

Coarse-to-Fine Encoding of Spatial Frequency Information Into Visual Short-Term Memory for Faces but Impartial Decay

Zaifeng Gao

Zhejiang University, Hangzhou, People's Republic of China,
and The Hebrew University of Jerusalem, Israel

Shlomo Bentin

The Hebrew University of Jerusalem, Israel

Face perception studies investigated how spatial frequencies (SF) are extracted from retinal display while forming a perceptual representation, or their selective use during task-imposed categorization. Here we focused on the order of encoding low-spatial frequencies (LSF) and high-spatial frequencies (HSF) from perceptual representations into visual short-term memory (VSTM). We also investigated whether different SF-ranges decay from VSTM at different rates during a study-test stimulus-onset asynchrony. An old/new VSTM paradigm was used in which two broadband faces formed the positive set and the probes preserved either low or high SF ranges. Exposure time of 500 ms was sufficient to encode both HSF and LSF in the perceptual representation (experiment 1). Nevertheless, when the positive-set was exposed for 500 ms, LSF-probes were better recognized in VSTM compared with HSF-probes; this effect vanished at 800-ms exposure time (experiment 2). Backward masking the positive set exposed for 800 ms re-established the LSF-probes advantage (experiment 3). The speed of decay up to 10 seconds was similar for LSF- and HSF-probes (experiment 4). These results indicate that LSF are extracted and consolidated into VSTM faster than HSF, supporting a coarse-to-fine order, while the decay from VSTM is not governed by SF.

Keywords: visual short-term memory, faces, spatial frequency, encoding, decay

Supplemental materials: <http://dx.doi.org/10.1037/a0023091.supp>

Working memory (WM) enables the on-line active maintenance and manipulation of a limited amount of information (Baddeley & Hitch, 1974). It is a critical component in almost any cognitive ability, such as visual and auditory perception, language processing, planning, reasoning, and the perception of motion (Baddeley, 2003; Hollingworth, Richard, & Luck, 2008; Wickelgren, 1997; Wood, 2007). Starting with the seminal work of Baddeley and his colleagues (e.g., Baddeley, 1981; 1983) and along with his model, it is well accepted that verbal and visual-spatial information rely on different WM components, the articulatory loop and the visuospatial sketch pad, which is conventionally labeled visual short-term memory (VSTM) (Phillips, 1974), respectively. Although the functional characteristics of both components have been extensively investigated (e.g., Gathercole & Baddeley, 1993; Logie, 1995), the principles governing the accumulation, maintenance,

and decay of visuospatial information in the VSTM are relatively less well understood (Baddeley, 1996, 2003; Jiang, Makovski, & Shim, 2009; Jonides et al., 2008; Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001; for a comprehensive review and state-of-the-art refer Luck, 2008). In the present study, we investigated the transfer of different spatial frequency (SF) ranges from visual representation to the temporarily maintained active storage in WM, focusing on the order at which relatively low and relatively high SF ranges are accumulated into the VSTM representation. In addition, we investigated whether different SF ranges decay at different pace while the visual representation fades from VSTM.

All visual images are composed of a range of SFs and our perceptual system analyzes the visual input via a number of SF channels (DeValois & DeValois, 1988; Ivri & Robertson, 1998; Westheimer, 2001). Indeed, spatial filtering is probably the most basic mechanism for extracting visual information from the retina (Ginsburg, 1986; Legge & Gu, 1989; Morgan, 1992). Therefore, many studies of visual object recognition and categorization addressed the question of how visual information is embedded in different SF ranges and how are SFs used during object recognition (e.g., Loftus & Harley, 2004, 2005; Schyns & Gosselin, 2003; Sowden & Schyns, 2006). Not surprisingly, these studies showed that different SF scales convey different types of information about objects: High spatial frequencies (HSF), which are fast luminance transitions in the image, convey the sharp, fine-scale details of the object; low spatial frequencies (LSF), which are slow (large-scale) luminance transitions in the image, convey the coarse, low-resolution aspects of the object (Schyns, 1998). In addition, once integrated in a perceptual representation, the visual system can use

Zaifeng Gao, Department of Psychology, Zhejiang University, Hangzhou, People's Republic of China and Department of Psychology, The Hebrew University of Jerusalem, Israel; and Shlomo Bentin, Department of Psychology, The Hebrew University of Jerusalem, Israel and Interdisciplinary Center for Neural Computation, The Hebrew University of Jerusalem, Israel.

This study was funded by NIMH grant R01 MH 64458 to Shlomo Bentin. We are grateful to Assaf Harel for insightful comments and Adi Moscovich, Alaa Watad for skillful research assistance.

Correspondence concerning this article should be addressed to Shlomo Bentin, Department of Psychology, The Hebrew University of Jerusalem, 91905 Jerusalem, Israel. E-mail: shlomo.bentin@huji.ac.il

the SF content in a flexible fashion depending upon the stimulus category and the perceptual intentions of the observer (Harel & Bentin, 2009; Morrison & Schyns, 2001; Schyns & Oliva, 1997; 1999). However, most of these studies addressed the question of SF integration during the formation of the visual representation at an early perceptual level (be it iconic memory or post-categorical informational persistence; e.g., Schyns, Petro, & Smith, 2007). Only a few studies explored whether and how the SF that have already been integrated as a perceptual representation are transferred into VSTM and none questioned if all SFs fade away from the VSTM at the same time. Here we addressed these questions using human faces as the stimulus-category of interest.

Faces are a class of objects that have received particular attention in studies of high-level visual perception in general and a focus on their SF content, in particular. This is because, beyond obvious ecological importance for human social interaction, face perception requires processing of both global and local components (Bentin, Golland, Flevaris, Robertson, & Moscovitch, 2006; Bruce & Humphreys, 1994) and, therefore, entails using several SF ranges almost by default. Although there are data suggesting that a midrange of SFs is optimal for face recognition (8–16 cycles per face (cpf) (Some authors report the SF characteristics of the image in cycles per degree [cpd]. However, since this measure varies with the distance between the eyes and the image we prefer using the more stable measure of cycles per face. So, in terms of cpd the best SF range for recognition of a 10×11 cm face seen from a distance of 70 cm would roughly be between 1 cpd and 2 cpd.); e.g., Bachmann, 1991; Costen, Parker, & Craw, 1996; Fiorentini, Maffei, & Sandini, 1983; Nasanen, 1999), other SF ranges are also used for different purposes. The homogeneous and prototypical global structure of faces, which makes them easily identifiable as a category, is conveyed primarily in a lower SF range, whereas specific details and their configurations that are needed for subordinate categorization at different levels are conveyed mostly by a combination of higher and midrange SFs (Goffaux, Hault, Michel, Vuong, & Rossion, 2005; Goffaux & Rossion, 2006; Halit, de Haan, Schyns, & Johnson, 2006; Hughes, Fendrich, & Reuter-Lorenz, 1990; Shulman & Wilson, 1987). Other studies suggested that face recognition in matching tasks is more strongly affected by the SF overlap (SFO) between the original (study) face and the probe (test) face than by the absolute SF content of the face (Collin, Liu, Troje, McMullen, & Chaudhuri, 2004; Liu, Collin, Rainville, & Chaudhuri, 2000).

In line with this general trend, most previous studies of visual processing of face images focused, on the one hand, on the order at which different SF are extracted from visual images while forming a visual percept of the face (e.g., Flevaris, Robertson & Bentin, 2008; Goffaux et al., 2011; Ruiz-Soler & Beltran, 2006) and, on the other hand, on the role of SF ranges in subordinate categorization and identification (Collin et al., 2004; Costen, Parker, & Craw, 1994a; Costen et al., 1996; Goffaux, Gauthier, & Rossion, 2003; Goffaux et al., 2005; Goffaux & Rossion, 2006; Halit et al., 2006). However, to our knowledge, no study investigated whether the encoding of the perceptual face representation into VSTM is an all-or-none process (cf. Sergent & Dehaene, 2004; Wang, 2001; Zhang & Luck, 2008) or mimics the sequence of integrating SFs in the perceptual representation. According to the latter alternative, different SF ranges are encoded into VSTM at different rates as determined either by a fixed, anisotropic order

from LSF to HSF as has been frequently found in a perception (Bar, 2004; Fiorentini et al., 1983; Hochstein & Ahissar, 2002; Hughes, Nozawa, & Kitterle, 1996; McSorley & Findlay, 1999; Parker, Lishman, & Hughes, 1997; Watt, 1987) or in an isotropic manner in which the firstly encoded SF range is determined by the kind of information which is diagnostic for the task at hand (Oliva & Schyns, 1997; Schyns & Oliva, 1997; Loftus & Harley, 2004). Furthermore, there is no study of how LSF and HSF decay from VSTM after a face representation has been already stored.

Addressing the maintenance and resistance to decay, several studies showed that all SFs are stored with high precision in VSTM over periods of at least 10 seconds (Magnussen, Greenlee, Asplund, & Dyrnes, 1990; Regan, 1985). Other studies suggested that the decay from VSTM involved a progressive loss of fine details (Gold, Murray, Sekuler, Bennett, & Sekuler, 2005; Harvey, 1986). Based on their findings, Gold et al. suggested that the decay from VSTM follows a deterministic sequence, for instance, from fine to coarse which mirrors the coarse-to-fine input. However, the objects that were used in these studies were simple gratings or meaningless textures. Other studies of VSTM also used simple, usually meaningless, stimuli (e.g., Awh, Barton, & Vogel, 2007; Vogel et al., 2001; Wheeler & Treisman, 2002; Xu, 2002). Only a few previous studies investigated the formation of more complex and meaningful VSTM representations (Curby & Gauthier, 2007; Curby, Glazek, & Gauthier, 2009), and no research explored how SFs comprised by meaningful visual stimuli decay from VSTM. Because most stimuli perceived outside the laboratory are informational—complex and meaningful, in the current study, we aimed at investigating VSTM encoding and decay of stimuli that are closer to real life.

