermott, & Chun, 1997; Puce, Allison, Gore, & McCarthy, 1995). The difference in findings between vision and haptics raises the possibility that different exploration strategies may recruit distinct feature integration processes, and that these processes may be lateralized to opposite hemispheres. We hypothesized that recognition requiring more simultaneous feature integration, such as visual face recognition, would preferentially activate the right FG, whereas recognition requiring more sequential feature integration, such as haptic face recognition, would preferentially activate the left FG. This hypothesis is consistent with other models of visual face processing, which suggest that representations of whole faces are the domain of the right hemisphere, whereas face parts are the domain of the left hemisphere (Rhodes, 1985).

To investigate the hypothesis that activation in the left and right FG is lateralized based on different feature integration processes, we developed an restricted-viewing task that mimicked important aspects of haptic exploration (Loomis, 1981). Subjects viewed stimuli through a small aperture that forced them to use a more sequential exploration strategy compared with when they were allowed to view whole stimuli. Experiment 1 contrasted aperture viewing of faces and novel objects (Greebles) and found no evidence for left lateralization under sequential exploration conditions. Instead, strong, sustained activation with aperture viewing was seen bilaterally in the FFA. Experiment 2 followed up this unexpected result and attempted to rule out alternative explanations for the results of Experiment 1 by testing two whole-viewing conditions that controlled for stimulus presentation time and the presence of dynamic image changes. Experiment 3 was designed to specifically assess face selectivity of aperture-viewing activation in the FFA.

2. Experiment 1

A restricted-viewing paradigm was developed to contrast sequential processing of faces and novel objects (Greebles). In the aperture-viewing conditions, subjects were allowed to explore the stimulus through a small aperture (Inui & Miyamoto, 1984; Jansen, Blackwell, & Marriott, 2003). Whole viewing was also included as a control condition. In the whole-viewing condition, subjects viewed faces and Greebles briefly. Based on the hypothesis that the right FG is more involved simultaneous feature integration and the left FG in sequential feature integration, we predicted that aperture viewing would show greater activation in the left FG than the right. Based on our previous findings of face-selective activation only in the left FG with haptic face recognition (James et al., 2006; Kilgour et al., 2005), it was hypothesized that aperture viewing would also produce greater face selectivity for faces over Greebles in the left FG than in the right.

2.1. Methods and materials

2.1.1. Subjects

Subjects were recruited from the undergraduate and graduate student populations at Indiana University. The study was approved by the IUB Institutional Review Board and Human Subjects Committee, and all subjects signed informed consent. Ten subjects (five male and five female) participated. One subject's data was not used because of head motion exceeding 1.5 mm. The final number of subjects analyzed was nine ($N = 9$).

2.1.2. Stimuli

Grayscale images of 24 female human faces and 24 novel objects (Fig. 1A) were used. Faces were taken from the Face Database of the Max Planck Institute for Biological Cybernetics (http://faces.kyb.tuebingen.mpg.de/). The novel objects were Greebles, which have been used in many previous studies on novel object recognition (for example, see Gauthier & Tarr, 1997; James, Shima, Tarr, & Gauthier, 2005). Faces were purposefully chosen as all female and Greebles were chosen as all the same gender to increase the item-to-item similarity of the stimulus set. The resolution of all images was 256×256 pixels.

2.1.3. Viewing procedures

For the whole-viewing condition (WH), subjects were presented with an image for 183 ms. When images were presented to subjects in the MRI on the rear-projection screen, they subtended approximately 6° of visual angle. For the aperture-viewing condition (AP), the image was obscured and only made visible through a small square aperture of 24×24 pixels size, or 35 min of visual angle. This size was determined based on two main considerations used in other research using restricted viewing (Inui & Miyamoto, 1984; Jansen et al., 2003). First, the aperture was small enough that it restricted the subjects to viewing to only a single internal face feature (i.e., eye, nose, or mouth) at one time. Second, the aperture was big enough that internal face features could be correctly categorized before learning as an eye, nose, or mouth 100% of the time when the aperture was centered on the feature (Fig. 1C). Subjects were able to view different parts of the image over a 12 s period

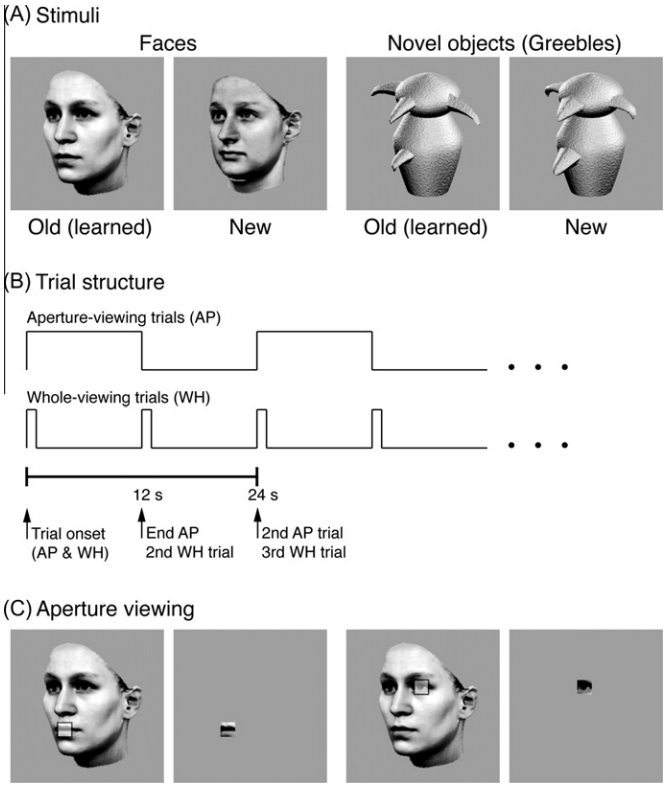

Fig. 1. Stimuli and experimental protocols for Experiment 1. (A) Stimuli were female faces and novel objects (Greebles). Subjects were presented with learned and unlearned stimuli. The task was to decide if a face or Greeble was old (learned) or new (unlearned). (B) Stimuli were presented using either aperture or whole viewing. Aperture-viewing trials consisted of 12 s of stimulation with a 12 s ITI. Whole-viewing trials consisted of 183 ms of stimulation with an 11.817 s ITI. (C) Example frames from an aperture-viewing trial. The first and third images show the position of the aperture on the whole face. The second and fourth images show the subject's view through the aperture.

of time (Fig. 1B) by moving the aperture in real time through the use of a MR-compatible button box that resembled a video game controller. Four buttons allowed movements up, down, left, and right; pressing combinations of buttons allowed diagonal movements. While a button was pressed, the aperture moved continuously. Restricting the participant's view of the object was intended to make them adopt a sequential exploration strategy when attempting to recognize the images (Loomis, 1981). The aperture moved with a velocity of 5 pixels every two frames, which equated to 150 pixels/s at the frame rate of 60 Hz or approximately 3.6° of visual angle per second. At that velocity, the subjects could move, for example, from the left eye feature to the right eye feature in approximately 500 ms.

2.1.4. Learning procedures

Before imaging, subjects were familiarized with the aperture-viewing procedure and exposed to half of the stimuli with both aperture and whole viewing. Learning was performed in a simulated MRI device with no magnetic field, but equipped with patient bed, bore, head coil, and visual presentation and response devices with similar dimensions to the real MRI. During learning, subjects associated names with 12 faces and 12 Greebles. Names for faces were women's first names and names of Greebles were pronounceable pseudowords. Six of the faces and Greebles were learned using aperture viewing and six were learned using whole viewing. Even though subjects did not need to be familiarized with the WH condition, the subjects were required to complete the same number of whole-viewing as aperture-viewing learning trials, such that learning exposure would not be a confound across conditions. Subjects were required to verify the names of faces and Greebles with 100% accuracy on two consecutive blocks of 12 trials to reach criterion. Blocks of face and Greeble trials were conducted separately and the criterion had to be met for both object categories before learning was complete. Mean learning time across subjects was approximately 90 min. The learning procedure was conducted either the day before or the day of testing.

