Behavioral/Cognitive

Signal Integration in Human Visual Speed Perception

Matjaž Jogan[1,2] and Alan A. Stocker[1,2]

[1]Department of Psychology and [2]Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6314

Object motion in natural scenes results in visual stimuli with a rich and broad spatiotemporal frequency spectrum. While the question of how the visual system detects and senses motion energies at different spatial and temporal frequencies has been fairly well studied, it is unclear how the visual system integrates this information to form coherent percepts of object motion. We applied a combination of tailored psychophysical experiments and predictive modeling to address this question with regard to perceived motion in a given direction (i.e., stimulus speed). We tested human subjects in a discrimination experiment using stimuli that selectively targeted four distinct spatiotemporally tuned channels with center frequencies consistent with a common speed. We first characterized subjects' responses to stimuli that targeted only individual channels. Based on these measurements, we then predicted subjects' psychometric functions for stimuli that targeted multiple channels simultaneously. Specifically, we compared predictions of three Bayesian observer models that either optimally integrated the information across all spatiotemporal channels, or only used information from the most reliable channel, or formed an average percept across channels. Only the model with optimal integration was successful in accounting for the data. Furthermore, the proposed channel model provides an intuitive explanation for the previously reported spatial frequency dependence of perceived speed of coherent object motion. Finally, our findings indicate that a prior expectation for slow speeds is added to the inference process only after the sensory information is combined and integrated.

Key words: **Bayesian model; model predictions; optimal integration; spatiotemporal frequency channels; speed prior; area MT**

Introduction

The relative movements of objects in our visual environment lead to complex patterns of spatiotemporal luminance changes in the retinal images. To form coherent motion percepts, the visual system must first detect and sense these changes at different spatial and temporal frequencies, and then combine the sensory information appropriately. Here, we investigated the computations that underlie this integration in the case of coherent motion.

The rich patterns of incoming visual information are decomposed into their basic spatiotemporal components in primary visual cortex (V1). These components are then appropriately combined and processed along a hierarchy of extrastriate cortical areas to represent more complex features (Felleman and Van Essen, 1991). Most neurons in V1 respond to moving stimuli and are tuned for a specific range in spatiotemporal frequency space (Movshon et al., 1985). The medial temporal (MT) area receives direct input from V1 and is considered the first extrastriate area that integrates visual motion information (Zeki, 1974). While it is relatively well understood how the responses of V1 neurons are combined to form the input to neurons in area MT (in particular with regard to their direction tuning; Adelson and Bergen, 1985; Simoncelli and Heeger, 1998; Perrone and Thiele, 2002; Rust et al., 2006; Solomon et al., 2011), it remains unclear how this neural integration relates to motion perception. What makes this question challenging but interesting is the fact that perceived motion depends on stimulus contrast and spatial frequency (Thompson, 1982; Smith and Edgar, 1991). Several studies have investigated the potential link between changes in motion percepts and the contrast and spatial frequency-dependent changes in the response characteristics of neurons in area MT (Churchland and Lisberger, 2001; Priebe and Lisberger, 2004; Liu and Newsome, 2005; Priebe et al., 2006; Stocker et al., 2009). Yet the results are at best not conclusive (for a more in depth discussion, see Krekelberg et al., 2006).

Figure 1 illustrates the conceptual framework within which we considered the problem of motion integration. We assumed a motion stimulus with a rich spatiotemporal frequency spectrum. For simplicity, we only considered coherent motion along a given motion direction (i.e., visual speed). We started with the assumption that stimulus motion is represented in a set of independent sensory channels (Campbell and Robson, 1968; Graham and Nachmias, 1971) each tuned for a specific spatiotemporal frequency band (Jogan and Stocker, 2011, 2013; Simoncini et al., 2012). We then asked the question how the visual system integrates the information provided by these channels and, potentially, combines it with prior expectations to form a coherent percept of motion. We formulated three Bayesian observer models (Stocker and Simoncelli, 2006) that differed only in the way they integrated information across the channels: optimally, by considering only the channel with the most reliable signal, or by forming an average percept based on each individual chan-

Received Nov. 24, 2014; revised May 9, 2015; accepted May 15, 2015.

Author contributions: M.J. and A.A.S. designed research; M.J. and A.A.S. performed research; M.J. and A.A.S. analyzed data; M.J. and A.A.S. wrote the paper.

This work was supported by the Office of Naval Research (Grant N000141110744). We thank the members of the Computational Perception and Cognition Laboratory and the reviewers for helpful feedback at various stages of the project.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Matjaž Jogan, Computational Perception and Cognition Laboratory, 3401 Walnut Street 304C, Philadelphia, PA 19104-6228. E-mail: mjogan@sas.upenn.edu.

DOI:10.1523/JNEUROSCI.4801-14.2015

nel. We performed a two-alternative forced-choice (2AFC) speed-discrimination experiment in which we selectively targeted four different spatiotemporal frequency channels. We validated the models against the data and found that only a Bayesian channel model with optimal signal integration can accurately predict the data both in terms of discrimination thresholds and perceived speeds.

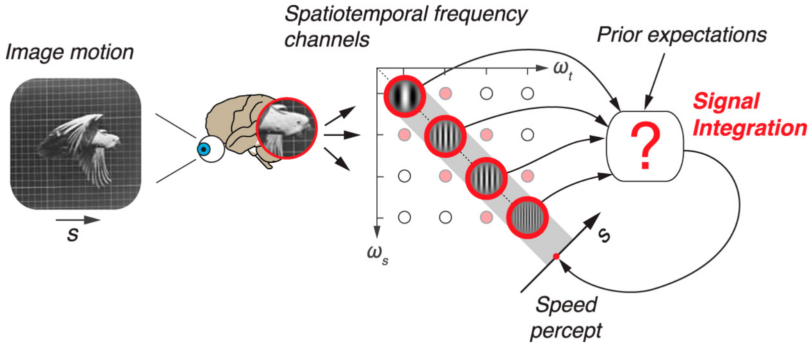


Figure 1. Signal integration across spatiotemporal frequency channels. Natural image motion typically exhibits a rich pattern of luminance changes over space and time, which is reflected in its broad spatiotemporal frequency spectrum. We assume that the early visual system, using channels that are each tuned to motion energy in a specific spatiotemporal frequency band, decomposes image motion into basic spatiotemporal signal components. Image motion with a coherent speed *s* along a given direction triggers responses in a set of spatiotemporal frequency channels whose preferred frequency tuning (ω_s, ω_t) is consistent with that speed (for simplicity, we illustrate only 1 spatial dimension). In this paper, we address the question how the visual system integrates the responses of these channels together with potential prior expectations to form a coherent percept of visual speed.

Materials and Methods

Four subjects participated in the speed-discrimination experiment (one female; three males). All but one subject were naive with regard to the purpose of the study at the time they were participating. Participants had normal or corrected-to-normal vision and all gave informed consent before the experiment. The study was approved by the University of Pennsylvania Institutional Review Board (protocol #813601). During the experiment, subjects were sitting in a darkened room and their head position was controlled with a chin rest. Stimuli were displayed at a distance of 60 cm on a Samsung Dell P992 CRT 17 inch computer display with 120 Hz refresh rate and 1024 × 768 pixel resolution. Gamma was corrected. The experiment was programmed in Matlab (Mathworks) using display routines from the MGL toolbox (http://justingardner.net/mgl), and was executed on an Apple Mac Pro computer with a 2.93 GHz quad-core Intel Xeon processor running OS X 10.6.8.

