

# CHANNEL SURFING IN THE BRAIN

Paul T. Sowden<sup>1</sup> and Philippe G. Schyns<sup>2</sup>

1. Department of Psychology, University of Surrey, Guildford, Surrey, GU2 7XH, UK
2. Centre for Cognitive Neuroimaging and Department of Psychology, University of Glasgow, 58, Hillhead Street, Glasgow, G12 8QB, Scotland

Corresponding authors: Sowden, P.T. (p.sowden@surrey.ac.uk), Schyns, P. G. (philippe@psy.gla.ac.uk)

## ABSTRACT

Vision provides us with an ever-changing neural representation of the world from which we must extract stable object categorisations. We argue that visual analysis involves a fundamental interaction between the observer's top-down categorisation goals and the incoming stimulation. Specifically, we discuss the information available for categorisation from an analysis of different spatial scales by a bank of flexible, interacting spatial-frequency (SF) channels. We contend that these channels' activity is not simply determined "bottom-up" by the stimulus. Instead, we argue that following perceptual learning a specification of the diagnostic, *object*-based, SF information dynamically influences the "top-down" processing of *retina-based* SF information by these channels. Our analysis of SF processing provides a case study that emphasises the continuity between higher-level cognition and lower-level perception.

For many, psychophysical research on spatial frequency (SF) conjures up images of dark laboratories and experiments that contribute little to cognition. Further, psychophysicists themselves have often not explicitly considered the use of SF analysis by cognition. Here, we build a bridge between these independent research traditions. We emphasise that SF analysis is a product of *both* the stimulus properties *and* the visual categorisation/recognition goals of the observer and we reconcile conflicting evidence regarding the observers' influence on SF analysis. Further, we argue that, as a function of perceptual learning, categorisation goals exert an increasing and *dynamic* influence on SF processing, even at early visual analysis stages. However, first, in order to understand the interaction between the top-down categorisation goals and the bottom-up stimulus information, we need to consider a distinction between the retinal stimulus SF specification and that for object recognition.

## **BACKGROUND: RECOGNITION AT MULTIPLE SPATIAL SCALES**

Imagine the following scene: you are sitting on a bench waiting for a friend and visually scanning the environment for his familiar face. The wait is not long before you spot a face, from a distance. As the face gets progressively closer, it looks familiar. In fact, your friend is now smiling in your direction.

This everyday event has involved the interaction between visual information from the outside world and recognition processes in the brain. Recognition starts with outside world information projected on the retina

according to the laws of optics. Importantly, retinal image size varies with viewing distance (Figure 1). Consequently, as observers, we are confronted with ever-changing retinal projections of outside world information to recognize. We will bring a book closer to our eyes to resolve its sentences, words or letters, whilst we might step back from a painting or a visual scene to appreciate its global impression. But then, we will also find that a happy face remains happy over many viewing distances, whereas other expressions, such as sadness, can only be recognized from closer up. Our visual system has evolved a bank of spatial frequency (SF) channels (Box 1) to handle the variety of scale information impinging on the retina. However, because the information carried by these channels varies with viewing distance, for stable recognition, the observer requires an additional, object-based, description of the SF's.

#### **BOX 1: SPATIAL FREQUENCY CHANNELS**

The visual system transforms the visual input into a description in terms of light-dark transitions repeating at various frequencies (e.g. see the stripes below Figure 2) across the retina. This is accomplished by a set of SF 'channels' (or filters), a few of which are depicted by the narrower individual curves in Figure 2. Each channel shows 'tuning'; that is preferential sensitivity to one particular frequency and declining sensitivity to increasingly different frequencies. The range of frequencies to which a channel responds defines its 'bandwidth' (indicated by the arrows across one of the channels in Figure 2). This is often

described in octaves, where an octave is a doubling of frequency (e.g. from 2 to 4 cycles-per-degree (c/deg) is a 1 octave step, from 4 to 8 c/deg is another 1 octave step). Channel bandwidths are mostly in the range of 1 – 1.5 octaves [38,39]. It is estimated that there are approximately 6 channels [39], which together define a Contrast Sensitivity Function (CSF, see broad curve in Figure 2). The CSF describes the relationship between SF and sensitivity to luminance contrast.

## **OBJECT-BASED vs. RETINA-BASED SPATIAL FREQUENCIES**

The frequency with which light-dark transitions repeat across an object image defines the *object-based* SF information in that image. These frequencies, measured in cycles-per-object, offer an absolute, stimulus-based measure of the cyclical information present in a stimulus. They represent one measure of the information that the observer might use for visual categorisation.

As demonstrated in Figure 3, a given number of cycles-per-stimulus can be rendered in a large image (e.g. A, B, ...) or in a smaller image (e.g. A', B', ...), depending on how many pixels represent each cycle (henceforth, we refer to cycles-per-object, cycles-per-image, cycles-per-face and cycles-per-stimulus interchangeably, because we will assume that the stimulus fills the image, as in Figure 3).

What happens when the external object cycles are projected onto the observer's two-dimensional retina of light receptors? Psychophysicists measure

these variations with the number of cycles-per-degree of visual angle impinging on the retina —i.e. the number of object cycles in the projected retinal image divided by the number of degrees that this projected image spans on the retina. This provides a *retina-based* SF image description.

The retinal projection of a face will change with viewing distance and correspond to one of the images of Figure 3 (e.g. A'–E'). Changes in retinal size vary the SF information available to the visual system. When a face moves towards the observer, its retinal projection size gradually increases. With size increase, two things happen: the retina-based HSF information initially represented in the smaller image shifts towards the lower retina-based SF's of the larger image; new retina-based HSF information is added and becomes available for recognition. When a face moves away from the observer, the opposite happens: its retinal projection size diminishes, retina-based HSF's are “peeled off” and the object information originally present in retina-based LSF's move towards retina-based HSF's.