2.1.5. fMRI testing procedures

Imaging was conducted at the Imaging Research Facility at Indiana University Bloomington with a 3T Siemens Magnetom TRIO. Subjects lay supine with their head secured in the 8-channel head coil with foam padding. Testing was divided into 12 5-min runs. Because aperture-viewing trials were twice as long as whole-viewing trials, eight aperture-viewing runs with 12 aperture-viewing trials per run, but only four whole-viewing runs with 24 whole-viewing trials per run, were required to produce the desired total of 96 trials for each of the aperture- and whole-viewing conditions, split evenly between face and Greeble stimulus categories (i.e., 48 trials per cell of the experimental design). On each trial, subjects performed an old/new recognition task, classifying faces and Greebles as either learned or unlearned. Subjects were instructed to make accurate decisions at the expense of speed.

For both aperture- and whole-viewing runs, hemodynamic responses were allowed to return to baseline for ~12 s between trials while the subjects fixated a small central cross. The start of a trial was signaled by the disappearance of the fixation cross 750 ms before stimulus onset. For whole-viewing trials, the whole image was flashed for 183 ms, with a fixed inter-stimulus interval of 11,817 ms (~12 s). For aperture-viewing trials, the aperture appeared in a random location on the part of the screen occupied by the stimulus image and subjects were allowed to explore the image for 12 s, followed by a 12 s inter-stimulus interval. The stimulus sets used for testing were comprised of the entire set of 24 faces and 24 Greebles, 6 aperture-viewing-learned, 6 whole-viewing-learned, and 12 unlearned for each stimulus category. Stimuli were always tested using the same viewing condition under which they were learned. Thus, the 48 trials for each combination of viewing condition and stimulus category were accomplished by presenting six learned and six unlearned stimuli (12 total) four times each. Faces and Greebles were presented in separate runs.

2.1.6. MRI acquisition

Functional imaging was done using a gradient echo EPI pulse sequence (TR = 2000 ms; TE = 25 ms; flip angle = 70°) with 33 axial slices oriented approximately parallel to the ACPC plane, 3.4 mm thick, for whole-brain imaging. In-plane resolution was 3.4 × 3.4 mm (field of view 220 × 220 mm, with 64 × 64 matrix) for an iso-metric voxel size of 3.4 mm. High-resolution T1-weighted anatomical volumes were acquired using Turbo-flash 3-D (TI = 1100 ms, TE = 3.93 ms, TR = 14.375 ms, flip angle = 12°) with 160 sagittal slices, and iso-metric voxel size of 1 × 1 × 1 mm.

2.1.7. Data analysis

Imaging data were pre-processed using Brainvoyager™ 3-D analysis tools. Anatomical volumes were transformed into the common stereotactic space of Talairach and Tournoux using an eight-parameter affine transformation. Functional data were aligned to the first volume of the functional run acquired closest in time to the anatomical series. Each functional run was then aligned to the transformed anatomical volumes using an intensity-based algorithm, transforming the functional data to the common stereotactic space. Before transformation, functional data underwent a linear trend removal, 3-D spatial Gaussian filtering

(FWHM 6 mm), slice scan-time correction, and 3-D motion correction.

2.2. Results

2.2.1. Behavioral data

Accuracies and reaction times are shown in Fig. 2. Each dependent measure was analyzed using a two-factor within-subjects ANOVA with viewing condition and stimulus category as factors. For accuracy, there was a significant two-way interaction ($F_{(1,8)} = 73.3$, $p < 0.001$). The interaction was due to very high accuracy with whole faces compared with the other three conditions. Post-hoc *t*-tests showed that whole faces were recognized more accurately than aperture faces ($t_{(8)} = 5.52$, $p < 0.001$), than whole Greebles ($t_{(8)} = 8.32$, $p < 0.001$), and than aperture Greebles ($t_{(8)} = 3.93$, $p = 0.004$). No other pair-wise post-hoc test neared significance using two-tailed tests with a false positive error rate of $p < 0.05$. There were also significant main effects of viewing condition ($F_{(1,8)} = 14.4$, $p = 0.005$) and stimulus category ($F_{(1,8)} = 8.6$, $p = 0.02$), however, these main effects could also be explained by the extremely high accuracy with whole faces. For reaction time, there was no significant interaction ($F_{(1,8)} = 1.91$, $p > 0.1$). There was a significant main effect of viewing condition ($F_{(1,8)} = 322.2$, $p < 0.001$), with aperture viewing requiring more time than whole viewing, as expected.

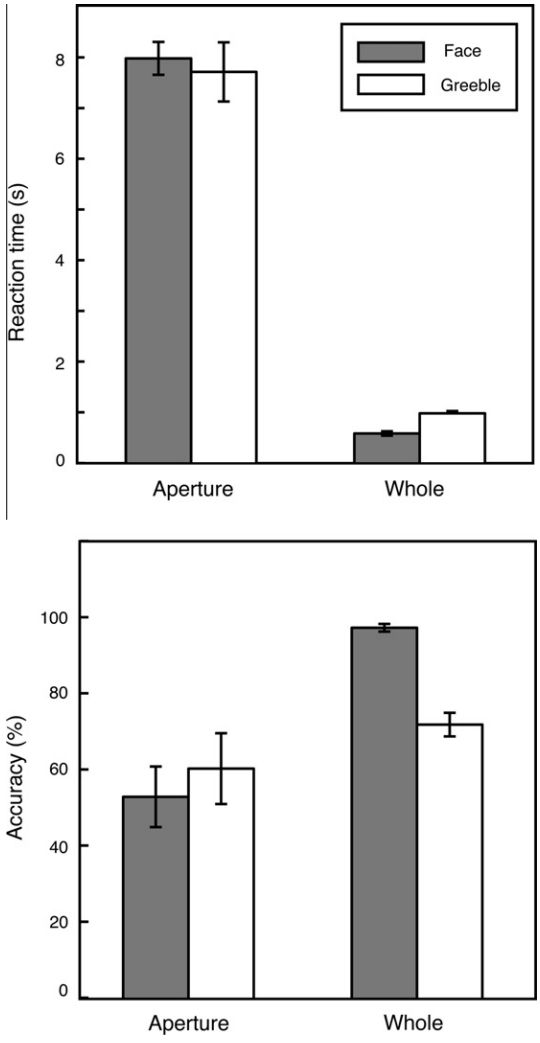

Fig. 2. Accuracy and reaction time as a function of stimulus class and viewing condition.

To summarize the behavioral findings, aperture viewing increased recognition time substantially for both faces and Greebles compared with whole viewing. One goal of the present study was to mimic the more sequential nature of haptic face recognition using aperture viewing. One of the most salient aspects of haptic face recognition compared to visual face recognition is the time required to haptically recognize faces. Mean recognition times with aperture viewing of faces were approximately equal to mean recognition times previously reported for haptic face recognition (James et al., 2006; Kilgour et al., 2005), suggesting that the aperture-viewing task mimicked that aspect of haptic face recognition. Aperture viewing also reduced accuracy for faces compared with whole viewing. Consistent with previous work on holistic processing of faces (McKone et al., 2001; Moscovitch et al., 1997; Rhodes, 1988; Rossion, 2008), the data suggest that whole image spatial integration processes are important for recognizing faces.