Stimuli. Stimuli were gratings, generated by taking one-dimensional bandwidth-limited random noise signals and replicating them along the second spatial dimension (Fig. 2). The random noise signals were created by inverse Fourier transforms (random phase). Each stimulus was defined by its spatial frequency spectrum with nonzero amplitudes only within narrow frequency bands ($b = 0.04\ \omega_s$) centered on four spatial frequencies $\omega_s \in \{0.5, 1, 2, 4\}$ cycles/° visual angle. The spectrum was uniform over the bands. Stimuli either had a single-band spectrum (Fig. 2*b*, single-channel conditions A–D) or a spectrum that consisted of various combinations of the single-band spectra (Fig. 2*c*, combined channel conditions AD, ABD, ABCD). Coherent motion stimuli were generated by rigidly translating the gratings at a given speed behind a static aperture. The aperture size was 4° and was smoothed with a circular cosine window of the same width. Stimulus intensity over time was modulated by a tapered cosine window (100 ms fade-in/fade-out; 600 ms total stimulus duration).

Stimulus calibration. Subjects first participated in a calibration procedure whose purpose was to individually adjust the spectral energies of the stimuli targeting single channels such that the subjects' discrimination thresholds for these stimuli were approximately within a desired range. The goal was to create single-channel stimuli that provided equally reliable sensory information. Subjects compared the speed of a test and reference stimulus pair that targeted the same spatiotemporal channel (same spatial frequency spectrum, balanced condition; Fig. 3*b*). The test stimulus was always moving at $s_t = 3°/\text{s}$ while the reference was moving at one of two fixed reference speeds, $s_{r1} < s_t < s_{r2}$ that were equally distant from the test (in the log-normalized space; see Eq. 1, below). Subjects were asked to select the stimulus that they perceived as moving faster and received feedback after each trial. We adaptively adjusted the amplitude of the spatial frequency spectrum of both the test and reference stimuli until the subject's discrimination thresholds approached a predefined target level. This adjustment was guided by the following procedure: we modeled a sequence of N recent trials at a particular reference speed s_r as a Bernoulli process with an unknown parameter Θ that describes the probability of subjects answering "reference faster." For $\Theta = \theta$, the probability of K "reference faster" answers in the past N trials is given by the binomial distribution $B(N, \theta)$. Given K and N, we were able to continuously infer the posterior probability of Θ by calculating the beta distribution, $\Theta \propto B(1 + N, 1 + K - N)$ (Bayes and Price, 1763). We formed a current estimate $\hat{\Theta}$ of the probability value by taking the mean of the posterior. We computed this estimate whenever the variance of the posterior distribution was below a certain threshold and reset the counter N. Based on this estimate, we then increased or decreased the spectral energies of the stimuli depending on some target probability values, assuming that increased energies lead to a decrease in threshold. The target probability values were $\Theta = 0.25$ and $\Theta = 0.75$, respectively, which correspond to a psychometric function with a slope of 0.6 (cumulative Gaussian in normalized log-units). This slope value is equivalent to a stimulus noise level of $\sigma = 0.6/\sqrt{2}$ according to signal detection theory (SDT; Green and Swets, 1966). Each staircase was terminated after 200 trials, leading to a total of 1600 trials per subject. The calibration procedure resulted in single-channel stimuli with individual spectral energies for individual subjects. Across all subjects, we found that the different channels had very different sensitivities. Specifically, the average stimulus power (integral over the power spectrum, scaled to represent displayed luminance values, averaged across subjects) for each channel was as follows: A, 0.9 cd/m^2; B, 1.9 cd/m^2; C, 2.5 cd/m^2; and D, 8.4 cd/m^2, corresponding to the following maximum contrast values (Michelson contrast): A, 3.0%; B, 8.7%; C, 12.7%; and D, 45.0%. These characterizations are in agreement with previous findings that reported decreased motion sensitivities at high (and very low) spatial frequencies (Chen et al., 1998).

Speed-discrimination experiment. Subjects performed a 2AFC visual speed-discrimination experiment (Fig. 3*a*). Each trial started with a fixation period (400 ms) that was followed by the presentation of a reference and a test grating on the left and right side of the fixation mark (600 ms). Positions were randomly assigned. Gratings were presented at 6° eccentricity. Both gratings were drifting in the same direction, either down-leftwards or down-rightwards randomly assigned. After the gratings disappeared, an indicator (white square) randomly appeared to the left or right of the fixation mark (duration, 300 ms; eccentricity, 0.6°; size, 0.3°), and the subject had to answer whether the grating on the indicated side was drifting faster or slower than the grating on the other side. The purpose of the indicator was to dissociate a subject's answer (yes/no) from the identity of the stimulus (faster/slower) as a precautionary measure to avoid potential decision biases. All experiments were self-paced, i.e., subjects had to push a button to start a new trial.

We characterized seven different stimuli from a total of 13 different 2AFC stimulus conditions. Seven balanced conditions (Fig. 3*b*) had a test and a reference stimulus with identical spatial frequency spectrum while in six unbalanced conditions (Fig. 3*c*) the test stimulus was compared with a reference stimulus that targeted all channels (ABCD). Stimuli were as described above and shown in Figure 2. Conditions were fully interleaved. Subjects did not receive feedback. The speed of the test grating

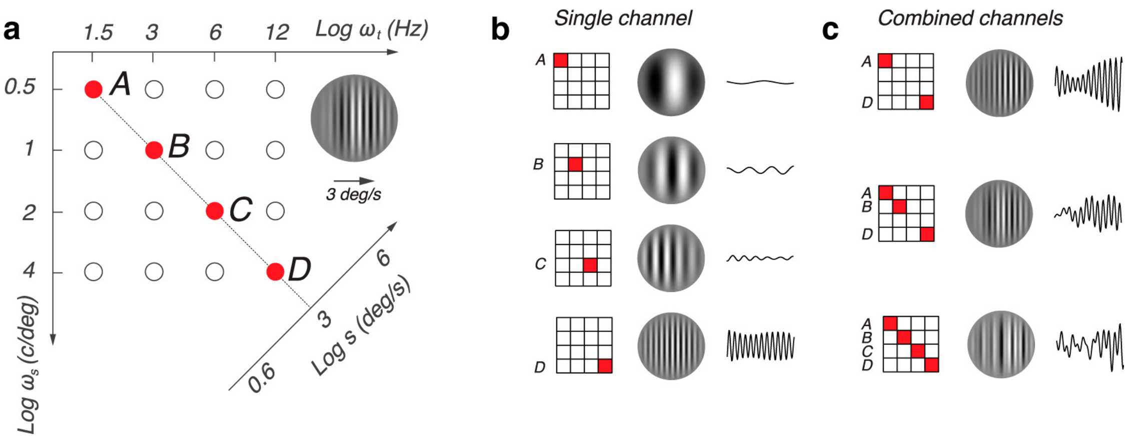


Figure 2. Motion stimuli targeting individual spatiotemporal frequency channels. ***a***, A stimulus that coherently moves with speed *s* has motion energy distributed along a line in the spatiotemporal frequency space (in log units), and therefore can be detected by spatiotemporal channels that are tuned for frequencies along this line. In our experiments we used grating stimuli drifting at 3° per second that had energy in narrow bands (0.04 ω_s) around four distinct spatial frequencies $\omega_s \in \{0.5, 1, 2, 4\}$ cycles/°, respectively. We assumed that these stimuli targeted four independent spatiotemporal frequency channels (A–D, red dots). ***b***, Synthesized example stimuli that target each of the four channels individually. ***c***, Stimuli that target two (A, D), three (A, B, D), or four (A–D) channels simultaneously. They are synthesized by combining the frequency spectra of the corresponding single-channel stimuli. This allowed us to compare perceptual performance for complex stimuli to that of their single-band components. Curves next to each stimulus represent the corresponding luminance profiles (before windowing).

was always 3° per second while the reference speed was governed by two adaptive staircases that each terminated after 100 trials. This led to a total of 2600 trials, which subjects completed in six sessions. At the beginning of each session, subjects performed a brief training run to familiarize themselves with the task (20 trials). We characterized the percepts of the test stimuli in each condition by extracting discrimination thresholds and matching speeds using a joint SDT analysis (Fig. 3). The confidence in the extracted values of both parameters was assessed by bootstrapping the data (100 iterations) and calculating the 95% sample intervals.

