

Running head: FLEXIBLE SCALE USE IN SCENE PERCEPTION

Top-down Attentional Modulation of Spatial Frequency Processing in Scene Perception

Emre Özgen, Paul T. Sowden, Philippe G. Schyns[†] and Christine Daoutis

Department of Psychology, University of Surrey, Guildford, Surrey, GU2 7XH, UK

[†]Department of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow, G12
8QB, UK

Please cite this article as:

Özgen, E., Sowden, P.T., Schyns, P. and Daoutis, C. (2005). Top-down attentional modulation of spatial frequency processing in scene perception. *Visual Cognition*, 12, 925-937.

Abstract

Recent evidence suggests that spatial frequency (SF) processing of simple and complex visual patterns is flexible. The use of spatial scale in scene perception seems to be influenced by people's expectations. However as yet there is no direct evidence for top-down attentional effects on flexible scale use in scene perception. In two experiments we provide such evidence. We presented participants with low and high pass SF filtered scenes and cued their attention to the relevant scale. In Experiment 1 we subsequently presented them with hybrid scenes (both low and high-pass scenes present). We observed that participants reported detecting the cued component of hybrids. To explore if this might be due to decision biases, in Experiment 2, we replaced hybrids with images containing meaningful scenes at uncued SF's and noise at the cued SF's (invalid cueing). We found that participants performed poorly on invalid cueing trials. These findings are consistent with top-down attentional modulation of early spatial frequency processing in scene perception.

Top-down Attentional Modulation of Spatial Frequency Processing in Scene Perception

The use of spatial scale information in the perception of scenes and other complex images shows flexibility. Evidence suggests that spatial frequency (SF) processing might be influenced by a variety of task-dependent factors such as type of categorization (Schyns & Oliva, 1999, Schyns, Bonnar & Gosselin, 2002), sensitisation through repeated exposure to spatial frequency filtered scenes (Oliva & Schyns, 1997), and attention (Sowden, Özgen, Schyns & Daoutis, 2003). For example, Schyns and Oliva (1999) used “hybrid” face stimuli, which contained both a low-pass (LSF) and a high-pass (HSF) SF filtered face in the same image. They found that participants reported seeing only one of the two components of a hybrid face, which depended on the type of face categorization they were asked to carry out—i.e. they reported seeing the LSF face when categorizing gender, and the HSF face when categorizing expressiveness. In a related vein, Oliva and Schyns (1997) repeatedly presented participants with scenes of one type of SF content (either LSF or HSF) combined with noise on the opposite scale, and asked them to categorize them as “city” or “highway”. Subsequently, and unknown to the participants, hybrid scenes were displayed, where both an LSF scene and an HSF scene (one highway, one city) were present. Participants who were sensitised with LSF scenes reported seeing the LSF component, and those sensitised with HSF scenes reported seeing the HSF component of hybrids, but not both.

The precise mechanisms involved in these phenomena are open to study. We propose that sensitisation (resulting from cognitive processes such as categorization, explicit top-down cueing, or restriction of information to a region of the SF spectrum) directs attention to those SF channels in early vision whose output is diagnostic for a given task. We found strong psychophysical evidence that a top-down attentional process can sensitise participants to specific SF channels (Sowden *et al*, 2003). In one experiment, participants made left-right tilt judgments on sinusoidal gratings presented at threshold contrast. We created SF

uncertainty (Davis & Graham, 1981; Davis, Kramer & Graham, 1983; Hübner, 1996a, 1996b) by presenting gratings at one of two possible SF's (e.g. 0.5 and 8 cycles/deg) intermixed randomly, and trained participants to attend to a symbolic sound cue acting top-down (cf. Hübner 1996a, 1996b), signalling the SF of each grating. Subsequently, we interleaved *plaid* stimuli, which consisted of two superimposed gratings (one at each SF) at opposite orientations, to draw an analogy to hybrids. On these trials participants typically reported the orientation of the plaid component corresponding to the cued SF and never perceived both components. This effect of sensitising participants to SF using sound cues is analogous to the sensitisation to spatial scale, resulting from categorization experience, reported by Schyns & Oliva (1999) and Schyns *et al.* (2002) where task cues the participant to attend to information at specific SF's. In further experiments we found that these effects of expectancy on grating detection were selective for SF in a manner similar to the SF channel tuning observed at early stages of visual analysis.