2.2.2. Viewing-time maps

For the aperture-viewing task, subjects were allowed to control the movement of the aperture. By calculating the amount of time that each pixel in the image was viewed on average across a trial, maps were created that provide evidence about the most salient or important features for performing the task. Viewing-time maps for each subject for both faces and Greebles are shown in Fig. 3. Although there was variability across subjects, the general pattern of viewing time for faces was similar across individuals. Most subjects spent a majority of the trial examining the top half of the face, with particular attention paid to the eyes and nose. There was more individual variability with Greebles than faces, but even with Greebles, subjects consistently spent more time viewing the upper half of the Greebles, and specifically the appendages rather than the other parts. Viewing times were also calculated separately for correct and incorrect trials to see if a subject's exploration pattern affected their performance. Viewing-time maps for correct and incorrect trials showed more variability than the maps in Fig. 3, presumably because they were calculated from half the trials on average. Viewing-time maps for correct and incorrect trials were similar to each other.

2.2.3. Region of interest (ROI) analysis

The FFA was functionally localized in each individual subject from a statistical parametric map generated using a general linear model (GLM) and a post-hoc contrast of whole-face and whole-Greeble conditions. Although, comparisons of faces with houses are used frequently to functionally localize the FFA, comparisons of faces with objects are also used routinely (Halgren et al., 1999; Kanwisher et al., 1997; Kung, Peissig, & Tarr, 2007), and there is evidence that faces activate the FFA more than novel objects such as Greebles (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). Even at a relatively liberal statistical threshold ($t = 2.5$), however, the right FFA was only found in five subjects and the left FFA in only one. A group-average map contrasting whole faces and Greebles, on the other hand, produced reliable clusters on the left and right FG (Fig. 4A). The map was generated using a fixed-effects general linear model and thresholded using the false discovery rate (FDR) technique ($q < .05$). The clusters were found at Talairach coordinates (left: −34, −50, −9; right: 35, −52, −11), highly consistent with coordinates previously reported for the FFA (Kanwisher et al., 1997). Timecourses were extracted from these ROIs for each of the four experimental conditions.

It should be noted that identifying a functional ROI using the same dataset that is subsequently analyzed could potentially represent a case of non-independent ROI analysis (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009; Poldrack & Mumford, 2009). The goal of the timecourse analysis, however, was not to compare whole faces and whole Greebles. Rather, the goals of the time-

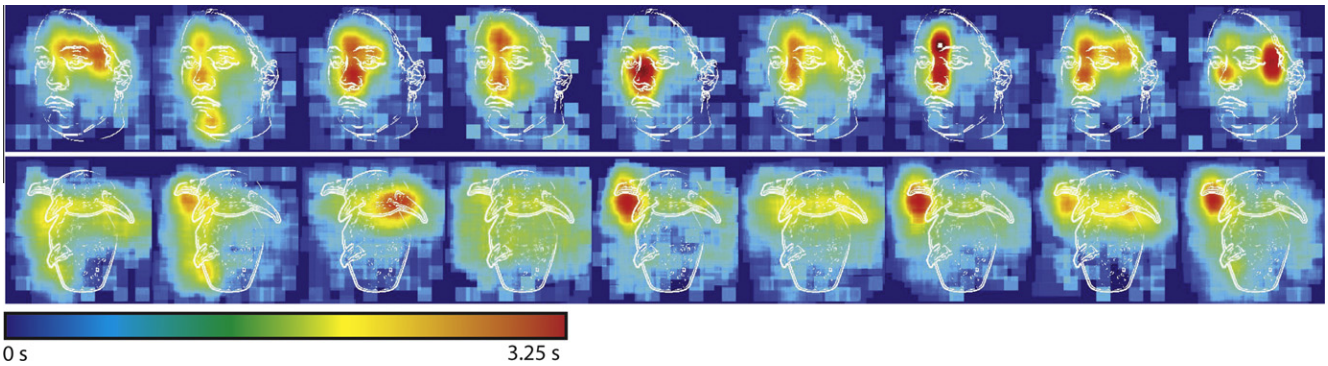

Fig. 3. Viewing-time maps for Experiment 1. The color of each pixel represents the mean viewing time across trials during the aperture-viewing condition. Each column is a different subject. Maps in the top row are viewing times for faces and the bottom row are viewing times for Greebles. The white outline is a representative stimulus from the set. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

course analysis were to compare aperture faces and aperture Greebles, to compare the pattern of activation across the left and right hemispheres, and to compare the widths of the activation functions with whole viewing and aperture viewing. Thus, the tests conducted as part of the timecourse analysis were statistically independent of the test used to functionally define the ROIs.

Fig. 4B shows the timecourses for the two stimulus classes and two viewing conditions averaged across subjects in the left and right FFAs. A summary graph of BOLD activation from the FFA with aperture viewing is shown in Fig. 4C. BOLD activation levels were calculated as the mean across a time window from 4 to 24 s. A 2×2 ANOVA conducted on the aperture-viewing data with hemisphere and stimulus class as factors showed no significant effects. Face stimuli produced slightly more activation than Greeble stimuli with aperture viewing, but the differences did not reach significance. There was no evidence of lateralization of the responses with aperture viewing, nor any evidence for the hypothesized lateralization of face selectivity.

The most striking result, however, was one that was not hypothesized. Aperture viewing produced a more sustained response than whole viewing in both the left and right FFAs (Fig. 4B). The whole-viewing timecourses peaked at 6 s post-stimulus-onset and returned to baseline by 12 s, which is the expected hemodynamic response function with a single briefly presented stimulus. Activation timecourses with aperture viewing, however, peaked later and required 24 s to return to baseline. The activation with aperture viewing was similar to what would be produced in a blocked design, in which whole faces were presented repeatedly for 12 s, suggesting that processing in the FFA was sustained throughout aperture-viewing trials.

2.3. Discussion

Aperture viewing produced several expected results. First, relative to whole viewing, aperture viewing increased the time necessary to recognize face and Greeble stimuli. The time required to recognize aperture-viewed faces was similar to what was previously found with haptic face recognition (James et al., 2006; Kilgour et al., 2005). Second, aperture viewing disrupted the ability to recognize faces. This finding is consistent with extensive research suggesting that faces are processed holistically (Bruce & Young, 1986; Farah et al., 1998; Gauthier & Tarr, 2002; Maurer et al., 2002; McKone et al., 2001; Moscovitch et al., 1997; Rhodes, 1988; Tanaka & Farah, 1993; Yovel & Duchaine, 2006). Previous aperture-viewing experiments, however, have used only static apertures (Haig, 1985). Thus, our findings contribute to the existing literature by showing that face processing is disrupted even when subjects have control over the information that they view through the aperture. Third, we found that subjects' exploration of faces through the aperture focused on the top half of the face, which is consistent with previous investigations of the relative importance of different internal face features for recognition (Haig, 1986; Schyns, Bonnar, & Gosselin, 2002; Sekuler, Gaspar, Gold, & Bennett, 2004; Yarbus, 1967).

