

Available online at www.sciencedirect.com



Vision Research 46 (2006) 1901-1910

Vision Research

www.elsevier.com/locate/visres

# Size-invariant but viewpoint-dependent representation of faces

Yunjo Lee<sup>a,\*</sup>, Kazumichi Matsumiya<sup>b</sup>, Hugh R. Wilson<sup>a</sup>

<sup>a</sup> Centre for Vision Research, York University, 4700 Keele Street, Toronto, Ont., Canada M3J 1P3 <sup>b</sup> Research Institute of Electrical Communication, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980–8577, Japan

Received 12 November 2004; received in revised form 16 December 2005

#### Abstract

The present study investigated the role of size and view on face discrimination, using a novel set of synthetic face stimuli. Face discrimination thresholds were measured using a 2AFC match-to-sample paradigm, where faces were discriminated from a mean face. In Experiment 1, which assessed the effect of size alone, subjects had to match faces that differed in size up to four-fold. In Experiment 2 where only viewpoint was manipulated, a target face was presented at one of four different views (0° front, 6.7°, 13.3°, and 20° side) and subsequent matches appeared either at the same or different view. Experiment 3 investigated how face view interacts with size changes, and subjects matched faces differing both in size and view. The results were as follows: (1) size changes up to four-fold had no effect on face discrimination; (2) threshold for matching different face views increased with angular difference from frontal view; (3) size differences across different views had no effect on face discrimination. Additionally, the present study found a perceptual boundary between 6.7° and 13.3° side views, grouping 0° front and 6.7° side views together and 13.3° and 20° side views together. This suggests categorical perception of face view. The present study concludes that face view and size are processed by parallel mechanisms.

Keywords: Face discrimination; Face size; Viewpoint; Synthetic faces

## 1. Introduction

A face is a three-dimensional (3D) object that is encountered from many directions and various distances and thus, face recognition requires representation of invariant aspects of facial structure across these changes. Despite these challenges, recognizing faces is usually an easy and effortless task for us. However, how the brain accomplishes face recognition under variations and changes in size and viewpoint still remains unclear. The present study addressed this question psychophysically and investigated how robust face perception would be under changes in size and view. Use of a psychophysical paradigm allows us to quantify these effects and also provides insight into the neural mechanisms that underlie them.

\* Corresponding author. *E-mail address:* ychlee@yorku.ca (Y. Lee).

Recent fMRI studies have provided evidence that the fusiform face area (FFA; Kanwisher, McDermott, & Chun, 1997) is involved in analyzing objects or faces independent of the visual cues defining their shape, such as size, but could respond selectively to different viewpoints (Andrews & Ewbank, 2004; Grill-Spector & Malach, 2001; Grill-Spector et al., 1999; Vuilleumier, Henson, Driver, & Dolan, 2002). Grill-Spector and her colleagues observed that the overall activation of the posterior fusiform gyrus (FG) is sensitive to different views of the same faces or cars but not to changes in size and position. The face-selective voxels remained strongly adapted to size and position changes even though a robust recovery from adaptation was observed when the stimulating face was rotated. Andrews and Ewbank (2004) found the same results in the FFA that corresponds to the posterior FG region from Grill-Spector et al. (1999). Adaptation to repeated presentations of the same face persisted in the FFA despite changes in size, but FFA responses were sensitive to the manipulation of viewpoint and emotional expressions. In contrast, they failed to

<sup>0042-6989/\$ -</sup> see front matter © 2005 Elsevier Ltd. All rights reserved. doi:10.1016/j.visres.2005.12.008

find any adaptation to repetitions of the same face in faceselective regions in the superior temporal lobe (STS). Also, Vuilleumier et al. (2002) showed similar results with objects presented in a long-term repetition priming design. The right FG showed priming-induced decreases by repetition of the same viewpoint, irrespective of size, suggesting viewpoint-dependent and size-invariant representations in neuronal populations in that region.<sup>1</sup> These results are consistent with data from single-cell recording of the primate inferotemporal cortex (IT) or the superior temporal sulcus (STS) and fMRI studies with human and primate subjects, which independently studied the effects of size and viewpoint changes (Desimone, Albright, Gross, & Bruce, 1984; Ito, Tamura, Fujita, & Tanaka, 1995; Logothetis & Pauls, 1995; Logothetis, Pauls, & Poggio, 1995; Lueschow, Miller, & Desimone, 1994; Malach et al., 1995; Op De Beeck & Vogels, 2000; Perrett, Mistlin, & Chitty, 1987; Perrett et al., 1985, 1991; Pourtois et al., 2005; Rolls & Baylis, 1986; Sáry, Vogels, & Orban, 1993; Schwartz, Desimone, Albright, & Gross, 1983; Wang, Tanaka, & Tanifuji, 1996).

Single-cell recording studies on monkeys have shown that IT cells are invariant to size changes and other shape properties. Schwartz et al. (1983) demonstrated that many IT cells are selective for shapes systematically varying in boundary curvature and maintain this selectivity over a two-fold increase in size, 3-5° of position changes, and contrast change. Sáry et al. (1993) also reported preserved shape selectivity over a four-fold size change and 4-5° of position changes. Ito et al. (1995) tested wide ranges of size with simply shaped images and found two types of cells: cells that are sharply tuned to particular ranges of size and those responsive to wide ranges of size. 43% of the anterior IT cells studied responded to ranges of linear size variation less than 2 octaves, but 21% responded to size ranges of more than 4 octaves. The selectivity for shape was mostly preserved throughout the entire range of size changes. On the other hand, Op De Beeck and Vogels (2000) found position-dependent size effects in the anterior part of the inferior temporal cortex (TE). They presented the four stimulus sizes over a four-fold range either in the foveal or peripheral position. When the general response pattern was separately plotted for each position, the average response increased with size at the periphery but the effect of size disappeared at the foveal position.

