Face Recognition Is Affected by Similarity in Spatial Frequency Range to a Greater Degree Than Within-Category Object Recognition

Charles A. Collin and Chang Hong Liu
McGill University

Patricia A. McMullen
Dalhousie University

Avi Chaudhuri
McGill University

Nikolaus F. Troje
Ruhr-Universität-Bochum

There is a great deal of evidence showing that face recognition is strongly affected by variations in spatial frequency content (Bachmann, 1991; Costen, Parker, & Craw, 1994; Nasanen, 1999). In contrast, object recognition may be relatively invulnerable to these sorts of manipulations (Biederman & Kalocsai, 1997). Some authors have argued that this is due to object recognition being initially and primarily based on features such as edges and lines (Biederman, 1987; Biederman & Ju, 1988; Biederman & Kalocsai, 1997), whereas face recognition relies on a broader range of information encompassing edges, textures, and shading (Bruce & Humphreys, 1994).

Many studies have shown that face perception is heavily influenced by the particular range of spatial frequency information available. Bachmann (1991), using a delayed match-to-sample task, found a rapid decrease in accuracy as coarse-quantized face images were reduced in resolution from 18 pixels per face width to 15. In related experiments, Costen et al. (1994, 1996), using a similar method to Bachmann’s, found abrupt drop-offs in performance as faces were low-pass filtered at cutoffs less than 8 cycles per face or high-pass filtered at greater than 16 cycles per face. In experiments comparing human delayed match-to-sample performance with that of an ideal observer algorithm, Nasanen (1999) found a reduction in the efficiency function of participants when frequencies around 8–13 cycles per face were masked by noise. Gold, Bennett, and Sekuler (1999) examined participants’ ability to recognize two-octave-wide bandpass-filtered images of letters and faces, finding that letters could be identified across a broad range of spatial scales whereas faces could be recognized at above-chance levels only when the band was at a harmonic mean of 8.8 cycles per face width. This was despite the fact that ideal observer algorithms solved the tasks with nearly equal ease at all bands. Together, these findings suggest the existence of a critical band of middle spatial frequencies, between 8 and 16 cycles per face, that is of greater importance to face recognition than outlying bands.

Studies examining recognition across image types have also indirectly suggested a disproportionate vulnerability of face perception to spatial frequency variations. Davies, Ellis, and Shepherd (1978) found that face images were difficult to recognize in line drawings relative to photographs. In contrast, early studies (Fraise & Elkin, 1963; Ryan & Schwartz, 1956) have shown that object recognition may actually be superior in line drawings relative to photographs. Biederman (1987; Biederman & Ju, 1988) showed that line drawings of objects could be recognized almost as well as photographs, whereas Bruce, Hanna, Dench, Healey, and Burton, (1992) showed that line drawings and two-tone images of faces were poorly recognized (see also Hayes, 1988). Biederman and Kalocsai (1997) have shown that object images may be recognized quite well from spatially complementary images but that participants have difficulty performing the same task with faces.
Liu, Collin, Rainville, and Chaudhuri (2000) found that face recognition is strongly affected by a factor they called spatial frequency overlap (SFO). They defined SFO as the range of spatial frequencies shared by two images. For example, an image low-pass filtered at eight cycles per degree shares one octave of the spatial spectrum with an image high-pass filtered at four cycles per degree. This is referred to as one octave of SFO. In their study, a single octave change in this factor was found to produce differences in accuracy on the order of 20% in some conditions, and manipulations in SFO elicited performance ranging from floor to ceiling.

Liu et al. (2000) suggested that SFO is a more important factor in determining how easily filtered faces are recognized than the spatial range occupied by the images. As evidence, they showed that recognition of filtered images was little affected by changes in the frequency cutoff when the learned images and tested images (or images to be matched) were filtered in the same way. Under these circumstances they found that a broad range of frequency bands was sufficient for recognition and that no particular band was more useful than any other.

On the basis of these findings, Liu et al. (2000) proposed that an understanding of spatial frequency effects on face perception could be enhanced by incorporation of the SFO factor in models of the recognition process. They pointed out that, in their experiments, performance did not drop off when learned and tested images contained only frequencies outside the critical band. For example, when participants had to learn a face low-pass at 2.4 cycles per face width and recognize the same face filtered in the same way, they did about as well as when the cutoff was moved up to 17.9 cycles per face width, on the other side of the critical range. The researchers suggested that this lack of a critical band effect in their data was based on the contributions of overlapping and interfering frequencies. They pointed out that studies such as those of Costen et al. (1994, 1996) used full-bandwidth pictures at learning, meaning that there was a large range of frequencies in the learned image that were not present in the tested image. These frequency elements may have interfered with recognition, as would have found no equivalents in the tested images. In Liu et al.’s control conditions, on the other hand, the learned and tested images were filtered in the same way. This meant that the frequency ranges overlapped completely and that there was no potential for interference from outlying frequencies.