To achieve this goal, we used a typical WM paradigm in which participants studied a positive set of faces for subsequent recognition and made old/new decisions about a probe-face presented after a varying stimulus-onset asynchrony (SOA). Because previous studies indicated that the storage capacity of VSTM is limited to three to four simple items (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001) or two to three faces (Curby & Gauthier, 2007; Eng, Chen, & Jiang, 2005), our positive set in each trial included two broadband (BB) faces. The probe face was filtered to include either a LSF or a HSF range. By manipulating the study-test SOA and assessing the ability of the participants to discriminate between “new” and “old” faces in each condition, we could evaluate the SF range that was available in the VSTM at different time points. Critical to our investigation, we selected the LSF and HSF ranges based on theoretical as well as empirically based criteria (given later).

General Methods

Participants

The participants were undergraduate students from the Hebrew University of Jerusalem who volunteered as part of course requirements or for monetary compensation and signed an informed written consent according to the institutional review board of the Hebrew University. All the participants had normal or corrected-to-normal vision and no history of psychiatric or neurologic disorders. Different participants were tested in four experiments: Sixteen (six males) in experiment 1, 54 (30 males) in experiment

2, 36 (24 males) in experiment 3, and 12 (seven males) in experiment 4.

Stimuli

The original stimuli consisted of 80 different BB images of Caucasian faces in front view, half female and half male. Each face was also presented in two spatially filtered forms, one preserving relatively low frequencies (LSF) and one preserving relatively high frequencies (HSF), yielding a total stimulus set of 240 faces (Figure 1). All images were 4 cm (width) \times 5.4 cm (height), which seen from a distance of 60 cm, subtended $3.8^\circ \times 5.1^\circ$ of visual angle. Mean luminance and RMS contrast were equated across images (Adobe Photoshop) without including the gray background in calculation. The stimuli were presented on a 19-inch monitor with a 100 Hz refresh rate and a grey background (International Commission on Illumination [CIE]: $x = 0.2891$, $y = 0.2926$; luminance = 49.33 cd/m^2 ; measured by SpectraScan 704, Photo Research).

The edited BB images were spatially filtered in MATLAB using a Butterworth filter with an exponent of four. The SF range for the LSF faces and HSF faces were 2–11.31 cpf and 11.31–64 cpf, respectively. These parameters were determined according to the following two criteria. The first was that the range of SFs which is typically used for face recognition (~ 8 –16 cpf) should be equally represented in the LSF and HSF ranges. This is because the tasks used in the present study required the recognition of individual faces. Furthermore, we considered that not including a psychologically real SF range would have considerably reduced the external validity of our findings. The second was that the HSF-BB faces SFO and the LSF-BB faces SFO should be equal (Collin et al., 2004; Liu et al., 2000). To meet these two criteria, we determined the LSF and HSF range as follows. First, we divided the 8–16 cpf range into two equal bandwidths of 0.5 octaves with 11.31 cpf as the middle point. (Note that this middle point is quite similar to the cutoff frequency used by Wenger & Townsend [12 cpf] on the

basis of an empirical calibration.) Next, we set the low limit of the LSF range to two cycles per image. This limit was set because previous research showed that information contained in SFs below two cycles per image is hardly useful for the recognition of grey images (Costen et al., 1994a; 1996; Hayes, Morrone, & Burr, 1986 for faces and Torralba, 2009 for non-face objects). As the range between 2 and 11.31 cpf comprises 2.5 octaves, in order to equalize the overlap between the LSF and the HSF range with BB, the HSF range also included 2.5 octaves, which determined the higher limit of this range to 64 cpf. Note, however, that information below 25 cycles per image is sufficient to represent almost all the fine edges which define the important image components (Hayes et al., 1986; Torralba, 2009).

The luminance of the original BB faces (mean luminance: 42.31 cd/m^2) and the filtered faces was fairly close (mean luminance is 45.13 cd/m^2 and 44.32 cd/m^2 for LSF and HSF faces, respectively). Moreover, because the nonlinear relationship between voltage input and luminance on cathode ray tube (CRT) monitor has an influence on the real SF content that the filtered stimuli contain (Peli, 1992), we determined the actual mean amplitude spectra of the stimuli in each condition post hoc. First, we measured the screen luminance using a MavoMonitor (GOSSEN Foto- und Lichtmesstechnik GmbH, Germany) at all 256 gray levels, which determined the gamma function of the CRT monitor. On the basis of these luminance values, we calculated the SF cut-off in all our BB, LSF, and HSF faces. Finally, the averaged SF amplitudes were averaged across the 80 faces in each condition. We found that the real SF cutoff for LSF face was around 2.8–13 cpf (2.2 octaves), whereas for HSF, it was around 13–64 cpf (2.3 octaves). Based on this analysis, we see that although the SF content of the images was slightly influenced by the nonlinear display of the CRT, the SF cut-offs in our study were roughly around the ones we intended to use (Supplementary Figure 1).

Experiment 1

In order to investigate the order in which different SF ranges are transferred from the perceptual representation to the VSTM, it is important to ensure first that the LSF and HSF are equally contained in the perceptual system at the end of the positive set display. To achieve this goal, in the present experiment, we used a perceptual task to determine the time course of SF accumulation in the perceptual representation formed during a face matching/identification task.

Method

Task, procedure, and design. Participants were seated in a sound-attenuated room at a distance of 60 cm from the monitor. One BB face and one filtered face were presented at the center of the screen simultaneously one above the other, with a vertical distance of 0.5° visual degree in between. Vertical alignment was used to prevent hemispheric differences in processing low and high SF ranges (Ivri & Robertson, 1998). The participants were instructed to indicate whether the two faces were the same identity or not by pressing keys that were labeled “yes” (“f”) and “no” (“k”) on the keyboard (on a Hebrew keyboard “f” represents the first letter of the word “yes” and “k” the first letter of the word “no”). We reasoned that because both the BB and the filtered face

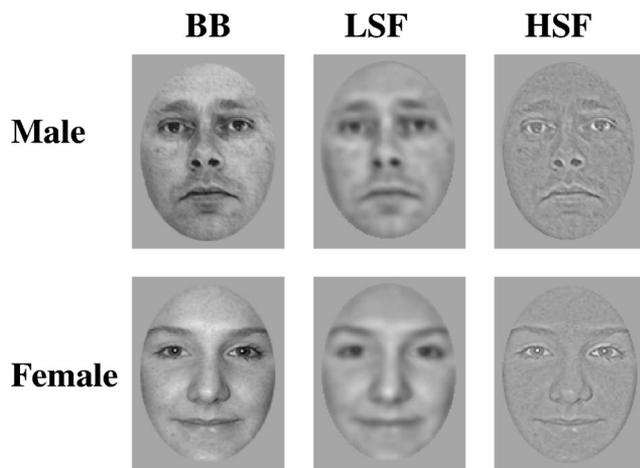


Figure 1. Examples of faces used in the experiment. Faces in the positive set were presented in Broad Band (BB - left column). The probes preserved only a low spatial frequency range (LSF - center column), or a high spatial frequency range (HSF - right column). See text for exact frequency range values.

were simultaneously presented, no memory component should affect performance in this task. Therefore, equal performance comparing BB faces with LSF and HSF faces should indicate that an equal amount of low and high SFs was extracted from the BB face during exposure time. Accuracy rather than speed was emphasized. In half of the trials, BB face was above the filtered face, whereas in the other half of the trials, BB face was below the filtered face. The gender of the two faces in a trial was the same.

As illustrated in Figure 2, each trial began with a red fixation cross exposed for 500 ms and followed by 500 ms blank interval. At the end of this interval, a BB face and a filtered face (either LSF or HSF) were presented for 250, 500, or 800 ms and the participant was allowed to respond during an interval of 2 seconds from stimulus onset. When a response was produced or the response interval was exhausted, the next trial started automatically.

A 2 (SF: LSF or HSF) \times 3 (Exposure Time: 250, 500, or 800 ms) within-subject design was used in this experiment. There were 32 trials in each condition, with a total of 192 trials which were fully randomized and presented in three blocks of 64 trials each. A break of 5 minutes was allowed between blocks so that the entire experiment lasted approximately 15 minutes. Prior to the experimental blocks, participants were trained with 24 practice trials including four trials from each experimental condition. The six faces used for practice were not included in the experimental trials.

Results and Discussion

Accuracy was determined by A_z and analyzed using the within-subjects analysis of variance (ANOVA) with SF and Exposure Time as independent factors. A_z is a monotonic transformation of d' bounded between zero and one, which has been shown to be more robust to influence of response bias on sensitivity and recommended by other researchers (Verde, Macmillan & Rotello, 2006; Richler, Tanaka, Brown, & Gauthier, 2008).

