Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli

Aude Oliva

Philippe G. Schyns

Dept. of Psychology

University of Glasgow

Glasgow G12 8QB

philippe@psy.gla.ac.uk

in press, Cognitive Psychology

Please send all correspondence concerning this manuscript to Philippe Schyns.

HEADING TITLE: Flexible scale usage in scene categorization

ABSTRACT

Efficient categorizations of complex visual stimuli require effective encodings of their distinctive properties. However, the question remains of how processes of object and scene categorization use the information associated with different perceptual spatial scales. The psychophysics of scale perception suggests a scenario in which recognition uses coarse blobs before fine scale edges, because the former is perceptually available before the latter. Although possible, this perceptually determined scenario neglects the nature of the task the recognition system must solve. If different spatial scales transmit different information about the input, an identical scene might be flexibly encoded and perceived at the scale that optimizes information for the considered task.

This paper tests the hypothesis that scale diagnosticity can determine scale selection for recognition. Experiment 1 tested whether coarse and fine spatial scales were both available at the onset of scene categorization. The second experiment tested that the selection of one scale could change depending on the diagnostic information present at this scale. The third and fourth experiments investigated whether scale-specific cues were independently processed, or whether they perceptually cooperated in the recognition of the input scene. Results suggest that a mandatory perception of multiple spatial scales promotes flexible scene encodings, perceptions and categorizations.

Efficient categorizations of complex visual stimuli require effective encodings of their distinctive properties. In the object recognition literature, scene categorization is often portrayed as the ultimate result of a progressive reconstruction of the input scene from simple local measurements. Boundary edges, surface markers and other low-level visual cues are serially integrated into successive layers of representations of increasing complexity, the last of which derives the identity of a scene from the identity of a few objects. For example, in Figure 1, combinations of fine-grained edge descriptors suggest the presence of cars, road panels, highway lamps and other objects which typically compose a highway scene. Precise classifications often require that the identification of component objects from such fine-grained cues precedes the identification of the scene.

However, there is data challenging such exclusive "object-before-scene" recognition scheme. Complex visual displays composed of many partially hidden objects are often recognized quickly, in a single glance--in fact, as fast as a single component object (e.g., Biederman, Mezzanotte, & Rabinowitz, 1982; Potter, 1976; Schyns & Oliva, 1994). This suggests that categorization processes can sometimes directly extract scene representations that allow "express," but comparatively less precise classifications of the input (Henderson, 1992). To illustrate the different routes to scene categorization, squint, blink or defocus while looking at Figure 1, another scene should appear (if this demonstration does not work, step back from the picture until you perceive a city).

INSERT FIGURE 1 AROUND HERE

Figure 1 (adapted from Schyns & Oliva, 1994) illustrates spatial scales, the perceptual output that might be used for precise and express visual categorizations. High Spatial Frequencies (HSF) represent the fine scale highway and Low Spatial Frequencies (LSF) encode the coarse scale city. Although it is now well established that the visual system generally operates at multiple scales, their selection for recognition is still a matter of on-going research. One possibility is that perception extracts the coarse scale before the fine scale and therefore coerces a mandatory coarse-to-fine recognition scheme (the *fixed usage* scenario). Alternatively, the information demands of a categorization task could bias recognition processes to operate at the most informative scale for task at hand (the *flexible usage* scenario). For example, while coarse scale information might be sufficient for an express categorization of Figure 1 as *city*, a more precise (e.g., *New York*) categorization of the same picture might require comparatively finer scale cues.

At an empirical level, the experiments presented in this paper seek to understand whether the fixed, or the flexible usage of spatial scales best accounts for scale-based scene recognition performance. In other words, they seek to understand the structure of the information that might support scene recognition “at a glance.” At a more theoretical level, scale-based scene recognition is used to exemplify another, more important phenomenon: That the way we categorize objects and scenes affects the way we perceive them.

The paper is organized as follows: We first review the long tradition of psychophysical, computational, and psychological studies of scale perception which all suggest a form of coarse-to-fine, fixed scale usage. We then argue that scale-based recognition might be more fruitfully framed as a flexible interaction between the information demands of specific categorizations and the perceptual availability of recognition cues at multiple spatial scales. Four experiments are conducted to test this hypothesis and its implications for recognition and scale perception. We believe that the evidence presented here might reshape conceptions of categorization, perception, and their interactions in future theories of everyday object and scene recognition.

Multi-scale processing in low-level perception. Following Fourier's theorem, linear systems analysis successfully demonstrated that any two-dimensional signal could be analyzed into a sum of sinusoids of different amplitudes, phases and angles, each of which represents the image at a different spatial scale. The 2-dimensional sinusoids composing visual signals are called Spatial Frequencies (SFs, as expressed by a number of cycles per degree of visual angle). A SF channel is a filtering mechanism; something which passes some, but not all, of the SF information it receives (de Valois & de Valois, 1990). A particular channel may transmit all the information that is below a particular spatial frequency (low-pass), above a particular frequency (high-pass), or within a restricted region of frequencies (band-pass). To illustrate, the two scenes composing each picture of Figure 1 roughly correspond to the information transmitted by a low- (the blobs) and a high- (the boundary edges) pass SF channel.

Evidence that perception filters the input with various SF channels originally arose from psychophysical studies on contrast detection and frequency-specific adaptation (see de Valois & de Valois, 1990, for an excellent review of spatial vision). In their seminal paper, Campbell and Robson (1968) demonstrated that the detection and the discrimination of a pattern could be predicted from the contrast of its individual frequency components. As this was only possible if perception was analyzing patterns with independent SF filters, the authors concluded that the visual system comprises groups of independent, quasi-linear band-pass filters, each of which is narrowly tuned to specific frequency bands (see also Graham, 1980; Pantle & Sekuler, 1968; Thomas, 1970; Webster & R. L. de Valois, 1985). Our visual system would "look at" an image through 4 to 6 overlapping SF filters (Ginsburg, 1986; Wilson & Bergen, 1979).

The underlying structure of these channels was the object of frequency-specific adaptation studies. The rationale of these experiments was that an adaptation to pattern X changing the appearance or the sensitivity to X, but not the appearance or sensitivity to pattern Y would indicate that the underlying structures were simultaneously processing independent aspects of the patterns (de Valois & de Valois, 1990). For example, Blackmore and Campbell (1969) showed that people exposed to a sinewave pattern oscillating at, e.g., 5 cycles/deg., exhibited a reduction in their ability to perceive contrast at this particular frequency. That is, adaptation to a SF selectively impaired sensitivity to this particular frequency, as if one channel was affected, but not the others (see also Pantle & Sekuler, 1968).

In summary of these early, but seminal results, there is considerable evidence that perception processes the visual input at different spatial scales, which are functionally described with SF channels. This description is functional because cells which are members of a SF channel are distinguished by their behavior--i.e., they are tuned to specific SFs. So *n* different channels may not necessarily be discrete and fixed structural entities like *n*-core cables. Instead, a channel maybe composed of whatever brain cells that participate in the transmission of information about an input with sufficient contrast within a range of particular SFs (de Valois & de Valois, 1990). Although recent research has shown that channels were interactive (e.g., Henning, Hertz & Broadbent, 1975) and nonlinear (e.g., Snowden & Hammett, 1992), there is little doubt that spatial filtering is *prior to* many early forms of human visual processing such as motion (Morgan, 1992), stereopsis (Legge & Gu, 1989; Schor, Wood & Ogawa, 1984), depth perception (Marshall, Burbeck, Ariely, Rolland & Martin, 1996) and saccade programming (Findlay, Brogan & Wenban-Smith, 1993). Hence, spatial scales are excellent materials to study the influence that higher-level processes such as scene categorization can exert on lower-level perception.

Multi-scale processing in computational vision. A question could be raised of the usefulness of multi-scale processing. If the original image contains all the information required for all its possible categorizations (e.g., “outdoor scene,” “city,” “New York”), what could be the purpose of spatial filters? Studies in computational vision have shown that recognition algorithms can hardly work with the raw pixel values of a digitized image; some qualitative description of the input must be first obtained. Ideally, this initial description should be compact and composed of elements closely corresponding to the important events of the outside world (Witkin, 1986). For scene and object recognition tasks, local minima of the input signal and its derivatives have been frequently used as the first descriptors of a layered reconstruction of the visual scene (e.g., Marr, 1982; Marr & Hildreth, 1980; Watt & Morgan, 1985; among many others). Local minima of the signal and its derivatives are particularly appropriate to describe objects because they can often be directly tied to important contours.

Scale, however, presents difficulties for obtaining good edge descriptions. The varieties of recognition tasks faced by our perception makes it very likely that there is not one particular scale of description that is universal, or intrinsically more interesting or important than any other (think, e.g., of recognizing a plane at the terminal, on the runway, or at 30,000 feet). Existence of important events at different scales led vision researchers to investigate multi-scale representations to organize and simplify the description of events (e.g., Burt & Adelson, 1983; Canny, 1986; Mallet, 1989; Marr & Hildreth, 1980; Watt, 1991; Witkin, 1986; among many others). For example, fine scale boundary edges are notoriously noisy, and they present many confusing details that would not appear in edges measured at a coarser resolution. However, fine details are often still necessary for a complete description of the object, for example to distinguish it from a similar object (see Norman & Erlich, 1987). Boundary edges that would coincide across resolutions could serve as a skeleton describing the coarse structure of an object which would later be fleshed out by the finer structures processed at higher resolutions (see, e.g., Canny, 1986; Mallet, 1991; Marr 1982; Watt, 1987). *Coarse-to-fine processing* summarizes the idea that it may be computationally more efficient to first derive a coarse (albeit imprecise) description of the image before extracting more detailed (but considerably noisier) information.

In summary, computational vision research suggests that no level of resolution is universally better or more important for object and scene recognition. For these reasons, it is often necessary to take multiple measurements of the input at different spatial scales. An efficient strategy starts with coarse measurements and converges on finer measurements.