In sum, there is now considerable evidence that a similar process takes place in the perception of our hybrid scenes and gratings. Further, our work on grating discrimination and detection suggests that attention may modulate the activity of early SF channels resulting in the selective perception of these stimuli. Related to this work, Bonnar, Gosselin and Schyns (2002) recently found evidence for bottom-up effects involving flexible use of spatial scale. They adapted participants to low-pass or high-pass dynamic noise presented over the entire display area, and subsequently presented them with an ambiguous image (Salvador Dali's painting of *Slave Market with Disappearing Bust of Voltaire*), which had different perceptions depending on whether the fine (HSF) or coarse (LSF) features were attended to. They found that adaptation with LSF noise resulted in the perception of the HSF features of the ambiguous image and vice versa. It appears therefore that a bottom-up adaptation of SF channels forces participants to use the unadapted channels, and determines which scale

information is perceived in the ambiguous image.

We have also found evidence that the likely locus of sensitisation to spatial scale and flexible scale use, might be early in the visual processing stream (Özgen, Sowden and Schyns, 2002). We repeatedly presented participants with LSF scenes in one part of the visual field, and HSF scenes in another. We observed flexible use of scale when hybrids were presented in line with sensitisation specific to each display location. We demonstrated location specificity across distances as small as about 4° visual angle. This evidence is consistent with an early locus as it is well-known that retinotopic mapping is preserved at early stages of visual analysis (Tootell, Silverman, Switkes & De Valois, 1982; De Valois & De Valois, 1988).

So far sensitisation effects have only been demonstrated with scenes using stimulus driven, or bottom-up designs. It is quite conceivable that when meaningful information is restricted to a band of SF's, the visual system (possibly early vision) sensitises to this band without the influence of cognitive factors. In a critical band noise masking paradigm, Solomon and Pelli (1994) found evidence that a single SF channel processes letters irrespective of the spatial scale of the noise. Recent work by Majaj, Pelli, Kurshan and Palomares (2002) demonstrated that the “*sole determinant*” of which scale observers attend to was stimulus size: a bottom-up process (i.e. they found that the SF channel used by the observer scaled to stimulus size). Consequently, they argue that “*observers are **not** free to choose which channels they use*”. These findings appear to suggest that although observers *can* attend to different spatial scales, this effect of attention is not top-down but driven by the physical properties of the stimulus. The present study addresses this issue by specifically investigating top-down attentional modulation effects on flexible scale use in scene perception.

In two experiments we explored whether top-down cueing can have the effect of

modulating the use of spatial scale much like our previous work with plaids and sine wave gratings (Sowden, Özgen and Schyns, 2003). Hübner (1996a) makes a distinction between iconic and symbolic cueing (similar to Posner's [1980] distinction between exogenous and endogenous orienting of attention). Iconic cues are usually same as the stimulus itself and drive attention bottom-up. Symbolic cues on the other hand are dissimilar to the stimulus, can be cross-modal, and drive attention top-down. Both types of cueing have been shown to reduce uncertainty effects in SF processing (Davis, Kramer and Graham, 1983). We therefore used sound cues to indicate the SF content of each scene, in order to drive a top-down attentional process, by simply manipulating participants' expectations about the stimulus. In both experiments we presented participants with a mixed block of LSF and HSF scenes (highways or cities), combined with noise at the opposite scale, and indicated the scale of each scene by the auditory cues "coarse" (LSF) and "fine" (HSF). The low and high-pass scenes were separated by 1.74 octaves along the SF spectrum ensuring they hit separate SF channels (De Valois & De Valois, 1988). We presented the scenes at threshold contrast, which ensured that the use of cues was necessary. Participants were required to categorise each scene as highway or city. After sufficient cue-dependent sensitisation to spatial scale, we started interleaving catch trials. In Experiment 1 these were hybrids which contained both a LSF and a HSF scene of opposite categories. In Experiment 2, addressing the issue of decision biases, we replaced hybrids with scenes at the opposite scale to that indicated by the cue. In both experiments we observed participants' category reports to indicate which scale they attended to under each cueing condition.