Some of the results, however, were not predicted. First, we hypothesized that aperture viewing would produce greater face selectivity between faces and Greebles in the left FG than the right. This hypothesis was based on our previous work with haptic exploration of faces and nonsense objects (James et al., 2006; Kilgour et al., 2005) and on work suggesting that the left hemisphere may be more involved in sequential, feature-based processing (Marsolek, 1999; Rhodes, 1985). However, we found no evidence for lateralization of face selectivity in the FFA with aperture viewing. Second, we hypothesized that, regardless of stimulus type, aperture viewing would recruit the left FG more than the right, because aperture viewing is less dependent on simultaneous integration of object features than whole viewing (Loomis, 1981; Rossion, 2008; Schiltz & Rossion, 2006). However, we also found no evidence for this hypothesis. In fact, in both the left and right FFA, aperture viewing produced more sustained activation than whole viewing. Together, these findings suggest that the functional organization of the FG is not lateralized based on the time scale of feature integration (i.e., simultaneous versus sequential). They also suggest that the FFA in both hemispheres is involved in simultaneous *and* sequential integration of isolated object features. One concern about basing these premises on the present experiment, however, is the discrepancy in stimulus presentation time between the whole-viewing and aperture-viewing conditions. This concern is addressed in Experiment 2.

3. Experiment 2

In Experiment 1, an attempt was made to equate behavioral performance between aperture viewing and whole viewing by using a brief stimulus presentation time for whole viewing and a long stimulus presentation time for aperture viewing. The long presentation time used with aperture viewing also mimicked the long exploration times required for successful haptic face recognition. The difference in stimulus presentation time in Experiment 1, however, may have produced qualitatively different effects on many perceptual and cognitive processes involved in face recognition, especially when brain activation was measured with a technique, like fMRI, that blurs neural responses over time. Another difference between aperture viewing and whole viewing in Experiment 1 was that the stimulus display was static with whole view-

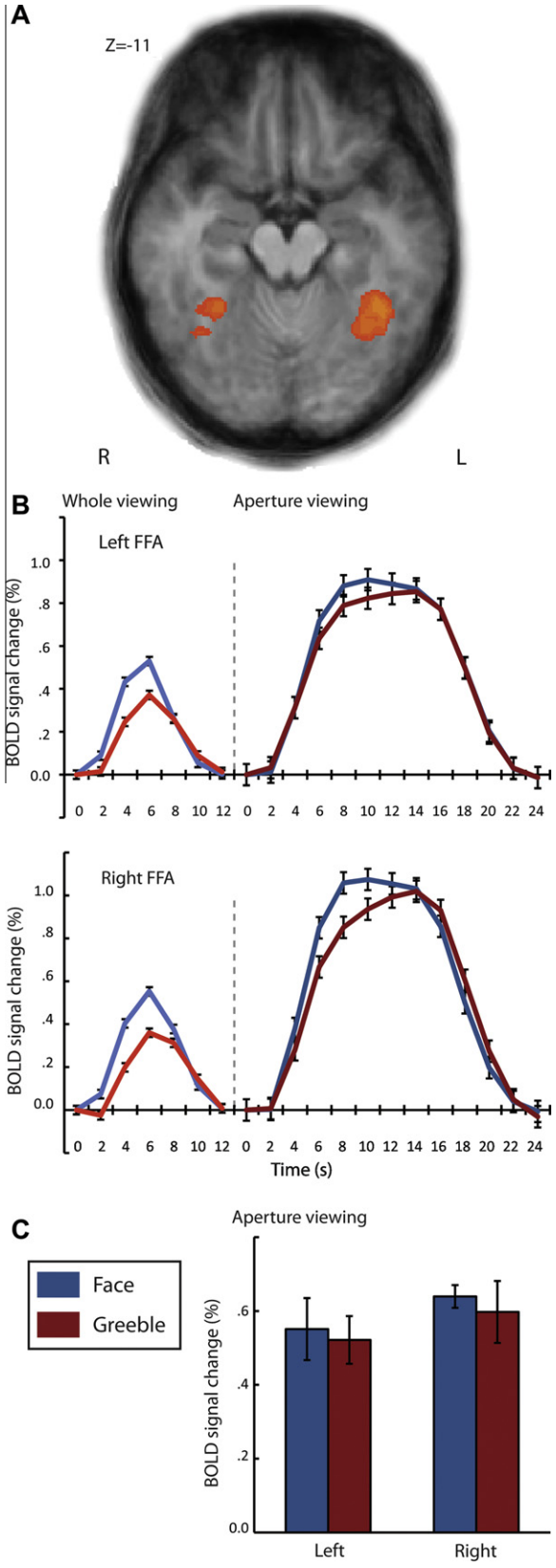

Fig. 4. Imaging results for Experiment 1. (A) A contrast of whole faces and Greebles showing left and right FFAs. (B) BOLD activation timecourses as a function of hemisphere, viewing condition, and stimulus class. Blue lines are faces and red lines are Greebles. (C) BOLD percent signal change with aperture viewing as a function of hemisphere and stimulus class. Error bars are standard error of the mean (SEM). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ing and dynamic with aperture viewing. Dynamic changes in the stimulus presentation could cause the recruitment of different perceptual processes involved in processing motion, and could also cause a release from habituation of the perceptual processes involved in analyzing the image.

The goal of Experiment 2 was to evaluate the effect of stimulus presentation time on activation in the FG using two control conditions that were not developed in Experiment 1. The first condition was called *sustained* whole viewing, where the whole face stimulus was presented continuously for 12 s, which was the same stimulus presentation time used for aperture viewing. The second condition was *repeated* whole viewing, where the whole face was briefly flashed six times, once every 2 s (12 s total), which again matched the stimulus presentation time of aperture viewing, but also introduced dynamic changes to the visual presentation. These two whole-viewing conditions were compared to the same aperture-viewing condition used in Experiment 1.

If the strong, sustained FFA activation with aperture viewing in Experiment 1 was an artifact of the long presentation time, then the whole-viewing conditions in Experiment 2 should produce greater activation than the aperture-viewing condition.

3.1. Methods and materials

To further investigate the nature of the unexpected results of Experiment 1, the methods used for the remaining experiments diverged considerably from initial experiment. First, familiarity was dropped as a factor of interest, and with it, the old/new recognition task. A two-alternative forced-choice (2AFC) task was adopted, because it allowed for better comparisons with existing research.

3.1.1. Subjects

Five undergraduate student subjects were recruited who were all new to the study. The study was approved by the IUB Institutional Review Board and Human Subjects Committee, and all subjects signed informed consent.

3.1.2. Stimuli

Twelve of the 24 female faces used in Experiment 1 were also used in this experiment. Twelve male faces were selected from the same database to bring the number of face stimuli to 24 (Fig. 5A). Greeble stimuli were not used in this experiment.

3.1.3. Viewing procedures

Aperture viewing (AP) was the same as Experiment 1. Two new whole-viewing conditions were developed (Fig. 5B). The first was sustained whole viewing (WS). For this condition, a whole face was presented continuously for 12 s, followed by a 12 s inter-trial interval. The second was repeated whole viewing (WR). For this condition, the same whole face was presented six times, each time for 300 ms with a 1700 ms inter-stimulus interval. This 12 s stimulus presentation block was followed by a 12 s inter-trial interval. The aperture-viewing, whole-sustained-viewing, and whole-repeated-viewing conditions all had the same trial length.

3.1.4. Learning procedures

Because familiarity was not an experimental factor in Experiment 2, the learning procedure was not as extensive as in Experiment 1. Although the learning tasks were the same as Experiment 1, the goal of the learning procedure was solely to acclimate subjects to the aperture-viewing procedure; therefore, there was no learning criterion. Subjects performed a pre-determined number of trials (48) with aperture viewing and whole viewing. Learning time was approximately 20 min.