As evident in Figure 3, accuracy was higher for LSF condition than HSF condition at 250 ms, but this difference almost vanished at longer exposure times. ANOVA showed that the main effect of exposure time was significant ($F(2, 30) = 20.286$, $MSe = 0.009$, $p < .001$; partial $\eta^2 = 0.575$). Paired-wise contrast revealed that

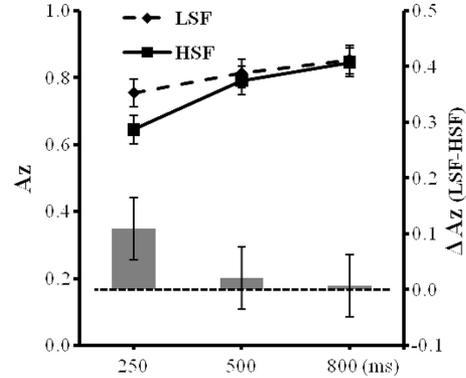


Figure 3. Mean A_z for the same/different classification of LSF-BB and HSF-BB pairs, and the difference between LSF and HSF, presented at different exposure times (250, 500, and 800 ms) in experiment 1. Error bars show 95% confidence intervals (Masson & Loftus, 2003).

performance at 800 ms (0.85) was higher than at 250 ms (0.70) and 500 ms (0.80) (95% confidence interval [CI] = [0.073, 0.227], $p < .001$; 95% CI = [0.003, 0.092], $p = .033$, respectively), and the performance at 500 ms was considerably higher than at 250 ms (95% CI = [0.034, 0.171], $p = .003$). Importantly, the interaction between SF and Exposure Time was significant ($F(2, 30) = 4.071$, $MSe = 0.006$, $p = .027$; partial $\eta^2 = 0.213$). Post hoc investigation of this interaction revealed that at 250 ms A_z was significantly larger in the LSF compared with the HSF condition (0.71 vs. 0.62, 95% CI = [0.022, 0.153]; $t(15) = 2.858$, $p < .015$). In contrast, the two SF conditions did not differ at 500 ms (0.76 vs. 0.74; 95% CI = [-0.053, 0.019], $t(15) < 1.0$) and at 800 ms (0.80 vs. 0.79, 95% CI = [-0.056, 0.045], $t(15) < 1.0$).

The current experiment showed that if two faces are exposed for 500 ms or longer, LSF and HSF are equally accumulated in the perceptual representation. Therefore, in the following experiments, we exposed the two-face positive sets for 500 ms and 800 ms while investigating the order of transferring high and low SFs contained in the face images from perception to VSTM and their decay from memory.

Experiment 2

In the present experiment, we investigated whether LSF and HSF contained in a perceptual representation are encoded in VSTM at a similar pace, as well as their decay speed from VSTM. To achieve this goal, we used a recognition memory task in which two faces formed the positive set in each trial and were followed by a probe which contained only the low or only the high range of SFs. The probe followed the positive set at varying SOAs. Given the outcome of experiment 1, we postulated that recognition differences between the LSF and HSF probes would reflect different order of encoding low and high SFs into the VSTM representation of the faces.

Method

Task, procedure, and design. A recognition memory test was designed. In each trial, a positive set of two faces was followed by a probe. The participants were instructed to indicate

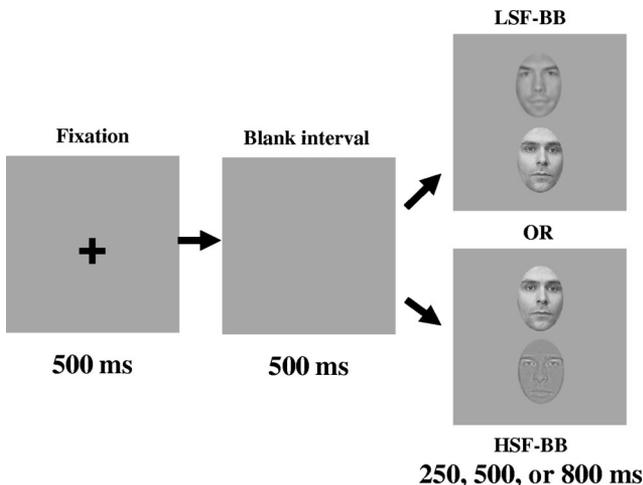


Figure 2. The time course of a trial in experiment 1.

by button press whether the probe was one of the two faces in the immediately preceding positive set (“old”), or not (“new”). Accuracy rather than speed was emphasized.

As illustrated in Figure 4, each trial began with a red fixation cross exposed for 500 ms and followed by a 500 ms blank interval. At the end of this interval two BB faces were displayed, randomly located within a radius of 5° around fixation, that is, in the parafovea. For one group of 30 participants, the positive set was exposed for 500 ms, whereas for other 24 participants the exposure time was 800 ms. After a blank inter-stimulus interval (ISI) a probe was presented. The length of the ISI was 600, 1,200, 1,800, or 2,400 ms for the 500 ms exposure condition, and 300, 900, 1,500, or 2,100 ms for the 800 ms exposure condition. The 300 ms shorter ISIs in the 800 ms exposure condition kept the SOAs between the positive set and the probe constant across the two exposure time conditions. The probe consisted of either a LSF or a HSF filtered face presented at the center of the screen until a response was initiated, or up to 2 seconds from onset. The next trial started automatically after response or at the end of the 2 seconds period.

Accuracy (A_z) was analyzed by a mixed-model ANOVA with Exposure Time (500 ms or 800 ms) as between-subjects factor, while SF (LSF or HSF) and SOA (1,100, 1,700, 2,300, or 2,900 ms) (The degrees of freedom were corrected whenever necessary using the Greenhouse-Geisser procedure) were within-subject factors. We set 2,900 ms as our upper-limit SOA because two previous studies showed that, the decay speed of VSTM representations of complex objects was faster than that of simple gratings with a “half-life” at around 3,000 ms (Cornelissen & Greenlee, 2000; Salmela, Makela, & Saarinen, 2010). There were 32 trials in each condition, yielding a total of 256 trials which were fully randomized and presented in four blocks of 64 trials each. A break of 8 minutes was allowed between blocks so that the entire experiment lasted approximately 40 minutes. Prior to the experimental blocks, participants were given 32 practice trials including four trials from each experimental condition.

Results

As evident in Figure 5a and 5b, probe recognition was fairly good but not at ceiling in both positive set exposure time conditions and was not dramatically affected by SOA. The most interesting outcome of this analysis was a putative interaction between the exposure time of the positive set and SF; that is, with 500 ms

exposure time there was a consistent advantage for LSF probes relative to HSF probes, however, with 800 ms exposure time this advantage was not evident.

These observations were corroborated by ANOVA. There was no main effect of Exposure Time ($A_z = 0.76$ and 0.79 for 500 ms and 800 ms conditions, respectively; $F(1, 52) = 2.6$, $MSe = 0.024$, $p = .113$; partial $\eta^2 = 0.05$), and a significant effect of SOA ($F(3, 156) = 4.3$, $MSe = 0.009$, $p < .01$, $G-GE = 0.92$; partial $\eta^2 = 0.08$), which did not interact with Exposure Time ($F(3, 156) = 1.5$, $MSe = 0.009$, $p = .213$; $G-GE = 0.96$, partial $\eta^2 = 0.03$). The pairwise contrasts showed that the accuracy at 2,900 ms SOA was lower compared with that at all other SOAs (95% CI = $[-0.05, -0.001]$, $p < .05$; 95% CI = $[-0.063, -0.022]$, $p < .001$; and 95% CI = $[-0.055, -0.009]$, $p < .01$ for the 1,100, 1,900, and 2,300 ms, respectively). A main effect of SF (95% CI = $[0.004, 0.039]$, $F(1, 52) = 6.1$, $MSe = 0.008$, $p = .016$; partial $\eta^2 = 0.10$) suggested that LSF probes were recognized better than HSF probes. However, the SF effect was qualified by a significant interaction with Exposure Time ($F(1, 52) = 5.12$, $MSe = 0.008$, $p = .028$; partial $\eta^2 = 0.09$), which was elaborated by separate ANOVAs for 500 ms and 800 ms exposure time.

For the 500 ms exposure time condition, ANOVA revealed higher accuracy for LSF probes ($A_z = 0.79$) compared with HSF probes ($A_z = 0.74$; $F(1, 29) = 10.9$, $MSe = 0.009$, $p < .005$; partial $\eta^2 = 0.27$). There was no main effect of SOA ($F(3, 87) = 1.8$, $MSe = 0.012$, $p = .158$) and no SF \times SOA interaction ($F(3, 87) < 1.0$).

For the 800 ms exposure time condition, there was no main effect of SF ($F(1, 23) < 1.0$), there was a significant effect of SOA ($F(3, 69) = 5.2$, $MSe = 0.007$, $p < .005$; partial $\eta^2 = 0.18$) and a slight tendency of SF \times SOA interaction ($F(3, 69) = 2.5$, $MSe = 0.06$, $p = .065$; partial $\eta^2 = 0.10$). Furthermore, t -test comparisons for independent samples showed that accuracy for the LSF-probes was equally high in both exposure time conditions (95% CI = $[-0.041, 0.031]$, $t(52) < 1.0$; see Figure 6a), whereas for HSF-probes condition, the A_z was significantly higher in 800 ms than 500 ms (95% CI = $[0.011, 0.077]$, $t(52) = 2.641$, $p < .025$; see Figure 6b). In other words, the addition of 300 ms to the exposure time of the positive set elevated the recognition of HSF-probes, but had no effect on LSF-probes.