Coarse-to-fine recognition. Although many psychophysical studies exist that are using simple sinewave stimuli to investigate issues of perceptual scales, comparatively fewer studies have dealt with scale-specific recognition of real-world visual displays (though see, e.g., Costen, Parker & Craw, 1994; Hayes, Morrone & Burr, 1986; Parker, Lishman & Hughes, 1992). In fact, many visual recognition studies (especially those using simple line-drawings of stimuli, e.g., Biederman, 1988; Boyce, Pollatsek & Rayner, 1989) implicitly assume that important processing mostly occurs at finer resolutions. But multi-scale recognition studies have demonstrated that coarse information is often processed before fine information, or that fine information takes longer to be processed. Evidence of coarse-before-fine processing was reported for face (e.g., Breitmeyer, 1984; Fiorentini, Maffei & Sandini, 1983; Sergent, 1982, 1986), object (e.g., Ginsburg, 1986), and scene recognition (Parker et al., 1992; Schyns & Oliva, 1994).

In a related vein, a phenomenon called *global-to-local* has been thoroughly studied since Navon's (1977) influential research (e.g., Kimchi, 1992, for a review; see also Hughes, 1996; Hughes, Nozawa & Kitterle, 1996; Lamb & Yund, 1996a, 1996b; Paquet & Merikle, 1988; Robertson, 1996; among many others). Navon used hierarchical letters similar to those presented in Figure 2. He demonstrated that whereas the global processing of F was not affected by the local Ls, the local processing of Ls was slowed down by incongruent global letters. This asymmetry which suggests that global structures are processed before local structures is called the *global precedence effect*.

INSERT FIGURE 2 AROUND HERE

Global precedence, as many other visual phenomena, has been grounded on coarse-to-fine processing (Badcock, Withworth, Badcock & Lovegrove, 1990; Lamb & Yund, 1996b; Shulman, Sullivan & Sakoda, 1986; Hughes et al., 1996). The claim is that a temporal delay between low and high spatial frequency channels would explain the precedence of global information (e.g., Breitmeyer, 1984; Ginsburg, 1986; Parker et al., 1992; Marr, 1982; among many others). However, the literature has often neglected an important difference between global and local processing, and scale perception.

Figure 2 illustrates this difference. The figure shows a hybrid stimulus composed of two letters, each represented at a different spatial scale. HSF represent F (for Fine), and LSF represent C (for Coarse). The capability to read F and C demonstrates that global processing may occur at the coarse *and* the fine scales. A closer look at Figure 2 should reveal that C and F are both composed of little Ls (for Local). The possibility of reading these shows that local processing can also be accomplished at both scales. Shortly put, Figure 2 shows that coarse-to-fine is a processing mode orthogonal to global-to-local. Global-to-local occurs in the two-dimensional image plane, but coarse-to-fine takes place in another, n-dimensional scale space. To picture the proper relation existing between coarse-to-fine and global-to-local, imagine a third axis orthogonal to the image plane. This axis represents n two-dimensional image planes (one per scale; in Figure 2, n = 2). Relevant recognition information may be extracted locally, or globally from these different scales.

In sum, as illustrated in figures 1 and 2, the point of hybrid stimuli, as opposed to Navon's letters, is that they explicitly multiplex different information in scale space, and therefore allow to study new classes of phenomena (the interactions between scale perception and recognition) that Navon letters were not designed to address. It is the main goal of our experiments to demonstrate that attention and recognition can selectively operate in scale space, on the third axis orthogonal to the 2D image plane of global or local processing. This has important implications for theories of attention and recognition to which we come back in the General Discussion.

Interactions of categorization and perception. Although there is considerable evidence that scale processing is prior to many low-level visual tasks (including motion, saccade programming, depth perception, edge detection, stereopsis and global-to-local) the question remains of how processes of scene and object categorization use the information associated with spatial scales. The reviewed evidence would suggest a fixed usage in which recognition uses coarse blobs before fine scale edges because psychophysics revealed that LSF are perceptually available before HSF (see, e.g., Breitmeyer, 1984; Ginsburg, 1986; Parker et al., 1992; Marr, 1982; among many others).

Although possible, this scenario neglects the nature of the categorization task (and its associated information requirements) the recognition system must solve. If different spatial scales transmit different information about the input, an identical scene might be flexibly encoded at the scale that optimizes the information demands of the classification at hand. Although this assumes that categorization processes can exert an influence on what is often considered to be relatively low-level, encapsulated processes (Fodor, 1983), recent studies on the interactions between categorization and perception have revealed that such influences could indeed exist (e.g., Schyns, Goldstone & Thibaut, in press; Schyns & Murphy, 1991, 1994; Schyns & Rodet, in press). For example, Schyns and Rodet (in press) demonstrated that different subject groups could orthogonally perceive identical stimuli as a result of categorizing and representing them. However, the determinant of these orthogonal perceptions was the creation of different feature vocabularies to represent the abstract shapes, not the usage of different spatial scales for the recognition of realistic scenes. Thus, although there is a growing body of evidence and arguments for the stance that categorization influences perception in very simple tasks (see also Goldstone, 1994), it remains an important empirical challenge to demonstrate that the high-level constraint of using diagnostic recognition information can change the perception of everyday objects and scenes (as is expressed by their spatial scale encodings).

Standard stimuli do not separate spatial scales and so one would never know which scale was used for a particular scene classification. However, as explained earlier, Schyns and Oliva's (1994) hybrid stimuli multiplex scene information in scale space and therefore authorize the investigation of scale-dependent recognition. Early studies (see Schyns and Oliva, 1994) revealed that hybrids (see Figure 1) were preferentially recognized in a coarse-before-fine sequence. In a matching task, brief (30 ms) presentations of these stimuli elicited matchings based on their coarse structures (*city* in Figure 1). Longer (150 ms) presentations of the same stimuli elicited the opposite matchings based on fine structures (*highway* in Figure 1). This effect was reproduced in a categorization task in which an animated sequence of two hybrids was preferentially categorized according to a coarse-to-fine sequence, although the animation simultaneously presented the fine-to-coarse sequence of another scene. However, because these experiments did not test the interactions between different categorization tasks and the perception of multiple scales, they could not distinguish between the perceptual vs. diagnosticity-driven scenarios of scale selection discussed earlier.

The studies reported here investigate whether scale usage is fixed and perceptually determined or whether it is flexible and diagnosticity-driven. The first experiment used hybrids for the visual priming of normal scenes to ensure that both the LSF and the HSF of hybrids were available shortly after the onset of visual processing. The second experiment changed the diagnosticity of the coarse (or the fine) scale of hybrids to test whether diagnosticity would affect the subsequent scale encodings of identical stimuli. The third and fourth experiments tested the perceptual implications of categorizing hybrids at a diagnostic scale. Experiment 3 tested whether the unattended scale of a scene was sufficiently processed that it could prime the explicit categorization of a subsequently presented scene *across* spatial scales. The nature (perceptual or conceptual) of this implicit processing of the unattended scale was the object of Experiment 4.

EXPERIMENT 1

Experiment 1 tests whether coarse and fine scales are both available at the onset of scene categorization. A perceptually determined coarse-to-fine recognition could arise, for example, because fine scale information is delayed in early vision. Even though SF channels supposedly operate in parallel on band-filtered images, speeded conditions of stimulation could affect the recording of fine scale information. In their control stimuli, Schyns and Oliva (1994) showed that the High Spatial Frequencies (HSF) of a scene stimulus presented for only 30 ms were successfully matched with a normal picture of the same scene. However, identical presentations of these HSF composed with Low Spatial Frequencies (LSF) in a hybrid induced a bias for LSF matches. Perhaps the LSF component of hybrids, and more generally the coarse spatial scale of natural scenes might prevent an adequate registration of HSF information--for example because of an inhibition across SF channels (Breitmeyer & Ganz, 1976), or because LSF tend to have a higher contrast than HSF in natural stimuli. Such low-level interferences could naturally promote a perceptual bias for perceiving coarse before fine.

As far as recognition (not low-level perception) is concerned, the important issue is whether a low-level perceptual bias would be so constraining that it would impose a mandatory coarse-to-fine recognition scheme. It is conceivable that the time course of scale perception has little or no influence on the initial scale that used for recognition. In other words, early biases in scale perception might not necessarily translate into the same biases in scale-based recognition.

Experiment 1 seeks to provide evidence that even a very short (30 ms), masked, exposure to one hybrid stimulus successfully primes the categorization of not one, but two scenes--the LSF scene *and* the HSF scene the hybrid represents. Similar LSF and HSF priming rates from the same hybrid images would suggest that both scales similarly constrained recognition processes. Although this data would not strictly rule out a very early perceptual advantage for coarse information, it would considerably diminish its impact on higher-level visual cognition--an aspect of the problem that is too often neglected in studies that generalize from the psychophysics of sine-wave gratings to the recognition of real-world pictures.

Methods

Subjects. Fifteen adult subjects from INPG (Grenoble) volunteered their time to participate to a priming task.

Stimuli. Twelve hybrid stimuli were composed from 4 scene pictures (a highway, a city, a living-room and a valley). Hybrids were composed by systematically mixing the LSF components of a particular scene (below 2 cycles/deg of visual angle) with the HSF components (above 6 cycles/deg of visual angle) of the remaining scenes. The exact procedure for computing the hybrids is presented in Schyns and Oliva (1994). Control stimuli were 4 LSF and 4 HSF, computed respectively by low- and high-passing the 4 Normal (N) scene pictures. Hybrids and control stimuli added to a total of 24 primes. Targets were always N pictures of the scenes. Stimuli subtended 6.38 x 4.38 degrees of visual angle on the monitor of an Apple Macintosh Quadra.

Procedure. The experiment was composed of 120 trials. A trial consisted of the presentation of a prime (either a LSF, HSF, N, LSF-Hybrid or HSF-Hybrid stimulus) for 30 ms, immediately followed by a full-spectrum mask for 40 ms, then by a N target picture of one of the 4 scenes. Subjects' task was to name the target scene as rapidly as they possibly could. It should be noted that a single hybrid stimulus can match with two different target scenes, depending on which component (LSF or HSF) accomplishes the match. A LSF-Hybrid (vs. HSF-Hybrid) denotes the situation in which the LSF (vs. HSF) of one hybrid matches with the target scene. We composed 60 related trials (the prime and the target were of the same scene) and 60 unrelated trials (the prime and the target did not match). We recorded subjects' categorization latencies (reaction times) with a Lafayette vocal key.