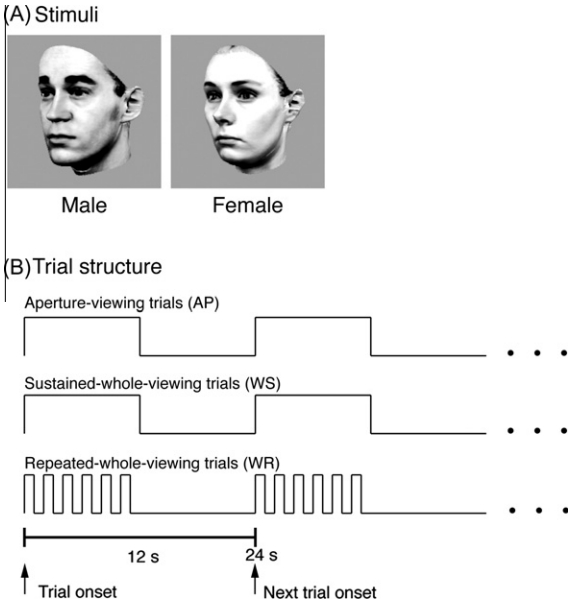

Fig. 5. Stimuli and experimental protocols for Experiment 2. (A) Stimuli were male and female faces. (B) Aperture-viewing trials consisted of 12 s of stimulation with a 12 s ITI. Sustained whole-viewing trials consisted of 12 s of stimulation with a 12 s ITI. Repeated whole-viewing trials consisted of a sequence of six 300 ms stimulation periods separated by 1700 ms inter-stimulus intervals. The sequences were separated by a 12 s ITI.

3.1.5. fMRI testing procedures

Testing was divided into six 5-min runs, two aperture-viewing runs, two sustained-whole-viewing runs, and two repeated-whole-viewing runs. Each run contained 12 trials, which led to a total of 24 trials for each viewing condition. Subjects performed a 2AFC task, indicating whether the stimulus face was male or female.

3.1.6. MRI acquisition and data analysis

Acquisition parameters and analysis steps were the same as Experiment 1.

3.2. Results

Accuracy with aperture viewing was better than in Experiment 1, suggesting that the male/female 2AFC task was easier with aperture viewing than the old/new recognition task. Response times were similar to Experiment 1. Aperture viewing had a mean response time of 9.33 s, whereas repeated whole viewing and sustained whole viewing had mean response times of 920 ms and 1140 ms, respectively. As expected, subjects' accuracy with aperture viewing was significantly worse than with either repeated whole viewing ($t_{(4)} = 3.72$, $p = .010$) or sustained whole viewing ($t_{(4)} = 3.49$, $p = .013$; Fig. 6A).

The rFFA was localized using an additional blocked-design functional localizer run comparing faces and buildings. Although this is a different contrast than what was used in Experiment 1 to localize the FFA (faces versus novel objects), it has recently been demonstrated that localization of the FFA is relatively invariant with regard to the stimulus class contrasted with faces (Berman et al., 2010). Mean timecourses extracted from the rFFA are shown in Fig. 6B. Aperture viewing produced a similar pattern of activation to the aperture-viewing condition in Experiment 1. The whole-viewing conditions, however, were different. In Experiment 1 (Fig. 4B), the brief whole-viewing condition showed a BOLD response function of approximately 12 s in width, whereas in Experiment 2 (Fig. 6B), the repeated and sustained whole-viewing conditions both showed a more sustained BOLD response function

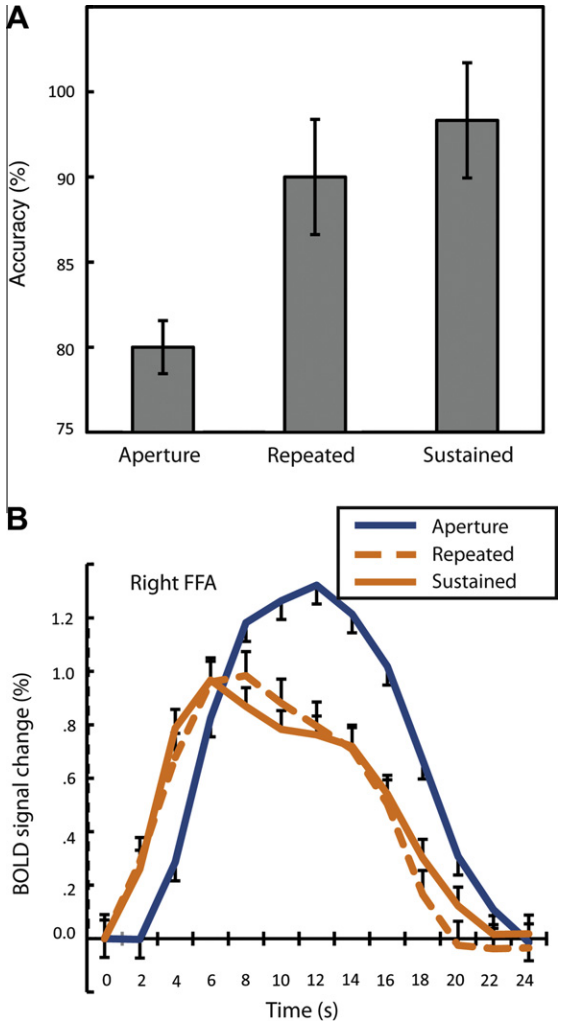

Fig. 6. Results for Experiment 2. (A) Accuracy with face stimuli as a function of viewing condition. (B) BOLD activation timecourses with face stimuli as a function of viewing condition.

of approximately 20 s and 22 s in width, respectively. In fact, the width of the whole-viewing functions was similar to the width of the aperture-viewing function.

3.3. Discussion

Similar to Experiment 1, aperture viewing produced strong, sustained activation in the rFFA. Repeated whole viewing and sustained whole viewing produced stronger and more sustained activation than the brief whole-viewing condition used in Experiment 1, but the level of activation was not significantly greater than aperture viewing. The robust activation function with aperture viewing cannot be explained simply by the presence of dynamic changes to the visual display, because the repeated whole-viewing trials contained transient changes to the visual stimulus presentation. These findings suggest that strong, sustained activation in the rFFA with aperture viewing represents the continuous recruitment of a feature integration mechanism that operates across multiple time scales.

Accuracy with aperture viewing was improved compared to Experiment 1, but still not at the level of whole viewing. Although stimulus presentation time was equated across the three viewing conditions, difficulty, as reflected by accuracy, was not equated. This suggests that the strong, sustained activation with aperture viewing may be the result of more effortful processing. This inter-

Fig. 7. Stimuli for Experiment 3.

pretation is consistent, however, with the premise that the FFA is involved in isolated feature integration and that those feature integration processes are being recruited as strongly, or even more strongly, during aperture viewing than during whole viewing.

Another possible explanation for the strong sustained activation with aperture viewing is the influence of attention. With our three whole-viewing tasks, subjects recognized the faces quickly, whereas with the aperture-viewing task, it took them considerable time. With their task completed quickly, subjects may have become less engaged with the stimulus, even in the repeated-whole-viewing condition where the face was repeatedly presented. Attention does influence the level of activation in the rFFA (O'Craven, Downing, & Kanwisher, 1999), and the size of the effect is significant, with unattended activation being 55–75% of attended activation. However, there are at least two reasons why an explanation based solely on differences in attention between whole and aperture viewing does not suffice. First, the BOLD response function for the aperture-viewing condition was 24 s in width, which reflects an underlying impulse function based on 12 s of sustained activity (Ashby & Waldschmidt, 2008; Boynton, Engel, Glover, & Heeger, 1996; Glover, 1999; Heeger & Ress, 2002). If subjects were less engaged once their task was complete, then the BOLD response function would reflect an underlying impulse function based on activity that was sustained only to the point of recognition. Second, although the effects of attention on BOLD activation are in the same range as the effect of aperture viewing compared to sustained whole viewing, a direct comparison of the effect of attention and the effect of restricted viewing assumes that the rFFA activation with a single face feature is of similar amplitude as activation with whole faces, which it is not (Tong, Nakayama, Moscovitch, Weinrib, & Kanwisher, 2000). In sum, attention was not equated in the aperture- and whole-viewing conditions, but this factor does not seem to fully explain the effect of aperture viewing on rFFA activation. Exploring the interaction of attention and sequential processing of object features is an interesting topic for future research.

