We used psychophysical reverse correlation to determine how directional signals are integrated across a time window of 300 ms. Directional tuning was time independent within the resolution of our measurements, as demonstrated by the fact that the perceptual filter was almost perfectly separable in its temporal and directional dimensions. The amplitude of the filter peaked very early (30–60 ms) and then quickly decreased to almost zero, after which it increased slightly again. We successfully modeled this bimodal behavior using a simple circuit where each directional filter normalizes its own output, with the normalizing signal delayed by ~100 ms.

Keywords: noise image classification, delayed normalization, separability


Introduction

Some aspects of neuronal responses take time to develop. For example, it has been shown that spatial frequency tuning becomes increasingly sharper over a period of ~200 ms after stimulus onset in V1 neurons (Bredfeldt & Ringach, 2002; Mazer, Vinje, McDermott, Schiller, & Gallant, 2002). In MT, some neurons take 50–100 ms to develop plaid-selective responses (Pack & Born, 2001; Smith, Majaj, & Movshon, 2005). Moreover, their directional selectivity undergoes fast changes depending on stimulus history (Perge, Borghuis, Bours, Lankheet, & van Wezel, 2005), a result that has also been demonstrated for orientation tuning in primary visual cortex (Dragoi, Rivadulla, & Sur, 2001; Dragoi, Sharma, Miller, & Sur, 2002; Dragoi, Sharma, & Sur, 2000; Felsen et al., 2002; Müller, Metha, Krauskopf, & Lennie, 1999) and for tangential cells in the fly lobula plate (Neri, 2007).

Our goal was to study directional tuning in human vision and to determine whether it undergoes substantial changes during the first two to three hundred milliseconds after stimulus onset. We derived time-varying directional tuning curves using motion noise that varied across time, similar to Neri and Heeger (2002) and Mareshal, Dakin, and Bex (2006) (see also Tadin, Lappin, & Blake, 2006). The task involved discriminating the direction of a circular field of moving dots, and human responses were analyzed using noise image classification (Ahumada, 2002). This technique exploits the correlation between trial-by-trial fluctuations in the noise samples and individual responses by the observers, allowing retrieval of the perceptual filter that underlies their behavioral choices (for a recent review, see Neri & Levi, 2006).

We found that directional selectivity was characterized by a tuning curve with a half-width at half-height of 30–50 deg, broadly consistent with electrophysiological studies in MT (Cook & Maunsell, 2004). This degree of tuning emerged very early and was invariant over time, at least within the resolution of our measurements. We verified this invariance by computing the expected direction–time surface filter that is obtained by multiplying the directional tuning curve averaged across time with the temporal impulse response averaged across direction (the assumption of separability) and by subtracting this prediction from the experimental filter. The resulting surface was virtually flat.

One interesting and unexpected feature of our data is that the amplitude of the filter peaked at the direction of the target as expected, reached its maximum value very early, then decreased, and finally showed a tendency to increase again. We were able to model this temporal evolution of amplitude modulation by using a minimal model involving linear directional filters followed by a self-normalization stage.

Methods

Stimuli

The stimulus consisted of a sequence of 10 frames like the one shown in Figure 1A, each lasting 30 ms (total...
Figure 1. (A) One frame of the stimulus. Dots moved only within the two circular regions to the sides of fixation (the two red circles were actually present in the stimulus). A fixed percentage of the dots (signal dots, indicated by red arrows) moved down on one side of fixation (left in the example) and up on the other side. Subjects had to indicate the side of the stimulus containing down motion (or up for one of the 3 subjects). (B) Black line shows number of dots moving in the direction indicated by the x axis on one motion frame (two successive frames) and within the left circular region. Red shows additional signal dots. (C) The stimulus contained 9 motion frames, defining a surface with dimensions "direction" (x axis) and "time" (y axis). (D) Model filters used to extract the response at directions \( \downarrow \) (left) and \( \uparrow \) (right). (E) The output from each filter normalized itself after a delay \( \tau \) of 90 ms. (F) The outputs from the two filters for opposite directions were subtracted from each other to generate the final response to the circled region on the left. The model processed the circled region on the right in a similar way and selected the region with largest output as target.