Results and Discussion

Priming rates were high in all priming conditions. Respectively N = 72 ms (Related 661 ms, Unrelated 733 ms), LSF = 37 ms (Related LSF 672 ms, Unrelated 709 ms), HSF = 32 ms (Related HSF 682 ms, Unrelated 714 ms), LSF-Hybrid = 25 ms

(Related 696 ms, Unrelated 721 ms), HSF-Hybrid = 47 ms (Related 674 ms, Unrelated 721 ms). Remember that LSF-hybrid and HSF-hybrid are not different stimuli, but the two priming conditions of the same hybrid. A two-way ANOVA (related/unrelated x prime type) revealed a main effect of priming, $F(1, 14) = 39.81$, $p < .0001$, a main effect of prime type, $F(4, 56) = 3.25$, $p < .02$, and a significant interaction, $F(4, 56) = 5.47$, $p < .001$. A significant priming effect was found for each type of prime stimulus (Newman-Keuls, $p < .05$, for all types of prime).

As explained earlier, hybrids have a different interpretation associated with each spatial scale. The results suggest that even a very short (30 ms), masked presentation of one hybrid systematically primed the recognition of not one, but two different scenes. To illustrate, the picture of Figure 1 successfully primed the recognition of *city* (when the target scene was a normal city) *and* the recognition of *highway* (when the target was a highway). An items analysis on the prime stimuli revealed a main effect of priming, $F(1, 11) = 77.97$, $p < .001$, and no effect of items (LSFs, HSFs, Ns, LSF-Hybrids and HSF-Hybrids). Importantly, priming rates were significant for all primes at the .05 level of significance. Moreover, a contrast showed that the priming rate of HSF and HSF-Hybrid was not globally different from that of LSF and LSF-Hybrids, $F(1, 44) < 1$, ns. Finally, a contrast between the LSF-Hybrid and HSF-Hybrid conditions (the two priming conditions of the same hybrid) indicated that the LSF component of an hybrid did not interfere with the HSF component, and *vice versa*, $F(1, 44) = 2.2$, ns.

Since no significant global difference was found between the positive priming rates of coarse and fine scale information presented in isolation, or added in an hybrid stimulus, we can confidently conclude that both spatial scales were available at the onset of visual processing. This suggests that the coarse-to-fine recognition scheme reported for face (e.g., Breitmeyer, 1984; Fiorentini, Maffei & Sandini, 1983; Sergent, 1982, 1986), object (e.g., Ginsburg, 1986), and scene recognition (Parker et al., 1992; Schyns & Oliva, 1994) might not necessarily result from one scale component being perceptually available before the other. Experiment 2 explores the determinants of scale selection for recognition.

EXPERIMENT 2

Even though Experiment 1 provides evidence against a perceptually-driven, coarse-to-fine recognition scheme, we still need to provide positive evidence that classification processes can independently operate with one scale, or the other. We hypothesized that recognition should preferentially use the scale at which task-dependent, diagnostic information is present. It is generally difficult to test this hypothesis on familiar scene categories because one does not know which scale information is diagnostic of which categorization. Furthermore, there is evidence that expertise with object categories can change the features that enter their representations (Schyns & Rodet, in press; Tanaka & Taylor, 1991). Consequently, different individuals might use different scales for an identical categorization. We side-stepped this general difficulty by aiming for an “existence proof” that flexible, diagnostic scale usage existed in visual cognition. Our strategy was simply to assign diagnosticity to the information content of one spatial scale and observe how this would influence the subsequent processing of full-scale hybrid stimuli.

The experiment was a two-phase design in which subjects were asked to categorize hybrid stimuli. In a sensitization phase, two groups of subjects (the LSF and the HSF group) were initially exposed to hybrids that were only meaningful at one scale (either LSF, or HSF), the other scale being structured noise. For example, the top picture of Figure 3 (that we call a LSF/Noise hybrid) shows a city in LSF with structured noise in HSF. The bottom picture (that we call a HSF/Noise hybrid) shows the same city in HSF to which LSF noise is added. We expected that these stimuli would sensitize categorization processes to seek scene cues at the diagnostic scale (either LSF or HSF, depending on the experimental group). The testing phase followed immediately, without any form of transition. Without subjects being aware, the two scale components of the test hybrids were both meaningful (as in Figure 1). If recognition processes flexibly adjusted to seek cues at the diagnostic scale, we should expect mutually exclusive categorizations (LSF vs. HSF) of the test hybrids in the experimental groups, without subjects being aware of the other meaningful scene. This result would provide evidence against a mandatory coarse-to-fine recognition scheme, and it would also provide positive evidence that diagnosticity can flexibly change the scale that is used for recognizing an identical visual input.

Methods

Subject. Twenty-four adult subjects with normal or corrected vision volunteered their time to participate to the experiment. They were randomly assigned to the LSF (vs. HSF) group with the constraint that the number of subjects be equal in each group.

Stimuli. Three types of hybrid stimuli were constructed (LSF/Noise, HSF/Noise and ambiguous) from different pictures of four scene categories (*city*, *highway*, *living-room* and *bed-room*). We synthesized a total of 6 LSF/Noise (vs. 6 HSF/Noise) sensitization stimuli by combining the LSF (vs. HSF) components of two distinct pictures of the categories with HSF (vs. LSF) structured noise (see Figure 3). Test stimuli were ambiguous hybrids, computed as explained earlier by combining the LSF and HSF components of two different scenes. We synthesized a total of 24 hybrids by systematically combining 2 different pictures of 4 distinct categories with the constraint that the two scenes composing each hybrid were of a different category. Hybrids subtended 6.4 x 4.4 deg of visual angle on the monitor of an Apple Macintosh.

INSERT FIGURE 3 AROUND HERE

Each hybrid (sensitization and test) was presented in a brief animation composed of three successive frames--at a rate of 45 ms per frame, to ensure that they fused on the retina (the total presentation time of the animation was 135 ms). The first, second and third frames presented the hybrid with low- and high-frequency cut-off points set at different cycles per/deg of visual angle. In Frame 1, LSF represented all spatial frequencies below 2 cycles/deg, and HSF represented all spatial frequencies above 6 cycles/deg. The cut-offs were changed in Frame 2 and 3 in such a way that they met in Frame 3--i.e., LSF and HSF cut-offs were respectively 3 and 5 cycles/deg in Frame 2, and 4 and 4 cycles/deg in Frame 3.

These animations are very important for different reasons. First, they present a full-spectrum stimulus to visual perception, a requirement that is often missing in studies which reported a coarse-to-fine process (e.g., Badcock et al., 1990; Hughes, 1986; Hughes et al., 1996). Secondly, the animations simultaneously presented a coarse-to-fine and a fine-to-coarse information sequence to the visual system (coarse-to-fine in LSF, and fine-to-coarse in HSF). Studies that directly fed visual cognition with such sequences *all* reported a coarse-to-fine bias (Parker et al., 1992; Schyns & Oliva, 1994). Lastly, testing revealed that this technique produces a brief motion in scale space which "locks" perception on the spatial scale that is selected at the onset of recognition (either LSF or HSF). Very few techniques exist in the literature that allow such locking on specific spatial scales (see, Olzak & Thomas, 1994, for a method using sine-wave gratings). Henceforth, when we refer to hybrids, we mean these brief animations.

Procedure. *Sensitization Phase*. LSF subjects were initially exposed to 6 LSF/Noise, and the HSF group saw 6 HSF/Noise. In a trial, subjects would see one hybrid for 135 ms on a CRT monitor. Order of trials were randomized with a 1.5 s interval between trials. Subjects' task was to categorize the hybrid by saying aloud one of four possible category names. As there was only one meaningful scene in LSF/Noise and HSF/Noise stimuli, subjects could only succeed by attending to the diagnostic scale (LSF or HSF).

Testing Phase. Test stimuli were presented immediately after the sensitization stimuli, without discontinuity in presentation. There are two ways to synthesize one hybrid from two scene pictures, depending on which picture is assigned to the LSF (or HSF) component. Half of the subjects of each group saw one version of the hybrids, and the other half saw the other version. For example, the first half saw LSF city1/HSF highway1 (block A) and the other half saw LSF highway1/HSF city1 (block B). There were 12 hybrids in each block. This strategy ensured a balanced design, without repetition of trials. Note that the pictures used for sensitization were not used for testing. The 12 hybrids of the testing phase were each presented as explained above, and the entire experiment lasted for about 2 minutes. Subjects were instructed to respond as fast and as accurately as they possibly could. We recorded the number of LSF (vs.) HSF categorizations of the 12 ambiguous hybrids in each condition.

Debriefing. After the experiment, we asked subjects several questions about the stimuli. One of these questions was particularly important for the interpretation of the results. Subjects were shown one hybrid stimulus composed of two meaningful scenes and were asked the following question: "Here is a stimulus composed of two scenes. (The experimenter would then point out to the two scenes.) Did you explicitly notice, or did you have the impression that there were such stimuli during the experiment?"

Results and Discussion

To ensure that the blocks A and B of test hybrids did not influence performance, we first ran an ANOVA taking the LSF- vs. HSF-group, block A vs. B and LSF vs. HSF categorizations as factors. As neither the block factor nor the interactions with LSF vs. HSF categorizations were significant, we collapsed the two blocks in each group. Subjects sensitized to the LSF scale categorized 73% of ambiguous hybrids according to their LSF component, while HSF subjects categorized 72% of the same stimuli on the basis of their HSF information. A t-test on the difference score between LSF and HSF categorizations showed a significant difference between the groups, $t(22) = 6.61$, $p < .0001$.

The data revealed mutually exclusive categorizations of identical stimuli[1]. There are at least two possible interpretations of the opposite categorizations. Subjects could simply notice that there were two meaningful scenes in the 12 hybrids, but strategically decide to report only the scale information congruent with their sensitization phase. Another, perhaps more interesting interpretation would propose that the sensitization phase influenced the way stimuli were encoded prior to categorization. That is, although lower-level perception would register the two spatial scales composing hybrids, diagnosticity would bias perceptual encodings towards the informative scale.