Whereas previous efforts in this domain have emphasized the importance of the specific band of frequencies held in common between learned and tested images, Liu et al. (2000) pointed out the importance of both the overlapping frequencies and those that are not shared between image pairs. The overlapping frequencies aid recognition, whereas the nonoverlapping ones potentially interfere with it. It should be noted that interference here does not mean low-level masking but refers to higher order spatial feature matching. The recognition process is interfered with when featural elements do not match between images. Throughout this article, we use the term interference in this sense.

Although Liu et al.’s (2000) data are useful, their work left an important question unanswered. That is, does the theory of overlapping and interfering frequency bands apply only to the recognition of face images? If object recognition is primarily edge based (Biederman, 1987; Biederman & Ju, 1988; Marr, 1982), one might expect that it would be less vulnerable to variations in SFO across learned and tested items. That is, although two face images filtered to include isolated bands of spatial frequencies might be difficult to recognize from one another, object images similarly filtered might be relatively easy to match. This relative invulnerability of object recognition is expected because edge features tend to transcend scale (Marr, 1982) and would therefore be available for matching across a wider range of spatial frequency bands than nonedge features such as shadows and textures. However, it should be pointed out that this implies a nonlinear higher level comparison of information across the frequency bands, as any two images that do not overlap in spatial frequency have a zero-order point-by-point correlation of luminance levels by definition.

The only previous study to directly compare the relative vulnerability of object and face recognition to variations in spatial frequency similarity is that of Biederman and Kalocsai (1997). They examined the ability of participants to recognize spatially complementary face or chair images. Their stimulus images were created by dividing the spatial frequency by orientation space into an 8 × 8 grid and filtering out every odd diagonal of cells in one image and every even diagonal of cells from the other image (i.e., by analogy to a chess board, the black squares were filtered out of one image and the white squares out of the other). This is analogous to the method used by Biederman and colleagues in previous studies (e.g., Biederman & Cooper, 1991), wherein segments of lines are removed from line drawings, except that the manipulation used by Biederman and Kalocsai took place in the frequency domain instead of the spatial domain. The end result of this manipulation was that there were no elements in the two images that shared either spatial frequency or orientation. This is somewhat similar to a condition of zero SFO, except that Biederman and Kalocsai’s manipulation affected both spatial frequency and orientation, whereas SFO manipulations affect spatial frequency in isolation.

Biederman and Kalocsai’s (1997) results showed that participants found it quite easy to match or recognize these spatially complementary images if they represented chairs but that they had difficulty doing the same task with face images. As an explanation of these findings, they proposed that the two types of images are stored in different ways. Objects, they suggested, are stored as geon structural descriptions, which are qualitative representations derived from line and edge features and which are free of spatial frequency information. Faces, on the other hand, are stored as a set of outputs of V1-type cells according to Biederman and Kalocsai, that is, a matrix of localized spatial frequency and orientation-selective filters. The difference in representation types is held to be responsible for the differences in sensitivity to spatial frequency content. Although Biederman and Kalocsai’s model may seem to be at odds with the fact that cells in the lateral occipital cortex have been shown to be largely size and position invariant (see, e.g., Grill-Spector et al., 1999; Grill-Spector & Malach, 2001), it may be the case that fusiform-gyrus cells are activated by a broad array of simple cells representing individual faces.

Although Biederman and Kalocsai’s (1997) results are intriguing, there are some limitations to their study. For instance, they examined spatial complementarity in terms of orientation and frequency at the same time, making it impossible to know what the effect of either factor alone might have been. Also, they examined only the extreme cases of complete spatial complementarity (roughly analogous to 0 octaves of SFO) and equivalence (roughly analogous to complete SFO), making it difficult to know what the continuous relationship was between similarity in spatial informa-
tion and the category of stimulus to be recognized (i.e., face vs. object). Finally, they applied filter profiles with sharp cutoffs, an unusual procedure that produces ringing, a pattern of low-frequency undulations in the luminance of the image. It is possible that this low-frequency interference affected face recognition more than object recognition, leading to an alternate explanation for their results.

The purpose of the present study was to directly compare the sensitivity of face and object recognition with variations in SFO across a range of intermediate levels. In Experiment 1, we compared faces with a single object category (chairs), similar to Biederman and Kalocsai (1997), but we defined similarity of spatial frequency content in terms of SFO, in a similar way to Liu et al. (2000), that is, as the range of spatial frequencies shared by two images. In Experiment 2, we examined whether the results found with chair stimuli generalized to other object categories. In Experiment 3, we examined what happened when the face stimuli were inverted and the degree of configural and holistic processing was thus reduced. All experiments used a same–different sequential matching task wherein different views of objects were shown at study and test phases. Recognition in this study was therefore defined as correctly matching two different object views. It should be pointed out that using different views at learning and testing is an important methodological consideration, as it ensures that genuine recognition of a particular class of visual objects is being investigated rather than general pattern matching. Our results suggest that differences in effects of SFO on face and object recognition help explain the vulnerability of face recognition to changes in image format and the lack of a middle-band superiority effect in object recognition.

**Experiment 1**

**Method**

Participants. Sixty-four undergraduates (16 male) from McGill University, Montréal, Québec, Canada, with ages ranging from 17 to 39 years (Mdn = 20 years), participated. Participants were randomly assigned to one of four groups: face test, face control, chair test, or chair control. Each group had 16 participants. All participants had normal or corrected-to-normal vision.