Experiment 1

In this experiment we first estimated participants' luminance contrast threshold for correctly categorising scenes as highway or city, for both low and high pass SF filtered scenes. We then sensitised participants to our sound cues across 80 trials. On each

sensitisation trial, we presented either a LSF scene combined with HSF noise or vice versa. A sound cue indicated the scale of each scene. Participants received feedback on their category responses throughout sensitisation. Subsequently we tested the effects of sensitisation in 64 trials with novel scenes and without feedback. On 25 per cent of the test trials (16 trials) we presented hybrid scenes where both a LSF and a HSF scene of opposite categories were present. We predicted that participants' reports of a scene's category would be biased in accordance with the cue; i.e. the category of the LSF component of a hybrid would be reported more often when the cue was LSF (the sound "coarse") and the category of the HSF component would be reported when the cue was HSF (the sound "fine").

Method

Participants

Twelve people with normal or corrected vision took part. They were paid a fee (5 GBP) for participation.

Stimuli and Apparatus

The stimuli were displayed on an EIZO FlexScan F980 CRT monitor driven by a Cambridge Research Systems 2/5 Visual Stimulus Generator, with a total display area subtending 14.1° (horizontal) by 10.6° (vertical) visual angle. Viewing distance was secured using a head and chin rest at 161cm. Stimuli were displayed in the centre of the screen, on a grey (mean luminance) background and measured 4.5 x 4.5°.

Two sets of 80 grey scale images were used: a "highway" set and a "city" set. In addition 64 noise patterns were constructed using Fourier transformation, by randomly shuffling the phase of SF's at all orientations of the scene images, and averaging a city and a highway image for each noise pattern. This meant that each noise pattern contained frequencies found in both a city and a highway. Image resolution was 250 by 250 pixels.

The scenes and noise patterns were SF filtered to form a low-pass and a high-pass

example of each. The absolute cut-off SF's for low and high-pass filters were 1.5 and 5 cycles/deg respectively. As described in the procedure, two types of stimulus were used in the experiment "scene + noise" (S+N) or *sensitisation* stimuli (see Figure 1a and 1b), and "hybrids" or *test* stimuli (Figure 1c). During the experiment, each stimulus of one of these two types was constructed at run-time by randomly selecting a scene and then combining it with either a noise (S+N) or another scene of the alternative category (hybrid). Run-time combination of images was achieved by alternating between two images at each frame of the monitor's refresh cycle (75Hz each at 150Hz refresh rate). The randomisation had the restriction that hybrids were constructed using novel scenes (i.e. those that have not been used for the S+N stimuli). For sensitisation, 64 scenes in each set were used, while the remaining 16 were used for test trials. For sensitisation trials, S+N stimuli were constructed by combining the 64 scenes in each set with the 64 noise patterns. The low-pass (or low spatial frequency – LSF) example of a scene was combined with a high-pass (HSF) example of a noise pattern or vice versa. This resulted in 128 highway and 128 city S+N stimuli. A hybrid was constructed by combining a LSF highway with a HSF city or vice versa, resulting in 16 hybrid stimuli (each hybrid was constructed with entirely new scenes). Noise patterns were additionally used as full-bandwidth backward masks.

As described in the procedure, stimuli were displayed at varying luminance contrast levels. Contrast manipulation was bipolar around the image and screen mean luminance, and was achieved at run-time by using built-in Visual Stimulus Generator functions.

Procedure

The experiment consisted of three phases: contrast threshold estimation phase, sensitisation phase, and test phase. In all three phases the task was the same; participants were presented with a single stimulus, and were required to categorise it as highway or city. Responses were made by pressing appropriate buttons on a game-pad. Participants were

shown LSF and HSF examples of highways and cities, and were shown 20 practice trials before they started the experiment.