stimulus duration was 300 ms). The dots were anti-aliased Gaussian blobs (standard deviation = 3.75 arcmin) with peak luminance of 74 (bright) or 0 (dark) cd/m² on a background luminance of 37 cd/m² (monitor was gamma-corrected). For regions where blobs overlapped, the pixel was assigned the luminance of the blob with largest absolute value at that position. Dots outside the two regions circled in red (diameter was 2.1 deg, distance of centre from fixation was 4.3 deg) were always present on the screen and never moved. Because the stimulus consisted of 10 frames, it contained 9 motion impulses. Half the noise dots moved from frame 1 to frame 2, the remaining half from frame 2 to 3 and so forth, meaning that on each frame half the noise dots had terminated their 2-frame motion while the other half was about to move (we adopted this configuration to avoid concurrent motion jumps for the whole display). At the end of each 2-frame motion, each dot was randomly relocated to a new starting position for the next 2-frame motion. Dot velocity was 4.1 deg/s. Within each region, a fixed percentage of the dots were signal dots, while the remaining dots were noise dots. Signal dots moved in opposite vertical directions within the two regions: If signal dots moved upward within the region on the right (as in Figure 1A), signal dots within the region on the left moved downward, and vice versa. Each noise dot could take any of 16 directions (nearby directions differing by 22.5 deg) with uniform probability. There were 16 dots in total for each motion impulse (total of 16 \( \times \) 9 = 144) in each circled region. Of these, 2 (S1 and S3) or 6 (S2) were signal dots.

Task

On each trial, observers were asked to press one of two buttons to indicate which of the two possible signal configurations were presented: either (1) signal dots moved upward within the region on the left and signal dots moved downward within the region on the right or (2) signal dots moved downward within the region on the left and signal dots moved upward within the region on the right. Their response triggered the next presentation. This task was worded in the following terms: press 1 for left or 2 for right to indicate the side where you saw predominantly downward motion, being aware that the other side always contains upward motion (for one subject S1, the target direction was upward). Observers were initially familiarized with the task by presenting them with 20–30 trials containing 100% signal dots. They learnt the task very quickly, after which we proceeded to reduce the % of signal dots until their performance was around threshold (75% correct responses). The data shown here were collected at a fixed threshold number of signal dots (detailed earlier). Percent correct for the task was 0.64 (S1), 0.78 (S2), and 0.77 (S3). Number of collected trials was 11,150 (S1), 10,100 (S2), and 5,000 (S3). Feedback (correct/incorrect) was provided after each response. S1 and S2 were naive regarding the methodology and purpose of the study and S3 was one of the authors (P. N.). For double-pass experiments, we ran blocks of 100 trials for which the second half of the block (50 trials) consisted of a random permutation of the trials presented during the first half. Number of collected trials was 3,100 (S1), 1,000 (S2), and 400 (S3).
Derivation of perceptual filters

The stimulus presented within each circled region on each trial $i$ can be described as $N_i(d, t)$, the number of noise dots that moved in direction $d$ at time $t$. There were 16 possible directions and 9 time points, defining the dimensionality of the surface. If $N_i^T(d, t)$ refers to the region containing signal dots moving in the target (T) direction and $N_i^{NT}(d, t)$ refers to the other region (where signal dots moved in the opposite nontarget direction), then $\Delta N_i = N_i^T - N_i^{NT}$, where $c = 0$ if the observer responded incorrectly on trial $i$ and $c = 1$ if the response was correct. The perceptual filter is computed as $PF(d, t) = \langle \Delta N_i \rangle - \langle \Delta N_i^b \rangle$ where $\langle \rangle$ is used to indicate the mean across trials $i$ (see Abbey & Eckstein, 2002). It should be pointed out that the distribution of the number of noise dots at each direction was not Gaussian, but Poisson with mean value smaller than 1 (a consequence of the uniform probability for assigning direction to the noise dots). Although this departure from a circular distribution may have introduced distortions in the recovery of the underlying filter if the system conformed to the linear (static)—nonlinear model (as is typically assumed in reverse correlation studies; Chichilnisky, 2001), our conclusions rely only marginally on this assumption because they are based on a full implementation of an observer model and associated classification images, and the model is not of the simple linear followed by static nonlinearity type anyway.