Using fMRI, Sawamura, Georgieva, Vogels, Vanduffel, and Orban (2005) found object adaptation and size invariance in macaque IT complex and human LOC though they failed to find complete size invariance unlike Grill-Spector et al. (1999). Their overall human data agree with Grill-Spector et al. (1999), but the extent of size invariance in Sawamura et al. was somewhat smaller. This discrepancy might result from the different ranges of sizes used in these two studies. Sawamura et al. used five discrete sizes of the same object over a four-fold range  $(2.3-9.2^{\circ})$ , while Grill-Spector et al. presented 30 different sizes over a range of  $10-30^{\circ}$ .

Invariance to size changes has also been observed with face stimuli. Rolls and Baylis (1986) found that the majority of neurons in the macaque middle and anterior STS showed invariant responses with respect to size as well as contrast alternation of the stimulus. The median size change that can produce greater than half of the maximal response was a factor of 12.0. Desimone et al. (1984) also demonstrated that IT and STS neurons maintained their selectivity to both monkey and human faces or hands over a wide range of size changes. In their previous studies, they found that macaque IT neurons have large receptive fields of median size  $26^{\circ} \times 26^{\circ}$  that invariably included the fovea (Desimone & Gross, 1979; Gross, Rocha-Miranda, & Bender, 1972). With these receptive fields larger than the stimulus, IT neurons responded to stimuli regardless of sizes changed within the receptive field (Desimone et al., 1984). In the human brain, activation of the LOC posteriorly including the lateral aspect of the posterior FG was not influenced by a four-fold change in visual size of both faces and objects (Malach et al., 1995).

While neuroimaging data have reported complete size invariance in the FFA and LOC (Andrews & Ewbank, 2004; Grill-Spector & Malach, 2001; Grill-Spector et al., 1999; Malach et al., 1995; Vuilleumier et al., 2002), most neurophysiological studies found size-invariant responses in a fraction of the neurons in the macaque IT that is homologous to the human LOC (Ito et al., 1995; Lueschow et al., 1994; Op De Beeck & Vogels, 2000; Schwartz et al., 1983). This surprising discrepancy would be attributed to the different approaches taken by these two techniques. Most single-cell studies report size-invariance in terms of stimulus selectivity, whereas fMRI assesses size-invariance on neuronal response levels (Sawamura et al., 2005). In single-cell studies, changes in size or position usually alter the absolute firing rate of the neuron, but the relative preference for a stimulus is maintained over changes within the receptive field. To this extent, IT neurons may exhibit size and position constancy (Desimone et al., 1984; Logothetis & Pauls, 1995).

Despite the size-invariant representation of faces and objects in the human and primate ventral stream, the majority of these cells appear to be viewpoint-dependent. In a study by Logothetis et al. (1995), a population of IT cells responded selectively to learned views of previously unfamiliar objects, while some of these view-selective cells exhibited response-invariance for changes in size or position. Six out of the nine view-selective cells tested showed size-invariant responses. Logothetis et al. (1995) found only a very small number of cells that showed viewpoint-invariance.

<sup>&</sup>lt;sup>1</sup> Additionally, Vuilleumier et al. (2002) found a hemispheric asymmetry: the left FG showed a generalization across views and sizes. The discrepancy between the results of Vuilleumier et al. and Grill-spector and her colleagues (1999, 2001) seem to arise from different presentation paradigms adopted. The immediate fMR adaptation paradigm used by Grill-Spector may favor perceptual stages of processing, whereas the long-term repetition priming used in Vuilleumier et al. is likely to assess more abstract, memory-based stages of processing (Pourtois, Schwartz, Seghier, Lazeyras, & Vuilleumier, 2005).

Most cells were tuned maximally to one view of an object, and their response fell gradually as the angle of rotation increased from the preferred view. Systematic view tuning curves were obtained when these cell responses were plotted as a function of rotation angle (Logothetis & Pauls, 1995).

With face stimuli, it was found that a majority of cells in the macaque superior temporal sulcus (STS) are viewercentered and exhibit unimodal tuning to one view (Perrett et al., 1985, 1987, 1991). Perrett et al. (1985, 1987) found that subpopulations of cells in macaque STS were selectively tuned to four characteristic views of the head: the front full face view, left and right profile views, and the back of the head. Desimone et al. (1984) also found faceselective cells in macaque IT and STS, some responding preferentially to the front view and some to the profile. Similarly, Wang et al. (1996) demonstrated with optical imaging study that IT cells selectively respond to faces, and the activation spot moved systematically in one direction across the cortex as the same face was rotated from the left profile to the right profile (that is, left profile, 45° left, front, 45° right, and right profile).

Using event-related fMRI during long-term repetition priming, Pourtois et al. (2005) found that the FFA in both hemispheres showed view-sensitive repetition effects, while the medial portion of the left FG showed repetition effects across all types of viewpoint changes. The left medial FG, which is adjacent to but not overlapped with the more lateral FFA, did not show any preferential category-selective responses for faces or houses. Their study suggests that both view-dependent and view-independent processes are embedded in the human temporal cortex, but full viewinvariant representation of faces is not achieved in the FFA.