In the first phase, each participant's luminance contrast threshold for correct categorisation of single (no noise added) scenes at a given scale was estimated using the ZEST method (King-Smith, Grigsby, Vingrys, Benes & Supowit, 1994), which converged on an estimate of threshold at 60% correct, with a Beta value of 3.5. Three independent ZEST runs were randomly interleaved, each lasting 48 trials, and the three resulting estimates were averaged to yield the final threshold estimate. One such block of ZEST was completed for each scale (LSF and HSF) separately. The stimuli were randomly selected from among 40 LSF and 40 HSF scenes in each category that were used in the subsequent sensitisation phase. The trial sequence for this procedure was as follows. Participants pressed the space bar to initiate each trial. This was followed by a 1s pause, then the stimulus was shown for 67ms, and finally a randomly selected full-bandwidth noise pattern was displayed as a backward mask with contrast fixed at 50%. Participants had unlimited time to respond.

Before the second phase participants were told that for the remainder of the experiment a sound cue would indicate what type of stimulus they would see in the following trial. Using examples, they were instructed that the sound "coarse" would precede LSF scenes, and the sound "fine" would precede HSF scenes. Participants were told that the images might appear quite faint (low contrast) and that these cues would help them in performing well on the task. However, they were also instructed that some of the time the cues may not be accurate. Following the instructions, 80 trials were completed: 40 LSF scenes (combined with HSF noise) and 40 HSF scenes (LSF noise). These were the same 40 scenes used in threshold estimation. The LSF and HSF components (scene or noise) of the images were presented at each participant's threshold contrast level relevant for that scale. The cues were accurate on all trials. In this phase, participants were presented with a feedback tone indicating their

accuracy. Again, participants initiated each trial by pressing the space bar which was followed by a 1s pause, then the sound cue, and another 1s pause from the onset of the sound cue. This was followed by the stimulus (67ms), and the backward mask (500ms). Participants heard the feedback tone immediately after they made their response.

Participants moved on to the subsequent test phase without a break. There were 64 trials in this phase: 48 S+N trials and 16 randomly interleaved hybrids. For the S+N trials 24 novel (not used in previous phases) scenes in each category were used. As described in stimuli, hybrids were constructed with another new set of 16 scenes in each category. The luminance contrast of the scenes and noise patterns at a given scale was set to each participant's threshold appropriate for that scale as in the previous phase. The trial sequence was same as the previous phase except no feedback was given in the test phase. At the end of the experiment participants were briefed on the nature of the experiment. Here we showed them a hybrid scene and asked them whether they had been explicitly aware of such stimuli.

Results

In the sensitisation phase, participants correctly categorised scenes at an average of 57.50% (2.6). This was significantly different from chance level of 50% ($t(11) = 2.9, p < 0.05$). There was a slight difference between performance on the LSF (55.2% [3.25]) and HSF (59.8% [2.45]) scenes but this was not statistically significant ($t(11) = 1.84, p = 0.094$).

Figure 2 summarises performance in the test phase. The x-axis shows the cue type. The filled bars show the percentage of correct category responses made to the single scene (S+N) stimuli; while unfilled bars show percentage of times when the category of the LSF component of a hybrid was reported. Note that the latter is complementary to per cent HSF reports (LSF reports = 100 - HSF reports). It can be seen that correct categorisation on the LSF and HSF scenes appear to be somewhat different. However, this difference was not statistically significant ($t(11) = 1.48, p = 0.17$). Crucially LSF reports were made more often

when the cue was LSF than when the cue was HSF (57.3% to 42.7%). This difference was statistically significant ($t(11) = 1.9, p < 0.05$). The average of with-prediction responses across the two cue conditions (LSF reports on LSF cue [57.3%], HSF reports on HSF cue [100 - 42.7 = 57.3%]) was significantly different from the no-bias level of 50% ($t(11) = 1.9, p < 0.05$).

Discussion

Just as we predicted, participants' perception of hybrid scenes was influenced by sound cues. When they were cued to attend to the lower end of the SF spectrum, they tended to report seeing the low-pass component scene, and when the cue directed their attention to high SF's, they reported the high-pass scene. This finding is consistent with our hypothesis that SF processing in scene perception can be influenced by top-down attentional modulation.

None of our participants reported being explicitly aware of the existence of hybrid stimuli in the experiment, although some suggested that at times they suspected that there might be two patterns of opposite scale in some images. It should be noted however that there were two patterns of opposite scale in all images throughout the experiment; the second pattern in sensitisation trials was noise and in test trials it was another scene. It is therefore impossible to know, based simply on self-reports, whether participants were in fact able to perceive both components of a hybrid scene and chose to report that which was signalled by the cue. This decision bias account is a strong counter-argument for our attentional modulation account. The next experiment deals with this issue by using invalid cueing on single-scene stimuli rather than hybrids.

