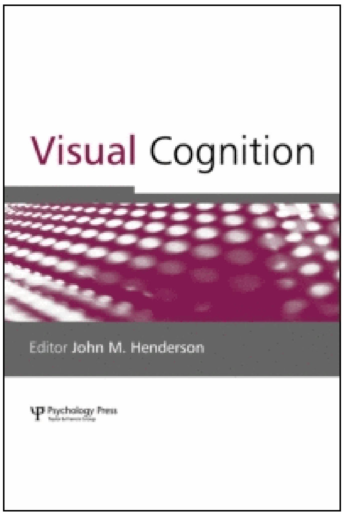

Visual Cognition
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683696

Learning influences the encoding of static and dynamic faces and their recognition across different spatial frequencies

Karin S. Pilz [a]; Heinrich H. Bülthoff [b]; Quoc C. Vuong [c]
[a] Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, ON, Canada [b] Max Planck Institute for Biological Cybernetics, Tübingen, Germany [c] Institute of Neuroscience & School of Psychology, Newcastle University, Newcastle upon Tyne, UK

First Published on: 01 September 2008

To cite this Article Pilz, Karin S., Bülthoff, Heinrich H. and Vuong, Quoc C.(2008)'Learning influences the encoding of static and dynamic faces and their recognition across different spatial frequencies',Visual Cognition,17:5,716 — 735

To link to this Article: DOI: 10.1080/13506280802340588

URL: http://dx.doi.org/10.1080/13506280802340588

VISUAL COGNITION, 2009, 17 (5), 716–735

Psychology Press
Taylor & Francis Group

Learning influences the encoding of static and dynamic faces and their recognition across different spatial frequencies

Karin S. Pilz

Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, ON, Canada

Heinrich H. Bülthoff

Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Quoc C. Vuong

Institute of Neuroscience & School of Psychology, Newcastle University, Newcastle upon Tyne, UK

Studies on face recognition have shown that observers are faster and more accurate at recognizing faces learned from dynamic sequences than those learned from static snapshots. Here, we investigated whether different learning procedures mediate the advantage for dynamic faces across different spatial frequencies. Observers learned two faces—one dynamic and one static—either in depth (Experiment 1) or using a more superficial learning procedure (Experiment 2). They had to search for the target faces in a subsequent visual search task. We used high-spatial frequency (HSF) and low-spatial frequency (LSF) filtered static faces during visual search to investigate whether the behavioural difference is based on encoding of different visual information for dynamically and statically learned faces. Such encoding differences may mediate the recognition of target faces in different spatial frequencies, as HSF may mediate featural face processing whereas LSF mediates configural processing. Our results show that the nature of the learning procedure alters how observers encode dynamic and static faces, and how they recognize those

Please address all correspondence to Karin S. Pilz, McMaster University, Department of Psychology, Neuroscience & Behaviour, 1280 Main Street West, Hamilton, Ontario L8S4K1, Canada. E-mail: pilzk@mcmaster.ca

This work was conducted while the first author was at the Max Planck Institute for Biological Cybernetics, Tübingen, Germany, and the Graduate School for Neural and Behavioural Sciences, Tübingen, Germany. We would like to thank Isabelle Bülthoff, Ian Thornton, and Christian Wallraven for helpful discussions on earlier drafts of the manuscript.

http://www.psypress.com/viscog DOI: 10.1080/13506280802340588

learned faces across different spatial frequencies. That is, these results point to a flexible usage of spatial frequencies tuned to the recognition task.

Keywords: Face recognition; Facial motion; Spatial frequency.

Face recognition is a remarkable human ability. We are not only able to judge the age or sex of another person but we can also recognize many individuals based on subtle differences in their facial configurations (for a review see Maurer, Le Grand, & Mondloch, 2002). This ability is especially impressive considering that, as objects, faces are highly similar. Several recent studies have shown that facial motion serves as a cue to person identification and that facial motion can lead to more robust encoding of information related to the identity of that person (e.g., Hill & Johnston, 2001; Lander & Bruce, 2003; Lander, Christie, & Bruce, 1999; Knight & Johnston, 1997; Lander & Chuang, 2005; Pike, Kemp, Towell, & Philips, 1997; Wallis & Bülthoff, 2001).

In our previous paper (Pilz, Thornton, & Bülthoff, 2006), we showed that nonrigid motion, the nonrigid deformations of a face when talking or making expressions, facilitates the encoding of unfamiliar faces using a detailed questionnaire during learning. We found that observers were faster at recognizing faces that had been learned from dynamic sequences than those that had been learned from static images in a visual search task. This dynamic advantage held across facial expressions and viewpoints, suggesting that the mental representation of dynamic faces is robust. As observers more naturally encounter faces as dynamic objects rather than static images, it is plausible to assume that the visual system has mechanisms that efficiently encode this common and behaviourally relevant dynamic aspect of faces.

But how does the dynamic advantage arise and how do representations derived from dynamic faces differ from representations derived from static images? Following previous work, a hypothesis we put forward is that observers develop representations for dynamically learned faces that are more robust against changes to viewpoint and expression. Due to the familiarity and behavioural relevance of facial motion, our visual system may have developed efficient mechanisms that facilitate encoding of dynamic over static facial information (Lander & Bruce, 2000, 2003; Pilz et al., 2006; Thornton & Kourtzi, 2002). Furthermore, this robust representation for dynamic faces may arise due to the encoding and representation of different visual information during learning. To assess possible differences between the mental representations for dynamically learned and those for statically learned faces, we tested whether observers show differences in recognizing the learned faces when these faces were degraded by different spatial filters.

Spatial filtering is thought to be an early step in visual processing. Luminance variability in the visual field, for example, is encoded by spatial filters. High spatial frequencies (HSF) represent fast variations in luminance and, therefore, emphasize fine details and edges. By comparison, low spatial frequencies (LSF) represent slow variations of luminance and emphasize the coarse cues (Morrison & Schyns, 2001).