4. Experiment 3

The results of Experiments 1 and 2 showed that aperture viewing produced strong, sustained activation in the FG, specifically the rFFA. This suggests recruitment of a mechanism that integrates isolated features over not only space, but also over time. What was not clear from the results, however, was whether or not the temporal integration process was face-selective. In Experiment 1, aperture viewing of both faces and novel objects (Greebles) produced sustained patterns of activation, but even though faces produced slightly more activation than Greebles, the difference between faces and Greebles was not statistically significant. Face selectivity of aperture-viewing activation, however, cannot be ruled out based on a single null finding. Previous research (Gauthier, Anderson, Tarr, Skudlarski, & Gore, 1997; Kung et al., 2007) suggests that the rFFA produces relatively strong activation with object sets such as Greebles that are categorized at the subordinate level and that share a common configuration of parts. Although faces produce more activation in the rFFA than Greebles, the effect size when comparing faces and Greebles may not be as large as when comparing faces with other control sets of objects that do not share a common configuration of parts. A commonly used control set for faces in studies of face selectivity is pictures of scenes and particularly, scenes with buildings (O'Craven & Kanwisher, 2000; Peelen & Downing, 2005; Tong, Nakayama, Vaughan, & Kanwisher, 1998). Experiment 3 was designed to further explore face selectivity of BOLD activation during aperture viewing by using scenes with buildings as control stimuli.

4.1. Methods and materials

4.1.1. Subjects

Eleven subjects (six male and five female) participated. Six of those subjects had also participated in Experiment 1 and the remaining five were recruited from the undergraduate and graduate student populations at Indiana University. The study was approved by the IUB Institutional Review Board and Human Subjects Committee, and all subjects signed informed consent.

4.1.2. Stimuli

The same 24 male and female face stimuli used in Experiment 1 were also used in this experiment. Twenty-four building stimuli were selected from photographs of houses and other buildings that were used in previous research on scene perception (Yi, Kelley, Marois, & Chun, 2006). Twelve pictures were specifically selected to be stereotypical of a "house that you live in", and the other twelve pictures were specifically selected to be stereotypical of a "building that you work in" (Fig. 7).

4.1.3. Viewing procedures

The viewing procedures were the same as Experiment 1.

4.1.4. Learning procedures

The learning procedures were the same as Experiment 2, except that subjects performed 24 trials with faces (instead of 48) and 24 trials with buildings.

4.1.5. fMRI testing procedures

Testing procedures were the same as Experiment 2. Subjects performed the male/female 2AFC categorization task with faces, but a 2AFC "lived in/worked in" categorization task with buildings. Face and building trials were intermixed for each run.

4.1.6. MRI acquisition and data analysis

Acquisition parameters and analysis steps were the same as Experiment 1.

4.2. Results

Behavioral results from the imaging sessions were unavailable due to a problem with the file output routine in the testing script.

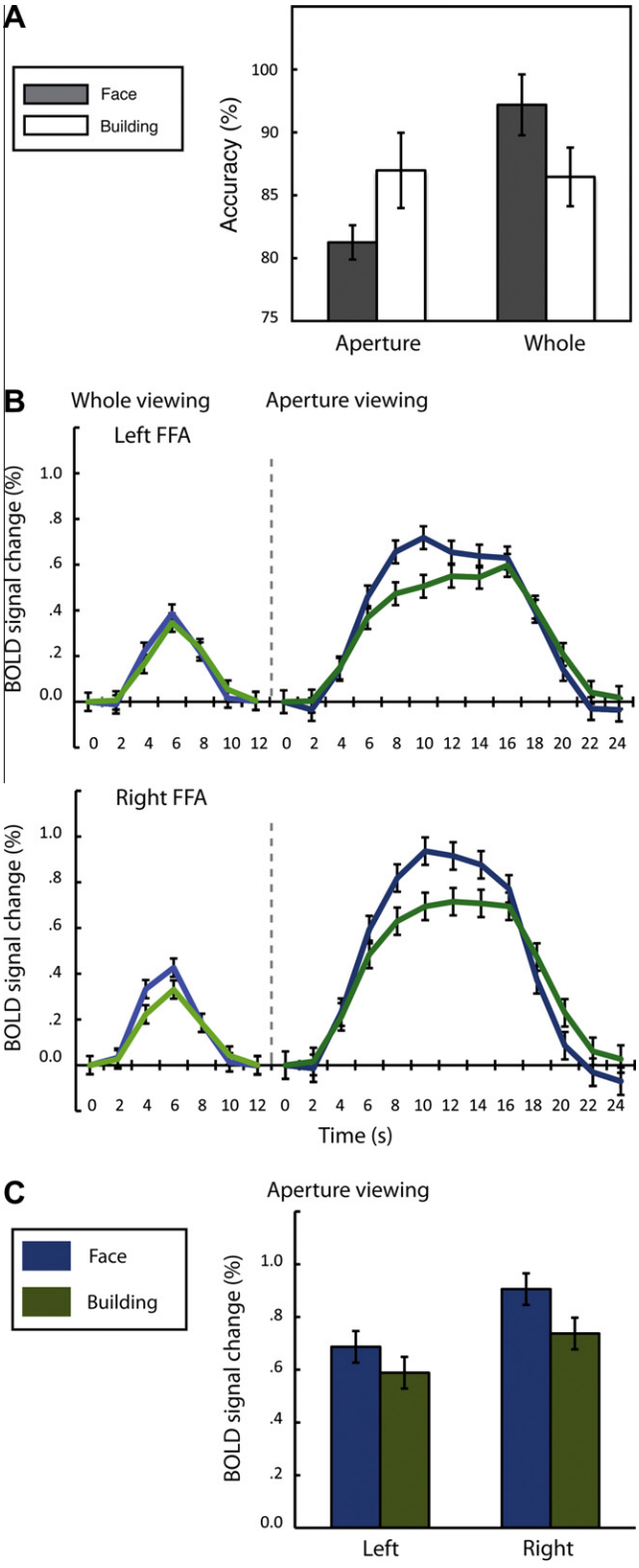

Fig. 8. Imaging results for Experiment 3. (A) Accuracy as a function of viewing condition and stimulus class. (B) BOLD activation timecourses as a function of hemisphere, viewing condition, and stimulus class. Blue lines are faces and green lines are buildings. (C) BOLD percent signal change with aperture viewing as a function of hemisphere and stimulus class. Error bars are SEM. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The problem was not discovered until all subjects had been tested. Observations made by the experimenters during testing, debriefing of subjects, and analysis of pilot testing data suggested that categorization performance was above chance and recognition times were similar to those found in Experiment 1. However, to confirm these observations, eight of the subjects ($N = 8$) were re-recruited to obtain behavioral measurements (i.e., they did not undergo another imaging session). For accuracy, there was a significant two-way interaction ($F_{(1,7)} = 11.1$, $p = .015$) between viewing condition and stimulus class (Fig. 8A). Post-hoc t-tests showed that accuracy was significantly greater for faces under conditions of whole-viewing than aperture-viewing ($t_{(7)} = 4.02$, $p = 0.0025$), whereas there was no significant difference in accuracy between whole buildings and aperture buildings ($t_{(7)} = 0.14$, *n.s.*). The difference in accuracy between whole faces and whole buildings was marginally significant ($t_{(7)} = 1.72$, $p = 0.065$). Response times with faces were very similar to those of Experiment 2, which used the same male/female 2AFC categorization task, and the pattern was similar to Experiment 1, even though a different task (old/new recognition) was used in that experiment.