Experiment 2

Just as in Experiment 1, we sensitised participants using sound cues and then tested them across 64 trials without feedback. The only difference was that in the test phase (again, 25% of the time) we randomly interleaved trials where the sound cue indicated the spatial

scale that was opposite to that of the scene component of a scene + noise stimulus. We predicted that when such invalid cues were provided participants would perform worse than when the cues were valid.

Method

Participants

Seventeen people with normal or corrected vision took part in the experiment, and were paid a participation fee (5 GBP).

Stimuli and Apparatus

The apparatus was the same as that used in Experiment 1. The same sets of highway, city and noise stimuli were used. However, in this experiment no hybrids were constructed, and only scene + noise (S+N) stimuli were used throughout. The composition of these stimuli followed the same rules. Separate sets of scenes were used in the sensitisation and test phases.

Procedure

The procedure was identical to that of Experiment 1 except for three points. First, through all three phases of the experiment, the noise components of the S+N stimuli were displayed at full contrast. To that end, the initial contrast threshold estimates were made using S+N stimuli (rather than single scenes as it was in Experiment 1) and by manipulating the contrast of only the scene component while holding the noise component at full contrast.

Second, the sensitisation phase consisted of 96 trials: 48 randomly selected filtered (24 LSF, 24 HSF) scenes in each category combined with noise at the opposite scale (these were also used in threshold estimation).

Third, although the test phase consisted of 64 trials as in Experiment 1, S+N stimuli were used throughout. This meant that 32 randomly selected novel scenes (16 on each scale) in each category were combined with noise at the opposite scale. However, on 16 randomly

selected trials in this phase, the sound cue indicated the opposite scale to what was actually shown (8 LSF, 8 HSF).

Results

Correct categorisation performance averaged across the sensitisation phase was at 66.9% (2.1). This was significantly different from 50% chance level ($t(16) = 7.96, p < 0.001$). The LSF and HSF performances in this phase were 70.96% (2.54) and 62.75% (3.5) respectively but this difference did not reach statistical significance ($t(16) = 1.8, p = 0.08$).

Figure 3 summarises performance in the test phase. On the x-axis, stimulus scale is indicated. The filled bars show trials with LSF cues and the unfilled bars show those with HSF cues. It can be seen that when the cue was congruent with stimulus scale participants' categorisation performance was higher than when it was incongruent. An ANOVA with factors "scale" (LSF and HSF) and "cue validity" (valid and invalid) revealed a significant main effect of cue validity ($F(1, 16) = 8.96, p < 0.01$). No other effects were significant in this analysis (largest $F = 0.25$). Performance was equivalent for LSF and HSF scenes on both valid and invalid cueing trials (smallest $p = 0.5$).

Discussion

Our prediction that invalid sound cues would result in poor performance on scene categorisation was supported. When the sound cue indicated the opposite spatial scale to the SF content of a scene, participants performed worse than when the cue was valid. Unlike in Experiment 1 participants were presented with only single scenes and therefore decision biases are less likely to account for the findings of this experiment. The evidence supports our claim that the perception of a SF filtered scene suffers if participants expect to see (and attend to) SF's that are not present in the scene.

General Discussion

In two experiments we explored whether attention channelled by expectations would

influence SF processing in scene perception. We presented participants with low and high-pass (1.74 octaves apart) SF filtered scenes and asked them to categorise them as highways or cities. Participants expected and attended to low or high SF's through the use of symbolic (top-down) cues; a sound cue indicated the spatial scale of each scene. In Experiment 1, we observed that faced with a hybrid stimulus (one that contains a low and a high-pass SF filtered scene) participants reported the component of the hybrid that corresponded to the cued and attended scale. However, it could be argued that this might reflect a decision bias and not attentional modulation of perception; participants might perceive both components but report the one indicated by the sound cue. To address this issue, we replaced hybrids with invalid cueing trials in Experiment 2. We presented participants, 25 per cent of the time, with trials where the sound cue signalled the scale opposite to that of the scene. We observed that categorisation performance dropped on invalid cueing trials.