Bayesian channel models with different forms of signal integration. We tested three different variations of a Bayesian observer model for speed perception, which we formulated with regard to a logarithmic speed space of the following form (Eq. 1): $s = \log(1 + s_{\text{linear}}/s_0)$, where $s_0 = 0.3$ is a small normalization constant. It has been shown that in this space, stimulus uncertainty is approximately constant over speed (Nover et al., 2005), which simplifies the Bayesian model formulation (Stocker and Simoncelli, 2006). We built on our previous modeling framework that allowed us to account for subjects' behavior in a 2AFC speed-discrimination task at the level of individual psychometric functions (Stocker and Simoncelli, 2006). The original model assumed that subjects estimate the speed of a stimulus based on a single likelihood function. Here, we augmented the model by assuming that the sensory information is distributed in the responses of independent spatiotemporal frequency channels. We assumed that a complex motion stimulus is driving the channels according to its motion energy in the corresponding frequency bands, eliciting a measurement vector $\vec{m} = [m_A, m_B, m_C, m_D]$. We parameterized the likelihood function $p(m_x \mid s)$ for each channel (channel likelihood) as a Gaussian (Eq. 2):

$$p(m_X|s) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{(s-m_X)^2}{2\sigma_X^2}\right).$$

The likelihood width σ_X depends on how strongly the channel is driven. Thus we assume that the likelihood function is uniform for nonactive channels.

In addition, we assumed that subjects' prior expectations follow a power-law function (Stocker and Simoncelli, 2006). In the logarithmic speed space, the logarithm of this prior can be expressed as a linear function $\log(p(s)) = as + b$, where a is the exponent of the power law. Finally, we assumed that perceived speed $\hat{s}$ equals the speed with maximal posterior probability. With these basic assumptions, we defined three Bayesian observer models that only differ in the way they integrate the signals across the individual channels. Note that all model formulations are expressed in the normalized log-speed space (Eq. 1).

Optimal integration. The "optimal model" integrates the information from all the channels (Fig. 4*a*). Assuming that the noise in the channels is independent, the model's likelihood function is the product of the individual channel likelihoods. With the above-described parameterizations of the channel likelihoods (Eq. 2) and the prior (Eq. 3) we can write the posterior as follows (Eq. 4):

$$p(s_{\text{opt}}|\vec{m}) = \frac{1}{\alpha} \exp\left[-\sum_X \frac{(s-m_X)^2}{2\sigma_X^2} + (as+b)\right],$$

with α a normalization factor. According to the chosen loss function, the percept (estimate) $\hat{s}_{\text{opt}}$ is then the value of s that maximizes the exponent of Equation 4; thus, the following equation (Eq. 5):

$$\hat{s}_{\text{opt}}(\vec{m}) = \text{argmax}_s\left[-\sum_X \frac{(s-m_X)^2}{2\sigma_X^2} + (as+b)\right].$$

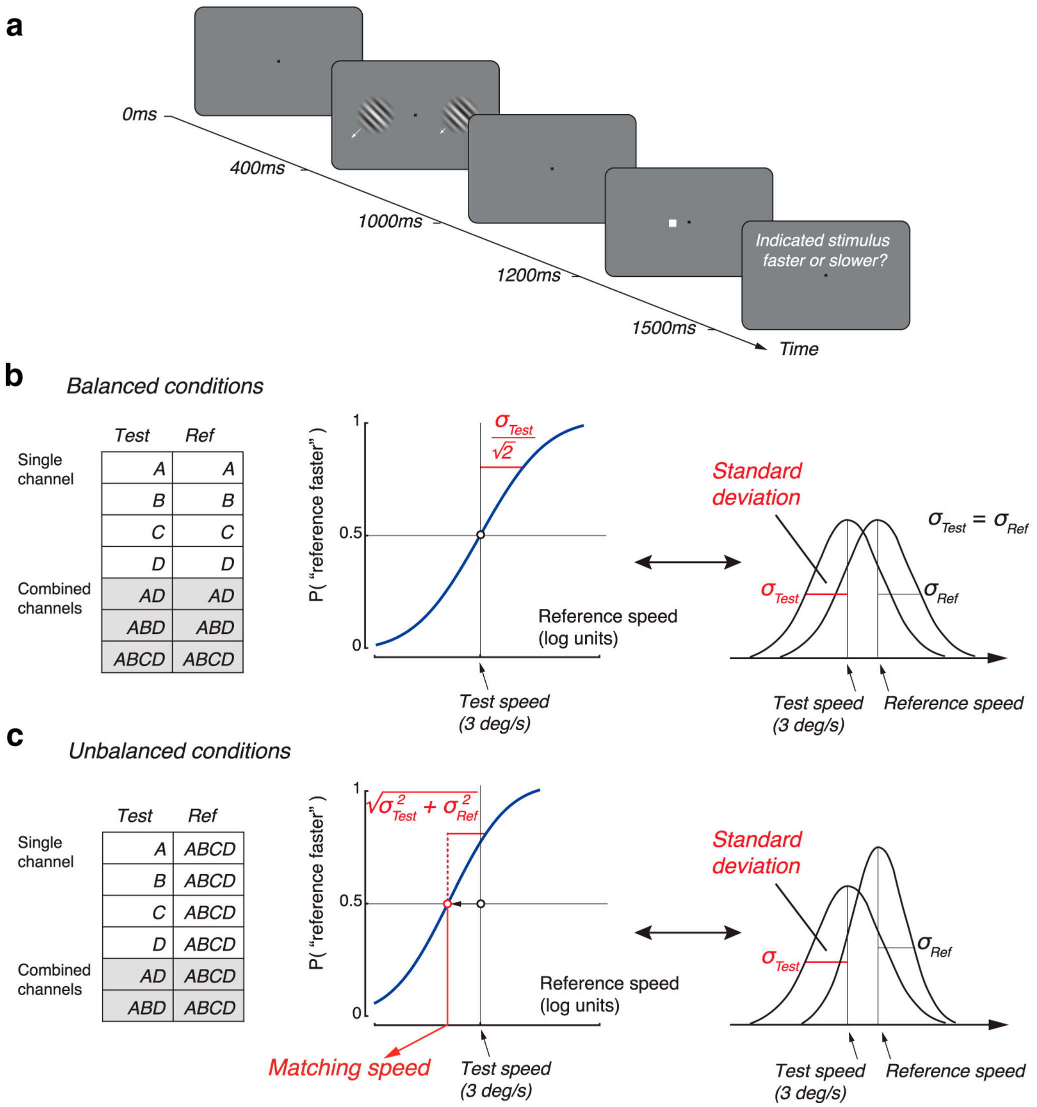
For example, for a stimulus that targets the two channels A and D (Fig. 2*c*), the observer model predicts a percept as follows (Eq. 6):

$$\hat{s}_{\text{opt}}(m_A, m_D) = \frac{\sigma_D^2}{\sigma_A^2 + \sigma_D^2} m_A + \frac{\sigma_A^2}{\sigma_A^2 + \sigma_D^2} m_D + a\frac{\sigma_A^2\sigma_D^2}{\sigma_A^2 + \sigma_D^2}.$$