Viewing-time maps for Experiment 3 are shown in Fig. 9. The exploration results for faces matched those of Experiment 1. Although there was some variability across subjects, the general pattern of viewing time for faces was similar. Most subjects spent a majority of the trial examining the top half of the face, with particular attention paid to the eyes and nose. The pattern of exploration with buildings was quite different from faces. With buildings, the viewing time was spread out, rather than being focused in particular locations. This was likely because the buildings did not share a common configuration of parts and subjects could not rely on specific locations to contain diagnostic features.

The left and right FFAs were functionally localized in each individual subject from a statistical parametric map generated using a GLM and post-hoc contrast of the whole-face and whole-building conditions. Using a FDR correction ($q = .05$) for multiple tests, both lFFA and rFFA were found in 8 of 11 subjects ($N = 8$). Timecourses for each of the four experimental conditions were extracted from these ROIs and are shown in Fig. 8B.

A summary graph of aperture-viewing BOLD activation from the left and right FFAs is shown in Fig. 8C. BOLD activation levels were calculated as the mean across a time window from 4 to 24 s. A 2×2 ANOVA was conducted on BOLD activation levels with hemisphere and stimulus class as factors. There was no significant 2-way interaction. The effect of hemisphere was not interpreted. There was a significant main effect of stimulus class ($F_{(1,7)} = 10.03$, $p = .015$). Planned comparisons between face and building conditions showed significant effects in the right ($t_{(7)} = 2.78$, $p = .015$) and left hemisphere ($t_{(7)} = 1.88$, $p = .05$), with face stimuli producing more activation than building stimuli.

4.3. Discussion

The results of Experiment 3 confirmed and extended the results of Experiment 1. Like Experiment 1, subjects preferentially viewed features at the top of face stimuli. Subjects did not, however, have a stereotypical pattern of exploration with buildings. Unlike faces and Greebles, building stimuli did not share a common configuration, suggesting that common configuration played a large role in determining the stereotypic exploration pattern. Also like Experiment 1, aperture viewing produced a strong, sustained pattern of activation in the right and left FFAs with both faces and buildings. Activation with aperture viewing was also found to be face-selective in both the right and left FFAs. Like Experiment 1, there was no evidence for lateralization of raw activation levels or face-selective activation of faces over buildings. The results suggest that the FFA is recruited bilaterally for integration of isolated features, and that recruitment is stronger for face features than for building features.

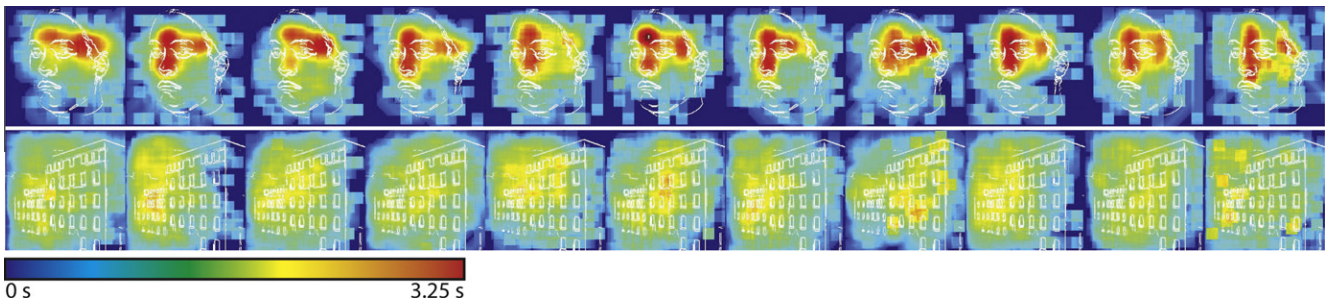

Fig. 9. Viewing-time maps for Experiment 3. The color of each pixel represents the mean viewing time across trials during the aperture-viewing condition. Each column is a different subject. Viewing times for faces are on top, with buildings on the bottom.

The 2AFC categorization task combined with the different presentations times led to similar accuracy with building stimuli across whole viewing and aperture viewing. Despite equated difficulty, as reflected by accuracy, aperture viewing with building stimuli still produced a more sustained activation function than whole viewing. This suggests that sustained activation with aperture viewing was not due to increased effort due to difficulty, but was more likely due to sustained effortful processing during sequential exploration of isolated features.

5. General discussion

The use of aperture viewing has provided insights into the functional specialization of the human occipito-temporal cortex for face, object, and scene recognition. Aperture viewing has been used previously with face, object, and scene stimuli to restrict viewing and study different perceptual and cognitive processes (Haig, 1985; Inui & Miyamoto, 1984; Jansen et al., 2003). Aperture viewing was used here to constrain the visual system to sequential, self-directed exploration of object features in much the same way that the haptic system explores objects (Loomis, 1981). Our initial hypothesis, derived from our study of haptic face recognition, was that simultaneous spatial feature integration would be biased to the right FG, whereas sequential feature integration would be biased to the left FG (Rhodes, 1985). Our findings did not support this lateralization hypothesis.

Instead, the left and right FG produced similar patterns of activation across most task and stimulus combinations. The most consistent finding was strong and sustained activation with aperture viewing in the right and left FFA. The sustained pattern of activation was consistent across stimulus types, including faces, novel objects (Greebles), and buildings. These findings suggest that the FG is involved in more than simultaneous spatial integration of features. Rather, our findings suggest that the left and right FG are involved in the spatial and temporal integration of features.

Aperture viewing has produced insights into the relative importance of different features for successful recognition (Haig, 1985; Inui & Miyamoto, 1984; Jansen et al., 2003), in the same way as eye tracking (Henderson, Williams, & Falk, 2005; Walker-Smith, Gale, & Findlay, 1977) and reverse correlation (Schyns et al., 2002; Sekuler et al., 2004). Viewing time results from the aperture-viewing condition confirmed previous work on face recognition suggesting that the eye features are important for successful recognition (Henderson et al., 2005; Sekuler et al., 2004). Exploration patterns with faces were highly consistent across subjects. Viewing-time patterns with buildings were qualitatively different than viewing-time patterns with faces. The exploration pattern with faces was focused, with time spent mainly on the eye features. The exploration pattern with buildings was diffuse, likely because of the lack of a common configuration among the building stimuli, and because they were presented from different viewpoints. Faces, on the other hand, all had a common configuration, and were presented from the same viewpoint. The pattern with Greebles was more similar to faces than buildings. Like faces, the Greebles shared a common configuration, and were also presented from the same viewpoint.

Aperture viewing is only one tool for examining the different contributions of spatial and temporal feature-integration operations. It is a tool that is well suited for comparing visual object recognition with haptic object recognition. The aperture-viewing task spatially restricted the amount of input, making visual exploration more similar to haptic exploration. Subjects were given control of the movements of the aperture, such that they could control the information input continuously through time, which is also similar to haptic exploration. Thus, the specific aperture-viewing procedure used in these experiments allowed a fair comparison with haptic recognition studies. But, compared to visual recognition studies, the aperture-viewing procedure sacrificed some control over the information made available to the subjects, because the subjects controlled the input rather than the experimenter. Other ways of presenting object features sequentially could offer more experimental control over the information made available to subjects from moment to moment and should be considered in future studies examining feature integration at different time scales. Continued investigation of sequential processing of faces, objects, and scenes using aperture viewing and haptic exploration will continue to increase our understanding of the cognitive and neural processes underlying human object recognition.