These findings are consistent with our attentional modulation account of the effects of expectancy on spatial frequency processing in both complex natural stimuli such as scenes and faces, and in simple psychophysical stimuli such as sine wave gratings and plaids. We argue that diagnosticity can act as an attentional filter in the processing of SF information. Particular types of spatial scale information can be diagnostic for a given type of face categorization (Schyns & Oliva, 1999) even at a specific location within the face stimuli (Schyns *et al.*, 2002), or according to what scale meaningful information is restricted to (Oliva & Schyns, 1997), even at a given retinal location (Özgen, Sowden and Schyns, 2002). In the present study the location of diagnostic information in the SF spectrum was signalled top-down with sound cues. A similar process of diagnosticity can also be found in psychophysical studies on uncertainty and expectancy. For each trial during a block of gratings varying in SF, the diagnostic information is located at the specific SF of the grating presented on that trial. Uncertainty about the SF of a sinusoidal grating elevates contrast

threshold for detection or discrimination (Davis *et al.*, 1983; Sowden *et al.*, 2003). However, crucially, a sound cue that indicates the diagnostic scale, makes detection/discrimination more accurate. In all these cases, it is essential to be attending to a particular scale, sometimes at a particular spatial location, in order to accomplish the task. Thus diagnosticity necessitates attention to these parameters. Evidence suggests that attentional modulation of SF channel processing at early stages of visual analysis may indeed be possible; (top-down) attention to the SF of a grating (through the use of symbolic cues) can be restricted to a 1.7 octave range of SF's, suggesting SF tuning such as that typical of early cortical SF channels (Sowden *et al.*, 2003). Task or categorization-based diagnosticity in the perception of complex patterns might thus drive attention to a particular band of SF's, modulating the activity of SF channels in a similar fashion.

These results on the top-down determination of spatial frequency processing can have far reaching consequences for human processes of scene recognition. As already discussed, there is evidence that different sorts of categorization information are available from, and used at different spatial frequency bandwidths and/or orientation and/or image locations (e.g., Oliva & Torralba, 2001; Schyns *et al.*, 2002). This, together with the findings reported here of flexible, retinotopic (Özgen *et al.*, 2002) tuning of spatial frequency channels to different spatial frequencies suggest a flexible account for the extraction of visual information in scene recognition.

It is now well established that the visual system samples little information in a given visual fixation (at least compared with what is available in the real world surrounding the organism, Archambault *et al.*, 1999; Henderson & Hollingworth, 1999; O'Regan, 1992; Rensink, O'Regan & Clark, 1997; Simons & Levin, 1997). This raises the generic question of how the visual system gets optimally tuned to extract the better cues to categorize the world from sparse information. We believe that the data reported here provide some elements of an

answer: The visual system is capable of tuning its information extraction systems (spatial frequency channels) to the spatial frequency bandwidth and image locations that are more informative for the task at hand.

References

- Archambault, A., O'Donnell, C. & Schyns, P. G. (1999). Blind to object changes: when learning the same object at different levels of categorization modifies its perception. *Psychological Science*, **10**, 249-255.
- Bonnar L., Gosselin, F. and Schyns, P. G. (2002) Understanding Dali's Slave Market with the Disappearing Bust of Voltaire: A case study in the scale information driving perception. *Perception*, **31**, 683-691.
- Davis, E. T. and Graham, N. (1981) Spatial frequency uncertainty effects in the detection of sinusoidal gratings. *Vision Research*, **21**, 705-712.
- Davis, E. T., Kramer, P. & Graham, N. (1983) Uncertainty about spatial frequency, spatial position, or contrast of visual patterns. *Perception & Psychophysics*, **33**, 20-28.
- De Valois, R. L. & De Valois, K. K. (1988) *Spatial Vision*. Oxford, UK: Oxford University Press.
- Henderson, J.M. & Hollingworth, A. (1999). High level scene perception. *Annual Review of Psychology*, **50**, 243-271.
- King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C. and Supowit, A. (1994) Efficient and unbiased modifications of the QUEST threshold method: Theory, simulations, experimental evaluation and practical implementation. *Vision Research*, **34**, 885-912.
- Hübner, R. (1996a) The efficiency of different cue types for reducing spatial-frequency uncertainty. *Vision Research*, **36**, 401-408.
- Hübner, R. (1996b) Specific effects of spatial-frequency uncertainty and different cue types on contrast detection: data and models. *Vision Res.*, **36**, 3429-3439.
- Majaj N. J., Pelli D. G., Kurshan P. & Palomares M. (2002) The role of spatial frequency channels in letter identification, *Vision Research*, **42**, 1165-1184.