Psychophysical results have supported the viewpointdependent representation in that when we recognize objects or faces from different viewpoints, there is a cost in reaction times (RTs) and sensitivity (Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992; Hill, Schyns, & Akamatsu, 1997; Logothetis, Pauls, Bülthoff, & Poggio, 1994; Shepard & Metzler, 1971; Tarr & Pinker, 1989). A viewpoint-dependent representation (viewer-centered coding) depends on the particular vantage point of a viewer relative to the object being recognized and captures how the object appears from a particular viewpoint. Thus, separate viewercentered coding is needed to enable recognition of the object from different views. If an object is represented in this manner, a change in orientation would degrade its recognition. Shepard and Metzler (1971), for example, measured the time required to recognize two perspective drawings of the same 3D objects and found that RT increased linearly with the angular difference in the portrayed orientations. Moreover, Edelman and Bülthoff (1992) found that generalization to novel views from a single trained view falls off with increasing angle of rotation. The availability of depth information through stereopsis did not change the dependence of error rate on viewpoint. With face stimuli, Hill et al. (1997) demonstrated that when

subjects learned one view for 1s and were immediately tested with different views, generalization from the learned front view was progressively poorer as the angle of rotation increased. Viewpoint-dependent recognition of objects was also observed with monkeys in animal psychophysical experiments (Logothetis et al., 1994). Monkeys trained with one view of an object performed best with that view and gradually worse as stimulus was rotated farther from the learned view. They failed to recognize views that differed more than 40° from the training view.

In the experiments reported here, we investigated how size and viewpoint affected face discrimination. The present study employed synthetic faces introduced by Wilson, Loffler, and Wilkinson (2002) in a face discrimination task. Unlike Hill et al. (1997), the discrimination task did not involve any learning or memory of faces. The goal of Experiment 1 was to assess the effect of size. Experiment 2 tested the effect of viewpoint, where viewpoint is expressed in terms of the angle of rotation in depth away from the front view. Experiment 3 investigated how face view interacts with size changes. We found that four-fold changes in size did not affect face discrimination, but that changes in viewpoint did. Threshold increased with angular difference from a front view. However, we found no interaction between size and view changes. These results suggest parallel processing of size and views in the FFA. Our results thus provide psychophysical support for fMRI findings (Andrews & Ewbank, 2004; Grill-Spector & Malach, 2001; Grill-Spector et al., 1999; Vuilleumier et al., 2002).

# 2. General methods

#### 2.1. Apparatus and calibration

All experiments were conducted on an iMac computer with  $1024 \times 768$  pixel spatial resolution, 75 Hz refresh rate and 8 bit/pixel gray scale. From a 1.31 m viewing distance, the screen subtended  $13.4^{\circ} \times 10.1^{\circ}$ , and each pixel was 47.0 arcsec in diameter. Mean luminance was  $38.0 \text{ cd/m}^2$ . Stimuli were generated in the Matlab environment and displayed using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). All experiments were conducted in a dimly lit room.

#### 2.2. Stimuli: Synthetic faces and synthetic face cubes

Human faces are extremely complex stimuli, which make it difficult to link a particular observed behavior to the underlying neural mechanisms. Synthetic faces significantly reduce the complexity of face stimuli based on 37 measurements of geometric information in the face eliminating the fine detailed texture of hair and skin, yet they provide subjects with sufficient geometric information for accurate discrimination (Wilson et al., 2002). Fig. 1 shows synthetic faces used in Experiment 2. Using fMRI, we have shown that synthetic faces generate 85% as large a BOLD signal in the FFA as do gray scale face photographs (Loffler, Yourganov, Wilkinson, & Wilson, 2005).

Synthetic faces were derived from frontal and 20° side views of individual face photographs. The 20° side view was chosen to avoid occlusion of one eye by the nose as well as protrusion of the nose beyond the head contour. Head shape and inner hairline are digitized relative to the bridge of the nose in polar coordinates. The shape of the head was represented by 16 radial measurements equally spaced around the head; the inner hair line was digitized at 9 additional radii on and above the horizontal meridian of the polar coordinates. The 16 points of the head shape were converted into



Fig. 1. Synthetic faces used in Experiment 2: from left to right, 0° front, 6.7°, 13.3°, and 20° side views. This example is a mean female synthetic face.

sums of radial frequencies (RFs) of 1–7 cycles (see Wilkinson, Wilson, & Habak, 1998; Wilson et al., 2002). Additionally, 14 measurements of facial feature positions were digitized: 4 for locations of eyes (x and y coordinates), 1 for height of brows, 3 for length and width of nose, 6 for location of mouth and thickness of lips. These points individuate only the location, width and length of facial features. For individual feature shapes, generic eye, nose, and mouth templates were used. Next, images were bandpass filtered with a 2.0 octave bandwidth difference of Gaussians (DOG) filter centered at 10.0 cycles per face width, which is optimal for face perception (Gold, Bennett, & Sekuler, 1999; Näsänen, 1999). Averaging over a set of 40 faces generated mean synthetic faces for frontal and 20° side view of each gender.

To assess the metric of synthetic face space, Wilson et al. (2002) devised the concept of face cubes that consist of groups of several synthetic faces made mutually orthogonal with respect to a given original face. One face serves as the origin of a local coordinate system, and four other faces define the axes that are mutually orthogonal and normalized to the same total geometric variation. The origin face was subtracted from the face vectors for each of four other faces and produced face-difference vectors from the origin on 37 measurements. The first vector A of these difference vectors was normalized to have a length of k based on a Euclidean norm. A second difference vector B was rotated to be perpendicular to A using Gram-Schmidt orthogonalization (see Wilson et al., 2002) and then normalized to length k. This process was repeated with two more vectors (for a total of four) for the experiments of the present study.