Visual information contained in different spatial frequencies has long been of interest in face recognition as different spatial frequencies are linked to different face recognition mechanisms and to different task demands. First, observers seem to rely on a critical frequency band for efficient face recognition. For example, Ginsburg (1978) was one of the first to show that portraits filtered to contain the frequency band between 8 and 32 cycles/face were the most recognisable. Since then, various other studies tested the effects of spatial filtering on face recognition performance. The results of several experiments to date converge on the conclusion that face recognition depends on critical spatial-frequency information that is contained in a band centred between 8 and 17 cycles/face (Costen, Parker, & Craw, 1994, 1996; Fiorentini, Maffei, & Sandini, 1983; Hayes, Morrone, & Burr, 1986; Näsänen, 1999; Tieger & Ganz, 1979). Second, the featural processing of a face may involve HSF (Costen et al., 1996; Goffaux, Hault, Michel, Vuong, & Rossion, 2005; Sergent, 1986), whereas the configural processing of a face—the processing of the global spatial relations among these features—may involve LSF (Fiorentini et al., 1983; Goffaux et al., 2005; Sergent, 1986; Tieger & Ganz, 1979). Several studies support the view that observers generally process faces configurally rather than feature-wise (e.g., Yin, 1969; for reviews see Valentine, 1988, and Maurer et al., 2002). Third, there is evidence that the visual system processes coarse (i.e., LSF) image information faster than fine (i.e., HSF) image information (for reviews see Morrison & Schyns, 2001, and Snowden & Schyns, 2006). Lastly, there is evidence that other factors may affect which spatial frequencies observers use when they encounter faces. For example, there is some evidence that observers process unfamiliar faces differently from familiar faces. Observers seem to rely more on featural information when they decide on the identity of a less familiar face, whereas they seem to rely more on configural information when they identify a familiar face (Megreya & Burton, 2006; but see Schwaninger, Lobmeier, & Collishaw, 2002). Such reliance on different spatial information may also underlie the representations of dynamic and static faces. Consequently, these differences in mental representations may lead to differences in the speed and/or accuracy of recognizing HSF and LSF for both types of learned faces. Overall, these prior findings point to a flexible usage of spatial information (Morrison & Schyns, 2001; Oliva & Schyns, 1997; Schyns & Oliva, 1999).

In addition to investigating differences in the encoding and representation of spatial frequencies when learning dynamic and static faces, we tested whether the dynamic advantage is due to different depths of encoding dynamic and static faces. Facial motion conveys important information about the intentions and emotions of a person (Bassili, 1978; Kamachi et al., 2001). Such motion may not only help observers when assessing a person's personality, but may also affect how observers encode his or her identity. That is, the in-depth processing of the face induced by the expressive facial movements may lead to a more robust representation compared to a static picture. Therefore, we manipulated the learning procedure so that observers had to process both dynamic and static faces during learning either at an in-depth level or at a superficial level (Bower & Karlin, 1974).

In the current study, we tested for possible differences in recognizing high and low filtered images of faces that were learned dynamically or statically in full spectrum and in colour. We used a delayed visual search paradigm as described in our previous paper (Pilz et al., 2006), which showed a robust dynamic advantage for face recognition. In this task, observers learned two faces: One face was presented as a dynamic video sequence, the other one as a static picture. In a subsequent visual search task, observers had to indicate whether either one of the learned faces was present or not. We believe that such a visual search paradigm resembles the natural situation of "finding a friend in a crowd" and therefore provides a behaviourally relevant task. All faces presented in the search array were presented as static images to equate the perceptual and response characteristics of the task. In addition, the images were either high-pass or low-pass filtered so that only high and low spatial frequency information was available for recognition. In Experiment 1, observers explicitly judged the personality and character traits of a statically and a dynamically presented face during learning. In Experiment 2, we investigate how the depth of learning, i.e., the observers' engagement in the learning task, affects the dynamic advantage, as well as their response characteristics to HSF and LSF images. Therefore, we used the same visual search procedure as in Experiment 1 but changed the learning procedure.

GENERAL METHODS

Apparatus

Both experiments were conducted on a Windows computer under the control of the PsychToolBox extension for MATLAB (Brainard, 1997; Pelli, 1997). Stimuli were presented on a 21-inch monitor with a resolution of 1024 pixels × 768 pixels and a frame rate of 75 Hz. Observers were seated 60 cm from the screen.

Participants

For both experiments, students from the MPI subject pool served as observers. All observers gave informed written consent and were paid 8€/hour. They had normal or corrected-to-normal vision and were naïve regarding the purpose of the study. Observers did not participate in more than one experiment.

Stimuli

In the current study, we used stimuli from the Max-Planck database of moving faces (Pilz et al., 2006; http://edb.kyb.tuebingen.mpg.de/). We chose five female and five male faces. Each face made two expressive gestures, which were surprise and anger. Two of the faces were randomly chosen as targets for each observer; the others served as distractor faces in the visual search task. One target face was female and the other was male. All observers saw the faces in surprise during learning and in anger during test. We chose different expressions to test whether the dynamic advantage generalizes to other facial expressions and to reduce the possibility of image matching during visual search. In addition, we did not counterbalance the expressions across the learning and search phases to stay in line with our previous work (Pilz et al., 2006).

Figure 1 shows an example of the last frame of two video sequences. The movie clips were 26 frames and presented at a frame rate of 25 frames per second for a total duration of 1040 ms. The clips started with a neutral expression and ended with the peak of the expression in the last frame. The stimuli presented during the learning phase subtended a visual angle of

Figure 1. Examples of the last frame (i.e., peak magnitude of the facial expression) of video sequences of the two expressive gestures used in the current study (left: anger, right: surprise). To view this figure in colour, please see the online issue of the Journal.

$15.8^\circ \times 15.4^\circ$, and the stimuli presented during the visual search phase subtended a visual angle of $6.0^\circ \times 5.8^\circ$. The radius of the visual search array subtended an angle of 18.8° with the midpoint of the images presented at equal distance from the fixation cross in the middle of the screen.

The static pictures shown during learning in both experiments were static pictures randomly selected from the video sequence on each presentation. Therefore, observers saw all the frames of the video sequence in both learning conditions. In addition, static pictures during the visual search phase were shown from three different viewpoints: Frontal (0°, as in the learning phase), facing 22° left, and facing 22° right. Similar to changing expressions, the purpose of this viewpoint manipulation was to test whether the dynamic advantage generalizes across different viewpoint and to reduce image-matching strategies (Pilz et al., 2006).