Stimuli and materials. Stimuli were spatially bandpass-filtered images of faces and chairs. The original face images were obtained from a database of 3-D laser-scanned head models provided by the Max-Planck Institute for Biological Cybernetics in Tübingen, Germany (http://faces.kyb.tuebingen.mpg.de/). This allowed photorealistic rendering of the faces from a wide range of angles. The models were originally created using a Cyberware laser scanner that records both surface shape and texture of 3-D forms. By mapping the texture onto the shape model, one could create an image of the faces from any angle. Further details regarding the face database may be obtained in Troje and Bülthoff (1996, 1998).

The object images were obtained by photographing a variety of chairs. Chair images were chosen to remain consistent with past studies (e.g., Biederman & Kalocsai, 1997) and because they, like faces, are a familiar and homogeneous object category that is typically seen upright. All faces and chairs were imaged from the frontal and three-quarter views. With faces, this meant simply that the 3-D model was placed at 0° and 45° relative to the virtual camera before an image was captured. With chairs, the front corners of the seat were placed equidistant from the camera for front shots, whereas the right front and rear left corners (and vice versa) were lined up relative to the camera for the three-quarter shots. The chair photographs were taken from eye level because chairs are commonly seen from above, whereas face images were taken from a 0° elevation because this is the most typical view of faces (i.e., eye to eye). Thirty individual faces and chairs were used. From each of these sets, 10 items were selected for use in practice sessions. All images were converted to 256 gray-level format, and the image backgrounds were replaced with a gray background before filtering. All were equated for mean luminance and root-mean-square (RMS) contrast. To equate the stimuli in overall dimensions, we changed the height and width of the chair images to match the aspect ratio of the faces. This did not result in any unusual appearance of the chairs, as the changes in aspect ratio were generally slight.

The spatial filtering of the images was done with MATLAB (Version 5.2; www.mathworks.com) software for Macintosh. To create the filtered versions, we passed the original full-bandwidth images through a bandpass filter created by multiplying together a low-pass and a high-pass Butterworth filter. Images were then inversely transformed into the spatial domain. Butterworth filters were used to avoid the ringing effects seen when images are filtered with abrupt spatial frequency cutoffs. Nonetheless, the functions were steep enough to provide good spatial frequency localization. The bandpass image was two octaves wide in all cases. That is, the high and low 50% cutoffs were two octaves apart. The low-pass filters were defined by

\[
1 + \frac{r}{c}
\]

and the high-pass filters by

\[
1 + \frac{r}{c}
\]

where \( r \) is the component radius and \( c \) the cutoff radius. All images had a bandwidth of two octaves.

Three SFO conditions were created by assigning cutoff values to pairs of bandpass filters such that they shared 0.33, 1.00, or 1.67 octaves of the spatial spectrum, respectively. Each of these overlap conditions was created at two regions in the spatial scale by having the filter pairs center around two different frequencies, 7.08 and 14.15 cycles per stimulus. Figure 1 illustrates the gain profiles and cutoff values of all the filters used. Also shown, at the top of each bandpass-filter function, is the harmonic mean frequency of that filter.

Figure 2 shows examples of both face and chair stimuli passed through the filters shown in Figure 1. The rows represent center frequency (7.08 and 14.15 cycles per stimulus), and the columns show overlap levels (0.33, 1.00, and 1.67 octaves). Each cell contains two chair images and two face images, with the lower frequency band on the right and the higher frequency band on the left. The term opposite filters is used throughout this article to refer to bandpass filters symmetrically positioned to either side of a given center frequency. For instance, at a center frequency of 7.08 cycles per stimulus in the one octave of overlap condition (see Figure 1, upper panel, dashed lines), the lower band ranges from 5.0 to 20.1 cycles per stimulus, whereas the higher band ranges from 10.0 to 40.0 cycles per stimulus. These are referred to as opposite filters to one another.

The experiment was run using a Macintosh G3/233. Stimuli were presented on a 21-in. AppleVision monitor that was calibrated to give a linear luminance profile. Images were 256 × 256 pixels, with the stimulus itself being approximately 136 pixels across, on average. The images were surrounded by a medium-gray background that filled the screen.

Design and procedure. The design of the experiment was a 2 × 2 × 3 mixed model. The factors were (a) stimulus type (face or chair), (b) task type (test or control), (c) center frequency (7.08 or 14.15 cycles per stimulus), and (d) SFO (0.33, 1.00, or 1.67 octaves). The first two factors were between participants, whereas the latter two were within participants.

The initial part of the experiment was a short practice session. Prior to beginning the practice session, participants were given written instructions on the computer monitor. On each trial of the practice session, participants saw two sequentially presented, spatially filtered images separated by a mask. The first image was presented for 1 s. This was followed by the
mask, which consisted of a scrambled version of the first stimulus created by dividing the image into an 8 x 8 grid and randomly shuffling the elements. The mask was also presented for 1 s. This was followed by the test image, which stayed on the screen until the participant's answer was given. Half the trials presented a 0° view of the stimulus first, followed by a 45° view, whereas, in the other half, this order was reversed. Angle order was randomized, and half of the participants saw 45° views to the right and half to the left.