A synthetic face cube consists of 21 faces: at the coordinate origin there is a mean face, and along each axis (a total of 5 axes including a diagonal) four faces are separated by the same increment step (see Wilson et al., 2002). For example, if the maximum geometric variation is 12%, along each axis, the four faces differed by 3%, 6%, 9%, and 12% from the mean.

#### 2.3. Procedure

All experiments used a 2AFC match-to-sample paradigm and the method of constant stimuli. In each trial, a target face was presented for 110 ms and followed by a wide field noise mask for 200 ms. The 110 ms of target presentation time has been reported to be sufficient for optimal face processing (Lehky, 2000). The position of the target face was randomly jittered by  $\pm 0.72^{\circ}$  from the center of the screen to prevent continuous fixation on any one feature of the face. The noise was a random dot pattern bandpass filtered with the same DOG filter used for synthetic faces with the same peak spatial frequency and bandwidth. Immediately after the mask, two comparison faces were presented side by side and remained on the screen until the subject made a decision. The subject responded with a mouse click on the face that appeared to be identical to the previously flashed target. The next trial was initiated by another mouse click by the subject.

Faces were discriminated from a mean face in order to calculate the face discrimination threshold in terms of geometric variation relative to the mean head radius. A mean face was always one of comparison faces, and it was also presented as the target face 20% of the time. However, subjects reported that they were not aware that one of the comparison faces was always a mean face. Also, a control experiment was conducted with

three comparison faces where the third face (as an additional distractor) came from an orthogonal axis but had the same increment distance as the other comparison face. In this control experiment, 20% of the time the mean face was presented as the target. The results have shown that subjects were able to choose the target face among three choices though their threshold increased by a factor of 1.37 in the same-view matching and 1.24 in the cross-view matching.

Each run used only one synthetic face gender and consisted of a total of 120 to 140 trials. Experiments were repeated at least four times alternating the gender of faces, and data were averaged across gender, as Wilson et al. (2002) found no statistical difference in discrimination thresholds between face genders. In each condition, the first experimental run was discarded as practice and then data from at least 3 runs were collected. In order to compute threshold, percentage correct responses were obtained along all five axes and averaged across axes for each increment value. Then, data were fit with a Quick (1974) function using maximum likelihood estimation, and the 75% correct point from a psychometric function was chosen as threshold.

## 3. Experiment 1: The effect of size

The effect of size on face discrimination was studied using a size change between the target and comparison faces. The size change was equal to a factor of four.

#### 3.1. Subjects

Five subjects, including an author, with normal or corrected-to-normal vision participated in this experiment. Four were female, and the mean age was 24 years.

## 3.2. Stimuli

Synthetic face cubes were generated as described in Section 2, except that all faces in the cube were generated at twice the linear size and bandpass filtered at half the frequency to produce large faces with a peak at 4.0 cpd. All faces in this cube were then subsampled by averaging each  $2 \times 2$  square of pixels to produce a cube of identical faces half as large with twice the peak frequency. Repetition of this procedure then produced a final face cube that contained faces 0.25 times as large as the original cube. In this experiment, the largest cube and smallest cubes were used to generate a factor of four linear size difference between flashed target and comparison faces. A large male mean face subtended approximately  $3.15^{\circ} \times 4.36^{\circ}$  of visual angle.

# 3.3. Procedure

There were four conditions by combination of target and comparison face sizes: both target and comparisons were small; target was small and comparisons were large; comparisons small and target large; and both target and comparisons large. Subjects were tested with same-view matching of these faces.

## 3.4. Results

Results of four observers for each face view in experiments employing the 4.0 linear size difference between flashed and comparison faces are plotted in Fig. 2. From left to right thresholds are plotted for the smallest size alone, flashed small followed by comparison large (S/L), flashed large followed by comparison small (L/S), and finally the largest size alone. Upper and lower graphs plot thresholds for front views and 20° side views, respectively. To determine whether discrimination across a four-fold size change was more difficult than discrimination with no size difference, the  $2 \times 2$ 



Fig. 2. Discrimination thresholds of faces differed in size (n = 4). Small condition is when both the target and comparison faces were small; S/L is for small target and large comparisons; L/S for large target and small comparisons; and Large is when both the target and comparisons were large. Error bars indicate 1 standard error (SE) above and below the mean.

repeated-measures ANOVA was performed separately for front and side views on the variables, Same Size (2 levels: small alone and large alone) and Different Size (2 levels: S/L and L/S). For the front view, none of the effects were significant: Same Size [F(1,3) = 0.923, p = 0.408,  $\eta^2 = 0.235$ ], Different Size [F(1,3)=2.672, p=0.201,  $\eta^2=0.471$ ], and Same Size × Different Size [F(1,3) = 0.496, p = 0.532,  $\eta^2 = 0.142$ ]. For the side view, there was a significant main effect of Same Size [ $F(1,3) = 45.850, p = 0.007, \eta^2 = 0.939$ ], but the main effect of Different Size was not significant [F(1,3) = 4.359, p = 0.128,  $\eta^2 = 0.592$ ], nor was the interaction between Same Size and Different Size [F(1,3) = 0.047, p = 0.842,  $\eta^2 = 0.016$ ]. Examining the individual side view data revealed that one subject (AD) had a higher threshold in the Small condition than in the Large condition (by a factor of 3.32), and so these results suggest that subject AD had difficulty with matching small faces rather than different sized faces. Moreover, in Fig. 2, thresholds were higher for the conditions where a small face was flashed as a target. The *t*-tests comparing thresholds from a small and large target showed a significant difference with side view [t(7)=2.645, p=0.033, d=0.935] but not with front view [t(7)=2.071, p=0.077, d=0.732]. It is important to note that the spatial frequency content of a small face subsampled from a large face is shifted to higher spatial frequencies and has less power. Hence, a small face flashed for 110ms should have a higher threshold than a large face flashed for the same duration.