In the debriefing phase, one of the questions specifically asked subjects whether they noticed that two meaningful scenes composed a large number of stimuli. 23 subjects (out of 24) reported seeing only one scene. These subjects were surprised to learn that two-thirds of the hybrids were composed of two scenes. Several subjects reported that the scene was perceived as a noisy picture--as if it was observed through a dirty window.

Together, the orthogonal, scale-based categorizations of identical pictures suggest that scale selection for recognition is best understood as flexible and "diagnosticity-driven." It is doubtful that categorizations could be arbitrarily maintained at a single spatial scale (when both scales were meaningful) if scale selection was mandatorily fixed in low-level vision. Instead, it appears that the processes of stimulus encoding were driven by the constraint of using diagnostic scale

[1] An independent control group of 12 subjects was exposed to the 12 ambiguous hybrids without a prior sensitization phase. No subject reported seeing two scenes in the hybrids. Of these subjects, 4 were "HF categorizers" (at least 70% of HF categorizations), 4 were "LF categorizers" (at least 80% of LF categorizations) and 4 subjects categorized equally at both scales. Across subjects, the averages were of 53% LF and 45% HF categorizations (the remaining 2% were errors), $t(11) = 0.47$, ns. These results indicate that without sensitization to a diagnostic scale, categorization processes can independently operate with one scale or the other.

information. Experiment 2 has far-reaching implications for the possible interactions between categorization and perception that we follow-up in experiments 3 and 4.

EXPERIMENT 3

Experiment 1 suggested that coarse and fine information were both registered early in processing. Experiment 2 provided evidence that diagnosticity could flexibly change the scale of stimulus encoding, and that subjects were not aware of the information present at the unattended scale. These results raise a number of new issues for the study of the interactions between categorization and processes of scale perception. For example, there is evidence that subjects' percepts of hybrid stimuli were limited to the content of the diagnostic scale in Experiment 2. However, when subjects categorized the diagnostic scale, did they perceptually register the other scale, or did diagnosticity block the perceptual registration of irrelevant information? If both scales were perceptually registered, could information at the irrelevant scale still influence the processing of the diagnostic scale?

Experiment 3 was designed to understand the nature of the influence of scale diagnosticity on the processes of scale perception. Hybrid stimuli are particularly well-suited to investigate this issue. Because they overlap two scenes at a different scale, it is always possible to sensitize categorization to one scale (as in Experiment 2) and to measure the influence of the unattended component across scales.

Experiment 3 used a priming situation (see Figure 4) in which subjects were asked to categorize a series of hybrid stimuli. Most of the stimuli presented a LSF meaningful scene to which structured noise was added and so we expected categorization to operate at the diagnostic scale, as in the LSF group of Experiment 2. To test the perceptual registration and the influence of the unattended scale, we interleaved a number of ambiguous hybrids in the series of LSF/Noise. Importantly, the HSF component of the ambiguous hybrids on trial *n* (see Figure 4) represented the same scene as the LSF component of the LSF/Noise on trial *n+1*. This identity of scenes (*highway* in the example) across scales was the gist of the priming situation. If subjects registered the HSF of hybrid *n* when they explicitly categorized its LSF, a facilitation (positive priming) effect might be observed for the LSF categorization on trial *n+1.* Such priming would indicate that although diagnosticity maintained categorization at one scale, the perceptual registration and the implicit processing of the other scale was not suppressed.

Methods

Subjects. Subjects were sixteen Grenoble University students who were paid to participate to the experiment.

Stimuli. Stimuli were the same as in Experiment 2: 6 LSF/Noise, 6 HSF/Noise, and 24 ambiguous hybrids generated by combining together 2 pictures of 4 categories (*city*, *highway*, *living-room* and *bed-room*), with the constraint that the two scenes composing a hybrid were of a different category. As in Experiment 2, three-frame animations of the hybrids were presented (at a rate of 45 ms per frame) with 2 and 6 cycles/deg, 3 and 5 cycles/deg and 4 and 4 cycles/deg cut-off points for the first, second and third frame, respectively. Each animation lasted for 135 ms.

Procedure. Stimulus presentation was similar to Experiment 2 with a few differences. Sensitization and ambiguous stimuli were not separated in two different sets but were instead interleaved throughout the experiment. Another major difference was that we explicitly trained subjects on two categorizations of all LSF/Noise hybrids prior to the experiment, to ensure stable base-line categorization latencies of the scenes.

INSERT FIGURE 4 AROUND HERE

Figure 4 details the priming situation. Stimulus *n-1* always was a LSF/Noise whose LSF (*city* in the example) component was identical to the LSF of stimulus *n*. This was meant to facilitate a LSF categorization of *n*, thereby reducing chances of its HSF categorization. The *n* hybrid was always ambiguous. Its HSF component (*highway* on Figure 4) always represented the same scene as the LSF component of the following LSF/Noise stimulus. This allowed the testing of a priming of the HSF of *n* on the LSF of *n+1*, across spatial resolutions (in the example, a priming of HSF *highway*, when people categorize the LSF *city*, on the subsequent LSF *highway* categorization). 1200 ms elapsed between the categorization of one hybrid and the presentation of the next hybrid. Each of the 24 hybrids served as stimulus *n* in composing 24 triples, using the appropriate *n-1* and *n+1* LSF/Noise.

Triples describe the organization of Related (R) trials. Triples were separated from one another with one LSF/Noise stimulus, with the constraint that the scene represented in LSF was different from the scenes composing the hybrids of the following triple. These 24 separators were used to compute Unrelated (UR) trials. UR trials measured the time required to categorize *n+1*, when *n* was a LSF/Noise stimulus--i.e., when there was no scene correspondence across resolutions. The entire experiment was composed of a total of 96 trials (24 triples plus 24 LSF/Noise).

The 96 hybrids were decomposed into three blocks, A, B and C and order of blocks were counterbalanced across subjects. After each block, subjects could rest. Subjects' task was to categorize stimuli (by naming them) as fast and as accurately as they possibly could. We recorded subjects' reaction times with a Lafayette vocal-key, as well as their categorization accuracy.

Debriefing. After the experiment, we asked subjects specific questions about the overall appearance of the stimuli. In one of these questions, subjects were shown a hybrid stimulus composed of two meaningful scenes and were asked the following question: "Here is a stimulus composed of two scenes. Did you explicitly notice, or did you have the impression that there were such stimuli during the experiment?" We also asked subjects how the stimuli looked like, in general.

Results and Discussion

Subjects' categorizations in Experiment 3 mirrored categorizations of the LSF group in Experiment 2, with minor differences. As there were repetitions of trials in

Experiment 3, the proportion of subjects who noticed two scenes in ambiguous hybrids grew to 4 in 16. Their data were discarded from the analysis. In the remaining data, we also removed the triples of hybrids for which a HSF categorization of the *n* stimulus was observed, to ensure that perceptual priming was only measured after explicit LSF categorizations of the ambiguous hybrids. An average of 2 triples (out of 24) were removed per subject.

A 91% average of LSF categorizations of the ambiguous hybrids indicates that subjects' categorizations were reliably maintained at the diagnostic scale. Although subjects systematically encoded and categorized LSF information, the unattended HSF scale of trial *n* primed LSF classification on trial *n+1*. Priming rates were high between R and UR trials (29 ms, R = 566 ms and UR = 595 ms), $t(11) = 4.37$, $p < .01$. It is worth emphasizing that these priming rates were gathered on triples for which (1) subjects denied seeing hybrids composed of two meaningful scenes and (2) for which subjects' categorization behavior controlled the scale at which hybrids were explicitly processed.

The results revealed that the HSF of a picture implicitly facilitated the explicit LSF categorization of the same scene across trials. This demonstrates that while scale diagnosticity controls scale selection for recognition, it does not block the perceptual registration of information at other scales. It is interesting to note that this implicit registration of the irrelevant scale took place when subjects explicitly categorized the relevant information of the other scale. Furthermore, this implicit registration is sufficient to influence the processing of a stimulus presented at another scale 1.2 s later. The nature of this implicit processing is the object of Experiment 4.

EXPERIMENT 4

Experiment 3 suggested that attention to, and explicit categorization of, a diagnostic scale did not prevent the perceptual registration and the implicit processing of the unattended scale. However, the question remains of the nature of this covert processing that was revealed by priming. Priming effects are typically dichotomized into perceptual and semantic (see Farah, 1989, for a review), and it is well-known that any of these can occur without explicit awareness of the prime stimulus (see Holender, 1986; Greenwald, Draine, Abrams, 1996 for reviews, and also Klinger & Greenwald, 1995; Marcel, 1983; Merikle, 1982 for word recognition; Carr, McCauley, Sperber & Parmelee, 1982 for object recognition; Ellis, Young & Koeken, 1993 for face recognition). Unlike objects and words, however, evidence of priming across spatial scales was never demonstrated before Experiment 3. Was this priming conceptual, or perceptual? In other words, did the covert processing of the unattended scale involve recognition, or was it "simply" of a perceptual nature?

Following Shiffrin and Schneider (1977), we might expect that all meaningful information in the two-dimensional visual field that reaches the eye should be covertly recognized. Subjects should then recognize (even if implicitly) the two scenes composing an ambiguous hybrid picture, because both of them were simultaneously present in the 2D visual field. However, the two scenes were represented at a different scale in a space orthogonal to the 2D visual field (cf. the example of Figure 2), and experiments 2 and 3 suggested that subjects only recognized one of these scales at a time. This raises the interesting possibility that explicit and implicit recognition might only use the information associated with one spatial scale, even when meaningful information at another scale co-exists in the 2D visual field.

It was the aim of Experiment 4 to understand whether or not implicit processing at an unattended scale involved recognition. To this end, we used a priming paradigm similar to Experiment 3, but the primes and targets were this time different scene pictures of the same category. Experiment 4 controlled similarity between primes and targets with two independent measures: The similarity judgments of human judges and the similarities measured by a perceptually-based metric. For each of the four experimental categories (*city*, *highway*, *bed-room* and *valley*), judges were instructed to decide which scene exemplars were most perceptually *similar* or *dissimilar* to a target scene. An objective, perceptually-based, translation invariant metric extracted the average energies of horizontal, diagonal and vertical orientations of the scene pictures at multiple scales. This metric was used to distinguish, across spatial scales, correlated and uncorrelated exemplars of the same category.