Cycles-per-object and cycles-per-degree of visual angle are complementary measures. They characterize two important aspects of a visual recognition task and two research traditions. The aim of visual cognition researchers is to understand the critical band(s) of cycles-per-object that represent the diagnostic information responsible for this or that categorisation. However, because retinal projections of diagnostic object SF's will vary in size, the critical band(s) of diagnostic SF will be “seen” by different SF channels (as

studied in psychophysics) with different contrast sensitivities. Thus, in a visual recognition task, an interaction will happen between the “top-down”, task-dependent, diagnostic, object-based information and the CSF of the visual system, whose “bottom-up” input includes the retinal projection of the diagnostic information. We next analyse how this varying retinal projection might support stable recognition.

### **THE RETINA-BASED SF INFORMATION USED FOR RECOGNITION**

Figure 3 suggests a simple mechanism for visual recognition across size changes. Assume a recognition task requiring SF information represented in a narrow band of cycles-per-object (e.g. one octave, from 4 to 8 object cycles). Given an image filtered to contain only this diagnostic band, because this information “passes over” different SF channels as the observer moves towards the object, channels could be “automatically” turned on (or off) by the presence (or absence) of the diagnostic information band within their tuning curve. Channel selection would then be “bottom-up,” determined by the laws of physical optics projecting the diagnostic object SF’s onto the retina. If SF channels scaled to image size like this, then identification would be invariant over many viewing distances. However, we will see that such a simple size-invariant recognition mechanism has been dis-confirmed.

Consider letter identification as a case study. Using identification of band-pass filtered letters, embedded in identically filtered noise, Parish and Sperling

[1] observed that letter identification performance was determined by the object SF's present and was invariant to changes in viewing distance (i.e. changes in retinal object size). Spatial frequency channels scaled with the retinal size of the letters to extract the same object SF's. However, subsequent work showed that letter identification is *not* size invariant. Both Majaj et al. [2] and Chung et al. [3] reported that changes in channel frequency did *not* scale with corresponding changes in letter size. Instead, with smaller letter sizes, the tuning peaks shift progressively towards lower object SF's [3]; observers identify large letters by their high frequency edges, but small letters by their low frequency gross strokes ([2]; see also [4] for scale dependent face detection). If changing letter size changes the SF band mediating their identification, and therefore the features subsuming identification, what is left of the simple model that scales channels to stimulus sizes to track the same letter information across different letter sizes? This question forces us to step back and reconsider how other factors, such as recognition task, might affect channel selection "top-down." Further, recent work underlines the potential for more flexible processing of SF (Box 2).

## **BOX 2: FLEXIBILITY OF SPATIAL FREQUENCY CHANNELS**

Early studies of SF channels assumed linearity of channel response to stimulus intensity and parallel (independent) processing of SF information across channels. These assumptions have been shattered with evidence of *collinear summation* [40,41] *spatial pooling* [42] *contrast-gain control* [43,44] and *perceptual*



*learning* (to be discussed later). Collinear summation suggests long-range lateral facilitation between channels of a similar SF and orientation, ruling out their independence. In a related vein, spatial pooling suggests integration of proximal elements, interfering with the independent processing of local elements implied if channels were independent. Global influence on local processing is also apparent where the perceived contrast of a stimulus fragment depends on the average contrast of surrounding fragments, implying a gain control mechanisms across channels.

From a conception of channels as fixed, oriented, SF-tuned filters that process the retinal image independently and in parallel, a modern conception arises in which channels are functional constructs emerging from a network of excitatory and inhibitory interactions in the brain. The properties of these channels are not fixed, because perceptual learning can modify the underlying neuronal network and adapt the visual system to the perceptual task at hand [27].

The neural structures underpinning SF channels are not yet fully elucidated. In primary visual cortex, visual area V1 has small receptive fields, mapping the fovea, susceptible to encode fine scale variations in the input. Away from the fovea, receptive fields increase to encode coarser scale variations. Neurons in area V2 and extrastriate visual areas have larger receptive fields than V1. Thus, foveal representations incorporate fine scales provided by V1 and coarser scales from different extrastriate areas. The access to a varying range of

spatial filters across the retina endows the network implementing the properties of SF channels with considerable flexibility and computational power.

## **EFFECTS OF RECOGNITION TASK ON INFORMATION USE**

Typically, visual events afford many categorizations resulting from the observer's cognitive requirements. For example, a face can be identified as George, a male, in his mid-thirties, often smiling, sometimes sad, and so forth. Further, any object can be identified at different levels of specificity, depending on expertise. For example, a bird will be categorized as a sparrow, or even as a dusky seaside sparrow by an expert [5,6]. An important empirical endeavour has therefore been to characterize the critical band(s) of object-based SF's that allow the multiple categorizations of a single visual event [7]. If, as discussed, diagnostic band varies across task and viewing distance, then size invariance comes into question making it important to understand how different tasks are performed over different viewing distances to better understand the ecology of recognition – the idea that some recognition tasks can be performed over a range of viewing distances, others not.

The diagnostic band of SF's of a given categorization task depends on (a) the distribution of category-specific diagnostic information in object frequencies and (b) the contrast sensitivity function of the observer, which will modulate the availability of the diagnostic information as a function of viewing distance.

Interactions between (a) and (b) could determine which SF channel(s) is chosen

as a function of (1) stimulus size and (2) which portion of the visual field is attended.

To illustrate, Gold, Bennett and Sekuler [8] compared how efficient humans are when identifying ten letters and ten faces. All stimuli were band-passed with various one and two-octave SF filters from low to high SF's. For both bandwidths, observers were able to identify letters over a wider range of frequencies than faces suggesting that the diagnostic band(s) of frequencies depends on object class. Further, our own work shows that observers use different object-based SF bands for the same object class depending on categorisation task. For example [9; see also 10-13, Box 3], upon seeing the "Hybrid Face" stimulus of Figure 4a, observers instructed to report its specific expression (task A) would tend to say "happy," using LSF's, but would tend to say "non-expressive", using HSF's, when asked whether the same stimulus was expressive or not (task B).