Acknowledgments

This research was supported in part by the Faculty Research Support Program through the IU Bloomington Office of the Vice President of Research, and in part by the Indiana METACyt Initiative of Indiana University, funded in part through a major grant from the Lilly Endowment, Inc. We thank Karin Harman James for insights throughout the project, Samantha Brandfon, Thea Atwood, and Rebecca Ward for help with data collection, and Tarlise Townsend and Maxim Bushmakin for help with data collection and analysis.

References

Ashby, F. G., & Waldschmidt, J. G. (2008). Fitting computational models to fMRI. *Behavior Research Methods, 40*(3), 713–721.

Berman, M. G., Park, J., Gonzalez, R., Polk, T. A., Gehrke, A., Knaffla, S., et al. (2010). Evaluating functional localizers: The case of the FFA. *NeuroImage, 50*(1), 56–71.

Biederman, I. (1997). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115–147.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience, 16*(13), 4207–4221.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305–327.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. W. (1998). What is "special" about face perception? *Psychological Review, 105*(3), 482–498.
Fox, C. J., Iaria, G., & Barton, J. J. S. (2009). Defining the face processing network: Optimization of the functional localizer in fMRI. *Human Brain Mapping, 30*(5), 1637–1651.
Gauthier, I. (2000). What constrains the organization of the ventral temporal cortex? *Trends in Cognitive Sciences, 4*(1), 1–2.
Gauthier, I., Anderson, A. W., Tarr, M. J., Skudlarski, P., & Gore, J. C. (1997). Levels of categorization in visual recognition studied using functional magnetic resonance imaging. *Current Biology, 7*(9), 645–651.
Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*(12), 1673–1682.
Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 431–446.
Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience, 2*(6), 568–573.
Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage, 9*(4), 416–429.
Haig, N. D. (1985). How faces differ – A new comparative technique. *Perception, 14*, 601–615.
Haig, N. D. (1986). Exploring recognition with interchanged facial features. *Perception, 15*, 235–247.
Halgren, E., Dale, A. M., Sereno, M. I., Tootell, R. B., Marinkovic, K., & Rosen, B. R. (1999). Location of human face-selective cortex with respect to retinotopic areas. *Human Brain Mapping, 7*(1), 29–37.
Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223–233.
Heeger, D. J., & Ress, D. (2002). What does fMRI tell us about neuronal activity? *Nature Reviews Neuroscience, 3*(2), 142–151.
Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory and Cognition, 33*(1), 98–106.
Inui, T., & Miyamoto, K. (1984). The effect of changes in visible area on facial recognition. *Perception, 13*(1), 49–56.
James, T. W., Servos, P., Kilgour, A. R., Huh, E., & Lederman, S. J. (2006). The influence of familiarity on brain activation during haptic exploration of 3-D facemasks. *Neuroscience Letters, 397*, 269–273.
James, T. W., Shima, D. W., Tarr, M. J., & Gauthier, I. (2005). Generating complex three-dimensional stimuli (Greebles) for haptic expertise training. *Behavior Research Methods, 37*(2), 353–358.
Jansen, A. R., Blackwell, A. F., & Marriott, K. (2003). A tool for tracking visual attention: The Restricted Focus Viewer. *Behavioral Research Methods, Instruments and Computers, 35*(1), 57–69.
Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*, 4302–4311.
Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B, 361*, 2109–2128.
Kilgour, A. R., Kitada, R., Servos, P., James, T. W., & Lederman, S. J. (2005). Haptic face identification activates ventral occipital and temporal areas: An fMRI study. *Brain and Cognition, 59*, 246–257.
Kilgour, A. R., & Lederman, S. J. (2002). Face recognition by hand. *Perception and Psychophysics, 64*(3), 339–352.
Klatzky, R. L., Lederman, S. J., & Reed, C. L. (1987). There's more to touch than meets the eye: The salience of object attributes for haptics with and without vision. *Journal of Experimental Psychology: General, 116*(4), 356–369.
Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience, 12*(5), 535–540.
Kung, C. C., Peissig, J. J., & Tarr, M. J. (2007). Is region-of-interest overlap comparison a reliable measure of category specificity? *Journal of Cognitive Neuroscience, 19*(12), 2019–2034.
Loomis, J. M. (1981). Tactile pattern perception. *Perception, 10*, 5–27.
Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: W.H. Freeman.
Marsolek, C. J. (1999). Dissociable neural subsystems underlie abstract and specific object recognition. *Psychological Science, 10*(2), 111–118.
Marsolek, C. J., & Burgund, E. D. (1997). Computational analyses and hemispheric asymmetries in visual-form recognition. In S. Christman (Ed.), *Cerebral asymmetries in sensory and perceptual processing*. Amsterdam: Elsevier.
Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 5*(6), 255–260.
McKone, E., Kanwisher, N., & Duchaine, B. C. (2006). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences, 11*(1), 8–15.
McKone, E., Martini, P., & Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology: Human Perception and Performance, 27*(3), 573–599.
Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience, 9*(5), 555–604.
O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). FMRI evidence for objects as the units of attentional selection. *Nature, 401*(6753), 584–587.
O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience, 12*(6), 1013–1023.
Peelen, M. V., & Downing, P. E. (2005). Within-subject reproducibility of category-specific visual activation with functional MRI. *Human Brain Mapping, 25*(4), 402–408.
Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature, 343*(6255), 263–266.
Poldrack, R. A., & Mumford, J. A. (2009). Independence in ROI analysis: Where is the voodoo? *Social Cognitive and Affective Neuroscience, 4*(2), 208–213.
Puce, A., Allison, T., Gore, J. C., & McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. *Journal of Neurophysiology, 74*(3), 1192–1199.
Rhodes, G. (1985). Lateralized processes in face recognition. *British Journal of Psychology, 76*, 249–271.
Rhodes, G. (1988). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception, 17*, 43–63.
Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica, 128*(2), 274–289.
Rossion, B., & Gauthier, I. (2002). How does the brain process upright and inverted faces? *Behavioral and Cognitive Neuroscience Reviews, 1*(1), 63–75.
Schiltz, C., & Rossion, B. (2006). Faces are represented holistically in the human occipito-temporal cortex. *NeuroImage, 32*, 1385–1394.
Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science, 13*(5), 402–409.
Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. *Behavioral Brain Science, 21*(1), 1–17.
Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology, 14*(5), 391–396.
Sergent, J., Ohta, S., & Macdonald, B. (1992). Functional neuroanatomy of face and object processing: A positron emission tomography study. *Brain, 115*(1), 15–36.
Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A, 46*(2), 225–245.
Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000). Response properties of the human fusiform face area. *Cognitive Neuropsychology, 17*(1–3), 257–280.
Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron, 21*(4), 753–759.
Walker-Smith, G. J., Gale, A. G., & Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception, 6*(3), 313–326.
Weiner, K. S., & Grill-Spector, K. (2010). Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *NeuroImage, 52*(4), 1559–1573.
Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.
Yi, D. J., Kelley, T. A., Marois, R., & Chun, M. M. (2006). Attentional modulation of repetition attenuation is anatomically dissociable for scenes and faces. *Brain Research, 1080*(1), 53–62.
Yovel, G., & Duchaine, B. (2006). Specialized face perception mechanisms extract both part and spacing information: Evidence from developmental prosopagnosia. *Journal of Cognitive Neuroscience, 18*(4), 580–593.
Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current Biology, 15*, 2256–2262.