As shown in Figure 5b and 5d, the pattern of the response bias was quite similar across the two exposure times. A significant

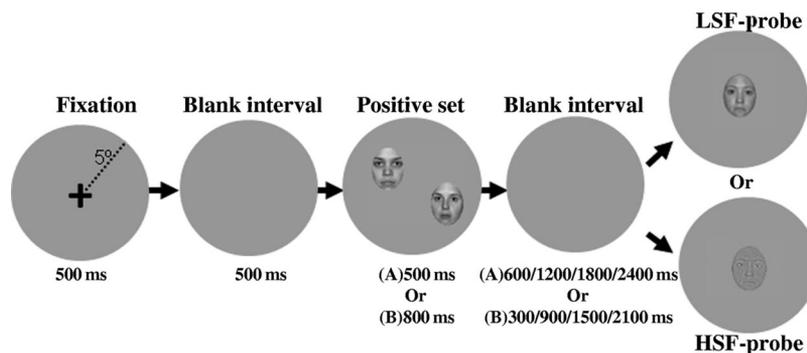


Figure 4. The time course of a trial in experiment 2. Note that the faces forming the positive set were placed randomly in the parafovea. This is conveyed here as a virtual circle, although the screen was entirely grey.

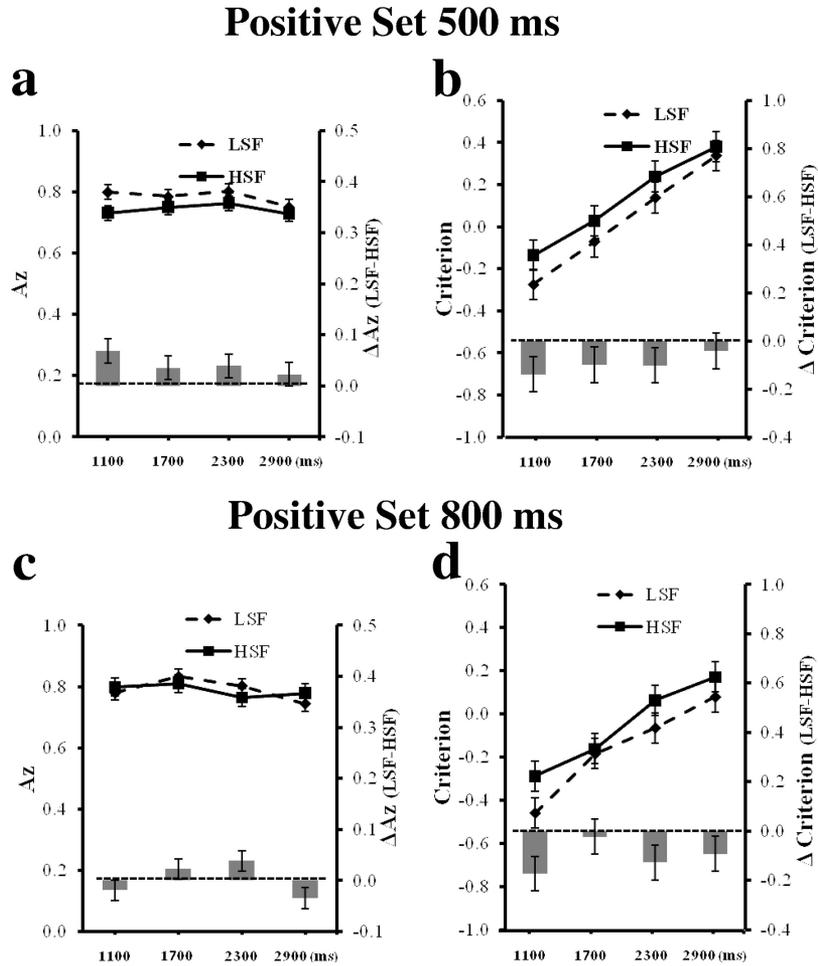


Figure 5. Experiment 2: (a, c) Mean A_z for the old/new classification of LSF-probes and HSF-probes, and the difference between LSF and HSF, presented at different SOAs (1,100, 1,700, 2,300, and 2,900 ms) in the 500 ms exposure time and 800 ms exposure time conditions, respectively. (b, d) The response criteria and the difference between LSF and HSF for each exposure time and SF conditions in the 500 ms exposure time and 800 ms exposure time conditions, respectively. Error bars show 95% confidence intervals (Masson & Loftus, 2003).

main effect of SOA ($F(3, 156) = 75.435$, $MSe = 0.085$, $p < .001$, $G-GE = 0.885$; partial $\eta^2 = 0.592$) showed that the tendency to categorize probes as “new” was augmented as SOA increased. A higher response bias was found in the HSF condition (0.037) than in the LSF condition (-0.062 ; 95% CI = [0.027, 0.172], $F(1, 52) = 7.659$, $MSe = 0.14$, $p = .008$; partial $\eta^2 = 0.128$). The response bias at 500 ms (0.080) was significantly higher than at 800 ms (-0.106 ; 95% CI = [0.082, 0.291], $F(1, 52) = 12.760$, $MSe = 0.036$, $p = .001$; partial $\eta^2 = 0.197$). The other factors were not significant (all F s < 1.0).

Discussion

When the positive set was exposed for only 500 ms, faces that contained only the relative lower ranges of SFs were classified more accurately as “old” or “new” compared with faces that contained only the relative higher SF ranges. Importantly, this effect was constant across retention intervals. Adding 300 ms more to the exposure time of the positive set abolished the LSF advantage;

in this condition, the recognition of HSF probes was elevated to the level of LSF probes. Since experiment 1 indicated that 500 ms is enough to equally represent LSF and HSF from two faces into the perceptual representation, the current LSF advantage cannot be explained by unbalanced accumulation of SF during perception. An alternative account for the LSF advantage is the possibility that it contains more information relevant to identification than HSF and thus lead to a response bias towards LSF-probes. Indeed, a recent study showed a higher tuning sensitivity for the SF between 4 and 11 cpf (Gaspar, Sekuler, & Bennett, 2008), which falls into our LSF range. However, would this account be true, the LSF advantage should have persisted in memory regardless of the exposure time of the positive set. Further, since the LSF advantage did not diminish with the increase in SOA, it cannot reflect a faster decay of HSF from VSTM, therefore supporting previous findings that information embedded in all SFs could be retained perfectly for a long time (Magnussen et al., 1990; Regan, 1985).

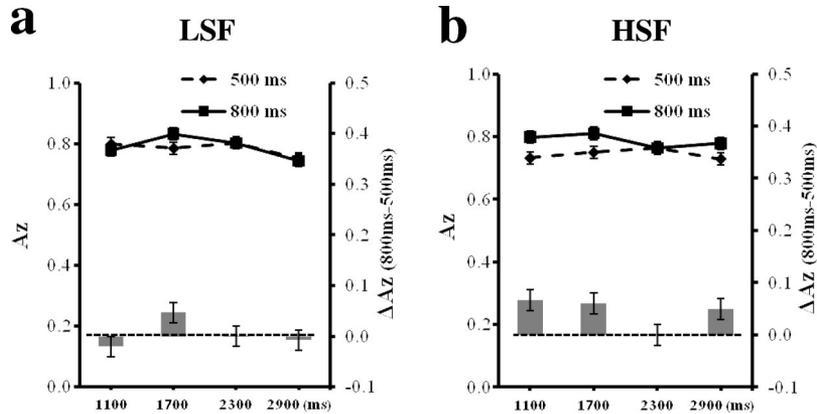


Figure 6. Accuracy (A_z) for discriminating “old” from “new” faces for (a) LSF-probes and (b) HSF-probes in the 500 ms condition and 800 ms condition of experiment 2. Note that SOA affects only the HSF-probes.

Hence, the current outcome showed, on the one hand, that the exposure time of the positive set determined the relative amount of higher and lower SF in the VSTM representation of faces beyond perception, prioritizing the LSF. This outcome indicates that the encoding of visual information in the VSTM is not an all-or-none process (cf. Zhang & Luck, 2008), but rather, at least for recognition, it follows a coarse-to-fine order so that LSF are extracted from the perceptual representation and/or consolidated into VSTM faster than HSF. Such a pattern would be in line with the Reversed Hierarchy Theory (Hochstein & Ahissar, 2002) which posits that human beings become aware of the global content of a visual image prior to its details.

On the other hand, the length of the ISI did not influence the LSF advantage. The absence of the ISI effect suggests that the transfer functions of SF from perception to VSTM are determined primarily by the accumulation of different SF ranges in the perceptual representation during the stimulus exposure time. This could happen if the transfer from perception to memory would occur in cascade with the formation of the perceptual representation, starting at some (yet to be specified) delay and proceeding at the similar sequence as the visual input. Because experiment 1 showed that LSF are accumulated in the perceptual representation faster than HSF, the transfer from perception to memory was initially dominated by LSF. At the end of 500 ms exposure time, although LSF and HSF had already been represented equally in perception, the visual information encoded in VSTM was still unbalanced, with more LSF than HSF, putatively reflecting the SF composite of the perceptual representation before 500 ms. Enhancing the exposure time allowed the accumulation of more HSF in VSTM abolishing the head-start of the LSF. This account is supported by the pattern of the SF \times Exposure Time interaction. Although the recognition accuracy of LSF-probes was not further improved by adding 300 ms exposure time, the recognition of HSF-probes following 800 ms exposure time of the positive set was higher than after 500 ms exposure time, reaching the level of accuracy of LSF-probes. This pattern suggests that an exposure time of 500 ms might be sufficient not only for extracting all the LSF from the image to perception but also to transfer this information into VSTM (although additional processing could continue after stimulus offset, see experiment 3). It also suggests that exposing the positive set image for 300 ms more allowed more HSF to be encoded into VSTM. Hence, these results

indicate that LSF are encoded faster into VSTM than HSF although they cannot determine whether this effect reflects prioritized transfer of low SFs, or dynamic changes in the SF constitution of the perceptual representation during gradual transfer of information from perception to memory.