- Oliva, A. and Schyns, P. G. (1997) Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, **34**, 72-107.
- Oliva, A. & Torralba, A. (2001). Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision*, **42**, 145-175.
- O'Regan, J. K. (1992). Solving the 'real' mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, **46**, 461-488.
- Özgen, E., Sowden, P. T., & Schyns, P. G. (2002). 'It's a highway – no, wait, it's a city': Retinal location specificity of flexible scale use. *Perception*, **31**(supplement), 111.
- Rensink, R. A., O'Regan, J. K. & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes on scenes. *Psychological Science*, **8**, 368-373.
- Schyns, P. G. and Oliva, A. (1999) Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, **69**, 243-265.
- Schyns, P. G., Bonnar, L. and Gosselin, F. (2002) Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, **13**, 402-409.
- Simons, D. J. & Levin, D. T. (1997). Change Blindness. *Trends in Cognitive Sciences*, **1**, 261-267.
- Solomon, J. A. & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature*, **369**, 395-397.
- Sowden, P. T., Özgen E., Schyns, P. G. and Daoutis, C. (2003) Expectancy effects on spatial frequency processing. *Vision Research*, *43*(26), 2759-2772.
- Tootell, R. B. H., Silverman, M. S., Switkes, E. & De Valois, R. L. (1982) Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, **218**, 902-904.

Author Note

Emre Özgen, Department of Psychology, University of Surrey; Paul T. Sowden, Department of Psychology, University of Surrey; Philippe G. Schyns, Department of Psychology, University of Glasgow; Christine Daoutis, Department of Psychology, University of Surrey. This work was supported by BBSRC Grant No's. 90/S13185 and S13186 awarded to Paul Sowden & Philippe Schyns.

Correspondence concerning this article should be addressed to Emre Özgen, Department of Psychology, University of Surrey, Guildford, Surrey, GU2 7XH, Great Britain or electronically to e.ozgen@surrey.ac.uk.

Figure Captions

Figure 1. Examples of stimuli used in the present study; **a**- low-pass highway combined with high-pass structured noise; **b**- high-pass city combined with low-pass structured noise; **c**- a *hybrid*: low-pass highway combined with high-pass city.

Figure 2. Summary of scene categorisation performance in the test phase of Experiment 1. The type of sound cue is shown on the x-axis. Filled bars show percent correct performance on the S+N trials. Unfilled bars the percentage of times when the LSF component of a hybrid was reported. Error bars show ± 1 standard error.

Figure 3. Average percent correct category responses across the four conditions in the test phase of Experiment 2. Stimulus scale is shown on the x-axis and percent correct performance is shown on the y-axis. Filled and unfilled bars show performance when the cue was LSF and HSF respectively. Error bars show ± 1 standard error.