In sum, recognition requires information from the visual stimulus. The evidence suggests that this information will vary with object class and categorization task (see Box 3). If these top-down task factors affect the relevant band of object information to extract, then do they affect early SF channel selection?

### BOX 3: THE RESEARCH AGENDA FOR RECOGNITION

If the bank of SF channels specifies one analysis of retinal information in terms of SF bands, the agenda for recognition is to understand how brain mechanisms extract, combine and utilise object information represented in different channels. For example, Figure 4c illustrates how facial expressions are recognised on the basis of different facial features, each represented by a specific combination of localized SFs. Such findings [11-13, 45] question the nature of the information used for recognition: one or several SF band(s) *per se*, or specific features represented across several SF bands, as in Figure 4c. This will largely set the research agenda and its associated methods. To illustrate, to determine the “preferred SF band(s) of X” the agenda should be to derive the SF parametric tuning curve for recognizing category X. That is, to split the stimulus into different conditions of SF bands, normalize power across conditions and compare responses – where these can be brain [46-49] or behavioural responses. However, visual stimuli in Fourier space really are 3D signals (the X and Y dimensions of the image plus the third Z dimension of SF bands) but SF tuning curves only estimate the third dimension of SFs. So, the methods of tuning curves, inherited from the psychophysics of simple sine-wave gratings, are not sufficiently powerful to determine, within any SF band, the specific information responsible for recognizing complex stimuli: this information could be any combination of features represented within this SF band. In addition, the emphasis on tuning to a preferred SF band tends to deemphasize the integration

across SF bands that could represent the feature combination in question. Thus, the research agenda for scale-based recognition must move on, from studying a complex 3D signal in one dimension, to embrace the study of scale processing in three dimensions. This Gabor analysis of human recognition information [11-13, 45] is consistent with evidence of top-down, retinotopic control of SF channels [19].

### TOP-DOWN CONTROL OF SF CHANNELS?

Our hybrid face and scene work suggests that the categorisation task might cue the observer to the SF channels from which to extract information. Analogous results in the psychophysical literature provide more direct evidence on the question of whether an observer could select SF channel(s).

Specifically, studies of SF *uncertainty* show that when an observer must detect a grating at a particular SF, they perform worse when it is randomly intermixed with other SF's than when it is presented in a block of trials at a constant SF [14,15]. Support for top-down channel selection arises because uncertainty effects are reduced or eliminated if the observer is symbolically cued (i.e. a top-down cue such as an auditory tone or a number) to stimulus SF on each, forced-choice, trial in an intermixed block [14,15]. Further evidence for top-down channel control is that when symbolic cues lead an observer to expect a grating at one SF their, forced-choice, detection of gratings at unexpected SF's

deteriorates with increasing distance from the cued SF [16], reminiscent of cortical SF channel tuning (Box 1).

Similar *expectancy* effects were observed with scene and face stimuli [9; see also 17]; observers were unaware of a face presented in the non-sensitised SF bandwidth of a face-hybrid (Figure 4a). Further, we found that, as with gratings, when symbolic auditory cues led an observer to expect a stimulus in a particular SF bandwidth, their categorisation of scenes was impaired in the unexpected bandwidth of a scene-hybrid (Fig 4b) or a scene + noise stimulus (see Figure 5 left panels) [18]. Finally, we observed that attention to particular SF bandwidths for scene categorisation can be specific to retinal location supporting the possibility that channel selection may involve early, retinotopically mapped, stages of visual analysis [19].

In sum, all these results imply that observers can select a specific SF channel(s) as a function of task information; a top-down effect. In contrast, critical-band noise-masking studies suggest that observers have little top-down control over SF channels [2,20-22; but see 23]. Figure 5 illustrates the logic of these experiments. Solomon and Pelli [20] used critical-band noise-masking to investigate letter identification. They found that the SF peak of the channel(s) used by an observer to identify letters shifted very little between low and high-pass noise conditions, even though letter information covers a broad part of the object frequency spectrum [2,20; but see 3]. This led them to conclude that

observers have little ability to select SF channels. So, how are we to resolve these opposing findings on top-down control of SF processing?

## **RECONCILIATION: LEARNING OF SIGNAL AND NOISE SF's**

We will argue that top-down channel selection depends on being able to learn the distribution of diagnostic signal SF information in noise and that this varies between SF uncertainty and cueing experiments on the one hand and those using critical-band noise-masking on the other.

Consider an observer faced with a novel object to categorise. The object representation will be partially obscured by various noise sources. The observer needs to learn, what diagnostic object information is available in the different retina-based SF bandwidths to support categorisation. An immediate problem confronting the observer is determining which elements of the stimulus are signal and which noise. Fortunately, the effects of noise will automatically reduce, without attention from the observer, because the noise averages out over time. Conversely, we argue that categorising the object requires the observer to attend the object-based SF information in different retina-based SF bands. This should drive a perceptual learning process about which SF's facilitate a particular categorisation of the object. Thus, over an extended period of everyday experience, an observer might learn that 8-32 cycles-per-face are diagnostic for face identification and how to extract this information from the relevant SF channels.

Now consider the effects of this prior knowledge on performance in experiments. In previous noise-masking studies [2,3,20-22], the observers were not explicitly told about the SF characteristics of the signal or noise. Thus, we might expect that performance will be strongly determined by observers' often extensive prior learning of the diagnostic SF bands for stimulus categorisation (e.g. if we assume one hour per day reading, we reach the age of 50 having gone through about a billion letters [2]). Consequently, learning to use other channels in an experiment will proceed slowly if at all. Conversely, in work on uncertainty and cueing, the cues specifically establish expectations about signal SF's and in the case of hybrid stimuli, because of their binary nature, about noise SF's as well. As suggested by the perceptual learning literature [24], this direct information on both signal and noise SF's might be crucial for observers to learn to select channels, especially when they must overcome prior learning and when the snapshot timeframe of an experiment is not sufficient to allow the full benefits of temporal averaging of noise to accrue.