Subtraction of separable prediction and separability index

Surfaces in Figures 3G and 4 (right column) were obtained by subtracting from $PF(d, t)$ the expression $PF^\ast (d, t) = (\sum_i PF(d, t) \cdot \sum_i PF(d, t)) / \sum_i PF(d, t)$. It is easy to verify that, for $PF(d, t) = f(d) \cdot g(t)$, $PF^\ast (d, t) = PF(d, t)$.

Modeling

The stimulus within each circled region in Figure 1A is $S_i^T(d, t) = N_i(d, t) + T(d, t)$ or $S_i^{NT}(d, t) = N_i(d, t) + NT(d, t)$ (depending on whether it contained target or nontarget). $T(d, t) = \delta (d - d_t) \cdot s$, where $d_t$ is the target direction (e.g., downward) and $s$ is the number of signal dots (T(d, t) is independent of time t, because signal dots were present at every motion transition throughout stimulus presentation). Similarly for NT(d, t), we convolved $S$ with a directionally selective receptive field defined by a two-dimensional Gaussian function $RF(d, t) = Gauss(\sigma_d, \sigma_t)$ with $\sigma_d = 40$ deg (broadly consistent with, e.g., Cook & Maunsell, 2004; Treue & Martinez Trujillo, 1999) and $\sigma_t = 1$ frame duration = 30 ms (broadly consistent with Bair & Movshon, 2004) to obtain $o_i^T(d, t) = S_i^T \ast RF$, and similarly for $o_i^{NT}$. We only used the two outputs at preferred and antipreferred directions $d_t$ and $d_{NT}$ for each convolution, i.e., $pref_i^T(t) = o_i^T(d_t, t)$ and antipref$_i^T(t) = o_i^T(d_{NT}, t)$. Similarly for $pref_i^{NT}$ and antipref$_i^{NT}$. The final response to the circled stimulus region containing signal dots moving in the target direction was $r_i^T = \langle \text{pref}_i^T \ast \text{antipref}_i^T \rangle$, (this is a scalar) and similarly for the response to the other region $r_i^{NT}$. Finally, on each trial $i$, the model selected the region...
associated with the largest response between $r_i^T$ and $r_i^{NT}$ as being the target region (i.e., it responded correctly if $r_i^T > r_i^{NT}$, and incorrectly otherwise). Before being temporally averaged and combined into $r$, filter outputs pref and antipref self-normalized their value as follows:

$$
\text{pref}(t) = \frac{\text{pref}(t)}{\text{pref}(t - \Delta t) + k}
$$

with $\Delta t = 3$ frames (90 ms) and $k = 2$ (this expression for normalization is similar to that used by Schwartz & Simoncelli, 2001). These parameter values were selected following pilot simulations. We used the same self-normalization for antipref. We challenged the model with 25 K trials taken directly from the data used with psychophysical observers. The black symbols in Figure 6 were obtained by simulating the response of a linear amplifier model (Murray, Bennett, & Sekuler, 2005) that uses the empirical classification image as a template. If $F(d, t)$ is the empirical classification image, the response corresponding to $S_i^T (d, t)$ on trial $i$ was simply obtained by template matching $r_i^T = \sum F(d, t) \cdot S_i^T (d, t)$, and similarly for $S_i^{NT} (d, t)$ to obtain $r_i^{NT}$. The linear amplifier model chooses the stimulus region associated with largest $r$ as target (correct when $r_i^T > r_i^{NT}$).