One-way ANOVAs and post-hoc comparisons were further conducted to compare four stimulus size conditions. With the front view, there was still no significant difference among four stimulus size conditions [F(3,12) = 1.025,p = 0.416,  $\eta^2 = 0.204$ ]. The four stimulus size conditions for the side view almost reached the *p*-level [F(3,12) = 3.310,p = 0.057,  $\eta^2 = 0.453$ ]. Tukey's HSD found a barely significant difference between the S/L and Large conditions of side view with a p value of 0.047, but a Scheffé test failed to find a significant difference in any of the comparisons. Fig. 2 (side view) shows that subjects had the largest threshold in S/L, but the other Different Size condition, L/S, did not have a significantly higher threshold than the Same Size conditions (small or large alone). Therefore, the overall results seem to indicate that size constancy operates on faces over a linear size range and peak spatial frequency range of at least 4.0.

## 4. Experiment 2: The effect of viewpoint

The effect of viewpoint on face discrimination was investigated. The viewpoint of the target and comparison faces was manipulated, while the size of faces was held constant.

#### 4.1. Subjects

Four subjects, one author and three naive volunteers, participated in this experiment. All had normal or corrected-to-normal vision, and three of them were female. The mean age was 26 years.

# 4.2. Stimuli

Four equally spaced face viewpoints were used: 0° frontal view, and 6.7°, 13.3°, 20° side views (see Fig. 1). All faces were drawn from a 4D face cube described in Section 2 for one gender and one view (either front or 20° side). For example, four front view faces were selected randomly from our database of 40 female and 41 male faces. These four front view faces were first orthogonalized and normalized, and four equal-distance increments were then created along each cube axis (total four) as well as the principal diagonal. The 20° side face corresponding to each front face was generated by adding the 20° mean side face  $(M_{20})$  and the difference from its mean ( $\Delta_{20}$ ). The new sets of 6.7° and 13.3° side views were morphed by adding a pair of corresponding frontal and 20° side views that were weighted appropriately. That is, a 6.7° side view was generated by giving 2/3 weight to the front view and 1/3 to the 20° side view on the corresponding axis and then adding these two differently weighted frontal and 20° side faces (i.e., 2/  $3M_0 + \frac{1}{3}M_{20} + \frac{2}{3}\Delta_0 + \frac{1}{3}\Delta_{20}$ ). In the case of a 13.3° side view, a 1/3 of frontal view and a 2/3 of corresponding 20° side view were added (i.e.,  $1/3M_0 + 2/3M_{20} + 1/3\Delta_0 + 2/3\Delta_{20}$ ). Orthogonalization and normalization procedures were performed in all face views to create synthetic face cubes and face spaces. Fig. 1 shows examples of these four face views.

# 4.3. Procedure

In each experimental run, the face view of a flashed target and two comparisons was manipulated. The target and comparison faces had the same view (same-view matching) or different views (cross-view matching). A novel set of synthetic face cubes was created for each run to minimize learning effects.

Same-view matching tasks were conducted to measure a baseline threshold to each face view. Cross-view matching measured a threshold for angular difference between flashed target and comparison views in two directions. In one direction (starting from 0°), the angular difference was based on a front view. Thus, the threshold measure of a 6.7° angular difference has two conditions: matching a flashed front view to a 6.7° side view and matching a flashed 6.7° side view to a front view. Data from these two conditions measuring threshold to the 6.7° angular difference were combined later because the two conditions were statistically indistinguishable [t(3) = 0.95, p = 0.41]. Also, in all subsequent experiments, the results of two conditions measuring threshold to the same angular difference were collapsed because there was no statistical difference between thresholds of those two conditions. Threshold to a 13.3° angular difference was measured in two conditions by matching a flashed front view to a 13.3° view and matching a flashed 13.3° view to a front view, and there was no statistical difference between the two conditions [t(3) = 0.005,p = 0.996]. But threshold to a 20° angular difference was tested in only one condition, matching a flashed front view to a 20° side view. On the other hand, the other direction (starting from 20°) based the angular difference on a 20° side view. That is, threshold to a 6.7° angular difference was measured by matching a 20° view to a 13.3° view and vice versa, and threshold to a 13.3° angular difference was by matching a 20° view to a 6.7° view and vice versa. These two conditions measuring thresholds for the same angular difference were not statistically different at both 6.7° and 13.3° angular differences, [t(3)=0.46, p=0.67] and [t(3)=0.35, p=0.75], respectively. Threshold to a 20° angular difference, however, was measured only by one condition, matching a flashed 20° side view to a front view.

## 4.4. Results

The 2 × 4 repeated-measures ANOVA was conducted on the discrimination threshold for the variables, Direction (2 levels: starting from 0° and starting from 20°) and Angular Difference (4 levels: 0°, 6.7°, 13.3°, and 20°). A main effect of Angular Difference was found highly significant [*F*(3, 9) = 15.450, p = 0.001,  $\eta^2 = 0.837$ ], so that there was an effect of viewpoint. Neither a main effect of Direction [*F*(1, 3) = 2.176, p = 0.237,  $\eta^2 = 0.420$ ] nor the interaction between Direction and Angular Difference [*F*(3, 9) = 2.818, p = 0.100,  $\eta^2 = 0.484$ ] was significant.