In the test conditions, the images to be matched consisted of pictures that had been filtered with opposite filters. For instance, in the 0.33-octave overlap condition at the lower center frequency (see Figure 1, top panel, dotted lines), the first face was filtered to contain only frequencies between 2.0 and 8.0 cycles per stimulus, whereas the second was filtered to contain only frequencies between 6.3 and 25.1 cycles per stimulus. In target trials, the two images were of the same face or chair. In distractor trials, the two images were of different items chosen randomly but with the constraint that distractor faces had to be of the same sex as the target face and distractor chairs had to be of the same type as the target chair (e.g., office chair, wooden chair, etc.).

In the control conditions, the images consisted of pictures that had been processed with the same filter. These self-match control conditions were run to ascertain whether overlap effects observed in the test conditions might be contributed to by differences in the efficacy of different bands. For instance, any increases in accuracy with rising overlap might be attributed either to (a) a rise in the range of spatial frequencies shared by the two images in a test pair or (b) the bands occupied by the two images in a test pair being individually more efficacious to the recognition process. To determine the magnitude of such effects, we tested the ability of participants to match two images occupying the same band of frequencies. Six conditions were run, corresponding to the six different filters used to produce images for the test conditions. In all conditions (test and control), participants were asked to answer as quickly as possible.

Data from the control conditions were used to derive baseline values against which to compare the various test conditions. For instance, the baseline value for the 0.33 octaves of overlap test condition at a center frequency of 7.08 cycles per stimulus was created by averaging the accuracy values from two control conditions: (a) the 2.0 to 8.0 cycles per stimulus condition and (b) the 6.3 to 25.1 cycles per stimulus condition.

Figure 1. Gain profiles and cutoff values of the filters used to create stimuli for all experiments. At the peak of each filter, the harmonic mean frequency of that filter is provided. oct. = octave; c/o = cycles per object.
That is, each baseline value was created by averaging accuracies for the two control conditions in which participants matched images occupying the two bands from the associated test condition. In all conditions, 12 practice trials were given, 2 from each of the six possible conditions. Trial order was randomized.

The experimental session immediately followed the practice session. The procedure here was the same as in the practice session, except that each condition was tested 40 times, for a total of 240 trials. The breakdown of these trials for the test conditions was 3 (overlap levels) × 2 (center frequencies) × 2 (target vs. distractor trial) × 20 (individual faces or chairs). The order of conditions was randomized with the constraint that the same face could not appear in sequential trials. In each condition, half the trials presented a 0° view of the stimulus first, followed by a 45° view, whereas, in the other half, this order was reversed. Angle order was randomized.

Following the presentation of each sequential image pair, participants responded as to whether the images were pictures of the same object or same face. This was done by means of buttons marked yes and no on the computer keyboard. Participant accuracy and reaction time (RT) were automatically recorded by the computer program.

**Results**

Accuracy data are shown in Figure 3. The graphs show means and standard errors. A four-way analysis of variance (ANOVA) was used to analyze the results. This showed a significant three-way interaction between the factors of stimulus type (face vs. chair), trial type (test vs. baseline), and overlap level (0.33, 1.00, and 1.67 octaves), $F(2, 120) = 3.2$, $p < .05$. Post hoc analysis using ANOVA and Tukey’s honestly significant difference (HSD; $\alpha = .05$) showed that this was due to a significant effect of overlap on face recognition in the test trials, an effect that was not evident in any of the other three Stimulus Type × Trial Type conditions. Thus, it appears that overlap had a strong effect in the face recognition condition, but not in the condition where chairs were being recognized.

In the test condition with faces as stimuli, the three overlap levels each differed from one another significantly at 67%, 72%, and 79% accuracy for 0.33, 1.00, and 1.67 octaves of overlap,

---

**Figure 2.** Example stimuli for Experiments 1 and 3. Each cell corresponds to one of the six Center Frequency × Overlap conditions tested. Each cell contains four images: the lower band and higher band versions of both a chair and a face stimulus. For instance, the upper left cell shows four stimuli from the lower center-frequency condition (7.08 cycles per stimulus) at the lowest overlap. These images thus correspond to the filters shown in the top panel of Figure 1 with dotted lines. The images on the left in each cell contain the lower of the two spatial frequency ranges compared in a given Overlap × Center Frequency condition, whereas the images on the right of each cell contain the higher of the two spatial frequency ranges compared in a given condition. c/o = cycles per object.
respectively. In the equivalent baseline conditions, little difference was seen between levels, with percentages of 77%, 80%, and 79% for the three levels in order. A similar analysis for objects showed that overlap had little effect on recognition in this case, with differences that were small and nonsignificant. Accuracy values in the test conditions were 67%, 70%, and 71% for the three overlap levels in order. These were not significantly different from one another. In the control conditions, a similarly flat function was seen, with associated values of 68%, 71%, and 70%.