To be able to compare the model's predictions to the data from the 2AFC experiment, we need a description of the distribution of percepts over repeated trials; thus, $p(\hat{s}_{\text{opt}}|s)$. In general, the full distribution is computed by mapping and marginalizing the estimation function $\hat{s}_{\text{opt}}(\vec{m})$ over the distributions of the sensory measurement vector $p(\vec{m}|s)$. With the assumptions that (1) the prior is smooth in the speed range we are considering (i.e., the exponent a is approximately constant; Eq. 3) and (2) the speed dependence of the likelihood width is weak (in log space), we have previously shown that the distribution is well approximated by a Gaussian (Stocker and Simoncelli, 2006). Mean and variance of this Gaussian can be computed for an arbitrary number of channels. For example, in the case of a stimulus targeting channels A and D, the mean is as follows (Eq. 7):

$$\begin{aligned} E\langle\hat{s}_{\text{opt}}|s\rangle &= \frac{\sigma_D^2}{\sigma_A^2 + \sigma_D^2} E\langle m_A|s\rangle + \frac{\sigma_A^2}{\sigma_A^2 + \sigma_D^2} E\langle m_D|s\rangle + a\frac{\sigma_A^2\sigma_D^2}{\sigma_A^2 + \sigma_D^2} \\ &= \frac{\sigma_D^2}{\sigma_A^2 + \sigma_D^2} s + \frac{\sigma_A^2}{\sigma_A^2 + \sigma_D^2} s + a\frac{\sigma_A^2\sigma_D^2}{\sigma_A^2 + \sigma_D^2} \qquad . \\ &= s + a\frac{\sigma_A^2\sigma_D^2}{\sigma_A^2 + \sigma_D^2} \end{aligned}$$

Similarly, we can approximate the variance as follows (Eq. 8):



	Test	Ref
Single channel	A	A
	B	B
	C	C
	D	D
Combined channels	AD	AD
	ABD	ABD
	ABCD	ABCD

	Test	Ref
Single channel	A	ABCD
	B	ABCD
	C	ABCD
	D	ABCD
Combined channels	AD	ABCD
	ABD	ABCD

Figure 3. 2AFC speed-discrimination experiment. ***a***, Subjects performed a 2AFC speed-discrimination experiment characterizing seven different motion stimuli. ***b***, Seven balanced conditions consisted of a test and a reference stimulus with identical spatial frequency spectrum. ***c***, Six unbalanced conditions had a reference stimulus with a spatial frequency spectrum that targeted all four channels (stimulus condition ABCD; Fig. 2*c*). Matching speeds of these psychometric functions indicate how fast the reference grating had to drift to be perceived as fast as the test (always drifting at 3° per second). It represents a relative measure of the perceived speed of the test stimulus (in units of the reference speed). Discrimination thresholds were obtained from a joint SDT analysis: for each test stimulus the corresponding pairs of balanced and unbalanced conditions were jointly fit to extract the SD σ_X of the test stimulus distribution. The SD σ_{Ref} was set to the value extracted from the fit of the balanced condition ABCD. All measurements are performed in a logarithmic space of visual speed. See Materials and Methods for more details.

$$
\begin{aligned}
var\langle \hat{s}_{\text{opt}}|s\rangle &\approx \left(\frac{\sigma_D^2}{\sigma_A^2+\sigma_D^2}\right)^2 var\langle m_A|s\rangle + \left(\frac{\sigma_A^2}{\sigma_A^2+\sigma_D^2}\right)^2 var\langle m_D|s\rangle \\
&= \left(\frac{\sigma_D^2}{\sigma_A^2+\sigma_D^2}\right)^2 \sigma_A^2 + \left(\frac{\sigma_A^2}{\sigma_A^2+\sigma_D^2}\right)^2 \sigma_D^2 \\
&= \frac{\sigma_A^2\sigma_D^2}{\sigma_A^2+\sigma_D^2} \quad .
\end{aligned}
$$

Having a description for the distributions of the model percepts over trials allows us to use SDT to directly generate model predictions for the full, experimentally measured psychometric functions (see Model predictions of the psychometric functions).

Maximally reliable channel. The "max model" only considers the channel that provides the most reliable sensory response (Fig. 4*b*). Its formulation is identical to the optimal model with the exception that the likelihood function equals the channel likelihood with smallest variance. Thus the mean and variance of the predicted percept are expressed as (Eq. 9)

$$E\langle \hat{s}_{max}|s\rangle = s + a\sigma_{\text{min}}^2,$$

and (Eq. 10)

$$var\langle \hat{s}_{max}|s\rangle = \sigma_{\text{min}}^2,$$

respectively.

Channel averaging. The "averaging model" assumes that an independent Bayesian estimate is performed for each channel (Fig. 4*c*). A posterior (Eq. 11)

$$p(s_X|m_X) = \frac{1}{\alpha}\exp - \left[\frac{(s-m_X)^2}{2\sigma_X^2} - (as+b)\right]$$

and subsequently an individual estimate (Eq. 12)