In sum, we argue that observers' ability to select channels is determined by an interaction between learning, knowledge and task-based expectations. From this two key, interrelated, issues for the future research agenda arise. The first concerns specifying the mechanisms by which categorisation drives learning of the diagnostic information and the SF channels carrying it and the second the neural locus of these effects.



## MECHANISMS AND LOCUS OF LEARNING

We noted earlier that the effects of categorisation and cueing on channel selection are SF tuned [16] and retinotopically mapped [19]. In further work, we found that categorisation increased perceptual sensitivity for cross (not within) category discrimination (a categorical perception effect) of Gabor patch stimuli in a way that was very highly tuned to stimulus orientation [25]. Similarly, research shows perceptual learning [26] that is also specific to features such as stimulus retinal position [26,27] and dependent on attention [24,28; but see 29]. This raises the possibility that top-down attention to category differences drives perceptual learning about them. The specificity of perceptual learning has been interpreted as evidence that it may involve early sensory analysis [24,26,27,30,31].

However, it has also been argued that if learning one task simply changed early sensory analysis this could interfere with the performance of other tasks dependent on the same analysis [26,31]; a version of the stability-plasticity dilemma. Helping stability, it has been argued that highly specific learning may be explained through re-weighting of *outputs* from early sensory analysis [32]. This may not be a complete explanation as there is converging physiological evidence [26,33] for modifications to early sensory analysis, although the interpretation of this remains controversial. An alternative possibility, summarised by Ahissar and Hochstein [24], is that *“learning.... begins at high level areas of the visual system, and when these do not suffice, progresses backwards to the*

*input levels*". By making high level modifications the first choice, stability is also helped.

Consistent with multi-level learning (see also [34]), we suggest that category learning may involve observers learning the settings of the visual filters at the stage(s) of analysis that optimises extraction of diagnostic information. This could account for the widely distributed neural correlates of visual category learning [35] as a function of task. In our tasks, the psychophysical evidence suggests observers learn the relevant settings of early SF filters. We further suggest that these settings could then be implemented dynamically, top-down, via feedback connections from the cortical regions holding the relevant category representations when the need arises to perform a particular task [26,31,33,34]. Dynamic control of filter settings would help to maximise the stability of perception for other tasks.

Evidence for the possibility of dynamic control arises from the growing literature showing that attention can modulate multiple stages of visual analysis including V1 [36]. Of particular interest is the recent application of linear pattern classification methods to the analysis of fMRI data to extend attentional modulation work beyond the 2D space domain by showing that attention can modulate orientation channel activity as early as V1 [37]. Such methods are beginning to pave the way to provide converging evidence on the link between our psychophysical observations of task driven selective processing of SF's and the underlying brain mechanisms.

## CONCLUDING REMARKS

A major challenge to our interaction with the world is the stable categorisation of visual events in conditions of ever-changing retinal stimulation. An important analysis of the available object information can be made in terms of SF's. Visual processing must establish the mapping between the viewing distance dependent retina-based SF information and the diagnostic object-based SF's that subtend stimulus perception and vary as a function of object class and categorisation task. To frame this complex interaction, we have made two critical points:

- (1) Channel selection is not simply stimulus driven and bottom-up; instead, observers can actively and dynamically change SF channel selection as a function of their cognitive task, a top-down control on the early, retinotopically-mapped selection of SF information.
- (2) When the observer explicitly attends the stimulus SF's supporting recognition this drives perceptual learning about the SF channels carrying diagnostic, object-based, SF information as a function of retinal image size.

In sum, top-down selection of diagnostic information, in retinotopically mapped SF channels, provides a bridge between the cognitive demands of the observer recognising visual events and the low-level constraints of the visual system which provides the available SF information.

## QUESTIONS

- Is attention drawn to SF channels, or to the features of an object represented across specific SF channels?
- How is SF information integrated across channels and time to perform a recognition task?
- How does the learning of diagnostic SF's and channels proceed?
  - How is the stability-plasticity dilemma addressed?
- How is top-down control of SF channels implemented following learning?
  - What is the anatomy of the top-down influences on earlier visual processing of SF? Which are the relevant neural circuits?
  - What are the mechanisms (e.g. priming of channels, selective attention to channels, or inhibition of channels)?

## Acknowledgements

This work was supported by the BBSRC (grants S13185 & S13186) and ESRC (R42200134352; R00032165). Thanks to Iona Alexander for comments on an earlier draft of this manuscript.

## References

1. Parish, D.H. and Sperling, G. (1991) Object spatial frequencies, retinal frequencies, noise, and the efficiency of letter discrimination. *Vision Res.* 31, 1399-1415.
2. Majaj, N. J., Pelli D. G., Kurshan P. and Palomares M. (2002) The role of spatial frequency channels in letter identification, *Vision Res.* 42, 1165-1184.
3. Chung, S.T.L., Legge, G.E. and Tjan, B. (2002) Spatial-frequency characteristics of letter identification in central and peripheral vision. *Vision Res.* 42, 2137-2152.
4. Ojanpää, H. and Näsänen, R. (2003) Utilisation of spatial frequency information in face search. *Vision Res.* 43, 2505-2515.
5. Tanaka, J. and Taylor, M. (1991) Object categories and expertise: Is the basic level in the eye of the beholder. *Cognit. Psychol.* 23, 457-482.
6. Gauthier, I., Skudlarski, P., Gore, J. C., Anderson, A.W. (2000) Expertise for cars and birds recruits brain areas involved in face recognition. *Nat. Neurosci.* 3, 191-197.
7. Schyns, P.G. (1998) Diagnostic recognition: Task constraints, object information and their interactions. *Cognition*, 67, 147-179.
8. Gold, J.M., Bennett, P.J. and Sekuler, A.B. (1999) Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Res.* 39, 3537-3560.