Center frequency had a significant main effect, $F(1, 60) = 11.62, p < .002$, but did not interact with other factors. Overall, the center frequency of 14.15 cycles per stimulus showed slightly greater accuracy, at 74%, than the center frequency of 7.08 cycles per stimulus, which yielded 71%. This is likely attributable to the greater amount of information in a given band of overlapping frequencies at higher points in the spectrum.

RT data are shown in Figure 4. As can be seen in the figure, RT data were quite flat with regard to SFO variations for both face and object conditions, with the largest difference between conditions being about 60 ms. Thus, the RT data failed to follow the accuracy data, showing that there was no speed-accuracy trade-off. This was confirmed by analysis with a four-way ANOVA, which showed no main effect of SFO.

**Discussion**

Accuracy data clearly indicated that face recognition is more sensitive to SFO variations than object recognition. Accuracy in
recognizing faces varied with similarity in spatial frequency range, whereas accuracy in recognizing objects did not. An explanation of these findings based on a greater homogeneity of face stimuli is ruled out by the control conditions, which showed that object recognition was overall more difficult than face recognition when spatial frequency range was identical.

Our findings have implications for a number of questions in the domain of visual recognition. First, they suggest that the roles of overlapping and interfering frequencies (Liu et al., 2000) are different in object and face recognition. It seems that nonoverlapping frequencies interfere little with the former task relative to the latter. Second, with regard to the inadequacy of line drawings in face recognition, the present results suggest that this effect is due, at least in part, to the vulnerability of this task to low SFO levels between a tested image and a learned image. Finally, concerning findings that show that middle frequencies are most important to face recognition, these results suggest an explanation for why such a middle-band advantage has not been found in object recognition. Each of these points is examined in more detail in the General Discussion.

Although the results of this experiment are useful, it is difficult to generalize from a single object class (chairs) to object recognition in general. To examine the generalizability of the findings of Experiment 1, we examined recognition of a variety of object types under similar circumstances in Experiment 2. If the findings of Experiment 1 were not due to particular characteristics of chair stimuli, we should have seen similar results in both experiments.

Experiment 2

This experiment was similar to Experiment 1 except that instead of presenting exemplars from the same category throughout the experiment, we presented exemplars from many different categories. The two exemplars presented on a given trial were always from the same category. We predicted that there would be little effect of SFO variations overall, replicating the results of Experiment 1 with chair images.

Method

Participants. Fourteen undergraduates (2 male) from Dalhousie University in Halifax, Nova Scotia, Canada, with ages ranging from 18 to 36 years (Mdn = 19 years), participated. Each participant performed both the test and control conditions in this experiment. All participants had normal or corrected-to-normal vision.

Materials and stimuli. The stimuli used in Experiment 2 were spatially bandpass-filtered images of various objects. Two exemplars from each of the following 12 object categories were used: hairclips, coffeemakers, forks, hammers, keys, padlocks, pens, purses, scissors, screwdrivers, staplers, and wrenches. The latter two categories were used in practice sessions only. These objects were chosen on the basis of availability of two visually similar exemplars. Figure 5 shows examples. The original images were obtained by digitally photographing the objects from an elevation of 20°. All objects were imaged from frontal (0°) and three-quarter (45°) views. All images were converted to 256 gray-level format before filtering. All images were equated for mean luminance and RMS contrast as well. The object images were equated in size at 192 pixels across the diagonal. This was found to be the best way to equate the sizes of the object images.

Figure 5. Example stimuli for Experiment 2. Each cell corresponds to one of the six Center Frequency × Overlap conditions tested. Each cell contains two images: the lower band and higher band versions of one example stimulus (a pair of scissors). For instance, the upper left cell shows stimuli from the lower center-frequency condition (7.08 cycles per object) at the lowest overlap. These images thus correspond to the filters shown in the top panel of Figure 1 with dotted lines. The images on the left in each cell contain the lower of the two spatial frequency ranges compared in a given Overlap × Center Frequency condition, whereas the images on the right of each cell contain the higher of the two spatial frequency ranges compared in a given condition. c/o = cycles per object.
which varied widely in aspect ratio. A 192-pixel diagonal corresponds to approximately 136 pixels in width when the aspect ratio is 1:1.

These images were filtered in the same way as the face images in Experiment 1. The experiment was run using a Macintosh G3/300. The experiment was presented on a 17-in. Studiovision monitor that was calibrated to provide a linear luminance profile.

**Design and procedure.** The design of the experiment was $2 \times 2 \times 2$. The factors were (a) task type (test or control), (b) center frequency (7.08 or 14.15 cycles per stimulus width), and (c) SFO (0.33, 1.00, or 1.67 octaves). All factors were within participants.

The procedure for this experiment was similar to that for Experiment 1 except that instead of face or chair images, participants were presented with pairs of images representing various objects. On each trial, two spatially filtered object images were presented. On match trials, these were two different images of the same exemplar. On unmatched trials, the images were of two different exemplars from the same category. For example, on a given match trial, a participant might see two images of the same coffeemaker; on an unmatched trial, a participant might see images of two different coffeemakers or two different wrenches. A second difference between this experiment and Experiment 1 was that here, all participants participated in both test and control conditions in this experiment, with the order of conditions being counterbalanced and the participants being assigned randomly to one order or the other (thus, 7 participants performed the experiment with each order).