Fig. 3 shows the mean discrimination thresholds of both cross-view matching and same-view matching (baseline) across four subjects. Discrimination threshold for matching different views was plotted as a function of angular difference between target and comparison views. Baseline threshold was plotted in the same graph to be compared with that for view change. Threshold for view change increased with angular difference from the front view. Threshold elevation, which is



Fig. 3. Mean discrimination threshold of cross-view matching (view changes) as a function of angular difference between target and comparison views (n = 4). Baseline threshold (same-view matching) is plotted as a function of face view. Threshold for view changes only includes the results of "starting from 0°" conditions where side view faces were matched with a front view. Error bars indicate 1 SE.



Fig. 4. Threshold elevation as a function of angular difference between target and comparison views. Threshold elevation was calculated by taking a ratio of threshold for view changes to its corresponding baseline.

defined as a ratio of threshold for angular difference to its corresponding baseline, is depicted in Fig. 4. In same-view matching, subjects tended to be most sensitive to the 0° front view face that is almost bilaterally symmetric. This may represent a perceptual advantage to a bilateral symmetric pattern.

Threshold elevations in Fig. 4 demonstrate intriguing results. There was only a slight increase of threshold at 6.7° angular difference, which is not statistically significant and has a small Cohen's d effect size [t(3) = 0.264, p = 0.809], d=0.132], but a significant increase in threshold was observed at the 13.3° and 20° angular differences, [t(3) = 4.092, p = 0.026, d = 2.046] and [t(3) = 3.750, d = 2.046]p = 0.033, d = 1.875], respectively. In addition, threshold elevation at the 20° angular difference was comparable to that of 13.3° angular differences. Four face views used in the present study are spaced with equal intervals of 6.7° between two adjacent views. Despite this equal interval, threshold elevation by view change did not follow a simple linear function. Instead, subjects' performance suggested that 0° and 6.7° views were treated similarly, and 13.3° and 20° views were treated similarly. These results might reflect a categorical perception of face views that 0° front views and 6.7° side views are perceptually grouped together, while 13.3° and 20° side views are separately grouped. Generalization over views was successful between front and 6.7° side views and then it fell off as the angular difference from the front view increased.

## 5. Experiment 3: Interaction between size and face view

Thus far, the effects of size and view change have been evaluated independently in experiments 1 (size) and 2 (view). Results revealed that size changes did not affect face discrimination whereas view changes did. The purpose of the present experiment is to assess whether an interaction exists between size and view change on face discrimination.

# 5.1. Subjects

Six subjects, including one author, participated in this experiment. Four of them were the same subjects participated in Experiment 2. All had normal or corrected-to-normal vision and five were female. The mean age was 25 years.

# 5.2. Stimuli

The synthetic face stimuli described in Section 2 were used. The linear dimension of comparison faces was reduced to half in the twofold condition, and to one fourth in the four-fold condition. A target face remained the same size as in Experiment 2. There were two face views tested, front and 20° side, and the flashed target face differed from the comparison faces in both view and size.

## 5.3. Procedure

Subjects participated in cross-view matching between front and 20° side views where the size of target and comparison faces were different. There were two conditions: matching a flashed front view to a 20° side view, and matching a flashed 20° side view to a front view.

## 5.4. Results

Fig. 5 compares the results of testing faces differing in both size and view with those of matching same-sized faces. Upper and lower graphs plot thresholds for matching a flashed front view to a 20° side view (F/S) and matching a flashed 20° side view to a front view (S/F), respectively. One-way ANOVAs comparing three different size conditions revealed no statistically significant difference: in F/S,  $[F(2, 11)=0.435, p=0.658, \eta^2=0.073]$ ; and in S/F,  $[F(2, 11)=0.315, p=0.736, \eta^2=0.054]$ . Therefore, view-matching thresholds were independent of size differences between target and comparison faces.

# 6. Discussion

The present study investigated the effect of size and viewpoint on face discrimination and explored how face view interacts with difference in size, using a novel set of synthetic face stimuli. The results were as follows: (1) size changes up to four-fold had no effect on face discrimination indicating almost perfect size constancy; (2) threshold for matching different face views increased with angular difference from frontal view; (3) 0° and 6.7° views were perceptually grouped together, as were 13.3° and 20° views; (4) no interaction between face view and size was found, as size differences across different views had no effect on face discrimination. These results are consistent with previous psychophysical findings on synthetic face adaptation (Anderson & Wilson, 2005): identity-specific face adaptation transfers across different sizes up to four-fold but not to a 20° different viewpoint. Moreover, the present study



Fig. 5. Interaction between face view and size. The results of testing faces differing in both size and view are compared with those of matching samesized faces. Same indicates discrimination threshold of no size change between target and comparison faces (n = 6). Two-fold indicates threshold for a two-fold change in size (n = 4). Fourfold shows threshold of a four-fold size change (n = 4). F/S is the condition that front faces (target) were matched to 20° side views (comparisons). S/F is when 20° side views (target) were matched to front views (comparison). Error bars indicate 1 SE.

illustrates psychophysical evidence consistent with fMRI and neurophysiological findings, which all indicate the sizeinvariant but viewpoint-variant characteristics of monkey IT neurons and human FFA responses (Andrews & Ewbank, 2004; Desimone et al., 1984; Grill-Spector & Malach, 2001; Grill-Spector et al., 1999; Ito et al., 1995; Logothetis & Pauls, 1995; Logothetis et al., 1995; Lueschow et al., 1994; Malach et al., 1995; Op De Beeck & Vogels, 2000; Perrett et al., 1985, 1987, 1991; Pourtois et al., 2005; Rolls & Baylis, 1986; Sáry et al., 1993; Schwartz et al., 1983; Vuilleumier et al., 2002; Wang et al., 1996).