9. 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.
10. Schyns, P.G. and Oliva, A. (1994) From blobs to boundary edges: Evidence for time and scale dependent scene recognition. *Psych. Sci.* 5, 195-200.
11. Smith, M.L., Cottrell, G.W., Gosselin, F. and Schyns, P.G. (2005) Transmitting and decoding facial expressions. *Psych. Sci.* **16**, 184-189.
12. Gosselin, F. and Schyns, P. G. (2001) Bubbles: A new technique to reveal the use of visual information in recognition tasks. *Vision Res.* **41**, 2261-2271.
13. Schyns, P.G., Bonnar, L. and Gosselin, F. (2002) Show me the features! Understanding recognition from the use of visual information. *Psych. Sci.* 13, 402-409.
14. Hübner, R. (1996) Specific effects of spatial-frequency uncertainty and different cue types on contrast detection: data and models. *Vision Res.* 36, 3429-3439.
15. Davis, E.T., Kramer, P. and Graham, N. (1983) Uncertainty about spatial frequency, spatial position, or contrast of visual patterns. *Percep. Psychophys.* 33, 20-28.
16. Sowden, P.T., Özgen, E., Schyns, P. and Daoutis, C. (2003) Expectancy effects on spatial frequency processing. *Vision Res.* 43, 2759-2772.

17. Oliva, A. and Schyns, P.G. (1997) Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognit. Psychol.* 34, 72-107.
18. Özgen, E., Sowden, P.T., Schyns, P. and Daoutis, C. (2005) Top-down attentional modulation of spatial frequency processing in scene perception. *Visual Cognit.* 12, 925-937.
19. Özgen, E., Payne, H.E., Sowden, P.T. and Schyns, P.G. (2006) Retinotopic sensitisation to spatial scale: evidence for flexible spatial frequency processing in scene perception. *Vision Res.* 46, 1108-1119.
20. Solomon, J.A. and Pelli, D.G. (1994) The visual filter mediating letter identification. *Nature* 369, 395-397.
21. Lu, A. and Doshier, B.A. (2004) Spatial attention excludes external noise without changing the spatial frequency tuning of the perceptual template. *J. Vision* 4, 955-966.
22. Talgar, C.P., Pelli, D.G. and Carrasco, M. (2004) Covert attention enhances letter identification without affecting channel tuning. *J. Vision* 4, 22-31.
23. Solomon, J.A. (2000) Channel selection with non-white noise masks. *J. Opt. Soc. Am. A* 17, 986-993.
24. Ahissar, M. and Hochstein, S. (2004) The reverse hierarchy theory of perceptual learning. *Trends Cogn. Sci.* 8, 457-464.

25. Notman, L., Sowden, P.T. and Özgen, E. (2005) The nature of learned categorical perception effects: a psychophysical approach. *Cognition* 95, B1-B14.
26. Fahle, M. (2004) Perceptual learning: A case for early selection. *J. Vision* 4, 879-890.
27. Sowden, P.T., Rose, D. and Davies, I.R.L. (2002) Perceptual learning of luminance contrast detection: specific for spatial frequency and retinal location but not orientation. *Vision Res.* 42, 1249-1258.
28. Sowden, P.T., Davies, I.R.L., Rose, D and Kaye, M. (1996) Perceptual learning of stereoacuity. *Perception* 25, 1043-1052.
29. Seitz, A. and Watanabe, T. (2005) A unified model for perceptual learning. *Trends Cogn. Sci.* 9, 329-334.
30. Fiorentini, A. and Berardi, N. (1980) Perceptual learning specific for orientation and spatial frequency. *Nature* 287, 43-44.
31. Schyns, P.G., Goldstone, R.L. and Thibaut, J.P. (1998) The development of features in object concepts. *Behav. Brain Sci.* **21**, 1-17.
32. Petrov, A.A., Doshier, B.A. and Lu, Z. (2005) The dynamics of perceptual learning: an incremental reweighting model. *Psych. Rev.* 112, 715-743.
33. Li, W., Piëch, V. and Gilbert, C.D. (2004) Perceptual learning and top-down influences in primary visual cortex. *Nat. Neurosci.* 7, 651-657.



34. Ullman, S. (1995) Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex* 5, 1-11.
35. Miller, E. K., Nieder, A., Freedman, D. J., and Wallis, J. D. (2003) Neural correlates of categories and concepts. *Curr. Op. Neurobiol.* 13, 198-203.
36. Posner, M. I., and Gilbert, C. D. (1999) Attention and primary visual cortex. *Proc. Nat. Acad. Sci.* 96, 2585-2587.
37. Kamitani, Y., and Tong, F. (2005) Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* 8, 679-685.
38. De Valois, R.L., Albrecht, D.G. and Thorell, L.G. (1982) Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res.* 22, 545-559.
39. Wilson, H.R. and Wilkinson, F. (1997) Evolving concepts of spatial channels in vision: from independence to nonlinear interactions. *Perception*, 26, 939-960
40. Polat, U. and Sagi, D. (1993) Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Res.* 33, 993 – 999.
41. Polat, U. and Sagi, D. (1994) The architecture of perceptual spatial interactions. *Vision Res.* 34, 73-78.
42. Wilkinson, F., Wilson, H.R. and Ellemberg, D. (1997) Lateral interactions in peripherally viewed texture arrays. *J. Opt. Soc. Am. A.* 14, 2057-2068.
43. Heeger, D.J. (1992) Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* 9, 181-197.

44. Foley, J.M. (1994) Human luminance pattern vision mechanisms: masking experiments require a new model. *J. Opt. Soc. Am. A*, **11**, 1710-1719.
45. Smith, M.L., Gosselin, F., and Schyns, P.G. (2006) Perceptual moments of conscious visual experience inferred from oscillatory brain activity. *Proc. Nat. Acad. Sci.* **103**, 5626-5631.
46. Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, **15**, 600-609.
47. Eger, E., Schyns, P. G., & Kleinschmidt, A. (2004). Scale invariant adaptation in fusiform face-responsive regions. *Neuroimage*, **22**, 232-242.
48. Goffaux, V., Gauthier, I., & Rossion, B. (2003). Spatial scale contribution to early visual differences between face and object processing. *Cognitive Brain Research*, **16**, 416-424.
49. Vuilleumier, P., Armony, J.L, Driver, J. & Dolan, R.J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience*, **6**, 624-631.
50. Burt, P., and Adelson, E.H. (1983) The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, **31**, 532-540.