From the outset, it is important to emphasize that we do not imply that the metric is supposed to capture everything there is to be captured about perceptual similarity in scene categories. Instead, the metric is used as an independent measure that serves to confirm that apparently similar (and dissimilar) scenes might effectively be perceptually similar (and dissimilar). Intuitions about perceptual similarity can be very misleading as the following example illustrates. Imagine your were asked which one of the b, c or d bed-rooms of Figure 5 is most similar to the a bed-room. You would probably select the b picture because both bed-rooms are composed of similar objects, the beds are similarly oriented, and so forth. However, it appears that in terms of the energy of orientations of simple spatial scale filters (to be presented below) the bed-room a is much more similar to c than to b.

INSERT FIGURE 5 AROUND HERE

Subjective and objective similarities were orthogonalized in the experimental design. The similarity relationships between prime and target could be one of four possibilities: judged similar and objectively similar (not shown in Figure 5), judged similar and objectively dissimilar (bed-room b in Figure 5), judged dissimilar and objectively similar (bed-room c in Figure 5), judged dissimilar and objectively dissimilar (bed-room d in Figure 5). Our reasoning was that a covert perceptual processing of the unattended scale--rather than a recognition--would be demonstrated if (1) the subjectively similar and objectively correlated condition elicited positive priming effects, and (2) the subjectively dissimilar and objectively uncorrelated condition did not prime. These two conditions, together with the fact that primes and targets were of the same category, would insure that the covert processing of the unattended scale was not of a conceptual nature--i.e., they did not involve recognition.

Methods

Similarity Metric. The similarity metric computes the correlation between two spatial scales. Typically, in computational vision, computation across spatial scales seeks to correlate the boundaries of blobs with the locations of fine-scale edges (see, e.g., Koenderink, 1984; Lindeberg, 1993; Watt, 1987, 1991; Witkin, 1986). Evidence of such spatial correlations indicates potentially useful object edges. However, a similarity metric based on such correlations would not have the desirable property of being shift-invariant (invariant to translations of the component objects in the image).

An alternative metric could directly use the overall intensity of the global and local orientations of the input signal at different spatial scales. Similarity between LSF and HSF would then be expressed in terms of how "vertical," how "horizontal," and how "diagonal" inputs are at different spatial resolutions, irrespectively of the precise locations of the objects in the scene. The main advantage of such a global, shift-invariant similarity metric is that it allows a direct comparison of different scenes from the same category--because component objects tend to change spatial location from the scene exemplar to the other. The main disadvantage of the metric is that is removes the important information of the spatial locations of major scene components. This metric, based on a Gabor rosace (Daugman, 1985) is detailed hereafter.

INSERT FIGURE 6 AROUND HERE

The computation of the metric starts with the amplitude spectrum of a 256x256 pixels scene (obtained after a Fast Fourier Transform). The amplitude spectrum represents the signal in terms of each component frequency at different orientations. The spectral density of the amplitude spectrum is then obtained by squaring each amplitude value. Spatial frequencies can be independently analyzed in the Fourier domain. That is, it is possible to understand how different spatial components at a different orientation contribute to the entire signal. These relative contributions were computed by covering the squared amplitude spectrum with a family of Gabor filters, to produce a Gabor rosace (Daugman, 1985, see Figure 6).

The Gaussian filters were centered one octave apart (which, from coarse to fine, translates into 1 cycle every 64, 32, 16, 8 , and 4 pixels). For each of the 5 frequency bands considered, four filters were located at 0, 45, 90, and 135 deg of orientation in the two-dimensional amplitude spectrum. Thus, a total of twenty filters (5 frequency bands x 4 orientations) covered the amplitude spectrum (see Figure 6). The convolution of each Gaussian filter with the spectral density computed the energy of the input for each spatial band and orientation. These measurements were summarized in a 20-dimensional vector that represented each scene. We then computed pairwise correlations between the scenes.

Subjects. Twenty-four students from the University of Glasgow were paid to participate to the experiment.

Stimuli. Objective similarity. Four scenes were chosen because they were most typical of each category of the experiment (*highway, city, bedroom* and *valley*). These scenes served as targets for priming. Primes were selected from a total of 120 pictures (30 per category). The selection operated as follows. For each target, the Gabor rosace was computed as explained earlier. The first 12 values of the 20 dimensional representation were extracted to represent the LSF of the targets--the frequencies below or equal to 4 cycles/deg of visual angle. For each potential prime, the last 12 values of the 20 dimensional vectors were extracted to represent the HSF--the frequencies above or equal to 4 cycles/deg of visual angle. A matrix was computed that correlated the LSF 12-dimensional vector of each target with the HSF 12-dimensional vector of each potential prime within the category. Overall, correlations ranged between .99 and -.05. In each category, we then sorted the primes into *highly correlated* (a correlation above .6) and uncorrelated (a correlation below .15).

Subjective similarity. For each category, in the highly correlated set, a group of 8 independent judges were asked to chose two pictures: The most similar, and the most dissimilar to the target. In addition, the judges were also requested to select a similar and a dissimilar scene in the uncorrelated set.

The conjunction of objective and subjective similarities authorized the construction of a contingency table composed of four scene primes per category: SC (similar and correlated), SU (similar and uncorrelated), DC (dissimilar and correlated) and DU (dissimilar and uncorrelated). The actual correlations for each condition of the contingency table for each category are presented in Table 2.

INSERT TABLE 2 AROUND HERE

Hybrid stimuli. Two types of stimuli (ambiguous and LSF/Noise hybrids) were computed, as in Experiment 3. The ambiguous hybrids were computed by systematically mixing the LSF component of one stimulus of the 4 categories with the HSF component of the SC, SU, DC and DU primes of the other categories. To illustrate, the LSF of a city was mixed with the HSF of all the primes of the other three categories. Hence, the LSF component of the city was mixed with 12 prime scenes. The HSF of these ambiguous hybrids on trial *n* served to prime the LSF of the LSF/Noise on trial *n+1* (see Figure 7). The systematic combination of one scene per category with the primes produced a total of 48 ambiguous hybrids. A total of 144 LSF/Noise were produced for this experiment.

Procedure. The priming structure of Experiment 4 was a triple of hybrids, as in Experiment 3 (see Figure 7). That is, the HSF component of the hybrid on trial *n* was a scene of the same category as the LSF component of the hybrid at *n+1*. For example, in Figure 7, the hybrid *n* mixes a bedroom in LSF with a HSF city. The hybrid *n+1* is formed with the LSF of the target city with noisy HSF. The priming condition between the HSF of *n* and the LSF of *n+1* could either be SC, SN, DC, DN or neutral (i.e., noisy

HSF on trial *n*). Hence, 60 triples composed the relevant trials of the experiment (12 SC triples, 12 SN, 12 DC, 12 DN and 12 Neutral). All others trials (180) were random presentations of the 144 different LSF/Noise stimuli. Experiment 4 was composed of a total of 360 trials.

INSERT FIGURE 7 AROUND HERE

In a sensitization phase, subjects were exposed to a total of 72 LSF/Noise hybrids. These were meant (1) to ensure that subjects would lock their categorizations to LSF, (2) to familiarize subjects with the target LSF of the experiment, and (3) to stabilize categorization Reaction Times (RT). In the experimental phase, subjects saw a total of 120 triples of hybrid stimuli that tested all conditions of similarity between HSF primes and LSF targets. Subjects could pause every 8 triples. Subjects were instructed to categorize the scenes as fast and as accurately as they possibly could, by naming the image aloud (possible names were "highway," "city," "room" and "valley"). A vocal key directly linked to a Power Macintosh 7500/100 recorded categorization latencies. To ensure that RT would not be too variable, instructions were given that subjects should keep a rhythm to their naming.

Debriefing: Debriefing was identical to Experiment 2 and 3.

Results and Discussion

Four subjects were removed from the analysis because they noticed that two scenes composed some of the items. Of the 20 remaining subjects, no subjects reported seeing an ambiguous hybrid in the experiment, even if they sometimes named HSF on trial *n* (on average, there were 4% of such HSF categorizations). LSF categorization accuracy was very high (95% correct for *n* trials, and 97% for LSF/Noise stimuli).

A general problem could be raised that in priming situations such as experiments 3 and 4, HSF could interfere when subjects process the LSF of *n* trials. This would result in slower processing of ambiguous hybrids with respect to the speed of processing of LSF/Noise hybrids. However, an analysis of RT did not reveal a significant difference between the average categorization times of ambiguous hybrids (612 ms) and their equivalent LSF/noise (606 ms), $t(19) < 1$, ns. Thus, we can confidently conclude that the meaningful HSF component did not interfere with LSF processing of ambiguous hybrids.

We can now turn to the main objective of Experiment 4, which was to understand the nature of the covert processing of the unattended HSF scale. This can be approached by comparing the priming rates obtained (1) when the similarity relationship changes between the HSF of trial *n* and the LSF of trial *n+1* and (2) when people explicitly categorize LSF (both on trial *n* and *n+1*). In other words, covert processing can be studied when categorization behavior controls the attended scale, and when the similarity between the unattended and attended scales is changed. Priming rate were computed between the Neutral and each of the similarity conditions (SC, SU, DC and DU). A one-way, within-subjects ANOVA with type of prime (Neutral, SC, SN, DC and DN) revealed a main effect of prime type, $F(4,76) = 7.277$, $p < .001$ (priming rates are shown in Table 3).

INSERT TABLE 3 AROUND HERE

Remember that we expected positive perceptual priming when the prime and target were similar according to the judges, and correlated according to the metric (the SC condition). A comparison between SC and neutral revealed a significant positive priming of 29 ms, $F(1,76) = 15.83$, $p < .001$. The other requirement to conclude that the priming is perceptual (as opposed to a conceptual) nature is that no priming should be observed when the judges and the metric found the prime and target from the same category to be dissimilar (the DU condition). A comparison between DU and neutral, revealed no significant effect of priming--12 ms, $F(1, 76) = 2.58$, ns. Hence, we can confidently conclude that the covert processing of HSF, when people explicitly categorized LSF, did not involve recognition--i.e., it was of a perceptual nature.