The static pictures in the visual search test phase always showed the peak of the expression, i.e., the last frame of the video sequence, and were spatially filtered faces. The two filtered versions of these images were created as follows. The original images were converted to greyscale images before they were Fourier transformed into the frequency domain. These Fourier transformed images were then multiplied with a high-pass and low-pass Gaussian filters with cutoff frequencies of 8 cycles per image width (c/iw) for the low frequency filter, which preserved frequencies below 8 c/iw, and 32 c/iw for the high frequency filter, which preserved frequencies above 32 c/iw.[1] Prior to filtering, all images were equated for contrast by equalizing the luminance histogram across the set of images. The filters are shown in Figure 2, and example images of filtered faces and a visual search array of LSF faces are shown in Figure 3.

Analysis

In both experiments, we examined the speed and accuracy of responses. For response times (RTs), we only analysed correct target-present trials (i.e., trials on which one of the learned target face was present). RTs and accuracy were separated into two blocks in both experiments to investigate effects of learning during the course of the visual search phase. Table 1 shows correct RTs and accuracy across all conditions in both experiments.

[1] Most previous studies used cycles per face width as they normalize faces to the same size and crop the images to that size. However, our faces were not normalized and some faces covered more or less area of the image. Furthermore, the viewpoint manipulation also changed the area of the image covered by the face (but not to a large extent). Thus, we have used cycles per image width for our filtering. For the purpose of the present study, the important point is that our filtered contained only high or low spatial frequencies (see Figure 3)—the precise frequencies were not critical to our conclusions.

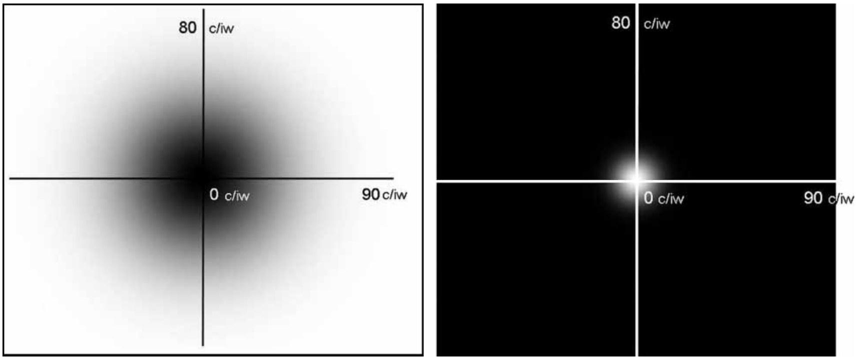

Figure 2. Example images of the Gaussian filters applied to faces, given in cycles per image width (c/iw). The left image illustrates the high-pass filter and the right image illustrates the low-pass filter. Lighter greyscale intensity indicates the frequencies that are preserved in the filtered faces.

We excluded observers who had less than 75% correct trials from the analyses. Because of the spatial degradation, the search task was very difficult. We therefore used a high performance threshold to ensure that observers had sufficiently encoded both dynamic and static faces. For both experiments, repeated measures analyses of variance (ANOVAs) were used to compare the factors of interest: Target type (dynamically learned face and statically learned face), filter (high spatial filter and low spatial filter), block (1 and 2), viewpoint (0°, 22° left, and 22° right), and set size (2, 4, and 6). A significance level of $p < .05$ was used in all analyses.

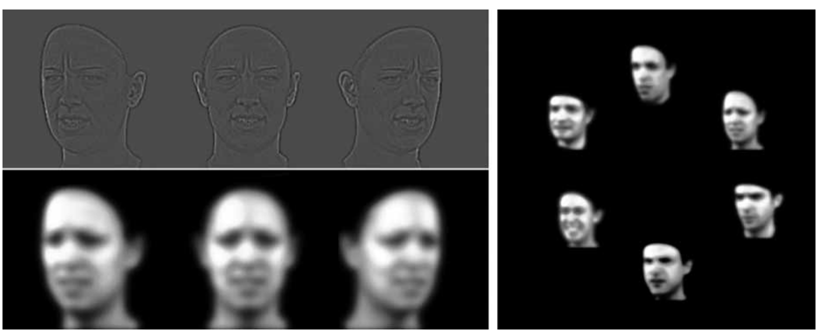

Figure 3. Left: Examples of the frequency filtered target images as used in the current experiments. Top row: High-pass filtered faces from three different viewpoints (from left to right: 22° from the left, frontal and 22° from the right). Bottom row: Low-pass filtered faces from the same three viewpoints. Right: Example picture of a visual search array with six LSF faces shown at 22° from the left.

TABLE 1
RT (ms) and accuracy (percentage correct) for dynamic, static, and absent trials for high spatial frequency (HSF) and low spatial frequency (LSF) conditions in Experiments 1 and 2 (the standard error of the means is in parentheses)

	Dynamic		*Static*		*Absent*	
	LSF	*HSF*	*LSF*	*HSF*	*LSF*	*HSF*
Experiment 1						
Accuracy						
Block 1	95.1 (2.6)	95.4 (2.2)	88.3 (5.3)	88.6 (1.8)	88.2 (5.2)	87.9 (4.9)
Block 2	96.0 (2.6)	96.9 (2.1)	92.0 (5.0)	94.4 (4.1)	95.8 (2.8)	94.4 (4.2)
RT						
Block 1	1755 (177)	1635 (142)	2149 (245)	1893 (182)	2898 (324)	2911 (331)
Block 2	1458 (126)	1297 (88)	1773 (186)	1672 (178)	2178 (158)	2171 (156)
Experiment 2						
Accuracy						
Block 1	97.3 (1.8)	85.1 (5.9)	93.1 (5.7)	86.6 (5.9)	82.2 (6.8)	77.6 (6.6)
Block 2	98.7 (1.0)	97.8 (1.9)	95.7 (3.0)	96.2 (3.1)	97.8 (2.0)	95.6 (3.1)
RT						
Block 1	2054 (239)	2161 (241)	2328 (406)	2018 (273)	3273 (394)	3178 (324)
Block 2	1492 (107)	1406 (133)	1814 (197)	1623 (141)	2269 (172)	2182 (137)

EXPERIMENT 1

In Experiment 1, we investigated whether we can replicate the dynamic advantage found by Pilz and colleagues (2006) with spatially degraded face images. More importantly, we tested whether observers in Experiment 1 show differences in the recognition of HSF or LSF targets for statically and dynamically encoded faces. Such differences may hint at differences in the robustness of the mental representations.