## Figure captions

**Figure 1.** Illustration of how viewing distance affects the retinal size of a given stimulus. Consider the three pictures of a smiling face. Imagine that the small picture corresponds to your friend standing far away (e.g. at 8 meters). The retinal projection of his face is small as illustrated. As he moves closer (e.g. at 4 meters, then at 2), the projection changes: it increases in size by a factor of two. Of course, as he walks towards you, the increase is effectively continuous and inversely proportional to the distance between your friend and your eyes. The retinal size, measured in number of degrees of visual angle on the X or Y dimension, of a distal stimulus is given by  $\text{visual\_angle} = (\text{atan}(\text{stimulus\_size}/\text{distance}))$ .

**Figure 2.** Spatial frequency channels and the contrast sensitivity function (CSF). The broad curve represents a typical CSF, which describes how the amount of contrast required to perceive a stimulus varies as a function of its SF. The effect of variation in stimulus SF on the appearance of a simple sinewave grating stimulus is shown below the figure. The CSF defines the envelope of a set of SF channels (represented by the narrow curves) each of which is responsive to (passes) a narrow band of SF's. The arrows shown at the bottom of the figure suggest that observers may attend to one (solid arrow) or more channels (dotted arrows) depending on their current task requirements.

**Figure 3.** Object-based spatial frequencies. An object represented in a digital image can be filtered to analyse its SF content – that is, its cyclical information. Here the left face is decomposed with five spatial filters. Each filter spans a one-octave bandwidth (see Box 1), without overlap between frequency bands. We can peel off each layer of object SF from the original stimulus and analyze the remaining SF information. Starting with the original face on the left, we peel off the layer of highest SF's (labelled HSF:  $X/2 \times Y/2$  in Figure 3) and obtain the face labelled A. The High Spatial Frequency (HSF) filter retains facial information associated with the most detailed light-dark transitions in the face (e.g the hairs, eyebrows and wrinkles of a face). If the image is of size X by Y, the most detailed variations are those represented by  $X/2$  and  $Y/2$  cycles-per-face (because it takes at least two pixels to express one cycle, i.e. one light-dark transition). Because all information expressed over 2 pixels has been removed in A, we can divide its size in two (to obtain face A') without *any* loss of face SF information between A and A': A and A' share the same number of cycles-per-face, despite the fact that A' is half the size of A. The process of peeling off HSF can be applied recursively [50]. The next spatial filter extracts a SF band between  $X/2$  and  $X/4$  cycles-per-face, the next filter SF band between  $X/4$  and  $X/8$  cycles-per-face, and so forth with decreasing details until the Low Spatial Frequencies (LSF) filter. Extracting the next level of HSF frequencies from A results in B, which has the same face SF information as B'. In the figure, the vertical arrows represent that change in

image size results from varying viewing distance; the horizontal arrow indicates increasing viewing distance. Images A .. E represent the availability of cycles-per-face information at a fixed viewing distance. Images A' .. E' represent the same cycles-per-face information with a varying viewing distance. This illustrates that cycles per faces can be identical across two stimuli when their cycles per degree of visual angle are not.

**Figure 4.** Face and scene stimuli. (a) A hybrid face stimulus (adapted from [10]) formed by superimposing a high-pass filtered image of a neutral expression woman with a low-pass filtered image of a smiling man (i.e. the images have been filtered to remove the LSF's or HSF's respectively). (b) A hybrid scene stimulus (adapted from [10]) containing a low-pass filtered highway image and a high-pass filtered city image. View the images from a distance, or defocus them, to reveal the low-pass component of each hybrid stimulus. (c) Illustration of the SF information used to recognize six basic facial expressions plus neutral (adapted from [11]). The data demonstrate that observers extract, combine and utilise SF information across bands to represent combinations of localized features that diagnose each expression (e.g. the mouth in 'happy', the two eyes in 'fear' and the mouth and the surroundings of the nose in 'disgust').

**Figure 5.** The critical-band noise-masking method. In the left panels a broadband city scene is shown combined with low (top) or high (bottom) pass

filtered noise (i.e. noise that has been filtered to remove the HSF or LSF components respectively). In the centre panels a function is plotted showing the relationship between contrast energy and object-based SF. The portion of the spectrum masked by the low and high-pass noise respectively is indicated by the overlaid noise samples. In the right hand panel the human CSF is shown together with the SF channels underlying it. The overlaid noise samples indicate the channels filled by the noise. So, take the case of a scene presented in low-pass noise (top row). For simplicity of explanation, we assume that the low object-based spatial frequencies of the noise correspond to lower retina-based frequencies. Accordingly, the lower SF channels will be polluted by noise. Researchers using this method suggest that if the observer can choose the channel(s) to monitor, and if there is available diagnostic information in the object frequency spectrum at higher SF's, then we might expect the observer to switch to higher SF channels in order to avoid the low-pass noise. The converse would be expected for high-pass noise.