The present study employed a perceptual discrimination task since no learning or long-term memory was involved, and results should be interpreted within this framework. Previous studies found that the effects of size changes could depend on task demands or the level of recognition. In Biederman and Cooper (1992), visual priming was not affected by size changes, but when the task required episodic memory, both RTs and errors increased with a change in object image size. Jolicoeur (1987), testing recognition memory in an old-new paradigm, found that objects presented at the same size as in the learning phase were more quickly and accurately recognized than those shown at a different size. These results suggest that images of an object may be stored in a size-specific manner in long-term memory, and recognition of these images would need extra time for size transformation (Ito et al., 1995).

In an earlier study of size effects on face recognition, Kolers, Duchnicky, and Sundstroem (1985) used a 5:1 linear size range of faces. After viewing faces at one size, they reported that performance was best when faces to be recognized were presented at the same size. Furthermore, performance fell off as the ratio between study and recognition faces increased. However, there is a major difference between this study and ours. Kolers et al. (1985) created their stimuli from large photographs that were reduced in size to make the smaller stimuli. Due to resolution limitations, therefore, the smaller stimuli contained less information than the larger stimuli. Thus, their results may be interpreted as indicating that recognition is best when the information content is identical but is degraded when comparing face pictures with different amounts of information. Due to bandpass filtering and incorporation of 37 geometric measurements, however, our synthetic face stimuli contained precisely the same amount of information at all sizes (although the peak spatial frequency band was shifted with size). Thus, our results indicate that when the quantity of information is constant, face discrimination exhibits size constancy even across spatial frequency bands.

For baseline measures in Experiment 2, all subjects were most sensitive to a frontal view, which is most nearly bilaterally symmetric. Previous studies using a perceptual discrimination task have found results similar to those of the present study. Wilson et al. (2002) found that subjects showed a lower threshold to a front view than a 20° side view face. In another study, subjects were more sensitive to patterns with a high degree of bilateral symmetry than to those with little or no bilateral symmetry (Wilson, Wilkinson, Lin, & Castillo, 2000). High sensitivity to a front view, which is almost bilaterally symmetric, is likely due to redundancy of information present in the front face: one side is almost a mirror of the other, so only one side of the face needs to be discriminated (Tarr & Pinker, 1990). However, there is a limitation in generalizing the results of same view matching to memory. When faces were memorized, this advantage of front view did not seem to matter in recognizing the same view faces. In a face recognition task by Hill et al. (1997), all the views learned (frontal, three-quarter, and profile) were equally well recognized.

Experiment 2 confirmed the viewpoint-dependency of face representation. More interestingly, it showed a clear discontinuity in the pattern of threshold elevation as a

function of angular difference. That is, a large elevation of threshold occurred between 6.7° and 13.3° angular differences while thresholds for 0° and 6.7° angular differences were comparable to each other, and those of 13.3° and 20° differences were similar. It is likely that subjects treated 0° front and 6.7° side views similarly, and 13.3° and 20° side views similarly. Thus, a perceptual boundary may exist between 6.7° and 13.3° views. Assuming approximate symmetry, this suggests that representations of specific face views have an effective bandwidth of about  $\pm 10^\circ$ . These results also appear to be consistent with a categorical representation: within a group, there seems to be a small perceptual difference but the perceptual difference is enhanced between groups.

This idea is consistent with the neurophysiological findings proposing population coding of faces and objects in monkey IT and STS. A number of studies have suggested that objects, including faces, are represented as a collection of specific views, each represented by a group of neurons selective to a combination of features within that view (Abbott, Rolls, & Tovee, 1996; Jolicoeur & Humphrey, 1998; Logothetis et al., 1995; Wallis & Bülthoff, 1999; Young & Yamane, 1992). Psychophysical studies also suggest that 3D objects are represented as collections of associated 2D images taken from multiple perspectives (Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992; Logothetis et al., 1994).

Viewpoint-dependent representations, however. appear to be observed with novel objects, but not with familiar ones. Booth and Rolls (1998) placed real plastic objects in the cages of monkeys destined for neurophysiological experiments for a period of time before testing, so that the monkeys became familiar with those objects. They then identified neurons in IT that responded equally to different views of the same familiar object but not to other objects even if the features or images were similar to the familiar one. These results imply that seeing an object from different views may build a viewinvariant representation (Booth & Rolls, 1998). The view-invariant responses seem to be formed by associating together the output of view-dependent neurons (Booth & Rolls, 1998; Rolls, 1992) and require only a small number of object views in training (Logothetis & Pauls, 1995). Neural network models also have proposed that a view-independent, object-centered recognition is achieved by 2D, viewer-centered representations that interpolate between the small number of stored views or templates (Poggio & Edelman, 1990; Rolls, 1992). Spatiotemporal order appears to play a crucial role in associating different views (Wallis & Bülthoff, 2001).

Our results provide psychophysical support for fMRI findings that the FFA appears to show size constancy but not orientation constancy (Andrews & Ewbank, 2004; Grill-Spector & Malach, 2001; Grill-Spector et al., 1999; Vuilleumier et al., 2002). In our study, Experiments 1 and 2 showed robust compensation for size change but not for view change. Furthermore, Experiment 3 demonstrated

that thresholds for matching different views are independent of size differences. We conclude that, in the human visual system, processing of face view and size may proceed in parallel.

# Acknowledgment

This study was supported by NIH Grant #EY002158 to H.R.W.