Methods

Participants. Fifteen naive observers (age range: 20–31; mean age: 25; 7 females, 9 males) with normal or corrected-to-normal vision participated in the experiment. Five observers were excluded from the data analysis since their performance was below 75%. The mean performance of observers who did not make criterion was 68.0%.

Design and procedure. We used the delayed visual search paradigm described by Pilz et al. (2006). Briefly, observers were familiarized with one male and one female frontal view face showing an expressive gesture of surprise. One face was presented as a short movie clip (dynamically learned target) and the other as a static picture (statically learned target). For each observer, the male and female targets were randomly assigned to the dynamic or static presentation condition. Both target faces were presented in full spatial-frequency spectrum and in colour. The two faces alternated 100 times on the screen, each time presented for 1040 ms with an interstimulus interval of 2000 ms. While watching the faces, observers filled out a questionnaire. They were asked to rate the attractiveness, age, kindness, aggressiveness, and intelligence of the two faces and had to describe the faces' prominent facial features and character traits in a few sentences.

After the learning phase, observers took a short break of approximately 3 min before continuing with the visual search phase. On each trial of this phase, two, four, or six static faces expressing anger were shown in a circular search array. Faces in the search array were either high-pass or low-pass filtered, and they were shown at 22° left, 22° right or frontal (0°) view (see Figure 3). Observers were asked to decide as quickly and accurately as possible whether either one of the learned faces was present in the search array or not. Observers responded "target present" by pressing the "s" key and "target absent" by pressing the "l" key. Auditory feedback was given for incorrect responses. Each trial started automatically after a response was given. The experiment consisted of 540 trials, in which each of the two target faces was present on 180 trials (hence, 360 target-present trials). In the remaining 180 trials, no target was presented (target-absent trials). All target type, filter, viewpoint, and set size trials occurred with equal frequency, and the order of the trials was randomized for each observer.

Results

Reaction time. The right column of Figure 4 and Table 1 show mean RTs over all observers. A repeated measures ANOVA revealed a main effect of target type: Observers were faster at recognizing dynamically learned targets than statically learned ones, $F(1, 9) = 9.8$, $p = .012$. There was also a main effect of filter: Observers were faster for LSF targets than HSF targets, $F(1, 9) = 8.6$, $p = .016$. In addition, observers were faster at finding targets in block 2 as compared to block 1, $F(1, 9) = 9.0$, $p = .015$. Lastly, there was a set size effect, $F(2, 18) = 78.4$, $p < .001$. There were no other main effects or interactions.

Accuracy. The left column of Figure 4 and Table 1 show mean accuracy data over all observers. A repeated measures ANOVA revealed a marginal effect of target type, $F(1, 9) = 3.9$, $p = .081$, and a main effect of block, $F(1, 9) = 5.1$, $p = .05$. However, these effects were modulated by a target type $\times$ block interaction, $F(1, 9) = 7.2$, $p = .025$. This interaction was due to observers' performance being worse for statically learned than for dynamically learned faces in block 1. This difference vanished in block 2. Therefore,

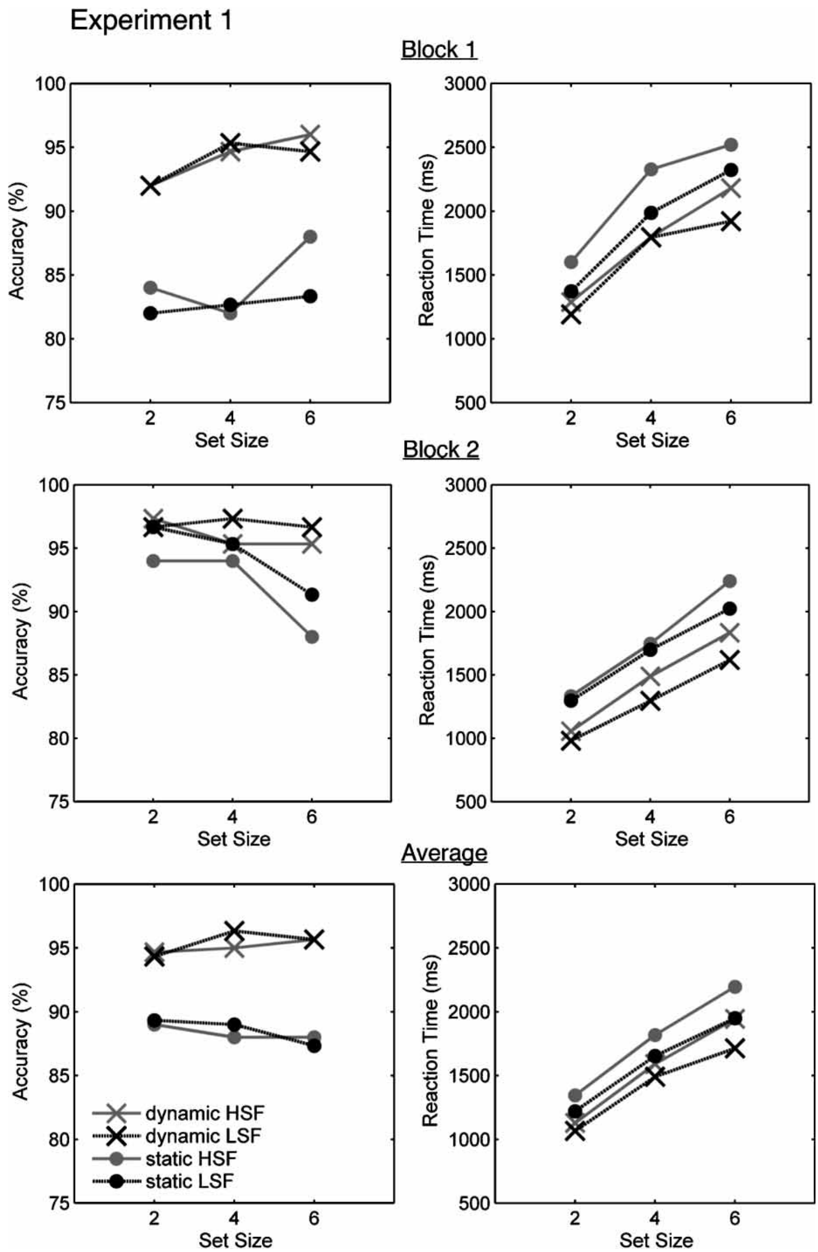

Figure 4. Accuracy (left) and RT (right) data from Experiment 1 for block 1 (top row), block 2 (middle row), and average (bottom row).