Results

General structure of the perceptual filter

Figure 2A shows the perceptual filter (averaged across the 3 subjects) as a function of both direction of motion (on the $x$ axis) and time (on the $y$ axis). As expected, there are positive modulations in the region of the target direction (downwards) and negative modulations in the region of the opposite direction. This is shown by the grey levels within red and blue outlines, respectively. We verified that the measured modulations within these two regions were symmetric within the precision of our measurements, which is not surprising given the nature of the task. We therefore combined the two regions into one region for the purpose of subsequent analysis, as shown in Figure 2B where the region in A indicated by the blue outline has been subtracted from the region indicated by the red outline. We further increased the signal-to-noise ratio of our measurements by symmetrically averaging the two regions indicated by the black outlines in B, based on the assumption that off-target directions would have symmetrical effects. This assumption is not only reasonable, but we also verified that the data conformed to it with the possible exception of subject 2 (S2 in Figure 4) who showed a slight bias for a direction immediately clockwise to target direction (however this bias does not affect the conclusions of this study). Symmetrical averaging of B results in C, which shows a clear positive peak at the target direction between 30 and 60 ms.

Panel D plots the average across time for the surface in C (black line), panel E the average across direction. The directional tuning curve in D shows that, as expected, the perceptual filter peaks at target direction. The half-width at half-height for this function is $36 \pm 7$ deg (mean $\pm SEM$). The temporal impulse response function in E shows that the perceptual filter reaches its maximum peak very early, then decreases, and finally increases again. These features of the perceptual filter are also visible in the Z score map (Figure 3A), where a statistically significant
modulation is observed both within the early and within the very late region of the surface. This late modulation was unexpected (see Discussion).

Separability of directional and temporal responses

An obvious and important question is whether directional tuning varies with time. This question can also be phrased in the following terms: Is the surface in Figure 2C separable across direction and time? In other words, is it well accounted for by simply multiplying D and E for every combination of direction and temporal event? We computed the difference between the measured perceptual filter and the equivalent surface that is obtained by multiplying the two marginal averages (D and E; for details, see Methods). A Z-score map of this difference is plotted in Figure 3B, which shows very little modulation (there is only 1 pixel (symmetrically on both sides) that meets $|Z| > 2$ (colored in Figure 3B), and none of the subjects shows a similar pattern). This is a clear indication that the perceptual filter is well described by a temporal modulation in the amplitude of a fixed directional tuning curve. We also attempted to measure half-width at half-height at different time points throughout the surface and were unable to find consistent changes across time (not shown).

As an additional test of the independence between directional and temporal responses, we computed a separability index based on singular value decomposition (for details, see Methods). This index ranges from 0.125 (not at all separable) to 1 (entirely separable). The value for the average human classification image was (mean ± SEM) 0.87 ± 0.04 (for individual subjects (S1–S3) 0.73 ± 0.07, 0.65 ± 0.08, 0.78 ± 0.06), confirming our earlier conclusion based on Figure 3B.
Data for individual subjects

Figure 4 plots the data of Figures 2C–2E and Figures 3A and 3B separately for each of the 3 observers. S1 and S3 are highly consistent with the average perceptual filter detailed in Figures 2 and 3. More specifically, both subjects show the same temporal modulation pattern with two positive regions, one early and one late. S2 generated a perceptual filter of very poor quality (see Z score scale to the right), so it is hard to establish whether the pattern is consistent with the other two subjects. It certainly showed a tendency towards larger modulation in the earlier part of the filter, and there is a hint of late positive modulation (see contour plot) even though it did not reach statistical significance. Overall, S2's data do not contradict the other two subjects, and it shows the same coarse trend, but the quality of this subject's data set was not sufficient to retrieve a detailed description of the perceptual filter. For this subject, we similarly attribute the lack of statistically significant modulations at target direction (downward) to the poor quality of the corresponding perceptual filter. The directional slice across the surface (red trace for S2 in Figure 4) shows that there was a positive modulation at target direction, but the associated error was large and the resulting Z score not significant.