The remaining conditions of similarity inform further the covert perceptual processing that took place. A positive priming effect was also found in the SU condition, when the prime was judged to be similar to the target, but uncorrelated according to the perceptual metric--26 ms, $F(1, 76) = 21.17$, $p < .001$. In conjunction with the previous results, this confirms that although our similarity metric captures some aspects of similarity, it does not account for all of them. As explained earlier, the metric is shift-invariant, but judges used the apparent correlation of edges in the image in their judgments of perceptual similarity. Our metric was not designed to capture these correlations, but others have argued that they were important in scale processing (e.g., Koenderink, 1984; Lindeberg, 1993; Watt, 1987, 1991; Witkin, 1986). However, our metric captures an aspect of low-level similarity that judges did not perceive at all, and that spatial correlations of edges across resolutions would not capture either. The DC condition (subjectively dissimilar, objectively correlated, see Figure 5, picture d) elicited a strong, but very counter-intuitive priming effect--34 ms, $F(1, 76) = 21.67$, $p < .001$. It should be noted that our metric can explain this perceptual similarity in terms of the global activation of spatial filters across spatial scales (i.e., how vertical, horizontal, and diagonal is the image at different spatial resolutions). This "global similarity" might have influenced explicit categorizations across scales. From a methodological standpoint, however, this condition constitutes a warning for studies which would infer a conceptual priming on the basis of similar evidence. Our data suggest that there are hidden, non-intuitive sources of perceptual similarities which might explain results that would otherwise be taken as evidence of covert processing involving recognition, or other semantic processing.

In sum, experiments 3 and 4 were designed to understand the nature of the influence of scale diagnosticity on the processing of spatial scales. Results suggested that although scale diagnosticity can flexibly change the scale that is used for scene categorization, the unattended component is still perceptually registered and covertly processed. Detailed investigations of this covert processing in Experiment 4 revealed that it was of a low-level, perceptual nature. We discuss the consequences of these results for the relationships between categorization and perception in the General Discussion.

GENERAL DISCUSSION

Scale perception is a low-level perceptual phenomenon which has been shown to occur prior to many early visual tasks such as motion (Morgan, 1992), stereopsis (Legge & Gu, 1989; Schor, Wood & Ogawa, 1984), depth perception (Marshall, Burbeck, Ariely, Rolland & Martin, 1996) and saccade programming (Findlay, Brogan & Wenban-Smith, 1993). Following the psychophysics of sinewave gratings, psychological and computational recognition studies have often assumed that coarse blobs should be recognized before fine boundary edges in complex visual stimuli such as faces, objects and scenes. However, this scenario neglects the information demands of the recognition task, and the influence this might have on "information picking" strategies in scale space. The main objective of this paper was to study an alternative scenario in which scale usage *for recognition* results from an interaction between the high-level constraint of locating diagnostic scale information and the mandatory perception of multiple spatial scales.

Results of Experiment 1 showed that contrary to the idea that scales are mandatorily recognized from coarse to fine, very brief exposures (30 ms) to hybrid stimuli equivalently primed the categorization of two scenes (the LSF and the HSF scenes composing the hybrids). This suggested that the time course of low-level scale perception might impose little constraint on the actual selection of the scale that is used for recognizing the input. Results of Experiment 2 demonstrated that the determinant of scale selection could be the presence of task-dependent, diagnostic information (coarse or fine) at a spatial scale. In Experiment 2, attention was selectively, and implicitly directed to the diagnostic scale, without subjects even being aware of the meaningful information presented at the other scale. Experiment 3 and 4 studied the implications of such "diagnosticity-driven" recognition scheme (Schyns, 1996) on the actual perception and processing of spatial scales. Experiment 3 revealed that subjects who categorized explicitly the diagnostic scale implicitly registered the irrelevant scale, which subsequently influenced explicit recognition. Experiment 4 demonstrated that covert processing at the irrelevant scale did not involve recognition, but was instead of a perceptual nature. Together, experiments 1 to 4 supported our proposal of a flexible, and diagnosticity-driven--rather than a fixed, perceptually determined--usage of spatial scales in visual cognition.

The idea that the information requirements of a categorization task can exert a strong influence on low-level processes such as scale perception raises a number of new issues in visual cognition. They all revolve around the precise nature of the interactions existing between high-level information demands and low-level information constraints in explanations of recognition performance (see also Schyns, 1996; Schyns, Golstone & Thibaut, in press; Schyns & Rodet, in press). The remaining sections discuss these new issues.

Implications for attentional research. Theories of visual attention generally oppose the "spotlight," and the "zoom lens" models. The spotlight model operates in the 2D visual field. Attention is characterized by a diameter (the region of the field that is attended) and it cannot be divided between two regions (see, e.g., Posner 1978, 1980; Posner, Snyder & Davidson, 1980; but see also Pylyshyn & Storm, 1988, for evidence that attention can be divided). The zoom lens model proposes that a wide field of view can be covered when resolution is poor, but that enhancement of resolution narrows down the field of view that is covered (Murphy & Eriksen, 1987). Shortly put, the spotlight model only operates in the 2D visual field and is therefore very similar to the characterization of global-to-local processing discussed earlier, while the zoom lens metaphor simultaneously operates in the visual field *and* in scale space, and is more similar to the conception of attention that emerges from our studies. However, our studies augment this conception of attention with new properties of recognition that are discussed in the next sections.

In experiments 3 and 4, we obtained positive priming effects in a paradigm overlapping stimuli which tends to elicit negative priming effects (e.g., DeSchepper & Treisman, 1996; Tipper, 1985; Tipper & Driver, 1988; Treisman & DeSchepper, in press). In negative priming paradigms, the unattended component of an overlapped stimulus on trial *n interferes* with the recognition of the attended component on trial *n+1*, even when instructions explicitly draw attention to only one component--e.g., "look at the the red objects, not the overlapping green objects", see Allport, Tipper & Chmiel, 1985; DeSchepper & Treisman, 1996; Rock & Gutman, 1981; Tipper & Driver, 1988; Treisman & DeSchepper, in press. However, our results show a facilitation effect in a similar priming situation.

In our studies, as opposed to negative priming studies, subjects were *never* instructed that some stimuli overlapped two meaningful components, and we did not explicitly instruct them to neglect some information. Instead, the implicit constraint of locating recognition information in scale space locked subjects' categorizations to a diagnostic scale. Furthermore, results of Experiment 4 suggested that the unattended scale was *not* recognized. It is therefore less surprising that we observed a facilitation rather than an interference. An interference (negative priming) implies that the neglected aspect of overlapped stimuli is at least implicitly recognized (e.g., Rock & Gutman, 1981; Tipper & Driver, 1988), but again, this was not the case in our studies. Together, results of experiments 2, 3 and 4 suggest (1) that attention can be implicitly drawn to a diagnostic scale in a space orthogonal to the 2D visual field, (2) that we are only able to attend to one scale at a time and (3) that we only recognize the scale we attend to, but that we nonetheless perceptually process the other scale.

These suggestions raise the interesting possibility that attention operates along two orthogonal dimensions. Along the first (and little studied) dimension, attention would be initially driven to the scale that is diagnostic of the recognition task; scale-specific cues would then serve as a basis to recognize the input stimulus. Along the second (and well-studied, see Treisman, 1987; Eriksen & St James, 1986; Eriksen & Yeh, 1985; Paquet & Merilke, 1988) dimension, the attentional window would specify the size of the processed area of the image at the selected scale. Recognition would here operate explicitly in the attentional window, and implicitly outside the attentional window.

Our emphasis on the prior perceptual selection of a spatial scale for recognition is in line with the recent proposal of Phillips and Singer (in press) that context (that we defined as a search for diagnostic, task-dependent information) could tune the selectivity and responsiveness of different spatial filters to modulate the neurophysiological filtering of relevant information. Relatedly, He, Cavanagh and Intriligator (1996, pages 334-335) suggested that "... spatial resolution is limited by an attentional filter acting beyond the primary visual cortex..." and that "... the attentional filter acts in one or more higher visual cortical areas to restrict the availability of visual information to conscious awareness." Experiments 2, 3, and 4 conclusively showed that subjects who categorized the diagnostic scale were unaware of information at the other scale. The actual testing of a diagnosticity-driven attentional mechanism in scale space clearly deserves further research in object and scene recognition.

Implications for object and scene categorization research. The idea that recognition could flexibly pick the scale information best suited to the task at hand has been neglected in recognition theories. Specifying a scale space for recognition, testing its plausibility and studying processing constraints within this scale space introduces new perspectives on visual processing. So far, most recognition theories assume that processing occurs at the finest scales with highly processed information, but our research demonstrated that processing could also use coarse scale, comparatively cruder information.

The structure of scale information for different categorizations of an identical face, object or scene should become an important topic of future recognition research. There is little doubt that different scales limit the nature of the information that can be extracted (e.g., Burt & Adelson, 1983; Lindeberg, 1993). However, it is much less clear how different categorizations of an identical object or scene could utilize this scale information. Research may reveal that very crude information is sufficient to distinguish, e.g., indoor from outdoor scenes, but that comparatively finer spatial cues would be required for a *city* (and even finer for a *New York*) categorization. The metric outlined in Experiment 4 could be used to start to map the hierarchical organization of categories with the hierarchical organization of scale information. When one scale would be shown to subsume a particular categorization, the scale-specific visual cues subserving this categorization could then be studied. It could also be tested how the addition of diagnostic chromatic cues facilitate recognition at different scales (see Oliva & Schyns, 1996). Chromatic cues are scale-specific (in fact, mostly LSF) information.