**Results**

Accuracy data are shown in Figure 6. The graphs show means and standard errors. A three-way repeated-measures ANOVA was used to analyze the results. This showed a significant main effect of overlap level, $F(2, 26) = 6.535, p < .01$. Post hoc testing with Fisher’s protected least significant difference ($\alpha = .05$) showed that this was due to the 1.67 octaves of overlap condition being more accurate than the 1.00 or 0.33 octaves conditions by 4%. This difference is seen in both test and control conditions, suggesting that it is not due to overlap effects per se but has to do with the characteristics of the individual images in the to-be-matched pair. Also, the effect is much smaller than the 12% increment seen in the test conditions of Experiment 1 under similar circumstances. There was no significant effect of center frequency. The interaction between overlap and trial type (test vs. baseline) was not significant, $F(2, 26) = 1.186, p = .32$.

RT data are shown in Figure 7. As can be seen in the figure, RT data were quite flat with regard to SFO variations for both face and object conditions. There was no evidence of a speed-accuracy trade-off. This was confirmed by analysis with a three-way ANOVA, which showed no main effect of SFO.

To further examine whether the results from this experiment showed a generalization of the effects seen in Experiment 1 to the various stimuli used in this experiment, we conducted an item analysis. That is, we conducted a three-way repeated-measures ANOVA with object exemplar number as the random factor to determine if the same effects would obtain when the source of error variance was the differences between objects rather than between participants. The results of this ANOVA were similar to that with participants as the random variable. There was a significant effect of overlap, $F(2, 38) = 9.52, p < .01$, but no significant interaction between overlap and trial type (test vs. baseline), $F(2, 38) = 2.32, p = .11$. This is compatible with our claim that objects show lesser overlap effects than faces, although the usual cautions about interpreting negative statistical results should be noted.

**Discussion**

These data support the generalizability of the results of Experiment 1 to other object classes. When participants were asked to match pairs of objects within the same basic category, they showed the same immunity to SFO effects with a wide variety of object categories. These results suggest that SFO is less of a factor in
object recognition than in face recognition. The small number of objects tested here is not sufficient to make a claim about all object recognition, of course. Indeed, there are almost certainly some objects that would show the same degree of overlap sensitivity as faces. However, it does seem that for a variety of common artifacts, at least, our results hold.

One possible explanation for the differences between recognition of faces and objects in Experiments 1 and 2 is that they are due to differences in the degree of configural or holistic processing involved in the recognition of these kinds of stimuli. If the recognition of face stimuli relies to a greater degree on configural encoding than object recognition and if face recognition requires a greater degree of SFO as well, then it is reasonable to hypothesize that one is related to the other. One reason to suspect that configural encoding might require greater SFO than encoding of parts is that the former is a higher order form of analysis than the latter, and higher order analyses may be more vulnerable to missing information simply because there are more steps involved in the analysis. That is, to do configural encoding, the visual system presumably has to first detect the parts of an object, next determine their relative distances from one another, and then calculate the ratios of these distances. On the other hand, part-based encoding requires only the first one or two steps. The more steps required, the more opportunities there are for the missing information to interfere with the analysis.

One way to test whether the greater need for SFO arises from the demands of configural processing would be to invert the faces. Upside-down faces have been shown to be processed in a less configural manner than upright faces. If configural processing was the source of SFO sensitivity in faces, this sensitivity should have been reduced in inverted faces.

**Experiment 3**

To test whether the need for configural encoding was the source of faces’ sensitivity to SFO, we tested the recognition of inverted faces under similar conditions to Experiments 1 and 2. Inverted faces have been shown to be processed in a less configural manner than upright faces. If configural processing was the source of SFO sensitivity in faces, this sensitivity should have been reduced in inverted faces.

**Method**

**Participants.** Thirty-six undergraduates (7 male) from McGill University, with ages ranging from 16 to 28 years (Mdn = 20 years), participated. Participants were randomly assigned to one of two groups: inverted face test or inverted face control. Each group had 18 participants. All participants had normal or corrected-to-normal vision.

**Materials and stimuli.** The stimuli used in Experiment 3 were the same filtered face images used in Experiment 1, except that they were presented inverted. The factors were (a) task type (test vs. control), (b) center frequency (7.08 or 14.15 cycles per stimulus), and (c) SFO (0.33, 1.00, and 1.67 octaves). The first factor was between participants, whereas the latter two were within participants. The procedure was identical to that used in Experiment 1, except that only face stimuli were presented, and these were presented inverted.
**Results**

Accuracy data are shown in Figure 8. The graphs present means and standard errors. A three-way ANOVA was used to analyze the results. This showed a significant two-way interaction between trial type (test vs. baseline) and SFO (0.33, 1.00, and 1.67 octaves), $F(2, 68) = 7.74$, $p = .0009$. Post hoc testing with ANOVA and Tukey’s HSD ($\alpha = .05$) showed that this was due to the overlap factor having a significant effect in the test conditions, where all levels were different from one another, but not in the control conditions, where none of the levels differed significantly. These results are nearly identical in shape and magnitude to those found with upright faces.