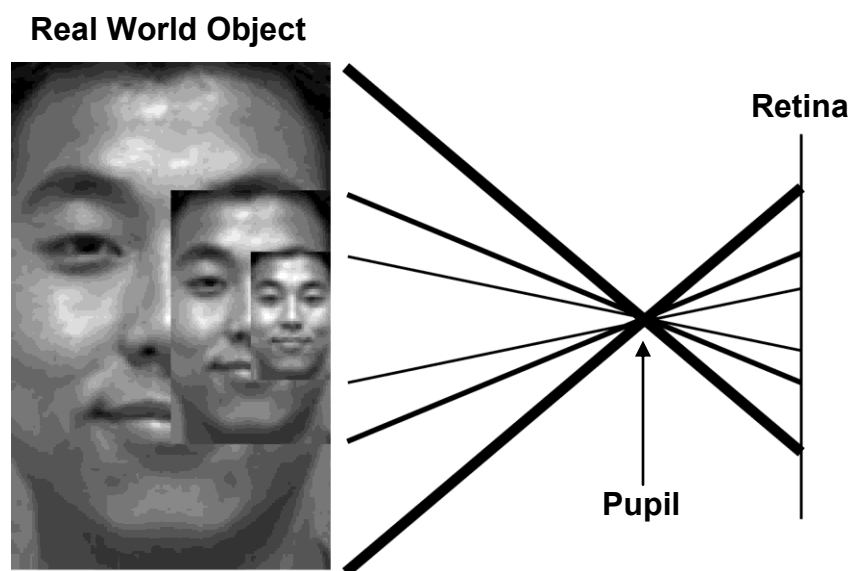


Figure 1

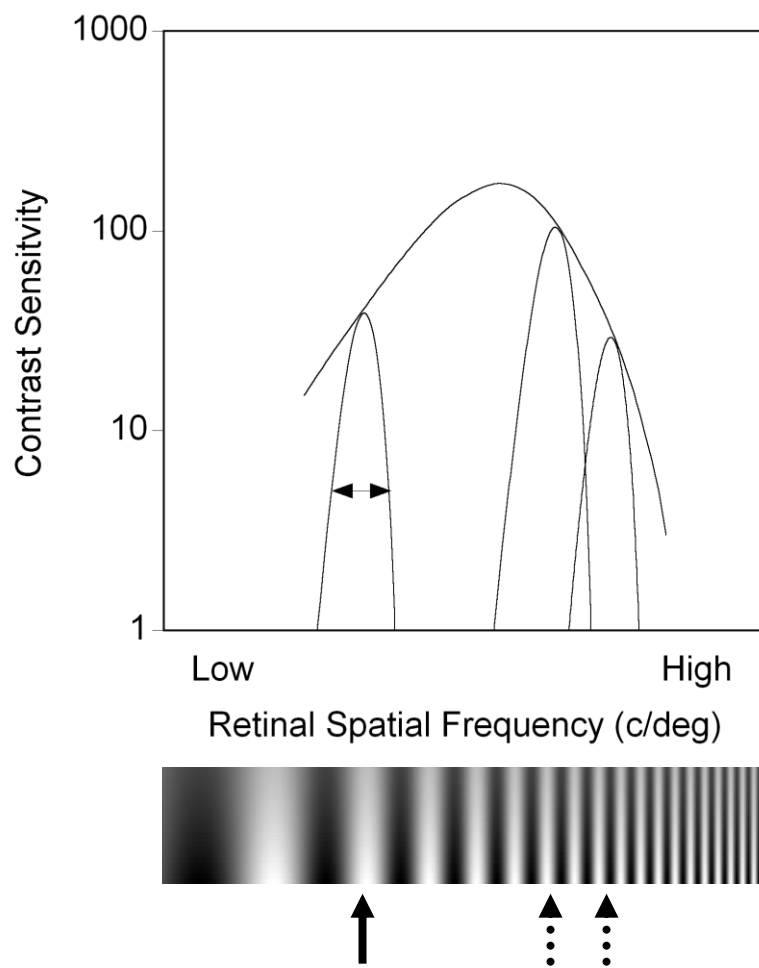


Figure 2



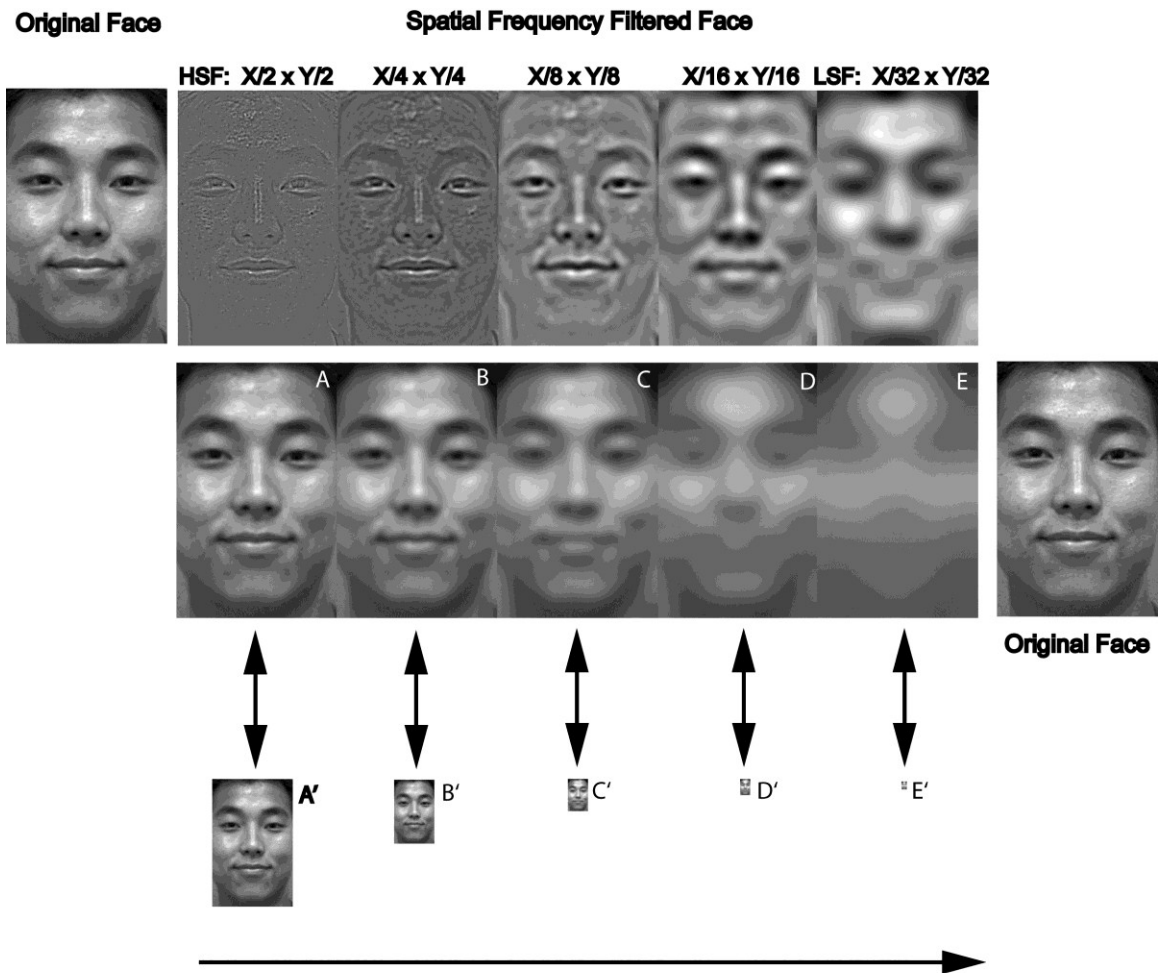


Figure 3

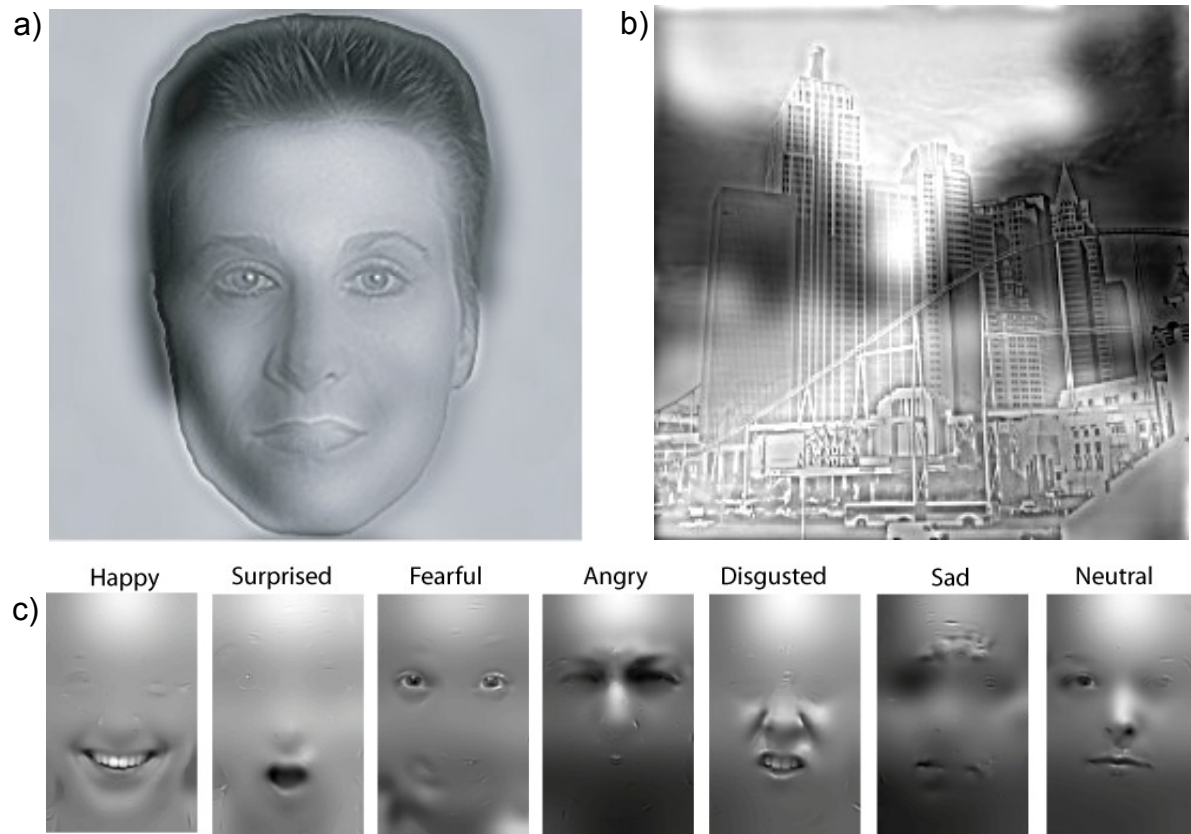