Figure 1

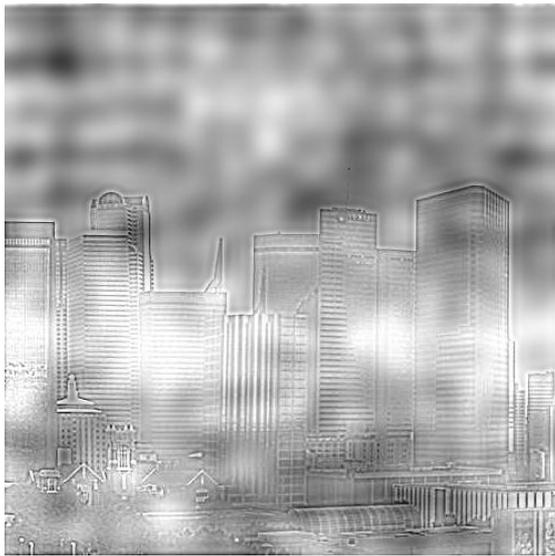
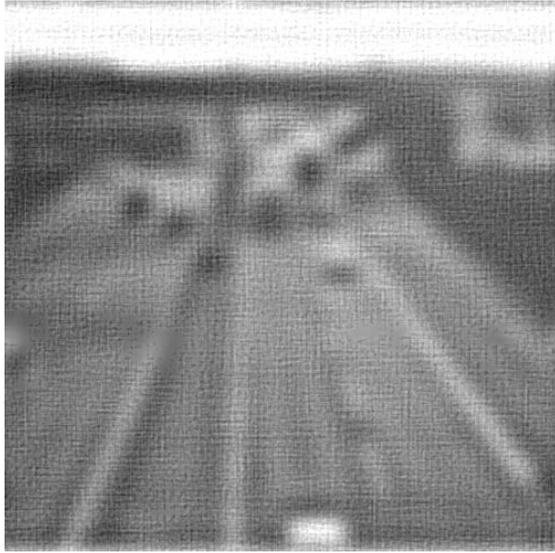


Figure 2

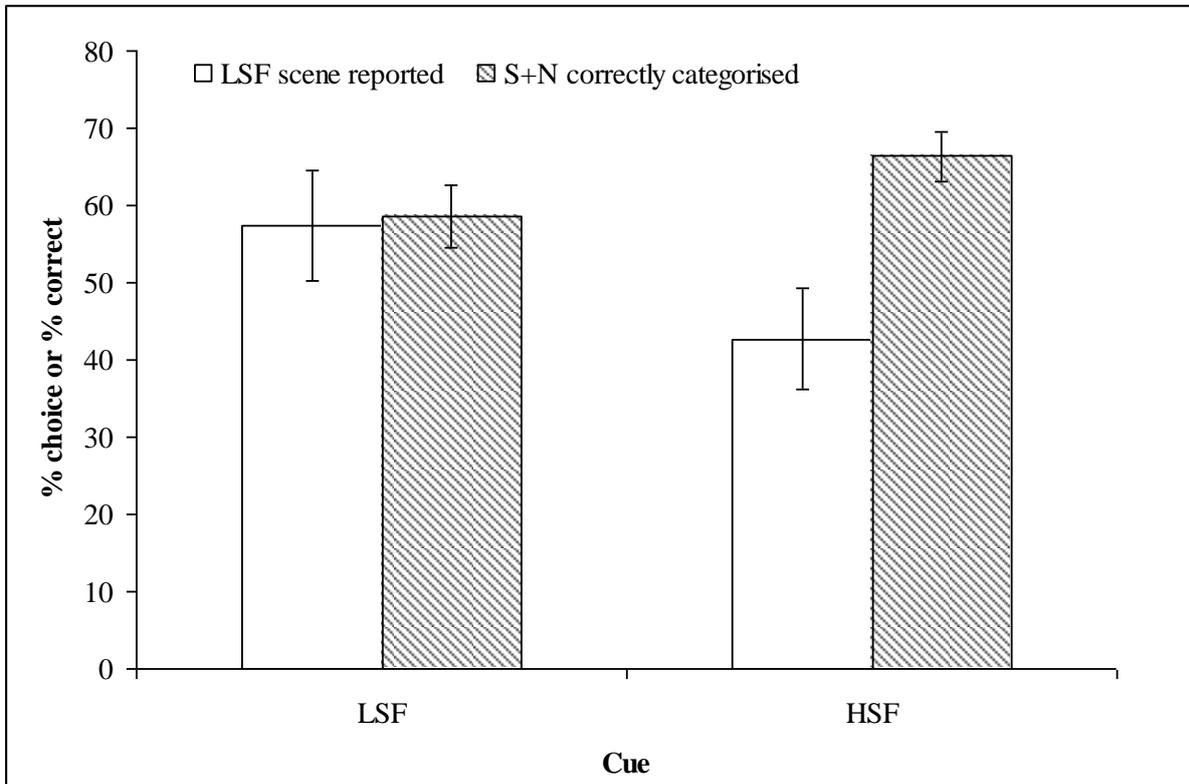


Figure 3