$$\hat{s}_X(m_X) = m_X + a\sigma_X^2$$

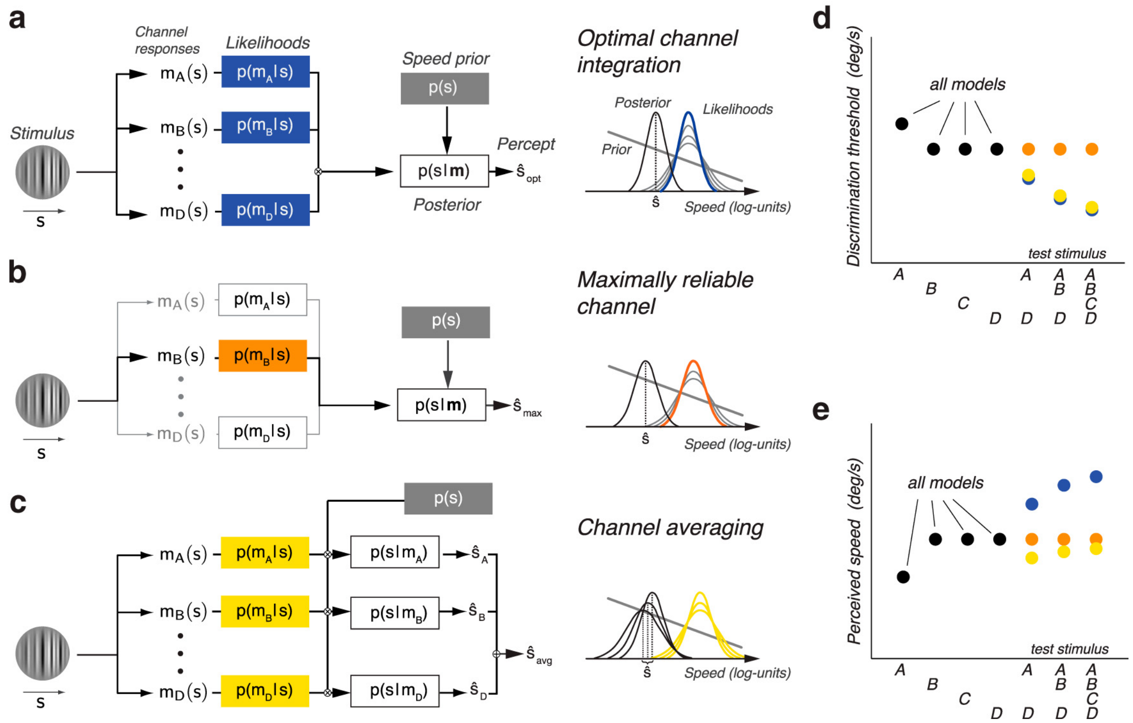


Figure 4. Observer models with different forms of signal integration. ***a***, Bayesian observer model performing optimal channel integration (optimal model). The model combines the sensory information contained in all channel responses (m_A, m_B, m_C, m_D) by multiplying the individual channel likelihoods. The combined likelihood function is then multiplied with a prior probability distribution over speed $p(s)$ according to Bayes' rule. The characteristic feature of the optimal integration model is that its likelihood function is always narrower than any of the individual channel likelihoods. ***b***, Bayesian observer model that only considers the most reliable channel response (max model). The likelihood function of this model is always identical with the likelihood function provided by the most reliable channel. ***c***, Observer model that averages independent estimates across all channels (averaging model). This model combines each channel likelihood with a prior, forms individual estimates $\hat{s}_X$, and then takes an average. ***d***, Model predictions for discrimination thresholds. We compared predictions for a given set of noise parameters and the prior exponent. All models, and therefore their predictions, are equivalent for stimuli that activate only one channel. For combined stimuli, the optimal (blue) and the averaging model (yellow) both predict a similar decrease in threshold while the max model (orange) predicts a threshold that is fixed at the level associated with the most reliable channel. ***e***, Assuming a slow-speed prior (i.e., a negative exponent in Eq. 3), all models predict the characteristic inverse relationship between threshold and perceived speed for single-channel conditions. However, only the optimal model predicts an increase in perceived speed for stimuli that activate multiple channels.

can be formulated for each channel X. The model percept then reflects the average estimate across all k channels; thus (Eq. 13):

$$\hat{s}_{\text{avg}}(\vec{m}) = \frac{1}{k}\sum(m_X + a\sigma_X^2).$$

Its mean and variance are (Eq. 14)

$$E\langle \hat{s}_{\text{avg}}|s\rangle = s + \frac{a}{k}\sum\sigma_X^2$$

and (Eq. 15)

$$var\langle \hat{s}_{\text{avg}}|s\rangle = \frac{1}{k^2}\sum\sigma_X^2,$$

respectively. Unlike the optimal integration model, we assume the averaging model to operate only on active channels, i.e., we implicitly assume that there is a thresholding mechanism that decides whether a channel is active or not.

Model predictions of the psychometric functions. The description of the models in terms of estimation mean and variance allows us to predict subjects' perceptual behavior in the 2AFC speed-discrimination task. As stated earlier, we assume that a subject's percept $\hat{s}$ of stimulus speed s over repeated trials follows a distribution $p(\hat{s}|s)$ that is well approximated by a Gaussian with mean and variance as derived above (e.g., Eqs. 7 and 8 for the optimal model). We can define this distribution for any model, stimulus type, and speed tested in our experiment. More specifically, we can define two distributions $p(\hat{s}_{\text{Test}}|s_{\text{Test}})$ and $p(\hat{s}_{\text{Ref}}|s_{\text{Ref}})$ for the test and the reference stimulus, respectively. According to SDT (Green and Swets, 1966), the probability that the reference is perceived to move faster than the test is as follows (Eq. 16):

$$P(\hat{s}_{\text{Ref}} > \hat{s}_{\text{Test}}) = \int_0^\infty p(\hat{s}_{\text{Ref}}|s_{\text{Ref}}) \int_0^{\hat{s}_{\text{Ref}}} p(\hat{s}_{\text{Test}}|s_{\text{Test}})\, d\hat{s}_{\text{Test}}\, d\hat{s}_{\text{Ref}}.$$

This represents a natural way to embed the Bayesian observer models in an SDT framework (Stocker and Simoncelli, 2006). It provides a description of subjects' perceptual behavior at the level of individual psychometric functions. Equation 16 also allows us to fit the individual models to the measured psychometric functions using a maximum likelihood optimization methods, as well as to quantify the accuracy of the model predictions in terms of their overall likelihood value in explaining the data ("goodness-of-prediction").

Results

We tested subjects in a 2AFC speed-discrimination experiment to measure their discrimination thresholds and matching speeds for all stimuli in our test set (Fig. 2*b*,*c*). The experiment consisted of 13 different stimulus conditions (test/reference pairs; Fig. 3). The power spectra for the single-channel stimulus components were chosen according to a calibration procedure. We then used the data of the single-channel conditions to predict subjects' perceptual behavior in the combined channel conditions (see Fig. 6) according to three Bayesian observer models with different forms of channel integration (Fig. 4).

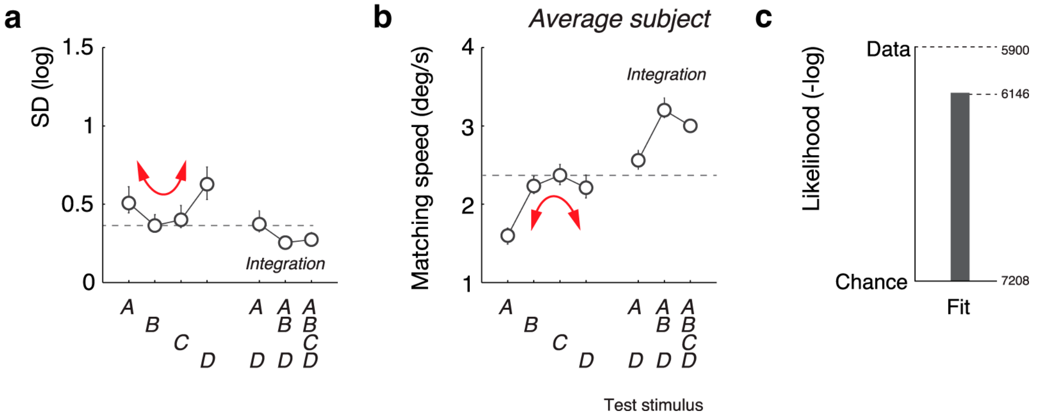


Figure 5. Measured discrimination thresholds and matching speeds. ***a***, ***b***, Discrimination thresholds (***a***) and matching speeds (***b***) for all stimulus conditions (shown for the average subject). Discrimination thresholds are indicated as SD σ_X according to a joint SDT analysis of the data from both the balanced and unbalanced conditions (Fig. 3*b*,*c*). Thresholds are lower for conditions with multiple active channels and perceived speeds are higher for lower discrimination thresholds (see data for individual subjects in Fig. 6). The inverse relationship between discrimination threshold and matching speed (red arrows) is a signature of a Bayesian observer model with a prior expectation for slow speed. It also suggests that dependencies of perceived speed on stimulus contrast (spectral energy) and spatial frequency can be reduced to the joint effect of these parameters on signal uncertainty. Error bars indicate the 95% confidence interval based on 100 bootstrapped samples of the data. ***c***, Goodness-of-fit as measured in negative log-likelihood. The value is presented relative to two reference values: the likelihood of a model observer providing random answers in the 2AFC task (Chance) and the likelihood of the data under a binomial distribution using the empirical probabilities of the measured psychometric function, i.e., data explaining itself (Data). The difference to the latter indicates the deviation of the data from a cumulative Gaussian description and thus reflects the amount of noise in the data.