Modeling

We modeled the shape of the perceptual filter using simple computational tools (see Methods). The general structure of the model is outlined in Figures 1D–1F. The stimulus (shown in C) is initially filtered by a brief directionally tuned operator (D). Only the two outputs from operators centered on target (↓) and opposite-to-target (↑) directions were used. Each output normalized itself after a delay of 90 ms (E). This normalizing signal functions as a servo-mechanism, but the presence of a temporal delay leads to an oscillatory pattern that mimics the bimodal behavior observed in the data. This is demonstrated by the simulation in Figure 5A, where the empirical data (left) are juxtaposed with the simulated perceptual filter (right), which shows two positive peaks at early and late points within the surface (the model surface (right) was obtained by simulating a psychophysical experiment where the model acted as observer). We selected a stimulus signal-to-noise-ratio (SNR) that generated a model performance of 70% correct responses, in close agreement with human performance (average across the 3 subjects was 73%). The corresponding model SNR was 10% of the human SNR (this translates into a noninteger number of noise dots; for the model, we could use noninteger numbers of noise dots because the input to our simulations did not consist of a stimulus image (for which the number of dots must take an integer value), but of a distribution like in Figure 1C).

The 10-fold difference in threshold SNR between model and human performance is very large. What could account for this discrepancy? A good candidate could be late internal noise. Human observers do not behave deterministically, i.e., there is a degree of intrinsic randomness to their responses that is independent of the stimulus. This fact can be demonstrated by performing a double-pass experiment, in which the same physical stimulus is presented twice. If the observer behaved deterministically, the same response should be generated on both passes of an identical stimulus. This is not always the case due to observer internal noise (Burgess & Colborne, 1988).

Figure 5. Results of model simulations. (A) Left half of the surface is taken from Figure 3A, right side shows the classification image corresponding to the simulations using the model outlined in Figures 1D–1F (for details, see Methods). Panel B shows the classification image obtained when the delayed normalization stage is removed from the model. Panel C shows the optimal signal-matched template. Panel D shows the separable prediction for the average human classification image in Figure 2C (for details, see Methods). Panel E plots the pixel-by-pixel correlation coefficient between the psychophysical classification image (left half of the surface in panel A) and each one of the four artificial classification images detailed above and depicted through panels A–D (the respective correlation coefficients are indicated by the same letters on the x axis in panel E).
We wished to establish whether internal noise could account quantitatively for the abovementioned discrepancy between our noiseless model simulations and human psychophysical data. We ran a series of double-pass experiments (see Methods) and determined human–human consistency (the % of trials on which the human observer is consistent with his/her own responses between two identical stimulus presentations) to be (mean ± SEM) 0.61 ± 0.01 (S1), 0.70 ± 0.01 (S2), and 0.80 ± 0.02 (S3). Human–human consistency can be used to establish upper and lower bounds on the best achievable human–model consistency, i.e., the % of trials on which the model response matches the human response (for details, see Neri & Levi, 2006; for related methods in the electrophysiological literature, see Haag & Borst, 1997; Neri, 2006; van Hateren & Snippe, 2001). This optimal range is indicated by the grey area in Figure 6. For two out of the 3 observers (S1–S2), the human–model consistency achieved by our model is satisfactory (points fall within the grey area). For one observer (S3), internal noise is not sufficient to explain the discrepancy between model responses and human responses. We do not have a ready explanation for this small discrepancy.