In contrast to the apparent coarse-to-fine order of encoding SF in VSTM, the decay from VSTM is probably not influenced by SF at least during the initial 2–3 seconds after encoding. This hypothesis is suggested by the absence of interaction between the SF and SOA across exposure time conditions even though the post-hoc analysis of the SOA effect showed that the recognition accuracy at 2,900 ms SOA was reduced relative to shorter SOAs. Hence, even when some drop in performance is already evident, all SF ranges are equally lost from memory.

Finally, the analysis of the response criteria showed that the participants adopted overall a conservative strategy tending to classify probes as “new” in case of uncertainty. For both LSF and HSF probes, the criterion for classifying a probe as “new” increased almost linearly with SOA and was larger following 500 ms than 800 ms exposure time of the positive set. To this end, it is interesting that for both exposure time conditions, the criteria were higher for HSF than for LSF, which goes along with the LSF advantage found in classification accuracy.

Assuming a simple time-lag between the onset of perceptual encoding and the onset transferring information from perception to VSTM, our model suggests that the transfer might continue even after the offset of the visual stimulus, perhaps from the iconic image which persists for a while after the stimulus offset. Because, in this experiment, the positive set image was not followed by a mask, we could not address this issue. Furthermore, we could not determine whether 800 ms are, indeed, enough to encode and consolidate all SF into VSTM or that some additional processing based on iconic memory persistence continued even after the image had been removed from vision. Experiment 3 was aimed at addressing this question.

Experiment 3

Experiment 3 had two goals. First, we investigated whether 800 ms exposure time is, indeed, enough to extract all the SFs into VSTM. An alternative hypothesis is that the process of encoding

continues after the offset of the stimulus using the iconic memory persistence (Vogel, Woodman, & Luck, 2006). To address this question, we presented a pattern mask at the offset of the positive set image. Such a mask should immediately erase the iconic memory and, therefore, terminate the processing of the two faces as soon as they are removed from the visual field. The second goal of this experiment was to replicate the absence of SF effects found in experiment 2 when the exposure time was 800 ms. Therefore, in half of the trials, the positive set was not backward masked.

Method

Stimuli. The faces used to form the positive sets and the probes were identical to those used in experiment 2. The masks were two faces presented upside-down in the exact spatial locations of the faces in the immediately preceding positive set. These faces were selected in each trial from a set of two male and two female faces to match the gender of the faces presented in the positive set.

We masked faces with inverted faces in order to ensure that the mask would evenly block all the SF channels activated by the faces as well as reduce the possible confounding of masking faces with those in the positive set. In addition, previous studies (e.g., Costen, Shepherd, Ellis, & Craw, 1994b; Loffler, Gordon, Wilkinson, Goren, & Wilson, 2005) as well as a pilot experiment that we conducted while preparing this study, showed that inverted faces as backward masks stop face processing more efficiently than other (meaningful or meaningless) patterns.

Design and procedure. In the present experiment, we included only the two shortest SOAs (1,100 ms and 1,700 ms). This change in the design was made for two reasons: First, the current experiment focused only on the encoding stage and, in experiment 2, we observed that in the 800 ms exposure time condition SF began to decay from VSTM after 1,700 ms. Second, adding the masking manipulation orthogonal to the SF manipulation and reducing the number of SOAs from four to two, we kept the total number of trials as well as the number of trials per condition (256 and 32, respectively) identical to the 800 ms exposure time condition in experiment 2.

As a result of these changes we ended up with a 2 (SF: LSF or HSF) \times 2 (SOA: 1,100 or 1,700 ms) \times 2 (Masking: masked or unmasked) within-subject design. The eight experimental conditions were randomized separately for each participant and were displayed using the same procedure as in experiment 2 except for

adding a mask in the masked conditions (see Figure 7). The positive set was presented for 800 ms and immediately followed by a 100 ms mask. Consequently, the ISIs between the mask and the probe was either 200 ms or 800 ms.

Results

As shown in Figure 8, when the positive set was backward-masked, the LSF advantage found with 500 ms exposure time was re-established despite the longer exposure time. In contrast, without masking, the SF effect apparently vanished, replicating the results of the 800 ms exposure time group in experiment 2 (data from individual participants are available in the Supplementary Figure 2).

Three-way ANOVA yielded a significant main effect of Masking revealing that old/new classification accuracy was reduced by masking ($A_z = 0.79$ and $A_z = 0.76$ for the unmasked and masked conditions, respectively; 95% CI = [0.015, 0.057], $F(1, 35) = 11.8$, $MSe = 0.008$, $p < .005$; partial $\eta^2 = 0.25$). A significant main effect of SF revealed that, across masking and SOA, accuracy was higher for LSF probes (0.79) than for HSF probes (0.76; 95% CI = [0.008, 0.053], $F(1, 35) = 7.3$, $MSe = 0.009$, $p < .01$; partial $\eta^2 = 0.17$). However, this effect was modulated by masking, as revealed by SF \times Masking interaction ($F(1, 35) = 6.4$, $MSe = 0.006$, $p < .025$; partial $\eta^2 = 0.15$). There was no significant effect of SOA (95% CI = [-0.013, 0.036], $F(1, 35) < 1.0$), and this factor did not interact with any other factors (all F s < 1.0). However, a second-order interaction suggested that the SF \times Masking interaction was modulated by SOA ($F(1, 35) = 4.4$, $MSe = 0.007$, $p < .05$; partial $\eta^2 = 0.11$).

To elaborate the above interactions, we analyzed the SF and the SOA effects by separate ANOVA in the unmasked and masked conditions. In the unmasked condition, neither the SF (95% CI = [-0.022, 0.034], $F < 1.0$) nor the SOA (95% CI = [-0.027, 0.032], $F(1, 35) < 1.0$) main effect was significant, and there was no interaction between the two factors ($F(1, 35) = 3.2$, $MSe = 0.01$, $p = .083$; partial $\eta^2 = 0.08$). This outcome replicates the results of the 800 ms exposure time condition of experiment 2. In contrast, in the masked condition, the main effect of SF was significant (95% CI = [0.024, 0.086], $F(1, 35) = 12.7$, $MSe = 0.009$, $p < .001$; partial $\eta^2 = 0.27$). The main effect of SOA was not significant even in the masked condition (95% CI = [-0.014, 0.055], $F(1, 35) = 1.472$, $p = .233$; partial $\eta^2 = 0.040$), and there was no SF \times SOA interaction ($F(1, 35) < 1.0$). This outcome is similar with the results of the 500 ms condition of experiment 2.

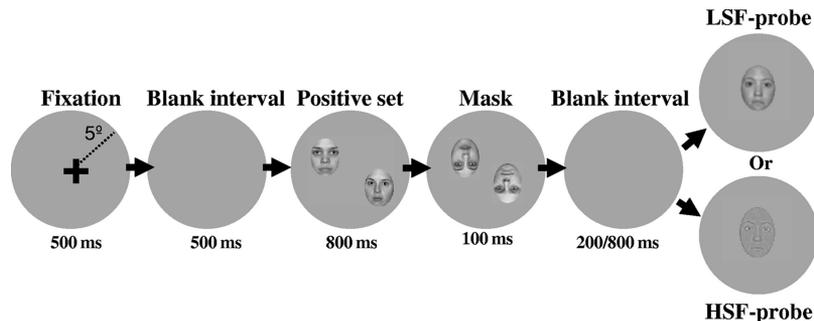


Figure 7. Time course of a trial with mask in experiment 3. Note that the faces forming the positive set were placed randomly in the parafovea. This is conveyed here as a virtual circle, although the screen was entirely grey.

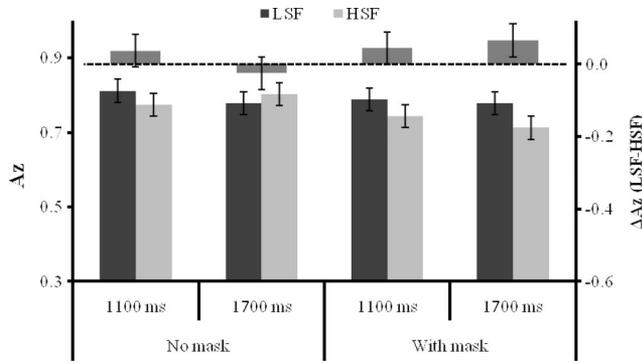


Figure 8. Mean A_z for the old/new classification of LSF-probes and HSF-probes, and the difference between LSF and HSF, presented at different SOAs (1,100, 1,700 ms) in experiment 3. Error bars show 95% confidence intervals (Masson & Loftus, 2003).