To determine if inverting the faces had any effect on their sensitivity to SFO variations, we compared the results of the upright face conditions from Experiment 1 with those of the inverted face conditions in Experiment 3. This was done by means of a four-way mixed ANOVA, with factors being face orientation (upright vs. inverted, or Experiment 1 vs. Experiment 3), trial type (test or baseline), SFO (0.33, 1.00, or 1.67 octaves), and center frequency (7.08 or 14.15 cycles per stimulus). The two-way interaction between SFO and trial type was significant, $F(2, 132) = 22.5$, $p < .0001$, replicating results from Experiments 1 and 3. The three-way interaction between face orientation, trial type, and SFO was not significant, $F(2, 132) = 1.65$, $p = .1964$, suggesting that the two-way interaction was unaffected by face orientation. That is, the effect of SFO seems to be equivalent for upright and inverted faces, despite the differences in the degree of configural processing in these two types of stimuli. However, the usual cautions about interpreting a negative result obviously apply here.

**Discussion**

With inverted faces, the accuracy function across SFO is virtually identical in slope to that seen in Experiment 1, where faces were upright. Though an expected overall drop in performance is seen, the shape of accuracy functions is unaffected by the rotation of the face images. This suggests that the reason for the difference in SFO effects between face and object recognition is not a result of dependency on configural or holistic mechanisms for face recognition. The usual cautions about interpreting negative results obviously apply here, but our results do seem to suggest that the sensitivity of face stimuli to SFO variations cannot be entirely attributed to the need for configural processing in these stimuli.

**General Discussion**

We showed in Experiment 1 that face recognition is affected by changes in SFO to a greater degree than the recognition of one object category (chairs). As SFO increased, participant accuracy in matching faces rose, whereas accuracy in matching the chair images stayed stable. Experiment 2 showed that the results for chair stimuli generalized to a variety of other objects. Experiment 3 found a similar pattern of results with inverted faces as upright ones. Taken together, these experimental results argue that face recognition is more sensitive to variations in spatial frequency range than within-category object recognition and that this sensitivity cannot be attributed entirely to the holistic or configural nature of face recognition.

Our results with face stimuli support Liu et al.’s (2000) findings, which showed that recognition of low-passed and high-passed faces was much more sensitive to SFO variations than to variations in the range of frequencies contained in the face images. We have expanded on their findings by showing that this sensitivity is not due to the necessity for configural processing in face images. Also, we have improved on their methodology by using bandpass-filtered images instead of low-pass and high-pass ones, which...
produced a potential confound whereby greater SFO was always accompanied by greater spatial frequency range in the learned and tested images. Another methodological improvement over this previous study is the use of different views in the pairs of images to be matched. Liu et al. used the same view of the two faces, leading to the possibility that their experiment was measuring image matching rather than face matching in particular. Our results alleviate this concern.

Although our results with face stimuli support Liu et al.’s (2000) findings, our results with object stimuli differ from their study’s results using face stimuli. Within-category object recognition does not show the same sensitivity to SFO that face recognition does. This suggests that Liu et al.’s findings do not apply to visual recognition in general and that high SFO sensitivity is limited to face recognition. The reason for this difference between objects and faces is difficult to determine with certainty from the present research, but some explanations can be ruled out. For instance, an explanation of our findings based on differences in the ease of recognition of the stimuli in the two categories is not compatible with the data. If the chair images were simply harder to match, for instance, one would expect to see lower accuracy levels, but the same pattern of findings. That is, there should have been the same two-way interaction between SFO and trial type in the object condition as there was with faces, but there was not. Instead, we saw flat accuracy functions relative to control conditions for object recognition.

Another explanation for the differences in SFO sensitivity between objects and faces that is not supported by our findings is one based on differences in configural processing. Faces are known to be processed in a more configural and holistic manner than objects, and this difference might account for our SFO findings. That is, one might hypothesize that configural analysis requires more information in common between learned and tested images than feature-based recognition. This would explain why faces are more sensitive to SFO variations than objects. However, our results in Experiment 3 fail to support this explanation, because inverting faces—and thus reducing the degree of configural processing involved in their recognition—did not reduce their sensitivity to SFO variations. It should be noted that inverting faces does not entirely remove configural processing from the recognition process, and so, it is possible that the sensitivity of inverted faces to SFO variations results from residual configural processing in the recognition of these stimuli. However, direct statistical comparisons between the data with upright and inverted faces showed no difference in the effect of SFO. We would expect at least some diminishing of the SFO effect with inverted faces if SFO sensitivity was related to configural processing, but no such diminishing was evident.