the rate of improvement across blocks was higher for the statically learned face. There were no indications of any speed–accuracy tradeoffs. In addition, as shown in Table 1, there were no effects of interest on absent trials in both Experiments 1 and 2, so these will not be discussed in further detail.

Discussion

The results from Experiment 1 showed an RT advantage for dynamically learned faces, which replicated the visual search advantage for dynamically learned faces found by Pilz et al. (2006). Observers were about 370 ms faster at finding the dynamically learned faces than they were at finding the statically learned faces in the search array. In addition, observers were faster at finding LSF than HSF targets in the search array regardless of whether faces were presented statically or dynamically during learning. This RT advantage for LSF faces suggests a configural processing of both target faces (Goffaux & Rossion, 2006). This advantage can also be explained by a coarse-to-fine bias for low spatial frequencies. Several studies have shown that observers process low spatial frequencies, i.e., the coarse details of an image, preferentially to high spatial frequencies, i.e., the fine details of an image (for reviews see Morrison & Schyns, 2001, and Snowden & Schyns, 2006). The set size effect in RT is typical for visual search tasks and shows that observers are slower at finding targets when more distractors are present in the visual search array (for a review see Wolfe, 1998).

In accuracy, observers showed a significant improvement for statically learned faces across block 1 and block 2. From this finding we infer that dynamic faces are encoded more robustly than static faces, for which the process of learning still seemed to occur during visual search. Interestingly, in our previous study, we did not find accuracy differences between dynamically and statically learned faces when these were tested in full spectrum (Pilz et al., 2006). However, the stimuli were degraded in this study which may have given rise to additional learning for statically but not dynamically learned targets during visual search. Furthermore, the feedback observers received may have facilitated additional learning of the degraded stimuli. Consistent with these possibilities, we found that there was no significant accuracy difference between dynamically and statically learned targets in block 2.

The present results suggest that the dynamic advantage is not driven by differences in the representation of spatial frequencies during the learning phase when observers processed faces in depth (Bower & Karlin, 1974), as observers were faster at searching for the LSF faces for both target types. Rather, the advantage may be due to a more robust representation of the dynamically learned face, as suggested by the accuracy data. Facial motion conveys important information about nonverbal communicational information, as well as a person's emotional state (Bassili, 1978; Kamachi et al., 2001). This motion may lead to a more robust representation. Thus, choosing a learning procedure that requires observers to explicitly judge a person's personality, a task observers usually do when encountering an unknown person in everyday life, may facilitate encoding of dynamically presented faces. For example, Bower and Karlin showed that faces judged for the deep characteristics of personality traits were recognized more accurately than faces judged for surface characteristics of sex. Thus, in Experiment 2, we used a learning procedure that only required observers to judge the faces superficially according to predefined traits on an online scale to investigate whether this would modulate the effect.

EXPERIMENT 2

To directly investigate the effect of learning on the dynamic advantage and how learning impacts the processing of HSF and LSF faces, we used a learning procedure that was less demanding to the observers but still engaged their attention on both dynamic and static target faces. Therefore, instead of answering a questionnaire on facial features and character traits, observers in Experiment 2 judged the faces according to predefined personality traits using an online rating scale (Bower & Karlin, 1974).

Methods

Participants. Fifteen naive observers (age range: 25–37; mean age: 28; 10 females, 5 males) participated in the experiment. Five observers were excluded from the data analysis since their performance was below 75%. The mean performance of observers who did not make criterion was 67.7%.

Design and procedure. The visual search phase was the same as the search phase described in Experiment 1. The only difference between the two experiments was in the learning phase. In Experiment 2, observers did not receive an in-depth questionnaire for the two faces during the learning phase but simply had to rate the faces according to age, intelligence, friendliness, extraversion, attractiveness, aggressiveness, nervousness, and happiness on an online scale from 1 to 5. Before a face appeared on the screen, observers were informed of which trait they would have to rate on that trial. Then the dynamic face appeared eight times on the screen. Afterwards a scale appeared on the screen, which observers used to rate the given character trait. After rating the dynamic face on one of the character traits, the static face appeared eight times on the screen and had to be rated by the same procedure. Eight different frames were shown for the static face randomly taken from the dynamic sequence. Each frame was shown for 1040 ms.

Results

Reaction times. The right column of Figure 5 and Table 1 show mean RT for all observers. In contrast to Experiment 1, there were no main effects of target type (dynamically or statically learned faces) or filter (high or low SFs). A repeated measures ANOVA revealed a main effect of block: Observers were faster at identifying targets in block 2 than block 1, $F(1, 9) = 15.3$, $p = .003$. There was also a main effect of set size, $F(2, 18) = 167.0$, $p < .001$, and a block × set size interaction, $F(2, 18) = 7.0$, $p = .007$. There were no other effects on RTs.

Accuracy. The left column of Figure 5 and Table 1 show accuracy data averaged over all observers. A repeated measures ANOVA revealed a main effect of filter, $F(1, 9) = 10.0$, $p = .011$, and block, $F(1, 9) = 14.3$, $p = .04$. However, these main effects were modulated by a filter × block interaction, $F(2, 18) = 10.0$, $p < .001$. This significant interaction was due to a larger filter effect in block 1 than block 2.