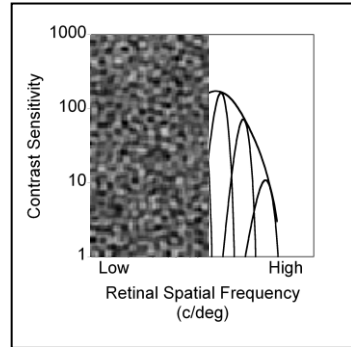
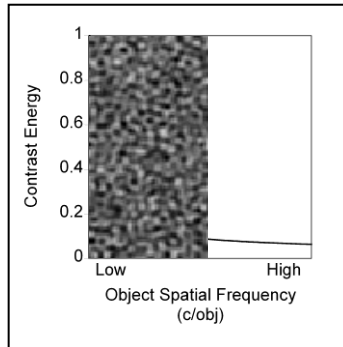


Figure 4

Low  
Passed  
Noise



High  
Passed  
Noise

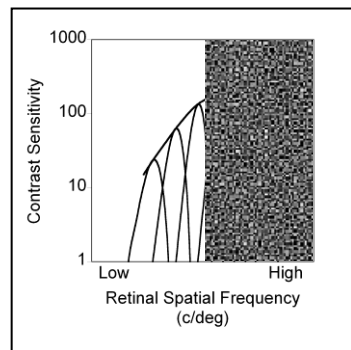
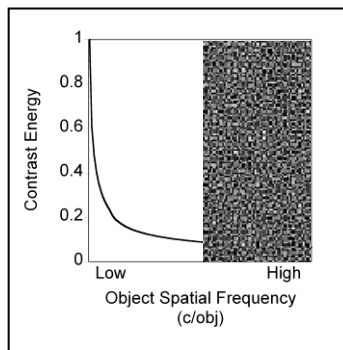


Figure 5