The research presented in this paper emphasizes the importance of explicitly studying recognition phenomena as interactions between categorization demands and perceptually available information. It borrows to categorization studies the notion of feature diagnosticity; the idea that specific visual cues are used for specific categorizations. Perception research reveal the perceptual materials with which categorization processes interact. The interactions between the information demands of a task and perceptually available information can explain the usage of image cues for object and scene recognition. This "diagnostic recognition" framework has been proposed as a generic approach to explain face, object and scene categorization performance (see Hill, Schyns & Akamatsu, in press; Schyns, 1996). We believe that diagnostic recognition is a necessary step forward in face, object and scene categorization studies.

REFERENCES

Allport, D. A., Tipper, S. P., & Chmiel, N. R. J. (1985). Perceptual integration and post-categorical filtering. In M. I. Posner and O. S. M. Marin (Eds.). *Attention and Performance, XI.* Hillsdale, NJ: Erlbaum.

Badcock, J.C., Whitworth, F.A., Badcock, D.R., & Lovegrove, W.J. (1990). Low-frequency filtering and the processing of local-global stimuli. *Perception*, 19, 617-629.

Biederman, I. (1988). Aspects and extensions of a theory of human image understanding. In Z. W. Pylyshyn (Ed.), *Computational processes in human vision: An interdisciplinary approach.* Norwood, NJ: Ablex.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, 14, 143-177.

Blackemore, C., & Campbell, F. W. (1969). On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology (London)*, 203, 237-260.

Boyce, S.J., Pollatsek, A., & Rayner, K. (1989). Effects of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance.* 15, 556-566.

Breitmeyer, B.G. (1984). *Visual masking: an integrative approach.* New York, Oxford, University Press.

Breitmeyer, B.G., & Ganz, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression and information processing. *Psychological Review*, 83, 1-35.

Burt, P., & Adelson, E. H. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31, 532-540.

Campbell, F. W., & Robson, J. G. (1968). Application of the Fourier analysis to the visibility of gratings. *Journal of Physiology London*, 88, 551-556.

Canny, J. F. (1986). A computational approach to edge detection. *IEEE Pattern Analysis and Machine Intelligence*, 8, 100-105.

Carr, T.H., McCauley, C., Sperber, R.D., & Parmelee, C.M. (1982). Words, pictures and priming: on semantic activation, conscious identification and the automaticity of information processing. *Journal of Experimental Psychology: Human Perception and Psychophysics*, 8, 757-777.

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

Daugman, J.G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of Optical Society of America*, A2, 1160-1169.

DeSchepper, B., & Treisman, A. (1996). Visual memory for novel shapes: implicit coding without attention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22, 27-47.

De Valois, R. L., & De Valois, K. K. (1990). *Spatial Vision*. Oxford University Press: New York.

Ellis, H. D., Young, A. W., & Koenken, G. (1993). Covert face recognition without prosopagnosia. *Behavioral Neurology*, 6, 27-32.

Eriksen, C. W., & St James, S. D. (1986). Visual attention within and around the field of focal attention : a zoom lens model. *Perception and Psychophysics*, 40, 433-458.

Eriksen, C.W., & Yeh, Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 583-597.

Farah, M.J. (1989). Semantic and perceptual priming: How similar are the underlying mechanisms? *Journal of Experimental Psychology: Human Perception and Performance*, 15, 188-194.

Findlay, J. M., Brogan, D., & Wenban-Smith, M. (1993). The visual signal for saccadic eye movements emphasizes visual boundaries. *Perception and Psychophysics*, 53, 633-641.

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

Fodor. (1984). *The modularity of Mind*. Cambridge, MA: MIT Press.

Ginsburg, A. P. (1986). Spatial filtering and visual form perception. In K. R. Boff, L. Kaufman and J. P. Thomas (Eds.). *Handbook of Perception and Human Performance, II: Cognitive Processes and Performance*. NY: Wiley.

Goldstone, R. L. (1994). influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, 123, 178-200.

Graham, N. (1980). Spatial frequency channels in human vision: Detecting edges without edges detectors. In C. S. Harris (Ed.), *Visual Coding and Adaptability.* Hilldsale, NJ: Lawrence Erlbaum.

Greenwald, A. G., Draine, S. C., & Abrams, R. L. (1996). Three cognitive markers of unconscious semantic activation. *Science*, 273, 1699-1701.

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

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, 383, 334-337.

Henderson, J.M. (1992). Object identification in context. The visual processing of natural scenes. *Canadian Journal of Psychology*, 46, 2.

Henning, G. B., Hertz, B. G., & Broadbent, D. E. (1975). Some experiments bearing on the hypothesis that the visual system analyzes spatial patterns in independent bands of spatial frequency. *Vision Research*, 887-899.

Hill, H., Schyns, P. G., & Akamatsu, S. (in press). Information and viewpoint dependence in face recognition. *Cognition.*

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal and visual masking: A survey and appraisal. *Behavioral and Brain Sciences*, 9, 1-66.

Hughes, H.C. (1986). Asymmetric interference between components of suprathreshold compound gratings. *Perception & Psychophysics*, 40, 241-250.

Hughes, H.C., Fendrich, R., & Reuter-Lorenz, P.A. (1990). Global versus local processing in the absence of low spatial frequencies. *Journal of Cognitive Neurosciences*, 2, 272-282.

Hughes, H.C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience*, 8, 197-230

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 604-610.

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: a critical review. *Psychological Bulletin*, 112, 24-38.

Klinger, M.R., & Greenwald, A.G. (1995). Unconscious priming of association judgments. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 569-581.

Koenderink, J.J. (1984), The structure of images. *Biological Cybernetics*, 50, 363-370.

Lamb, M.R., & Yund, E.W. (1993). The role of spatial frequency in the processing of hierarchically organized picture. *Perception & Psychophysics*, 54, 773-784.

Lamb, M.R., & Yund, E.W. (1996a). Spatial frequency and attention: Effects of level-, target- and location-repetition on the processing of global and local forms. *Perception & Psychophysics*, 58, 363-373.

Lamb, M.R., & Yund, E.W. (1996b). Spatial frequency and the interference between global and local levels of structure. *Visual Cognition*, 3, 401-427.

Legge, G. E., & Gu, Y. (1989). Stereopsis and constrast. *Vision Research*, 29, 989-1004.

Lindeberg, T. (1993). Detecting salient blob-like images structures and their spatial scales with a scale-space primal sketch: a method for focus-of-attention. *International Journal of Computer Vision*, 11, 283-318.

Mallet, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Pattern Analysis and Machine Intelligence*, 11, 674-693.

Mallet, S. G. (1991). Zero-crossings of a wavelet transform. *IEEE Information Theory*, 37, 1019-1033.

Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*, 15, 197-237.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Marr, D., & Hildreth, E. C. (1980). Theory of edge detection. *Proceedings of the Royal Society of London, Series B*, 207, 187-217.

Marshall, J. A., Burbeck, C. A., Ariely, J. P., Rolland, J. P., & Martin, K. E. (1996). *Journal of the Optical Society of America A*, 13, 681-688.

Merikle, P.M. (1982) Unconscious perception revisited. *Perception and Psychophysics*, 31, 298-301.

Morgan, M. J. (1992). Spatial filtering precedes motion detection. *Nature*, 355, 344-346.

Murphy, T. D., & Eriksen, C. W. (1987). Temporal changes in the distribution of attention in the visual field in response to precues. *Perception & Psychophysics*, 42, 576-586.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9, 353-383.

Norman, J., & Ehrlich, S. (1987). Spatial Frequency filtering and target identification. *Vision Research*, 27, 87-96.

Oliva, A. & Schyns, P. G. (1995). Mandatory scale perception promotes flexible scene categorization. *Proceedings of the XVII Meeting of the Cognitive Science Society*, 159-163, Lawrence Erlbaum: Hilldsale, NJ.

Oliva, A. & Schyns, P.G. (1996). Color influences fact scene categorization. *Proceedings of the XVIII Meeting of the Cognitive Science Society*. Lawrence Erlbaum: Hilldsale, NJ.

Olzak, L. A., Wickens, D., & Thomas, J. P. (1993). Why we can't see the forest for the trees: Serial processing of disparate spatial frequency bands. *Perception*, 22, 7.

Pantle, A., & Sekuler, R. (1968). Size detecting mechanisms in human vision. *Science*, 162, 1146-1148.

Paquet, L., & Merikle, P.M. (1988). Global precedence in attended and nonattended objects. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 89-100.

Park, J., & Kanwisher, N. (1994). Negative priming for spatial locations: Identity mismatching, not distractor inhibition. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 613-623.

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

Phillips, W.A., & Singer, W. (in press). In search of common foundations for cortical computation. *Behavioral and Brain Sciences*.

Posner, M. I. (1978). *Chronometric explorations of mind.* Lawrence Erlbaum, Hilldsale, N.J.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3-25.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109, 160-174.

Potter, M. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 509-522.

Pylyshyn, Z.W. & Storm, R.W. (1988). Tracking multiples independent targets: evidence for a parallel tracking mechanisms. *Spatial Vision*, 3, 179-197.

Robertson, L.C. (1996). Attentional persistence for features of hierarchical patterns. *Journal of Experimental Psychology: General*, 125, 227-249.

Rock, I., & Gutman, D. (1981). Effects of inattention of form perception. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 275-285.

Schor, C. M., Wood, I. C., & Ogawa, J. (1984). Spatial tuning of static and dynamic local stereopsis. *Vision Research*, 24, 573-578.

Schyns, P.G. (1996). Diagnostic recognition: Task constraints, object information and their interactions. Submitted for publication.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (in press). The development of features in object concepts. *Brain and Behavioral Sciences*.

Schyns, P. G., & Murphy, G. L. (1991). The ontogeny of units in object categories. *Proceeding of the XIII Meeting of the Cognitive Science Society*, 197-202, Lawrence Erlbaum: Hilldsale, NJ.

Schyns, P. G., & Murphy, G. L. (1994). The ontogeny of part representation in object concepts. In Medin (Ed.). *The Psychology of Learning and Motivation*, 31, 305-354. Academic Press: San Diego, CA.

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. & Rodet, L. (in press). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory & Cognition.*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.

Sergent, J. (1982). Theoretical and methodological consequences of variations in exposure duration in visual laterality studies. *Perception and Psychophysics*, 31, 451-461.