Also, in contrast to Experiment 1, there was a main effect of viewpoint, $F(1,9) = 3.6$, $p < .07$, and a block × target type × viewpoint interaction, $F(2, 18) = 4.0$, $p = .036$. This finding points to a less robust representation as search accuracy did not generalize to novel viewpoints. There were no indications of speed–accuracy tradeoffs.

Discussion

In Experiment 2, dynamically and statically learned faces were recognized equally quickly during visual search after observers superficially rated the target faces' personality during a learning phase.

We also did not find a difference between the dynamically and statically learned faces in recognition accuracy. However, the accuracy data showed an effect of filter, which varied across blocks. These effects were due to observers being more accurate at identifying HSF than LSF targets on block 1. This HSF advantage nearly vanished on block 2, which suggests that observers learned to identify the LSF targets as accurately as HSF faces by block 2. There are two possible reasons for these filter effects. First, observers may rely more on high-frequency featural processing for both target faces (Costen et al., 1996; Goffaux et al., 2005; Sergent, 1986). The learning procedure used in this experiment may encourage encoding strategies that do not exploit the dynamic aspects of faces during learning (Bassili, 1978; Kamachi et al., 2001). Second, the results may be due to coarse-to-fine processing strategies during

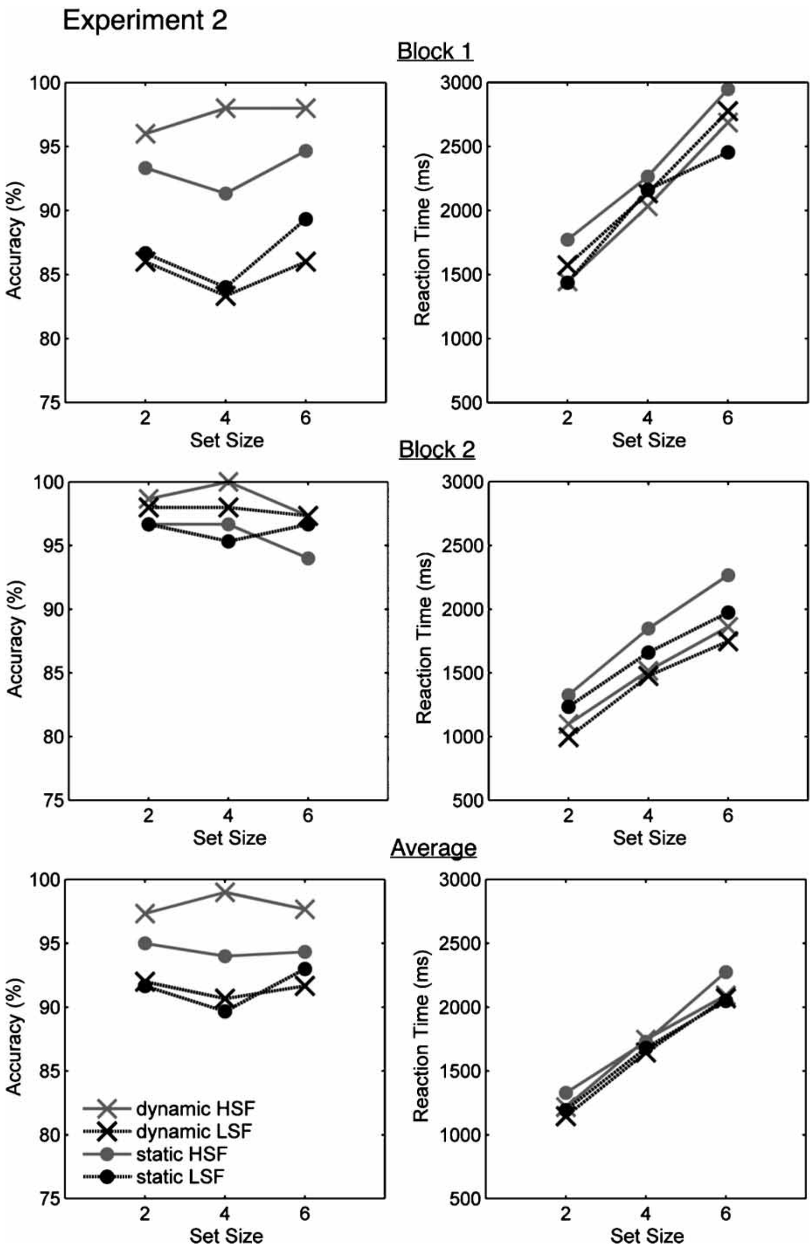

Figure 5. Accuracy (left) and RT (right) data from Experiment 2 for block 1 (top row), block 2 (middle row), and average (bottom row).

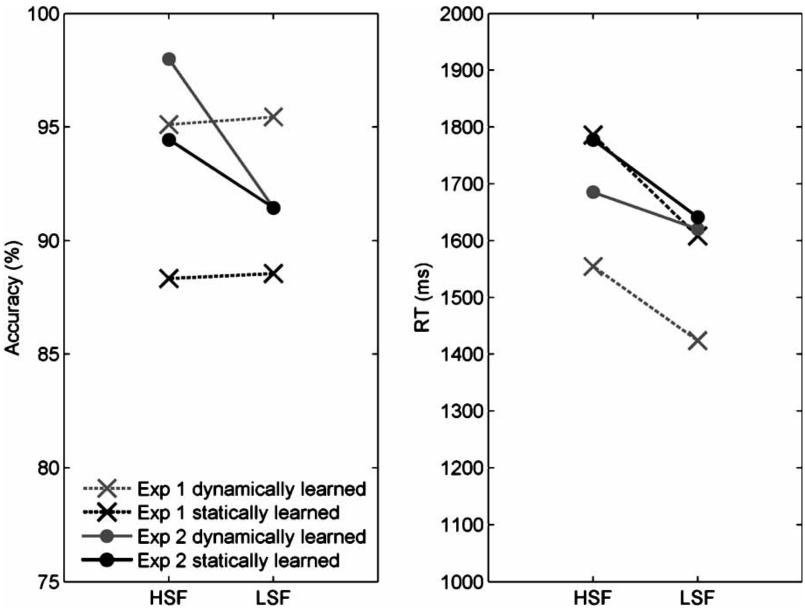

Figure 6. Accuracy (left) and RT (right) data from Experiments 1 and 2 for high-spatial frequency (HSF) and low-spatial frequency (LSF) targets, and for dynamically and statically learned faces.

the recognition stage. Previous results from a number of psychophysical studies have suggested that our visual system has a coarse-to-fine bias in the processing of sinusoidal gratings (Breitmeyer, 1975; Gish, Shulman, Sheehy, & Leibowitz, 1986) as well as natural scenes and faces (Parker, Lishman, & Hughes, 1992, 1997; Schyns & Oliva, 1994, 1999). Because the superficial learning procedure used in Experiment 2 may lead to less robust face representations for both dynamic and static targets compared to the more in-depth learning procedure used in Experiment 1, observers may generally need more information to process both target faces in Experiment 1. Therefore, they may be better with HSF than LSF faces because HSF faces contain more visual information.