#### References

- Abbott, L. F., Rolls, E. T., & Tovee, M. J. (1996). Representational capacity of face coding in monkeys. *Cerebral Cortex*, 6, 498–505.
- Anderson, N. D., & Wilson, H. R. (2005). The nature of synthetic face adaptation. *Vision Research*, 45, 1815–1828.
- Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *Neuroimage*, 23, 905–913.
- Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. Journal of Experimental Psychology: Human Perception and Performance, 18(1), 121–133.
- Booth, M. C. A., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral Cortex*, 8, 510–523.
- Brainard, D. H. (1997). The psychophysics toolbox. Spatial Vision, 10, 433–436.
- Bülthoff, H., & Edelman, S. (1992). Psychophysical support for a twodimensional view interpolation theory of object recognition. Proceedings of the National Academy of Sciences of the United States of America, 89, 60–64.
- Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulusselective properties in inferior temporal neurons in the macaque. *Journal of Neuroscience*, 4(8), 2051–2062.
- Desimone, R., & Gross, C. G. (1979). Visual areas in the temporal cortex of the macaque. *Brain Research*, 178, 363–380.
- Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, 32, 2385–2400.
- Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research*, 39, 3537–3560.
- Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychologica*, 107, 293–321.
- Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, 24, 187–203.
- Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology*, 35, 96–111.
- Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition*, 62, 201–222.
- Ito, M., Tamura, H., Fujita, I., & Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of Neurophysiology*, 73(1), 218–226.
- Jolicoeur, P. (1987). A size-congruency effect in memory for visual shape. Memory and Cognition, 15(6), 531–543.
- Jolicoeur, P., & Humphrey, G. K. (1998). Perception of rotated two-dimensional and three-dimensional objects and visual shapes. In V. Walsh & J. Kulikowski (Eds.), *Perceptual constancy: Why things look as they do* (pp. 69–123). Cambridge: Cambridge University Press.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in extrastriate cortex specialised for face perception. *Journal of Neuroscience*, 17, 4302–4311.

- Kolers, P. A., Duchnicky, R. L., & Sundstroem, G. (1985). Size in the visual processing of faces and words. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 726–751.
- Lehky, S. R. (2000). Fine discrimination of faces can be performed rapidly. Journal of Cognitive Neuroscience, 12, 848–855.
- Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature Neurosicence*, 8(10), 1386–1390.
- Logothetis, N. K., & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex*, 3, 270–288.
- Logothetis, N. K., Pauls, J., Bülthoff, H. H., & Poggio, T. (1994). View-dependent object recognition by monkeys. *Current Biology*, 4(5), 401–414.
- Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, 5(5), 552–563.
- Lueschow, A., Miller, E. K., & Desimone, R. (1994). Inferior temporal mechanims for invariant object recognition. *Cerebral Cortex*, 4, 523–531.
- Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 92, 8135–8139.
- Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research*, 39, 3824–3833.
- Op De Beeck, H., & Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *Journal of Comparative Neurology*, 426, 505–518.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Perrett, D. I., Mistlin, A. J., & Chitty, A. J. (1987). Visual neurons responsive to faces. *Trends in Neuroscience*, 10(9), 358–364.
- Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J., et al. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Experimental Brain Research*, 86, 159–173.
- Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., et al. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London B*, 223, 293–317.
- Poggio, T., & Edelman, S. (1990). A network that learns to recognize threedimensional objects. *Nature*, 343, 263–266.
- Pourtois, G., Schwartz, S., Seghier, M. L., Lazeyras, F., & Vuilleumier, P. (2005). Portraits or people. Distinct representations of face identity in the human visual cortex. *Journal of Cognitive Neuroscience*, 17(7), 1043–1057.

- Quick, R. F. (1974). A vector-magnitude model of contrast detection. *Kybernetik*, 16, 1299–1302.
- Rolls, E. T. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philo*sophical Transactions of the Royal Society of London, Series B: Biological Sciences, 335, 11–21.
- Rolls, E. T., & Baylis, G. C. (1986). Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Experimental Brain Research*, 65, 38–48.
- Sáry, G., Vogels, R., & Orban, G. A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science*, 260, 995–997.
- Sawamura, H., Georgieva, S., Vogels, R., Vanduffel, W., & Orban, G. A. (2005). Using functional magnetic resonance imaging to assess adaptation and size invariance of shape processing by humans and monkeys. *Journal of Neuroscience*, 25, 4294–4306.
- Schwartz, E. L., Desimone, R., Albright, T. D., & Gross, C. G. (1983). Shape recognition and inferior temporal neurons. Proceedings of the National Academy of Sciences of the United States of America, 80, 5776–5778.
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. Science, 171, 701–703.
- Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21(2), 233–282.
- Tarr, M. J., & Pinker, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science*, 1(4), 253–256.
- Vuilleumier, P., Henson, R. N., Driver, J., & Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature Neuroscience*, 5(5), 491–499.
- Wallis, G., & Bülthoff, H. (1999). Learning to recognize objects. Trends in Cognitive Sciences, 3(1), 22–31.
- Wallis, G., & Bülthoff, H. (2001). Effects of temporal association on recognition memory. Proceedings of the National Academy of Sciences of the United States of America, 98, 4800–4804.
- Wang, G., Tanaka, K., & Tanifuji, M. (1996). Optical imaging of functional organization in the monkey inferotemporal cortex. *Science*, 272(14), 1665–1668.
- Wilkinson, F., Wilson, H. R., & Habak, C. (1998). Detection and recognition of radial frequency patterns. *Vision Research*, 38, 3555–3568.
- Wilson, H. R., Loffler, G., & Wilkinson, F. (2002). Synthetic faces, face cubes, and the geometry of face space. *Vision Research*, 42, 2909–2923.
- Wilson, H. R., Wilkinson, F., Lin, L.-M., & Castillo, M. (2000). Perception of head orientation. *Vision Research*, 40, 459–472.
- Young, M. P., & Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex. *Science*, 256(29), 1327–1331.