Discrimination thresholds and matching speeds

Figure 5 shows the extracted discrimination thresholds (indicated as the SD of the noise from a joint SDT analysis; Fig. 3) and the matching speeds for the average subject (i.e., the psychometric functions were computed across trial data from all subjects). The discrimination thresholds for stimuli targeting channels A–D in isolation were comparable yet generally higher for the stimuli with the highest and lowest spatial frequency bands. The thresholds for the combined stimuli targeting multiple channels, however, are without exception all lower than the individual thresholds for each of their stimulus components. In addition, the threshold generally decreases for stimuli that target an increasing number of channels. Both effects are a clear indication of channel integration, signaling that the uncertainty of the overall sensory representation decreases by combining information across independent channels. The pattern is exactly reversed for the extracted matching speeds. The matching speeds for the combined stimulus conditions are all higher than the matching speeds for any of their individual single-channel stimulus components. Likewise, the matching speed is generally higher for stimuli targeting multiple channels. Matching speed is a relative measure of the perceived speed of the test stimulus in units of the reference stimulus. Thus this inverse relationship between perceived speed and discrimination threshold soundly supports the prediction of a Bayesian model with a prior expectation of slow speeds: the higher the signal uncertainty (thus the higher the discrimination threshold), the stronger the effect of the prior and thus the slower the perceived speed. This behavior is well preserved across all stimulus conditions tested.

Optimal channel integration best predicts the data

The above qualitative comparison of the measured discrimination thresholds and matching speeds (Fig. 5) with the model characteristics (Fig. 4*d*,*e*) already indicates that the optimal model may best capture the characteristics of the observed perceptual behavior. To perform a more quantitative model comparison, we tested how well each model can predict the subjects' psychometric functions for the combined stimulus conditions based on the data from the single-channel conditions. We fit each model to the data from both the balanced and unbalanced single-channel stimulus conditions. Note that because the models are equivalent with regard to single-channel stimulus conditions, their fit model parameters (i.e., the channel likelihood widths σ_A, σ_B, σ_C, and σ_D, and the local prior exponent α) should be identical as well. Thus, we constrained the reference likelihood σ_{Ref} in the unbalanced conditions to be the empirical values extracted from the balanced stimulus condition ABCD to obtain identical fits of the different models for the single-channel conditions in the 2AFC experiment. Fit parameter values for all subjects are listed in Table 1. The likelihood widths directly reflect the stimulus noise levels. Their fit values lie within a reasonable range of the target value of the calibration procedure (≈0.42). The fit prior exponents are similar to previously found values (Stocker and Simoncelli, 2006; Hedges et al., 2011; Sotiropoulos et al., 2014). The data from the balanced and unbalanced single-channel conditions fully constrained these parameters.

Table 1. Values of the fit model parameters (identical for all models)

Subject	σ_A	σ_B	σ_C	σ_D	σ^a_{Ref}	Prior exponent α
#1	0.53	0.42	0.42	0.39	0.24	−2.70
#2	0.39	0.31	0.24	0.23	0.20	−5.00
#3	0.67	0.52	0.54	0.65	0.39	−2.71
#4	0.50	0.39	0.33	0.45	0.21	−1.50
Average	0.59	0.44	0.43	0.48	0.27	−1.97

[a] σ_{Ref} Was not actually fit but was set to the empirical value of σ_{ABCD} obtained from the balanced condition ABCD.

We then used the fit model parameters to predict perceptual behavior for the combined stimulus conditions according to each of the three models. Predictions consisted of the full psychometric functions from which thresholds and matching speeds were extracted. Figure 6 shows the extracted thresholds and matching speeds for all stimulus conditions and all subjects (plus the average subject) together with the model predictions. While the models (equally) well fit the data for the single-channel conditions, only the optimal model also well predicted the data for the joint-channel conditions. The optimal model is the only model that can account for the increase in matching speed for stimuli that target multiple channels. We further quantified this by computing a goodness-of-prediction measure for each model, which we de-