GENERAL DISCUSSION

In two experiments, we investigated whether differences in the encoding of dynamically and statically presented faces lead to differences in recognizing HSF and LSF target faces in a delayed visual search task (Pilz et al., 2006). In addition, we examined how different learning procedures affected the encoding process and the subsequent recognition of spatially degraded faces.

To summarize, the two main results from these experiments are as follows. First, observers showed an overall RT advantage for dynamically learned faces in Experiment 1 but not in Experiment 2 (Pilz et al., 2006). Second, observers in Experiment 1 showed an RT advantage for LSF targets, whereas in Experiment 2 they show an accuracy advantage for HSF targets. Figure 6 provides an overview of these two results to allow a direct comparison of the effects of dynamically and statically learned target faces and spatial filter.

Interestingly, observers continued to learn targets across the visual search blocks. In Experiment 1, this improvement was found for statically learned but not dynamically learned faces. In Experiment 2, this improvement across block was found for both dynamically learned and statically learned faces and was dependent on the spatial frequency. These noticeable block effects may be due to the fact that we degraded the stimuli, which made it possible for observers to encode additional information during visual search. The additional learning may also have been enhanced by the feedback observers received.

Overall, given that identical visual search tasks were used in both experiments, the only explanation that can account for these results lies in the learning procedure used. That is, these different procedures lead to different encoded mental representations for dynamically and statically learned faces in Experiments 1 and 2.

Bower and Karlin (1974) showed that faces judged for the deep characteristics of personality traits were recognized more accurately than faces that were only judged for surface characteristics of sex. The questionnaire used in Experiment 1 required observers to explicitly write down their impressions of the targets' personality and facial features in their own words. By comparison, observers in Experiment 2 only had to tick marks on an online scale to judge the faces according to predefined character traits. Using a questionnaire that requires observers to explicitly assess personality, which is a judgement observers frequently make when encountering unfamiliar people, may have induced observers to adopt a strategy that was more sensitive to study the role of motion in face recognition. Moreover, observers in Experiment 1 may have encoded the dynamically learned face more robustly than the statically learned one because the visual system may integrate the facial expressions over time into a representation of the facial identity for the dynamic target. This integration may account for the large RT advantage for dynamically learned faces in Experiment 1 (see also Pilz et al., 2006). Such integration was absent with a more superficial learning procedure used in Experiment 2. Future research is needed to investigate how the learning context affects the temporal integration of behaviourally important facial motion for recognition purposes.

Our results further suggest that the nature of the representation encoded for the different learning procedure (in-depth vs. superficial) and different target types (dynamically learned vs. statically learned) lead to processing strategies that exploit different spatial frequencies during recognition. It is evident that different spatial frequencies convey different kinds of information. HSF information is known to mediate the processing of fine details such as bars and edges. By comparison, LSF information is known to mediate the processing of coarse cues like pigmentation and shape-from shading information (for reviews see Morrison & Schyns, 2001, and Snowden & Schyns, 2006). Importantly, several studies have suggested a coarse-to-fine processing bias bias for the extraction of different spatial frequency information. This bias was first demonstrated in studies that showed that the time to detect an onset or offset of a sinusoidal grating increases monotonically with spatial frequency (Gish et al., 1986). Further experiments were able to extend the coarse-to-fine processing bias to natural scenes. In an experiment by Parker et al. (1992), for example, observers had to rate the image quality of natural scenes. Those scenes were presented in sequences of three filtered versions over an interval of 120 ms. Performance was significantly better when the order of spatial information in a sequence moved from coarse to fine detail than when the order moved from fine to coarse. The results of our study provide direct evidence that the nature of the representation can lead to different recognition strategies using a single paradigm. The RT results from Experiment 1, which showed that observers are faster at identifying LSF targets for both dynamic and static targets, may be explained by coarse-to-fine processing. By comparison, in Experiment 2, this effect may vanish because observers did not learn the faces robustly enough and need more "fine" featural information to identify faces.

The depth of encoding across the two experiments may also lead to different strategies in terms of the facial information that observers use during visual search. In particular, observers may use configural face recognition strategies in Experiment 1, whereas they may use featural strategies in Experiment 2. Goffaux and colleagues (Goffaux et al., 2005; Goffaux & Rossion, 2006) suggested that the configural processing of faces is mediated by LSF, whereas featural processing is mediated by HSF. According to this assumption, the results from Experiment 1 suggest that observers recognize faces configurally rather than featurally, as they are faster at recognizing LSF than HSF targets. But why is this bias for configural face processing absent in Experiment 2? It has been suggested that the relationship between featural/configural and HSF/LSF processing of faces is far from unequivocal (Palmer, 1993). For instance, configural information can be conveyed by both LSF and HSF (Peyrin, Chauvin, Chokron, & Marendaz, 2003). Indeed, early studies testing for featural and configural face processes used line drawings of faces which contain predominantly high spatial frequencies (e.g., Tanaka & Farah, 1993). Again, future work is necessary to determine how learning interacts with subsequent configural and featural processing of dynamic and static faces.