In addition, we also analyzed the effect of Masking separately in the LSF and the HSF probe conditions. This analysis revealed that LSF-probes had not been influenced either by Masking (95% CI = $[-0.012, 0.036]$, $F < 1.0$) or by SOA (95% CI = $[-0.013, 0.056]$, $F(1, 35) = 1.6$, $MSe = 0.010$, $p = .21$; partial $\eta^2 = 0.045$), and there was no Masking \times SOA interaction ($F(1, 35) < 1.0$). In contrast, for HSF-probes, performance was better in the unmasked condition (0.79) than in the masked condition (0.73; 95% CI = $[0.028, 0.093]$, $F(1, 35) = 14.0$, $MSe = 0.009$, $p < .001$; partial $\eta^2 = 0.289$). There was no main effect of SOA (95% CI = $[-0.040, 0.042]$, $F(1, 35) < 1.0$) but a significant Masking \times SOA interaction ($F(1, 35) = 5.5$, $MSe = 0.006$, $p < .025$; partial $\eta^2 = 0.13$). Post hoc pairwise comparisons showed that the masking effect on HSF-probes was significant at 1,700 ms SOA (95% CI = $[0.046, 0.13]$, $t(35) = 4.096$, $p < .001$), but not at 1100 ms SOA (95% CI = $[-0.008, 0.069]$, $t(35) = 1.566$, $p = .126$).

Discussion

When the positive set was not masked, the current results replicated the 800 ms exposure time condition in experiment 2; that is, the old/new classification accuracy for HSF-probes was as high as for LSF-probes. However, when the faces in the positive set were backward-masked by inverted faces, the current results replicated the advantage of LSF-probes over HSF-probes which was found when the exposure time of the positive set was only 500 ms. Moreover, our analysis showed that masking affected performance only for HSF-probes. Hence, the addition of 300 ms to the exposure time of the positive set was insufficient to fully represent the HSF in the VSTM.

These data suggest that some processes which do not depend on the existence of the visual information in the visual field are essential for the formation of reliable representations in VSTM. Since performance with LSF probes was not affected by masking, it stands to reason that these are processes that start and may be accomplished while the visual stimulus is on display, but may also continue after it disappears. Whereas the extraction of SF from the visual image into perception is completed by this time, the consolidation of SFs in VSTM probably continues on the basis of visual persistence. Additional parametric studies are necessary to unveil the exact time course of this process. However, psycho-

physical studies indicated that visual persistence cannot last more than about 300 ms (Coltheart, 1980; Irwin & Yeomans, 1986).

Finally, like in the previous experiment, across SF ranges, there was no evidence for selective decay of SF from VSTM. However, the range of SOAs used so far might have not been sufficient to reveal such selectivity. Therefore, in the next experiment, we explored the SF decay up to 10 seconds.

Experiment 4

In the previous experiments, we did not find significant decay from VSTM up to 2,900 ms SOA and even in that condition recognition memory was still significantly above chance ($p < .001$). In fact, previous research suggested that SF can be maintained in VSTM even for 10 seconds (Magnussen, Greenlee, Asplund, & Dyrnes, 1991; Regan, 1985). However, these studies tested the VSTM using simple Gabor gratings composed from only one frequency. Therefore, although they clearly demonstrated that the longest ISI used in experiment 2 might have been too short, the results of these studies cannot be directly generalized to VSTM for complex stimuli comprising a large range of SFs. To this end, in the next experiment, we investigated patterns of decay for LSF and HSF extending the ISI to 10 seconds.

Method

Procedure stimuli and design. This experiment was identical in all to the 800 ms condition of experiment 2 except that the ISI between the positive set and the probe was randomly selected in each trial to be 1, 5, or 10 seconds, and each face was used six times. The 800 ms exposure time (of the positive set) was used to ensure that both SF ranges would be equally represented in VSTM at least before decay might occur.

Results and Discussion

As illustrated in Figure 9, recognition accuracy dropped dramatically from 1 to 5 seconds without further change. Importantly,

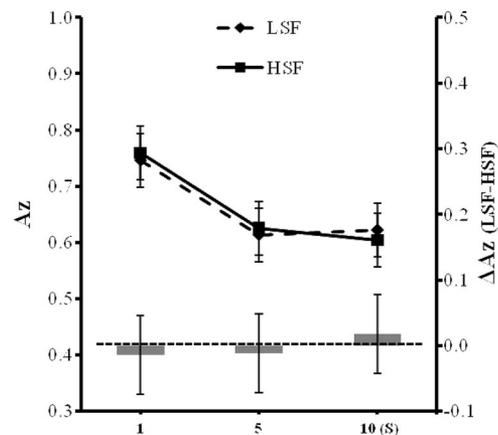


Figure 9. Mean A_z for the old/new classification of LSF-probes and HSF-probes, and the difference between LSF and HSF, presented at different SOAs (1, 5, and 10 seconds) in experiment 4. Error bars show 95% confidence intervals (Masson & Loftus, 2003).

there was no difference between recognition accuracy of LSF and HSF probes at any ISI. ANOVA confirmed that the main effect of SF was not significant ($F(1, 11) < 1.0$) and did not interact with ISI ($F(2, 22) < 1.0$). The main effect of ISI was significant ($F(2, 22) = 19.063$, $MSe = 0.008$, $p < .001$; partial $\eta^2 = 0.18$). Pairwise contrasts showed that the A_z was larger when the SOA was 1 seconds (0.75) than either 5 seconds (0.62; 95% CI = [0.052, 0.213], $p < .005$) or 10 seconds (0.613; 95% CI = [0.081, 0.198], $p < .001$), and there was no difference between 5 seconds and 10 seconds (95% CI = [-0.068, 0.081], $p = 1.0$).

Despite significant reduction in face recognition with ISI, the decay in VSTM was almost identical for LSF and HSF probes. Therefore, these data provide converging evidence against a differential speed of decay for LSF and HSF from VSTM.

General Discussion

In the present study, we investigated the order of encoding high and low SFs comprised in face images from the perceptual representation into the VSTM. We also investigated whether HSF and LSF decay from VSTM at different rates during a storage period lasting up to 10 seconds. The results indicated a coarse-to-fine order of encoding. Specifically, we found that like the extraction of SFs from the visual image into the perceptual representation (e.g., Goffaux et al., 2011; Ruiz-Soler & Beltran, 2006), the encoding of SFs into VSTM is faster for LSF than HSF. The current data suggest that the transfer of information from perception to visual memory begins while the image is still being encoded in perception and continues after the actual image disappears, probably based on the iconic memory persistence (cf. Loftus & Hanna, 1989). Additionally, we found that the speed of decay was similar across SF ranges.

This study added to visual face processing research in several ways. First, previous studies of visual face processing focused on extracting SFs from the visual image during the formation of perceptual representations (e.g., Goffaux et al., 2003, 2011; Costen et al., 1996; Flevaris et al., 2008), on the role of different SF ranges during early face processing (e.g., Goffaux & Rossion, 2006) or on the selective usage of SF ranges in different tasks (e.g., Schyns & Oliva, 1999; see Ruiz-Soler & Beltran, 2006 for a review). In contrast, here we focused on the transfer of SF from perception to VSTM and on its decay from VSTM. Second, most previous studies of memory for SFs focused on their maintenance in VSTM using simple stimuli such as single-frequency Gabor patches (Lee & Harris, 1996; Magnussen & Greenlee, 1999; Regan, 1985; for a review see Magnussen, 2000). Therefore, although these studies demonstrated that single LSF and HSF are represented in VSTM with equal precision, they were not designed to tap the encoding rates for different frequency ranges. In contrast, we presented stimuli which include a wide spectrum of SFs and used a more traditional recognition memory design in which a probe had to be matched with two members of a positive set stored in VSTM. This allowed us to tap into the VSTM encoding stage.

The assumption that the order of encoding visual information about faces in VSTM follows a coarse-to-fine order seemingly contradicts the conclusion reached in one of the few previous studies that investigated how different SFs map into different types of physiognomic information during face perception and memory tasks (Wenger & Townsend, 2000). Assessing the effect of the SF

ranges in discrimination and delayed face-matching tasks coupled with similarity ratings, Wenger and Townsend rejected a heuristic based on a general LSF dominance while matching a probe to a pre-studied face (as well as one assuming that, as a rule, configural information is contained in the LSF range whereas the face details are contained in the HSF range). Alternatively, these authors suggested that the type of SF activated in the VSTM representation is flexible and determined by the task. Specifically, they showed that HSF were best suited for discrimination tasks while LSF were best suited for recognition tasks. However, note that the procedures used by these authors were very different from ours. Particularly, since in their study (experiment 2), the first face was exposed for 1.5 seconds and not backward masked, it is plausible that all the SF contained in the BB “study face” was encoded in the VSTM representation. Hence, although their study elegantly addressed the question of how SFs are used in VSTM after being encoded, it did not investigate the process of encoding per se. To this end, we are fairly confident in claiming that, at least when the faces are encoded in VSTM for subsequent recognition, visual information in memory is transferred from the perceptual representation to the VSTM in a coarse-to-fine order.