The human–model consistency plotted on the y axis in Figure 6 was obtained by matching the human responses with model responses to identical stimuli, i.e., at the SNR used with human observers. For those SNR regimes, our model performed correctly on every trial or nearly so. This means that the result in Figure 6 is not specific to our model but applies to any model that correctly detects the target on every trial. For comparison, we report human–model consistency for a linear amplifier model (Murray et al., 2005; Neri & Levi, 2006) that uses the classification image itself as template for matching (black symbols in Figure 6; see Methods). It can be seen that this model is equally (or possibly slightly more) consistent with the human responses. In summary, Figure 6 only allows us to conclude that the high performance of our model is not inconsistent with the human data when one takes into account internal noise, but the results in Figure 6 do not provide specific evidence in favor of our model over many other potential candidates (e.g., an ideal template that only integrates motion energy at the direction of the target for the entire duration of the stimulus, which also correctly detects the target on every trial).

More specific evidence in support of our model is provided by Figure 5E, where we show that it is superior to a few others in its ability to generate classification images that resemble the human data. As a measure of how closely the simulated classification image matches the average human classification image, we simply took the pixel-by-pixel correlation between the two. The correlation for the model discussed so far (classification image shown in Figure 5A) is 0.77 (plotted in Figure 5E at abscissa label “A” with 95% confidence interval). For comparison, we consider the classification images generated by (1) the same model but without the delayed normalization stage (classification image in Figure 5B), (2) a linear amplifier model that uses the optimal signal-matched template (classification image in Figure 5C), and (3) a linear amplifier model that uses the separable prediction for the human classification image as template (Figure 5D).

The corresponding correlation coefficients are plotted in Figure 5E, which shows that our model provides a better account for the empirical classification image than both models 1 and 2 and is close to the separable model. The high accuracy of the separable model is not surprising given the result in Figure 3B, where we showed that the human classification image is separable. However, the separable model is simply a descriptive, not a mechanistic model. The self-normalizing model has the advantage that it is mechanistic (i.e., it potentially informs us about the mechanisms underlying the observed data) and at the same time accounts for the classification image almost as well as the separable one. The importance of the delayed normalization stage is emphasized by the drop in correlation that is observed when this stage is removed from the model (for models A and B, compare the two correlation coefficients in Figure 5E).
Discussion

Temporal dynamics of neuronal selectivity

Neuronal changes in directional tuning with time have been investigated in both macaque (Perge et al., 2005) and cat (Vajda, Borghuis, van de Grind, & Lankheet, 2006) cortex. Although some of these effects appear to involve complex adaptive mechanisms acting at very fast time scales, recent work on fly tangential cells has shown that they can be explained using simple static nonlinearities (Neri, 2007). In general, these studies do not provide conclusive evidence for adaptive changes in directional selectivity with time. Related work in MT has shown that some aspects of directional tuning may change over time scales of \( \sim 100 \) ms, like the emergence of plaid selectivity: some MT neurons initially respond to the direction of the component gratings in a plaid, and only later develop plaid selectivity (Pack & Born 2001; Smith et al., 2005). This result is only marginally related to the present study, but it provides an example of the notion that directional tuning may indeed vary on a time scale that is accessible to psychophysical measurements.

Other aspects of neuronal selectivity change dynamically on a similar time scale (Ringach, Hawken, & Shapley, 1997). For example, some V1 neurons are initially responsive to low spatial frequencies, but within the span of \( \sim 200 \) ms their preferred spatial frequency shifts to higher values (Bredfeldt & Ringach, 2002; Mazer et al., 2002). Primary visual cortex also shows dynamic shifts in orientation tuning that appear to depend on fast-scale adaptive phenomena, whereby neurons shift their preferred orientation away from the orientation of an immediately preceding stimulus (Dragoi et al., 2000, 2001, 2002; Felsen et al., 2002). It is still unclear what mechanisms underlie these dynamic phenomena, but their temporal evolution is sufficiently slow to expect that it may be possible to measure its characteristics at the perceptual level.