More generally, HSF faces may contain more or additional information than the information contained in LSF images (e.g., HSF faces have both configural and featural information). In Experiment 2, observers may have encoded a less robust representation of the target faces because of the superficial learning procedure used. Therefore, they may need more information to recognize target faces during visual search, and consequently, they were more accurate with HSF target faces but did not show an RT advantage for LSF targets. In a related paper, Megreya and Burton (2006) suggested that less familiar faces are represented more featurally, whereas familiar faces are represented more configurally. This dissociation may be due to the fact that HSF faces contain both featural and configural information and, therefore, observers are better able to recognize HSF faces when they are less familiar with target faces. As noted earlier, however, other studies have shown that observers may rely on both configural and featural information to process familiar and unfamiliar faces (e.g., Schwaninger et al., 2002). Our present findings help resolve some of these earlier contradictory findings. In conjunction with previous work, the present results support a flexible usage hypothesis in which task demands bias the recognition process to operate at the most informative scale for the task at hand (Morrison & Schyns, 2001; Oliva & Schyns, 1997; Schyns & Oliva, 1999).

The results of the current study have two major implications for face recognition. First, we replicated the dynamic advantage found by Pilz et al. (2006). This dynamic advantage emphasizes the hypothesis that the visual system uses mechanisms that mediate the processing of behaviourally relevant dynamic information, such as facial expressions and movements. Second, we found that different learning procedures may tune these mechanisms to those spatial frequency scales that convey the most relevant information to solve a certain task. Moreover, these mechanisms may be affected by additional factors such as the depth of encoding during learning, rather than facial motion per se, which ultimately allows the visual system to flexibly optimize visual spatial information for face recognition in a dynamic environment.

REFERENCES

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *Journal Experimental Psychology: Human Perception and Performance*, *4*, 373–9.

Bower, G. H., & Karlin, M. B. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology*, *103*, 751–757.

Brainard, D. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Breitmeyer, B. G. (1975). Simple reaction time as a measure of the temporal response properties of transient and sustained channels. *Vision Research*, *15*, 1411–1412.

Costen, N. P., Parker, D. M., & Craw, I. (1994). Spatial content and spatial quantization effects in face recognition. *Perception*, *23*, 129–146.

Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception and Psychophysics*, *58*, 602–612.

Fiorentini, A., Maffei, L., & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception*, *12*, 195–201.

Ginsburg, A. P. (1978). *Visual information processing based on spatial filters constrained by biological data*. Unpublished doctoral dissertation, Cambridge University, Cambridge, UK.

Gish, K., Shulman, G. L., Sheehy, J. B., & Leibowitz, H. W. (1986). Reaction times to different spatial frequencies as a function of detectability. *Vision Research*, *26*, 745–747.

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception*, *34*, 77–86.

Goffaux, V., & Rossion, B. (2006). Faces are "spatial": Holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1023–1039.

Hayes, A., Morrone, M. C., & Burr, D. C. (1986). Recognition of positive and negative band-pass filtered images. *Perception*, *15*, 595–602.

Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology*, *11*, 880–887.

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perception*, *30*, 875–87.

Knight, B., & Johnson, A. (1997). The role of movement in face recognition. *Visual Cognition*, *4*, 265–273.

Lander, K., & Bruce, V. (2000). Recognizing famous faces: Exploring the benefits of facial motion. *Ecological Psychology*, *12*, 259–272.

Lander, K., & Bruce, V. (2003). The role of motion in learning new faces. *Visual Cognition*, *10*, 897–921.

Lander, K., Christie, F., & Bruce, V. (1999). The role of movement in the recognition of famous faces. *Memory and Cognition*, *27*, 974–985.

Lander, K., & Chuang, L. (2005). Why are moving faces easier to recognize? *Visual Cognition*, *12*(3), 429–442.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Science*, *6*(6), 255–260.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory and Cognition*, *34*(4), 865–876.

Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects and scenes. *Psychonomic Bulletin and Review*, *8*, 454–469.

Näsänen, R. (1999). Spatial-frequency bandwidth used in the recognition of facial images. *Vision Research*, *39*, 3824–3833.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, *34*, 72–107.

Palmer, S. (1993). Modern theories of gestalt perception. In G. W. Humphreys (Ed.), *Understanding vision* (pp. 39–70). Oxford, UK: Blackwell.

Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, *21*, 147–160.

Parker, D. M., Lishman, J. R., & Hughes, J. (1997). Evidence for the view that temporospatial integration in vision is temporally anisotropic. *Perception*, *26*, 1169–1180.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Peyrin, C., Chauvin, A., Chokron, S., & Marendaz, C. (2003). Hemispheric specialization for spatial frequency processing in the analysis of natural scenes. *Brain and Cognition*, *53*, 278–282.

Pike, G. E., Kemp, R. I., Towell, N. A., & Phillips, K. C. (1997). Recognizing moving faces: The relative contribution of motion and perspective view information. *Visual Cognition*, *4*, 409–437.

Pilz, K. S., Thornton, I. M., & Bülthoff, H. H. (2006). A search advantage for faces learned in motion. *Experimental Brain Research*, *171*(4), 436–447.

Schwaninger, A., Lobmeier, J. S., & Collishaw, S. M. (2002). The role of featural and configural information in familiar and unfamiliar face recognition. *Lecture Notes in Computer Science*, *2525*, 643–650.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time and spatial-scale-dependent scene recognition. *Psychological Science*, *5*, 195–200.

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, *69*, 243–265.

Sergent, J. (1986). Microgenesis of face perception. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A.M. Young (Eds.), *Aspects of face processing* (pp. 17–33). Dordrecht, The Netherlands: Martinus Nijhoff.

Snowden, P. T., & Schyns, P. G. (2006). Channel surfing in the visual brain. *Trends in Cognitive Science*, *10*(12), 538–545.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, *46*(4), 225–245.

Thornton, I. M., & Kourtzi, Z. (2002). A matching advantage for dynamic faces. *Perception*, *31*, 113–132.

Tieger, T., & Ganz, L. (1979). Recognition of faces in the presence of two-dimensional sinusoidal masks. *Perception and Psychophysics*, *26*, 163–167.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology*, *79*, 471–491.

Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences*, *98*, 4800–4804.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–73). Hove, UK: Psychology Press.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141–145.

Manuscript received August 2007
Manuscript accepted June 2008
First published online September 2008