It is worth noting that a similar coarse-to-fine encoding sequence was also found in face perception, particularly at early stages of basic-level categorization (e.g., Goffaux et al., 2003, 2011; for a more general perspective see Busey & Loftus, 1994 and Loftus & Irwin, 1998). In experiment 1, we replicated this pattern and found that it takes about 500 ms for the HSF to reach the same level of representation as LSF. Because the processing of LSF relies primarily on the quicker magnocellular pathway in contrast with HSF processing which relies primarily on the slower parvocellular pathway, the faster accumulation of LSF in the perceptual system can be easily explained (Bar, 2004; Howard & Reggia, 2007; Merigan, Byrne, & Maunsell, 1991a; Merigan, Katz, & Maunsell, 1991b). However, we do not know whether the faster accumulation of LSF compared with HSF in the VSTM is accounted for by same mechanism. In fact, not everybody agrees that the transfer of information from perception to memory is gradual. Several studies proposed that by the time the storing of an object into VSTM starts, all perceptual processes related to that object are already completed (see for example Awh et al., 2007; Chun, & Potter, 1995; Eng et al., 2005; Scolari, Vogel, & Awh, 2008; Vogel et al., 2001; Wheeler & Treisman, 2002; Xu, 2002). Within this theoretical framework an “all-or-none” encoding processing was implicitly assumed (Sergent & Dehaene, 2004; Zhang & Luck, 2008). The gradual accumulation of visual information in VSTM demonstrated in the current study does not support the “all-or-none” theory. Moreover, the current results do not support the hypothesis that the transfer from perception to VSTM starts only after the perceptual representation is fully established. Would that be so, exposure time should have not affected the speed of encoding in VSTM in experiment 2 since LSF and HSF were equally represented in perception 500 ms after the stimulus onset and the SOA was the same in the 500 ms and 800 ms exposure time conditions. Furthermore, if the completion of the perceptual representation would require more than 500 ms, or would encoding in VSTM start only after the offset of the visual image, we should have observed larger frequency-range effects in the 800 ms condition because the ISI to the probe was shorter. Alternatively, the similar coarse-to-fine pattern found in perception and VSTM sug-

gests that encoding in VSTM might take place in cascade with perception. That is, visual information begins to accumulate in VSTM soon after it is perceived, even if the perceptual integration is not completed. Hence, the process of VSTM consolidation probably mimics the encoding order in perception. Further support to this model comes from the masking effect in experiment 3. As we found there, preventing additional encoding or consolidation in VSTM at the offset of the stimulus reduced the recognition of HSF-probes without changing performance with LSF-probes. This indicates that by 800 ms the LSF had been transferred into VSTM, whereas HSF were still being processed.

This sequential face-encoding mechanism may fit well into the framework of a more general model of VSTM which has been suggested on the basis of neuroimaging studies (Xu & Chun, 2006, 2007, 2009). According to that model, visual information is accumulated and represented in VSTM in two stages. First, the visual system selects a fixed number of up to four objects from a crowded scene based on their spatial information. Crucially, at this stage objects are represented only at a low resolution which does not reveal their complexity. During a second stage, high resolution information is added to the VSTM representations, flexibly changing its capacity according to the objects' complexity. In a similar vein, two recent studies showed that highly distinguishable simple features, which have already been segmented into individual low-resolution objects by a parallel perceptual process, are encoded into VSTM automatically. High-resolution information is fed subsequently enriching the coarse representations with details, as a result of a serial, resource-consuming perceptual process under top-down control (Gao, Shen, Gao, & Li, 2008; Gao et al., 2009). These data are congruent with the current findings and, in concert, support a coarse-to-fine construction process which might be characteristic to all visually complex stimuli. To this end, it is possible that the "all-or-none" model reflects the first stage of encoding simple stimuli (defined by single features) that can be sufficiently and distinctly represented at a coarse level. Indeed, VSTM studies on which the all-or-none model was based actually used low-resolution information which is resolved at the parallel processing stage of perception (Wang, 2001; Zhang & Luck, 2008; see also the discussion in Bays, Catalao, & Husain, 2009).

Turning our attention to the second goal of this study, we found no evidence for a differential decay of high- and low-frequency ranges. Although decay from memory was found in experiment 2 and 4, the reduction in performance did not interact with SF. This outcome is in line with earlier research which demonstrated that all SFs, once encoded in memory, are similarly robust to decay (Lee & Harris, 1996; Magnussen, 2000; Magnussen & Greenlee, 1999; Magnussen et al., 1990; Regan, 1985). Our results extend these findings from Gabor gratings as used in those studies to more complex visual stimuli such as faces. Hence, we suggest that in contrast to the coarse-to-fine order of encoding, regardless of whether the representation fades gradually (e.g., Harvey, 1986; Gold et al., 2005), randomly (Kinchla & Smyzer, 1967) or in an all-or-none fashion as suggested by the recently proposed "sudden death" model (Zhang & Luck, 2009), the decay from VSTM is not governed by SF. By implication, this hypothesis could be extended to claim that although it makes sense to assume that details are lost from memory first, the decay from VSTM is not determined by the level of visual resolution represented in the memory representation.

Before concluding we should consider two caveats. The first is that we used a categorical design to distinguish between LSF and HSF, whereas, in reality, SFs are, indeed, a continuous variable. Although we used a theoretical approach to select the ranges of LSF and HSF, the boundaries were arbitrary to some extent. Moreover, as our post hoc analysis of amplitude spectra demonstrated that the spectra of the HSF and LSF stimuli were not distinguished categorically (cf. Peli, 1992; Pointon, 1993). To address this caveat future extension of the current outcome should probably take a parametric manipulation (e.g., Gaspar et al., 2008). The second caveat is that the generalization of the present data to other stimulus categories should be attempted with caution since the selection of filters was based on the range most relevant for face identification. Furthermore, the faces used in the current research did not contain outer facial features, such as hair or a natural contour. Although such stimuli allowed a better control, they are different from faces that we usually remember in the normal life. Moreover, there is evidence that inner and outer face features might be processed differently in the brain (e.g., Zion-Golumbic & Bentin, 2007).

Notwithstanding these caveats, the outcome of this study leads to three important points: First, information is transferred from perception to VSTM gradually starting before the perceptual representation is fully assembled and continuing after the visual stimulus is removed from sight. Second, probably as a result of this cascade-type processing, the spatial frequencies comprising the perceptual representation are encoded into VSTM gradually, with low- spatial frequencies preceding high -spatial frequencies. Third, the decay from VSTM occurs in parallel for HSF and LSF, at least during the first 10 seconds after to-be-stored visual images are removed from the visual field. Additional parametric studies are needed to unveil the exact timing of these processes.

References

- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*, 622–628.
- Bachmann, T. (1991). Identification of spatially quantised tachoscopic images of faces: How many pixels does it take to carry identity? *European Journal of Cognitive Psychology*, *3*, 85–103.
- Baddeley, A. (1981). The concept of working memory: A view of its current state and probable future development. *Cognition*, *10*, 17–23.
- Baddeley, A. (1996). The fractionation of working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 13468–13472.
- Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829–839.
- Baddeley, A. D. (1983). *Working memory. Philosophical Transactions of the Royal Society of London, B*, *302*, 311–324.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (pp. 647–667). New York, NY: Academic Press.
- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*, 617–629.
- Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*, 7, 1–11.
- Bentin, S., Golland, Y., Flevaris, A., Robertson, L. C., & Moscovitch, M. (2006). Processing the trees and the forest during initial stages of face perception: Electrophysiological evidence. *Journal of Cognitive Neuroscience*, *18*, 1406–1421.

- Bruce, V., & Humphreys, G. W. (1994). Recognizing objects and faces. In V. Bruce & G. W. Humphreys (Eds.), *Object and face recognition* (pp. 141–180). Hove, England: Erlbaum.
- Busey, T. A., & Loftus, G. R. (1994). Sensory and cognitive components of visual information acquisition. *Psychological Review*, *101*, 446–469.
- Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 109–127.
- Collin, C. A., Liu, C. H., Troje, N. F., McMullen, P. A., & Chaudhuri, A. (2004). Face recognition is affected by similarity in spatial frequency range to a greater degree than within-category object recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 975–987.
- Coltheart, M. (1980). Iconic memory and visible persistence. *Perception and Psychophysics*, *27*, 183–228.
- Cornelissen, F. W., & Greenlee, M. W. (2000). Visual memory for random block patterns defined by luminance and color contrast. *Vision Research*, *40*, 287–299.
- Costen, N. P., Parker, D. M., & Craw, I. (1994a). Spatial content and spatial quantisation 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.
- Costen, N. P., Shepherd, J. W., Ellis, H. D., & Craw, I. (1994b). Masking of faces by facial and non-facial stimuli. *Visual Cognition*, *1*, 227–251.
- Curby, K. M., & Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychonomic Bulletin & Review*, *14*, 620–628.
- Curby, K. M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 94–107.
- DeValois, R. L., & DeValois, K. K. (1988). *Spatial vision*. New York, NY: Oxford University Press.
- Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review*, *12*, 1127–1133.
- Fiorentini, A., Maffei, L., & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception*, *12*, 195–201.
- Flevaris, A. V., Robertson, L. C., & Bentin, S. (2008). Using spatial frequency scales for processing face features and face configuration: An ERP analysis. *Brain Research*, *1194*, 100–109.
- Gao, T., Shen, M., Gao, Z., & Li, J. (2008). Object-based storage in visual working memory and the visual hierarchy. *Visual Cognition*, *16*, 103–106.
- Gao, Z., Li, J., Liang, J., Chen, H., Yin, J., & Shen, M. (2009). Storing fine detailed information in visual working memory—Evidence from event-related potentials. *Journal of Vision*, *9*, 17, 1–12.
- Gaspar, C., Sekuler, A. B., & Bennett, P. J. (2008). Spatial frequency tuning of upright and inverted face identification. *Vision Research*, *48*, 2817–2826.
- Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Essays in cognitive psychology. Hove, UK: Lawrence Erlbaum Associates.
- Ginsburg, A. P. (1986). Spatial filtering and visual form perception. In K. R. BofL, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 34–1-34–44). New York, NY: Wiley.
- Goffaux, V., Gauthier, I., & Rossion, B. (2003). Spatial scale contribution to early visual differences between face and object processing. *Cognitive Brain Research*, *16*, 416–424.
- 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., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2011). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral Cortex*, *21*, 467–476.
- 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.
- Gold, J. M., Murray, R. F., Sekuler, A. B., Bennett, P. J., & Sekuler, R. (2005). Visual memory decay is deterministic. *Psychological Science*, *16*, 769–774.
- Halit, H., de Haan, M., Schyns, P. G., & Johnson, M. H. (2006). Is high-spatial frequency information used in the early stages of face detection? *Brain Research*, *1117*, 154–161.
- Harel, A., & Bentin, S. (2009). Stimulus type, level of categorization and spatial-frequencies utilization: Implications for perceptual categorization hierarchies. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1264–1273.
- Harvey, L. D. (1986). Visual memory: What is remembered? In F. Klix & H. Hagedorf (Eds.), *Human memory and cognitive capabilities* (Vol. 1, pp. 173–187). The Hague, The Netherlands: Elsevier Science.
- Hayes, T., Morrone, M. C., & Burr, D. C. (1986). Recognition of positive and negative bandpass-filtered images. *Perception*, *15*, 595–602.
- Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*, 791–804.
- Hollingworth, A., Richard, A. M., & Luck, S. J. (2008). Understanding the function of visual short-term memory: Transsaccadic memory, object correspondence, and gaze correction. *Journal of Experimental Psychology: General*, *137*, 163–181.
- Howard, M. F., & Reggia, J. A. (2007). A theory of the visual system biology underlying development of spatial frequency lateralization. *Brain and Cognition*, *64*, 111–123.
- Hughes, H. C., Fendrich, R., & Reuter-Lorenz, P. A. (1990). Global versus local processing in the absence of low spatial frequencies. *Journal of Cognitive Neuroscience*, *2*, 272–282.
- Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statics of natural images. *Journal of Cognitive Neuroscience*, *8*, 197–230.
- Irwin, D. E., & Yeomans, J. M. (1986). Sensory registration and informational persistence. *Journal of Experimental Psychology: Human Perception and Performance*, *12*, 343–360.
- Ivri, R. B., & Robertson, L. C. (1998). *The two sides of perception*. Cambridge, MA: MIT Press.
- Jiang, Y., Makovski, T., & Shim, W. M. (2009). Visual memory for features, conjunctions, objects, and locations. In J. R. Brockmole (Eds.), *The visual world in memory* (pp. 33–65). Hove, UK: Psychology Press.
- Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, *59*, 193–224.
- Kinchla, R. A., & Smyzer, F. (1967). A diffusion model of perceptual memory. *Perception & Psychophysics*, *2*, 219–229.
- Lee, B., & Harris, J. (1996). Contrast transfer characteristics of visual short-term memory. *Vision Research*, *36*, 2159–2166.
- Legge, G. E., & Gu, Y. C. (1989). Stereopsis and contrast. *Vision Research*, *29*, 989–1004.
- Liu, C. H., Collin, C. A., Rainville, S. J., & Chaudhuri, A. (2000). The effects of spatial frequency overlap on face recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 956–979.
- Loffler, G., Gordon, G. E., Wilkinson, F., Goren, D., & Wilson, H. R. (2005). Configural masking of faces: Evidence for high-level interactions in face perception. *Vision Research*, *45*, 2287–2297.