Absence of directional tuning changes

We did not find any obvious dynamic change in directional tuning within our stimulus window of 300 ms. The surface describing directional tuning as a function of time was separable in these two dimensions (Figures 3B and 4), implying that it was well described by a simple temporal modulation in the amplitude of a fixed directional tuning function (not changing in width). The half width at half height of the tuning function was between 30 and 50 deg (S1: 32 ± 9; S2: 44 ± 22; S3: 47 ± 14; direct measurement of half width at half height on linearly interpolated tuning function), which is broadly consistent with related measurements in motion-sensitive single units (e.g., Cook & Maunsell, 2004) but slightly narrower than reported by previous reverse correlation measurements in humans (estimated at \( \sim 70–90 \) degrees from Figure 7 in Murray, Sekuler, & Bennett, 2003).

Our failure to observe dynamic modulations of directional tuning in human vision parallels recent findings on the dynamics of spatial frequency tuning. Mareschal et al. (2006) used a reverse correlation technique that shares many similarities with ours (and with Neri & Heeger, 2002). They masked a Gabor target using spatiotemporal noise and derived temporally modulated perceptual filters for spatial frequency. In contrast to related findings in single neurons (Bredfeldt & Ringach, 2002), they observed no change in spatial frequency selectivity across time, similar to our results on directional selectivity. Interestingly, orientation selectivity in single neurons does not appear to change with time (Mazer et al., 2002).

As a word of caution, it should be emphasized that the above statements only apply within the context of the resolution of our measurements. We sampled directional tuning every 22.5 degrees, and at this resolution we were unable to find convincing evidence for changes in directional tuning across time. We cannot exclude the possibility that, had we sampled the directional tuning curve more finely, we may have observed significant changes.

Temporal modulation of amplitude

The temporal impulse response associated with the perceptual filters in this study shows an early peak between 30 and 60 ms, which declines to 0 modulation between 120 and 180 ms (Figure 2E). A unimodal behavior of this kind is not particularly surprising, as it is commonly observed in single neurons (see Figure 1 in Bair & Movshon, 2004; see also Cook & Maunsell, 2004) and has been explicitly or implicitly demonstrated by various psychophysical investigations (e.g., Ringach, 1998), although never in the context of directional tuning. A more interesting and unexpected feature of the impulse response in Figure 2E is that, after dropping to 0, it rises again towards the very end of the stimulus (between 240 and 300 ms). We observed this effect in at least 2 out of 3 subjects (the third subject showed a hint of this effect, but it was not statistically significant, see Figure 4). Interestingly, although they do not comment about it, close inspection of the data of Tadin et. al (2006) reveals strikingly similar behaviour (their Figures 1C, 2, 3A, 5, and 6A–6C).

This curious feature of the data prompted us to model it using a simple self-normalization scheme, mediated by a minimal feedback circuit. As shown in Figure 1E, the output from the sensory filter in the model is fed back to itself after a temporal delay of roughly 1/10 s. This scheme is analogous to a simple self-correcting unit (e.g., a thermostat), but the temporal delay in the feedback signal causes the impulse response to oscillate with time, generating a second peak in the later portion of the modeled perceptual filter that captures the corresponding modulation in the human data (Figure 5A).
The normalization stage used here shares similarities with previously proposed models of cortical processing in MT (Heeger, Simoncelli, & Movshon, 1996; Simoncelli & Heeger, 1998), the main differences being (1) that the normalizing signal in our model is only between units with same directional preference, whereas MT normalization models use nondirectional normalization; and (2) that our model incorporates a delay component within the normalizing circuit, whereas previous MT normalization models do not specify temporal dynamics. The first difference is not substantial, as self-normalization can be thought of as a variant of pooled normalization but using a weighted pooling function that emphasizes same direction over different directions. The second difference (the temporal delay) is an issue that has not received much attention in the physiological literature, so at present it is not possible to say much about its potential applicability/plausibility in the context of cortical processing.