Liu et al. (2000) suggested that the sensitivity of face recognition to reductions in SFO was due to the effects of nonoverlapping frequencies. They asserted that when there was spatial frequency information in a learned image that was not in a tested image (or vice versa), that information interfered with recognition. If that is the case, then the information in nonoverlapping frequencies in object images does not seem to interfere with recognition, at least not to the same degree. A possible explanation for this involves greater similarity of information at different spatial frequency bands in object images than in face images. That is, perhaps any two filtered versions of an object image are more similar to one another than equivalently filtered face images. However, because any two images that do not overlap in spatial frequency have a zero-order point-by-point correlation of luminance levels by definition, the differences in similarity across spatial bands would have to be at a higher order, perhaps featural, level. One reason to believe that object images might have a greater featural similarity across scale is that objects are better recognized by their sharp edges (Biederman, 1987; Biederman & Ju, 1988) than faces (Bruce et al., 1992). Sharp edges are composed of harmonics that transcend scale (Marr, 1982), meaning that elements of them are present in a wide range of spatial frequency bands. The visual system could exploit this characteristic of sharp edges in matching images with different spatial frequency content by using nonlinear analyses of spatial frequency information.
Our finding that objects are less sensitive to SFO variations than faces is compatible with Biederman and Kalocsai’s (1997) results showing that spatially complementary images of objects are more easily recognized than similarly filtered face images. However, our study differed from theirs in a number of important ways. Firstly, their stimuli were created such that the learned and tested images had spatial frequency elements from across the same spatial range. Our stimuli were created to have different spatial ranges, thus isolating the factor of spatial frequency from that of phase. Another key difference is that their study did not allow estimates of the quantitative effect of SFO because it used only two conditions: complete overlap and complete complementarity. By using varying levels of SFO, we were able to examine how intermediate levels of this factor affected recognition performance. Finally, we used a wider variety of objects than Biederman and Kalocsai, allowing for greater generalizability of findings.

A number of studies have shown that face recognition is poor when the image is presented in various nonphotographic image types, for example, line drawings (Bruce et al., 1992) and two-tone images (Hayes, 1988). In contrast, objects are well recognized in line drawings (Biederman, 1987; Biederman & Ju, 1988). Liu et al. (2000) suggested that the vulnerability of face recognition to changes in image format may be due to the lack of SFO between the learned images (typically photographs) and the tested images (e.g., line drawings). If so, then our results help explain the relative invulnerability of object recognition to these same changes in format. That is, because object recognition requires less SFO than face recognition, the low SFO levels introduced by changes in image format cause relatively little difficulty.

A number of studies have shown that a band of middle spatial frequencies, about 8–16 cycles per face width, is more useful for face recognition than spatial frequencies outside this range (Costen et al., 1994; Gold et al., 1999; Nasanen, 1999). However, Liu et al. (2000) did not find any advantage for the middle band when both learned and tested images were filtered in the same way (see also Boutet, Collin, & Faubert, 2003), although there were some effects of spatial frequency filtering whereby higher bands produced better recognition. They pointed out that studies finding a middle-band superiority tended to use unfiltered images at learning and filtered images at testing. It seems that the middle-band advantage arises only when the learned image is unfiltered. The results of the present experiment are further evidence of this. There was no advantage of any spatial frequency range in the control conditions of Experiments 1 and 3. Experiment 2 showed a slight increase in recognition rates as the bands moved up the spectrum, but not an advantage of the middle band. This suggests that the middle-frequency advantage for faces arises because of effects of interfering frequencies outside the range of the filtered test images (Liu et al., 2000). In this context, our finding that object recognition has little sensitivity to SFO may explain why there has been no similar middle-band advantage found for objects. That is, there seems to be less sensitivity of object recognition to spatial frequency variations overall, and this may be due to a lesser degree of interference from frequencies outside the range of the test image.

One issue that remains to be examined by future research in this domain concerns how different forms of face and object recognition are affected by spatial frequency manipulations. The studies that have shown a middle-band advantage for face recognition have primarily used a match-to-sample or sequential matching procedure, as we have here. The SFO hypothesis is particularly amenable to this kind of methodology, but it is not clear how it would apply to other methodologies, such as the recognition of famous faces. Further research is planned to examine this issue.

To summarize, we found that although face recognition is highly sensitive to variations in shared spatial frequency range between learned and tested images, object recognition is not. Our results with faces support and extend those of Liu et al. (2000), but our results with objects suggest that there are differences in how spatial frequency information is utilized in these two tasks. Our results are largely compatible with those of Biederman and Kalocsai (1997) in that they support the hypothesis that object recognition is less sensitive to spatial frequency variations and may be more feature based than face recognition, although it should be emphasized that we did not test the specific recognition-by-components model they espoused. The lesser effects of SFO on object recognition may be part of the reason why this task is less vulnerable to variations in image format than face recognition. The relative vulnerability of face recognition to SFO variations may in part explain the middle-frequency advantage in face recognition and the absence of such an advantage in object recognition.

References


Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Identification of band-

Received August 18, 2003
Revision received February 6, 2004
Accepted April 14, 2004

---

**Low Publication Prices for APA Members and Affiliates**

**Keeping you up-to-date.** All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor.* High School Teacher and International Affiliates receive subscriptions to the *APA Monitor,* and they may subscribe to the *American Psychologist* at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

**Essential resources.** APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the American Psychological Association,* and on dozens of new topical books each year.

**Other benefits of membership.** Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

**More information.** Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.