- Loftus, G. R., & Hanna, A. M. (1989). The phenomenology of special integration: Data and models. *Cognitive Psychology*, *21*, 363–397.
- Loftus, G. R., & Harley, E. M. (2004). How different spatial-frequency components contribute to visual information acquisition. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 104–118.
- Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic Bulletin & Review*, *12*, 43–65.
- Loftus, G. R., & Irwin, D. E. (1998). On the relations among different measures of visible and informational persistence. *Cognitive Psychology*, *35*, 135–199.
- Logie, R. H. (1995). *Visuo-spatial working memory*. Essays in cognitive psychology. Hove, UK: Lawrence Erlbaum Associates.
- Luck, S. J. (2008). Visual short-term memory. In S. J. Luck & A. Hollingworth (Eds.), *Visual memory* (pp. 43–85). New York, NY: Oxford University Press.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.
- Magnussen, S. (2000). Low-level memory processes in vision. *Trends in Neurosciences*, *23*, 247–251.
- Magnussen, S., & Greenlee, M. W. (1999). The psychophysics of perceptual memory. *Psychological Research*, *62*, 81–92.
- Magnussen, S., Greenlee, M. W., Asplund, R., & Dyrnes, S. (1990). Perfect visual short-term memory for periodic patterns. *European Journal of Cognitive Psychology*, *2*, 345–362.
- Magnussen, S., Greenlee, M. W., Asplund, R., & Dyrnes, S. (1991). Stimulus-specific mechanism of visual short-term memory. *Vision Research*, *31*, 1213–1219.
- Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, *57*, 203–220.
- McSorley, E., & Findlay, J. M. (1999). An examination of a temporal anisotropy in the visual integration of spatial frequencies. *Perception*, *28*, 1031–1050.
- Merigan, W. H., Byrne, C. E., & Maunsell, J. H. (1991a). Does primate motion perception depend on the magnocellular pathway? *Journal of Neuroscience*, *11*, 3422–3429.
- Merigan, W. H., Katz, L. M., & Maunsell, J. H. (1991b). The effects of parvocellular lateral geniculate lesions on the acuity and contrast sensitivity of macaque monkeys. *Journal of Neuroscience*, *11*, 994–1001.
- Morgan, M. J. (1992). Spatial filtering precedes motion detection. *Nature*, *355*, 344–346.
- Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin & Review*, *8*, 454–469.
- Nasanen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research*, *39*, 3824–3833.
- Oliva, A., & Schyns, P. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, *34*, 72–107.
- 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.
- Peli, E. (1992). Display nonlinearity in digital image processing for visual communications. *Optical Engineering*, *31*, 2374–2382.
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*, 283–290.
- Pointon, C. A. (1993). “Gamma” and its disguises: The nonlinear mappings of intensity in perception, CRTs, film, and video. *SMPTE Journal-Society of Motion Picture & Television Engineers*, *102*, 1099–1108.
- Regan, D. (1985). Storage of spatial-frequency information and spatial-frequency discrimination. *Journal of the Optical Society of America A*, *2*, 619–621.
- Richler, J. J., Tanaka, J. W., Brown, D. D., & Gauthier, I. (2008). Why does selective attention to parts fail in face processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1356–1368.
- Ruiz-Soler, M., & Beltran, F. S. (2006). Face perception: An integrative review of the role of spatial frequencies. *Psychological Research*, *70*, 273–292.
- Salmela, V. R., Makela, T., & Saarinen, J. (2010). Human working memory for shapes of radial frequency patterns. *Vision Research*, *50*, 623–629.
- Schyns, P., & Oliva, A. (1997). Flexible diagnosticity, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception*, *26*, 1027–1038.
- Schyns, P. G. (1998). Diagnostic recognition: Task constraints, object information, and their interactions. *Cognition*, *67*, 147–179.
- Schyns, P. G., & Gosselin, F. (2003). Diagnostic use of scale information for componential and holistic recognition. In M. E. Peterson & G. Rhodes (Eds.), *Perception of faces objects and scenes: Analytic and holistic processes* (pp. 120–145). Oxford: Oxford University Press.
- 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.
- Schyns, P. G., Petro, L. S., & Smith, M. L. (2007). Dynamics of visual information integration in the brain for categorizing facial expressions. *Current Biology*, *17*, 1580–1585.
- Scolari, M., Vogel, E. K., & Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychonomic Bulletin & Review*, *15*, 215–222.
- Sergent, C., & Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychological Science*, *15*, 720–728.
- Shulman, G. L., & Wilson, J. (1987). Spatial frequency and selective attention to local and global information. *Perception*, *16*, 89–101.
- Sowden, P. T., & Schyns, P. G. (2006). Channel surfing in the visual brain. *Trends in Cognitive Sciences*, *10*, 538–545.
- Torralba, A. (2009). How many pixels make an image? *Visual Neuroscience*, *26*, 123–131.
- Verde, M. F., Macmillan, N. A., & Rotello, C. M. (2006). Measures of sensitivity based on a single hit rate and false alarm rate: The accuracy, precision and robustness of d' , A_z and A' . *Perception & Psychophysics*, *68*, 643–654.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 92–114.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1436–1451.
- Wang, X. J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in Neurosciences*, *24*, 455–463.
- Watt, R. J. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *Journal of the Optical Society of America A, Optics and Image Science*, *4*, 2006–2021.
- Wenger, M. J., & Townsend, J. T. (2000). Spatial frequencies in short-term memory for faces: A test of three frequency-dependent hypotheses. *Memory and Cognition*, *28*, 125–142.
- Westheimer, G. (2001). The Fourier theory of vision. *Perception*, *30*, 531–541.
- Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*, 48–64.
- Wickelgren, I. (1997). Getting a grasp on working memory. *Science*, *275*, 1580–1582.
- Wong, J. H., Peterson, M. S., & Thompson, J. C. (2008). Visual working memory capacity for objects from different categories: A face-specific maintenance effect. *Cognition*, *108*, 719–731.

- Xu, Y. (2002). Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 458–468.
- Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*, 91–95.
- Xu, Y., & Chun, M. M. (2007). Visual grouping in human parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 18766–18771.
- Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in Cognitive Sciences*, *13*, 167–174.
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235.
- Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science*, *20*, 423–428.
- Zion-Golombic, E., & Bentin, S. (2007). Dissociated neural mechanisms for face detection and configural encoding: Evidence from N170 and Gamma-band oscillation effects. *Cerebral Cortex*, *17*, 1741–1749.

Received March 4, 2010

Revision received September 7, 2010

Accepted September 14, 2010 ■