The delayed self-normalization scheme we used to model our results may also relate to the "self-inhibition" mechanism that has been proposed by previous investigators to implement redundancy reduction in the temporal domain: "The delayed temporal inhibition suppresses the sustained response that would otherwise be evoked by a steady light. Here we formulate the hypothesis that temporal inhibition mediates prediction in the time domain, just as lateral inhibition mediates prediction in the spatial domain" (Srinivasan, Laughlin, & Dubs, 1982, p. 438). In these models, the temporal delay associated with self-inhibition is comparable with the temporal integration characteristics of the mechanism itself. In relation to this specific notion, it is interesting that the delay we observed here (~100 ms) is comparable with the temporal integration window for early motion detectors in human vision (Burr, 1980). It is possible, though very speculative at this stage, that the self-normalization scheme we propose in Figure 1 may, in some conditions, serve the functional role of maximizing temporal information transmission via redundancy reduction in directionally selective motion-sensitive mechanisms in human vision.

Our model is far more efficient than the human observers. This superiority can, at least partly, be attributed to the fact that the model does not incorporate any source of internal noise, whereas human observers are inherently noisy. We quantified human internal noise and showed that its magnitude can account reasonably well (2 out of 3 subjects) for the difference in performance between human and model responses (Figure 6). Nevertheless, the fact remains that human efficiency in our experiments was remarkably low. This result may seem hard to reconcile with the high efficiencies that have been reported by some previous studies of motion detection/discrimination; for example, Barlow and Tripathy (1997) report efficiencies up to 44%.

There are several differences between our stimuli and those used by these authors (e.g., ours involved multiple pulses of motion pairs in temporal succession while Barlow & Tripathy, 1997, used one 2-frame motion pulse), but the most significant difference is perhaps the type of noise that was used by these different studies. Barlow and Tripathy (1997) used random dynamic noise, where dots changed position on every frame and motion pairs occur as "spurious pairs", when dots on successive frames happen to be in spatiotemporal proximity by chance. This type of noise is ideal for targeting the earliest stages of motion detection supported by spatiotemporal correlators and places the bottleneck for performance at this early stage.

In our experiments, this stage is almost entirely bypassed by the stimulus, in that both signal and noise dots consisted of motion pairs. The type of noise we used is likely to place the bottleneck for performance at a later integrative stage, after local motion detection has been completed. Different amounts of internal noise may be associated with accessing these different stages, which may account for the difference in reported efficiency between previous studies and ours. Further work will be necessary to establish whether this explanation can quantitatively account for the discrepancy. It should also be noted that although Barlow and Tripathy (1997) were indeed able to measure efficiencies around 40% in some conditions, for most of their experiments efficiency was much lower, in some conditions as low as 0.1% (see their Figure 9).

Conclusions

Directionally selective mechanisms in human vision display a characteristic temporal modulation that peaks within the first 100 ms, decreases to virtually zero gain within the second 100 ms, and rises slightly again during the third 100 ms. Directional tuning does not change throughout this temporal range, i.e., the directional and temporal responses are separable. We demonstrated this result using two different methods (subtraction of separable prediction and a singular value separability index). We modeled the late modulation using a simple delayed normalization scheme that shares similarities with existing models of MT (Simoncelli & Heeger, 1998). Finally, our measurements indicate that observer internal noise was not negligible in our experiments, possibly accounting for most of the discrepancy between human and ideal performance.

Acknowledgments

Supported by NIH grant RO1EY01728. P.N. is currently supported by the Royal Society (University Research Fellowship). We thank three anonymous reviewers for their useful comments.

Commercial relationships: none.
Corresponding author: Peter Neri.
Email: pn@white.stanford.edu.
Address: School of Optometry, University of California at Berkeley, Berkeley CA 94720, USA.
References


Pack, C. C., & Born, R. T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual...