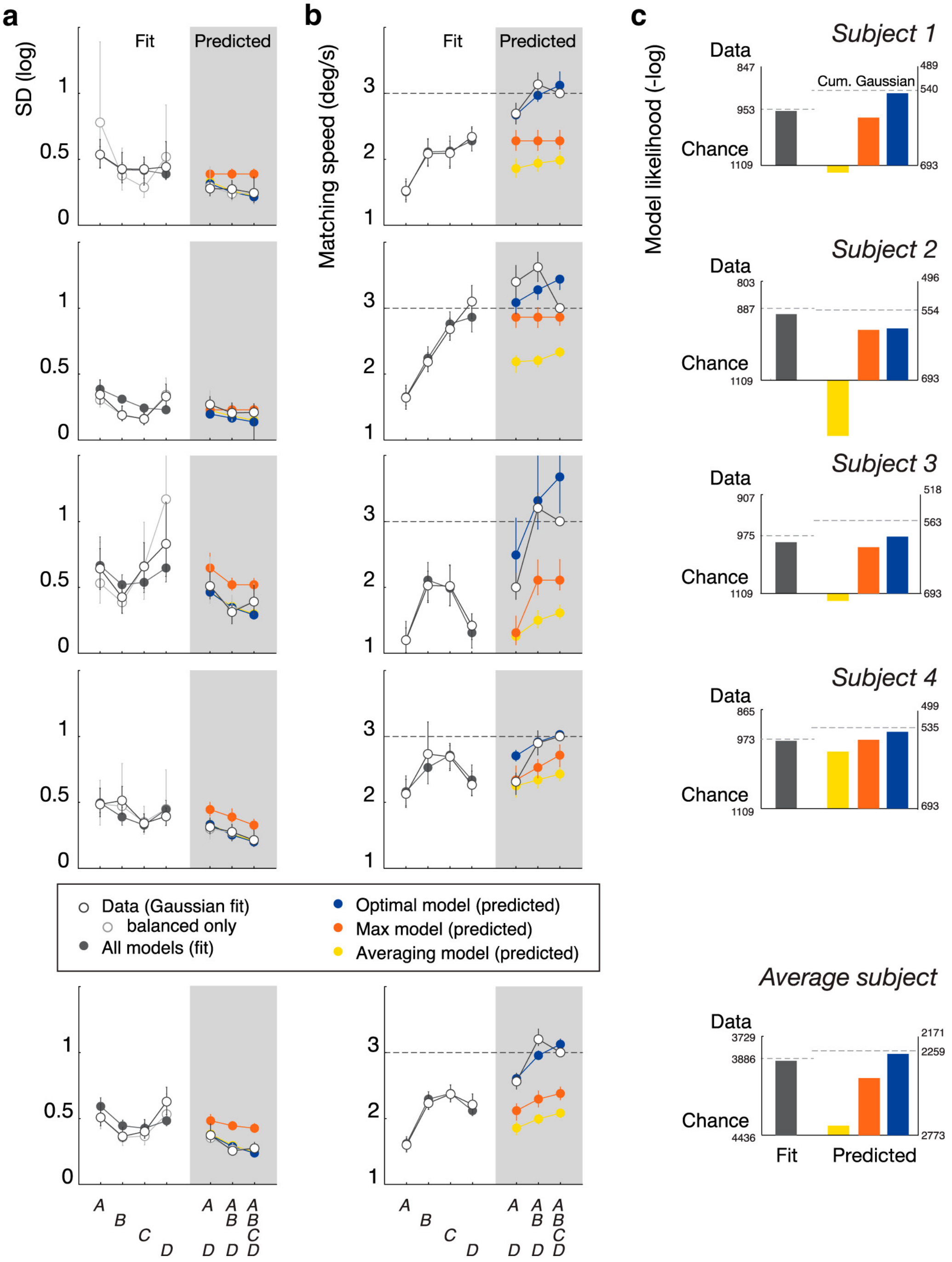


Figure 6. Model predictions of discrimination thresholds and matching speeds. We evaluated how well discrimination thresholds and matching speeds for combined channel stimuli can be predicted from model fits to single-channel stimulus conditions. Predicted conditions are those with gray background. Model fit and predictions were based on embedding the observer model within the SDT framework as described in Materials and Methods. Error bars indicate the 95% confidence interval based on 100 bootstrapped samples of the data. ***a***, Panels in this column show the measured and predicted discrimination thresholds for all four subjects (and the average subject). Light gray data points represent the measured thresholds for the balanced stimulus conditions only. Model fits are identical for the single-channel conditions because we constrained the noise parameter of the reference stimulus σ_{Ref} to be the value directly obtained from the balanced condition ABCD (Fig. 3). Predictions of the optimal and the average model are similar and close to the data, while the thresholds predicted by the max model are in general higher. ***b***, Predictions for matching speed, however, clearly suggest that only the optimal model can accurately account for the increased matching speeds in the combined stimulus conditions. ***c***, Goodness-of-predictions: negative log-likelihood values for the predicted psychometric functions, relative to the range set by the likelihood of a coin-flip model (Chance) and the likelihood of the data itself (Data; for details, see Fig. 5). Dashed lines represent the likelihood values provided by individual cumulative Gaussian fits to the measured psychometric functions. The optimal model best predicts the data for all subjects. Its predictions almost reach the likelihood levels set by individual cumulative Gaussian fits to the data (average subject). We think this is quite remarkable, particularly given that the predictions were for 5 of 13 stimulus conditions. Thus ∼40% of the data were predicted from the other 60%. For completion, the likelihood values for the fits to single-channel conditions only are also shown (black bar).

fined as the log-likelihood of the data being predicted. For every subject, the optimal model outperforms both the max and the averaging model (Fig. 6*c*). The fact that the values for the optimal model are close to the values obtained by fitting individual cumulative Gaussians to each stimulus condition further indicates that the optimal model is not only outperforming the other models but is also accurately predicting perceptual behavior.

Discussion

We have demonstrated that human visual speed perception can be accurately described as the result of a Bayesian inference process that optimally integrates visual speed information across different spatiotemporal frequency channels. We experimentally measured human subjects' speed-discrimination performance using a set of synthesized visual motion stimuli that specifically targeted four independent spatiotemporal frequency channels. Stimuli either targeted each channel individually or various combinations of channels simultaneously. Throughout all stimulus conditions, the data showed a distinct inverse correlation between discrimination thresholds and matching speeds. This correlation is a signature of a Bayesian observer model with a prior belief for slow speeds. In addition, we were able to successfully predict individual subjects' perceived speeds for the combined stimuli based on their data from the single-channel conditions. We compared the predictions of a novel Bayesian channel model that optimally integrates speed information across all channels with those of a model that assumed only the most reliable channel (max model) or performed a weak form of integration (Landy et al., 1995; Yuille and Bülthoff, 1996; averaging model). The optimal model clearly outperformed both alternative models in terms of their measured goodness-of-prediction value (log-likelihood). Its predictions almost as well accounted for the measured psychometric functions as the fits of those functions with individual cumulative Gaussians (average subject). The model comparison is particularly fair given that the different models have exactly the same model parameters and are computationally equivalent for single-channel stimulus conditions. A model analysis based on goodness-of-prediction rather than goodness-of-fit circumvents the problem of overfitting, a problem often associated with Bayesian observer models because of their relative large power for the typically small amount of data available (Jones and Love, 2011).

The presented work extends the model and results of previous work (Stocker and Simoncelli, 2006). It provides further experimental evidence for the notion that perceived visual speed is the result of Bayesian inference with a prior expectation for slow speeds (Weiss et al., 2002). It introduces an augmented model formulation that is a step toward a more biophysically detailed Bayesian observer model. The new model assumes that inference is based on a distributed and implicit representation of visual speed. It allows us to incorporate known aspects of the neural organization of the visual motion pathway without giving up the rigor of a normative modeling approach that can explain perceptual behavior at the level of individual psychometric functions. Our model also provides a new interpretation of the traditional concept of "channels" (Campbell and Robson, 1968; Graham and Nachmias, 1971) by embedding it within a Bayesian estimation framework. Finally, the optimal Bayesian channel model provides a unifying explanation for the reported dependencies of perceived speed on stimulus contrast (Thompson, 1982; Stone and Thompson, 1992) as well as spatial frequency (Smith and Edgar, 1991; Priebe and Lisberger, 2004; Brooks et al., 2011). In our model, the influence of these attributes is reduced to their effect on the uncertainty of the channel signals. The uncertainty depends on the amount of sensory drive (contrast) and channel identity (different spatial frequencies target different channels with different sensitivities).

Implications for neural processing of visual speed

The results of our computational/behavioral study have some implications with regard to the underlying neural processing of visual speed. The fact that the optimal model well explained the data and clearly outperformed the averaging model suggests that the integration of the sensory information happens before prior expectations enter the inference process (Fig. 4). Given that most electrophysiological studies did not find any signs of truly speed-tuned neurons in the motion pathway earlier than area MT (and even there, their fraction within the whole population of MT neurons seems rather small; Priebe et al., 2003, 2006), this implies that the combination of the sensory information with the prior belief is likely to occur downstream of area MT. This might explain why previous studies have found it difficult to agree on how the response characteristics of MT neurons are linked to behavioral measures of perceived speed (Priebe and Lisberger, 2004; Krekelberg et al., 2006). However, this does not automatically imply that prior information is also represented downstream of area MT as has been proposed (Yang et al., 2012). Yet, it suggests that at least some read-out mechanism or mapping of the MT neural population is required to get a signal that is a direct representation of perceived stimulus speed. Some evidence indeed exists that a labeled-line readout of MT neural responses to broadband grating motion stimuli with different contrasts can reproduce the perceptually measured bias toward slow speeds with decreasing contrast (Stocker et al., 2009). Interestingly, some recent theoretical studies suggest that Bayesian inference can be well approximated by these type of decoders if the prior information is embedded in the tuning characteristics of the neural population being decoded (Wei and Stocker, 2012; Ganguli and Simoncelli, 2014). Thus prior information can be implicitly embedded in the population tuning characteristics yet only becomes effective during the read-out process of the population. Such implicit representation would also explain the results of a recent fMRI study that showed that the contrast dependence of perceived speed is already reflected in the BOLD signal of V1 when decoded appropriately (Vintch and Gardner, 2014).