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

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a theory. *Psychological Review*, 84, 127-190.

Shulman, G.L., Sullivan, M.A., Gish, K., & Sakoda, W.J. (1986). The role of spatial-frequency channels in the perception of local and global structure. *Perception*, 15, 259-273.

Shulman, G. L., & Wilson, J. (1987). Spatial frequency and selective attention to local and global information. *Perception*, 16, 89-101.

Snowden, R. J., & Hammett, S. T. (1992). Subtractive and divisive adaptation in the human visual system. *Nature*, 355, 248-250.

Tanaka, J., & Taylor, M. E. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology,* 15, 121-149.

Tipper, S. P. (1985). The negative priming effect : inhibitory priming by ignored objects. *The Quartely Journal of Experimental Psychology*, 37 A, 571-590.

Tipper, S. P., & Driver, J. (1988). Negative priming between pictures and words in a selective attention task: Evidence for semantic processing of ignored stimuli. *Memory & Cognition*, 16, 64-70.

Thomas, J. P. (1970). Model of the function of receptive fields in human vision. *Psychological Review*, 77, 121-134.

Treisman, A. (1987). Preattentive processing in vision. *Computer Vision, Graphics and Image Processing*, 31, 156-177.

Treisman, A., & DeSchepper, B. (in press). Object tokens, attention and visual memory. In T.Inui & J.McClelland (Eds.), *Attention and Performance XVI: Information integration in perception and communication.* Cambridge, MA: MIT Press.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97-136.

Ward, L.M. (1982). Determinants of attention to local and global features of visual forms. *Journal of Experimental psychology: Human Perception and Performance*, 4, 562-581.

Watt, R. J. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *Journal of Optical Society of America*, A 4, 2006-2021.

Watt, R. J. (1991). *Understanding vision*. Academic press, London.

Watt, R. J., & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research*, 25, 1661-1674.

Webster, M. A., & De Valois, R. L. (1985). Relationship between spatial frequencies and orientation tuning of striate-cortex cells. *Journal of the Optical Society of America*, *A*, 2, 1124-1132.

Wilson, H.R., & Bergen, J.R. (1979). A four mechanism model for spatial vision. *Vision Research*, 19, 1177-1190.

Witkin, A. (1986). Scale-space filtering. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, 1019-1022. Los Altos, CA: Morgan Kauffman.

Table 1: Percentage of LSF and HSF categorizations of the hybrid stimuli in the two experimental conditions of Experiment 2.

		Categorizations	
Conditions	LF	HF	Error
LF-group	73%	24%	3%
HF-group	24%	72%	4%

Table 2: Correlation values between LSF encodings of the target scene and the four HSF encodings of the primes (SC, SU, DC and DU) in the categories of Experiment 4.

			Categories		
Prime	*City*	*Highway*	*Bed-room*	*Valley*	**Means**
SC	.71	.60	.98	.89	.79
SU	.15	-.07	.13	.14	.09
DC	.83	.67	.71	.71	.73
DU	.06	.07	-.1	- .1	-.02

Table 3: Categorization latencies (in ms) and error rate (in %) in the four conditions (SC, SU, DC, DU and N) of Experiment 4.

			Conditions		
Prime	*SC*	*SU*	*DC*	*DU*	**N**
RT (ms)	585	588	580	602	614
Error (%)	1.25	1.25	0.94	0.94	0.31

ACKNOWLEDGMENTS

The authors wish to thank Pierre Demartines, Jeanny Hérault and Christian Jutten from LTIRF at INPG in Grenoble for useful discussions about stimulus computation. Many thanks to Martha Farah, Gregory Murphy and two anonymous reviewers for helpful comments on an earlier version of this manuscript. Experiments 1 and 2 were presented at the Annual Meeting of the Cognitive Science Society, in Pittsburgh, 1995. Part of this research was funded by an FCAR grant awarded to Philippe G. Schyns.

FIGURE CAPTIONS

Figure 1: This figure (adapted from Schyns & Oliva, 1994) shows an example of a hybrid stimulus used in our experiments. The picture mixes the fine information (High Spatial Frequencies) of a highway with the coarse information (Low Spatial Frequencies) of a city. To perceive the city in Low spatial Frequencies, squint, blink, defocus, or step back from the picture. Hybrid stimuli (see Schyns & Oliva, 1994) are unique because they multiplex information in scale space.

Figure 2: This figure illustrates the important difference existing between coarse-to-fine and global-to-local processing. LSF represent C (for coarse), and HSF represent F (for Fine). The reading of C and F shows that global processing can occur at both the coarse and the fine scales. The reading of the small Ls (for Local) composing the larger letters reveal that local processing is also possible at both scales. Hence, coarse-to-fine and global-to-local occur in different spaces. Global-to-local occurs in the two-dimensional visual field, while coarse-to-fine occurs on a third dimension, orthogonal to the image plane. On this third axis, each resolution of the image represents a different spatial scale (in the figure, there are two such spatial scales).

Figure 3: This figure illustrates the stimuli used in the sensitization phase of Experiment 2. The top picture is a LSF/Noise hybrid composed of the LSF of a city added to structured Noise in HSF. The bottom picture is a HSF/Noise resulting from the addition of the HSF of the same city and LSF structured noise.

Figure 4: This figure illustrates the design of Experiment 3. *n-1* and *n+1* are LSF/Noise hybrids; *n* is an ambiguous hybrid. The succession of these three hybrids illustrates the gist of the cross-frequency priming situation. The HSF component of *n* was the same scene as the LSF component of *n+1*, and so we expected a facilitation (a positive priming) of the LSF *n+1* categorization.

Figure 5: This figure illustrates four of the pictures used to compute the hybrids of Experiment 4. Picture a is the target whose LSF representation was named in a priming task. The three other pictures served to compute the HSF primes in a design

similar to the one presented in Figure 4. Human judges found picture b to be similar to a, but an objective metric rated them as dissimilar. Judges found a and c to be similar, but the metric rated them as dissimilar. Both the judges and the metric found d to be dissimilar to a. These conditions of similarity were among those used as priming conditions in Experiment 4

Figure 6: This figure illustrates the Gabor filtering procedure that served to compute the similarity between scene pictures in Experiment 4. The two axis represent Fourrier space. The circles represent the Gaussian masks at different orientations and scales that were used to convolve the squared amplitude spectrum. The output of this process was a 20-dimensional vector that was used to compute correlations of scenes across spatial scales.

Figure 7: This figure illustrates the priming situation of Experiment 4. Note that the HSF of *n* represent a different exemplar from the category represented in the LSF of *n+1.*


60

(n-1)

(n)

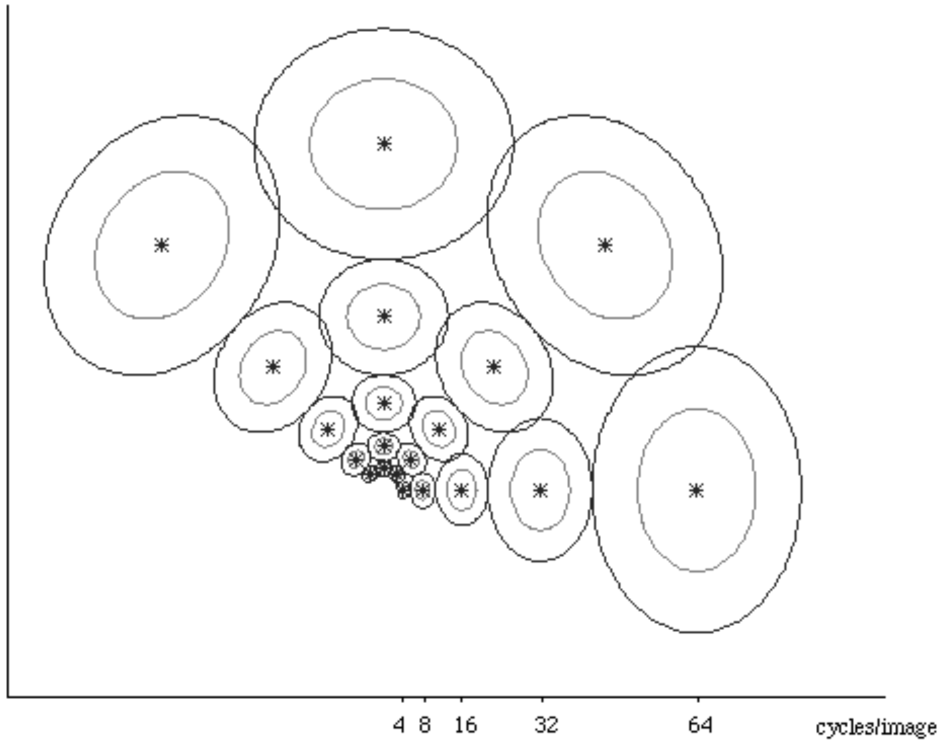
(n+1)

a

b

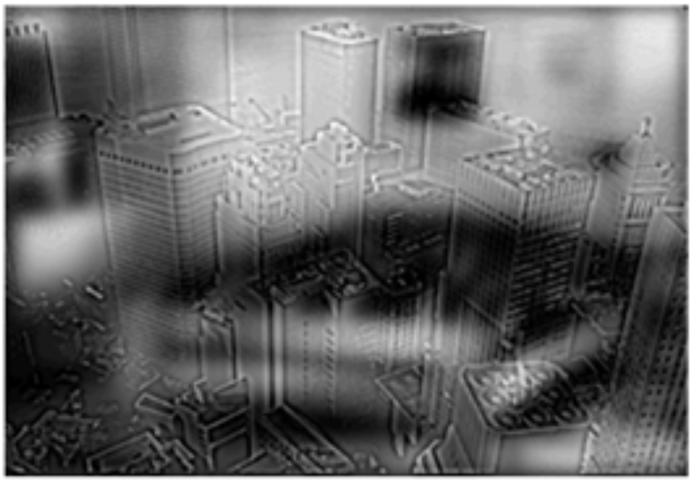
c

d


4
8
16
32
64
cycles/image

(n-1)

(n)



(n+1)