Limits of optimal signal integration

While the literature reports many instances of optimal combination of sensory evidence from different sensory pathways (Ernst and Banks, 2002; Hillis et al., 2004; Landy et al., 2011), our results suggest that similar computations may also occur within a single pathway. However, there are limits to optimal integration. If the sensory information originates from different sources, then clearly the right strategy is not to integrate the information (Körding et al., 2007; Knill, 2007). This may explain some of the differences between our results and the results of a recent study by Simoncini and colleagues (Simoncini et al., 2012). In their study, Simoncini and colleagues measured how sensory integration across spatiotemporal frequency channels may differ with regard to visuomotor behavior compared with perception. In contrast to our results, their results showed no evidence of an increase in perceptual sensitivity for stimuli with broader spatiotemporal frequency spectra (i.e., channel integration). We believe that the difference in results is mainly because the noncoherent motion stimuli (motion clouds) used in their study may have led the visual system to segregate rather than integrate the sensory information. Other crucial stimulus parameters were also different, which could further explain the difference in findings. Among these differences were most notably the stimulus speed at which integration was tested (20 vs 3°/s), stimulus size (27 vs 4°), and stimulus location (foveal vs 6° eccentricity). However, our results do not rule out the possibility that there may be limits also to optimal integration for coherent motion stimuli. Channel interdependencies induced by, for example, suppressive mechanisms

(Cui et al., 2013), divisive normalization processes (Carandini and Heeger, 2012), or noise correlations (Huang and Lisberger, 2009; Ponce-Alvarez et al., 2013) may limit the amount of information that is conveyed by individual channels. A careful exploration of such potentially limiting mechanisms will require targeted experiments that must include more complex and possibly stronger stimuli targeting an even larger number of channels. In any case, the results of these experiments will allow us to further refine the presented observer model by incorporating additional details of the underlying neural processing into its Bayesian formalism.

Note that there is a more concrete explanation for the slight increase in threshold for the ABCD compared with the ABD stimulus seen for some of the subjects (Fig. 6*a*,*b*) than assuming channel interdependencies. Because stimulus ABCD simultaneously served as test and reference stimulus and thus was present in every trial of the unbalanced conditions, it was substantially over-represented in the total stimulus ensemble. This over-representation likely produced some form of habituation effect. For example, it might have induced perceptual adaptation that resulted in a mild sensitivity reduction for the reference stimulus, which would explain the deviations both in terms of threshold and matching speed (Ledgeway and Smith, 1997; Stocker and Simoncelli, 2009).

Last, the results of our study present an important step toward a better understanding of human visual speed perception. We believe that we have presented a general model framework that potentially allows us to account for the perceived speed of individual subjects for arbitrary motion stimuli.

Notes

Part of this work has been presented at the annual Vision Science Society meeting in 2011, the Computational and Systems Neuroscience meeting in 2012, and the annual meeting for Advances in Neural Information Processing Systems in 2013.

References

Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. J Opt Soc Am A 2:284–299. CrossRef Medline

Bayes Mr., Price Mr. (1763) An essay toward solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS. Philosophical Transactions (1683–1775), 370–418.

Brooks KR, Morris T, Thompson P (2011) Contrast and stimulus complexity moderate the relationship between spatial frequency and perceived speed: Implications for MT models of speed perception. J Vis 11(14):19. CrossRef Medline

Campbell FW, Robson JG (1968) Application of Fourier analysis to the visibility of gratings. J Physiol 197:551–566. CrossRef Medline

Carandini M, Heeger DJ (2012) Normalization as a canonical neural computation. Nat Rev Neurosci 13:51–62. CrossRef Medline

Chen Y, Bedell HE, Frishman LJ (1998) The precision of velocity discrimination across spatial frequency. Percept Psychophys 60:1329–1336. CrossRef Medline

Churchland MM, Lisberger SG (2001) Shifts in the population response in the middle temporal visual area parallel perceptual and motor illusions produced by apparent motion. J Neurosci 21:9387–9402. Medline

Cui Y, Liu LD, Khawaja FA, Pack CC, Butts DA (2013) Diverse suppressive influences in area MT and selectivity to complex motion features. J Neurosci 33:16715–16728. CrossRef Medline

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433. CrossRef Medline

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1–47. CrossRef Medline

Ganguli D, Simoncelli EP (2014) Efficient sensory encoding and Bayesian inference with heterogeneous neural populations. Neural Comput 26: 2103–2134. CrossRef Medline

Graham N, Nachmias J (1971) Detection of grating patterns containing two spatial frequencies: a comparison of single-channel and multiple-channel models. Vis Res 11:251–259. CrossRef Medline

Green D, Swets J (1966) Signal detection theory and psychophysics. New York: Wiley.

Hedges JH, Stocker AA, Simoncelli EP (2011) Optimal inference explains the perceptual coherence of visual motion stimuli. J Vis 11(6):1–16. CrossRef Medline

Hillis JM, Watt SJ, Landy MS, Banks MS (2004) Slant from texture and disparity cues: optimal cue combination. J Vis 4(12):967–992. Medline

Huang X, Lisberger SG (2009) Noise correlations in cortical area MT and their potential impact on trial-by-trial variation in the direction and speed of smooth-pursuit eye movements. J Neurophysiol 101:3012–3030. CrossRef Medline

Jogan M, Stocker AA (2011) Optimal signal integration across spatiotemporal frequency channels accounts for perceived visual speed. J Vis 11(11):699. CrossRef

Jogan M, Stocker AA (2013) Optimal integration of visual speed across different spatiotemporal frequency channels. In: Advances in neural information processing systems 26 (Burges CJC, Bottou L, Welling M, Ghahramani Z, Weinberger KQ, eds), pp 3201–3209. Lake Tahoe, NV: Annual Conference on Neural Information Processing Systems.

Jones M, Love BC (2011) Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. Behav Brain Sci 34:169–188; discussion 188–231. CrossRef Medline

Knill DC (2007) Robust cue integration: a Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. J Vis 7(7):5. Medline

Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L (2007) Causal inference in multisensory perception. PLoS One 2:e943. CrossRef Medline

Krekelberg B, van Wezel RJ, Albright TD (2006) Interactions between speed and contrast tuning in the middle temporal area: implications for the neural code for speed. J Neurosci 26:8988–8998. CrossRef Medline

Landy MS, Maloney LT, Johnston EB, Young M (1995) Measurement and modeling of depth cue combination: in defense of weak fusion. Vis Res 35:389–412. CrossRef Medline

Landy MS, Banks MS, Knill DC (2011) Ideal-observer models of cue integration. In: Sensory cue integration (Trommershauser J, Körding K, Landy MS, eds), pp 5–29. New York: Oxford UP.

Ledgeway T, Smith AT (1997) Changes in perceived speed following adaptation to first-order and second-order motion. Vis Res 37:215–224. CrossRef Medline

Liu J, Newsome WT (2005) Correlation between speed perception and neural activity in the middle temporal visual area. J Neurosci 25:711–722. CrossRef Medline

Movshon J, Adelson E, Gizzi M, Newsome W (1985) The analysis of moving visual patterns. Exp Brain Res 11:117–151. CrossRef

Nover H, Anderson CH, DeAngelis GC (2005) A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. J Neurosci 25:10049–10060. CrossRef Medline

Perrone JA, Thiele A (2002) A model of speed tuning in MT neurons. Vis Res 42:1035–1051. CrossRef Medline

Ponce-Alvarez A, Thiele A, Albright TD, Stoner GR, Deco G (2013) Stimulus-dependent variability and noise correlations in cortical MT neurons. Proc Natl Acad Sci U S A 110:13162–13167. CrossRef Medline

Priebe NJ, Lisberger SG (2004) Estimating target speed from the population response in visual area MT. J Neurosci 24:1907–1916. CrossRef Medline

Priebe NJ, Cassanello CR, Lisberger SG (2003) The neural representation of speed in macaque area MT/V5. J Neurosci 23:5650–5661. Medline

Priebe NJ, Lisberger SG, Movshon JA (2006) Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. J Neurosci 26:2941–2950. CrossRef Medline

Rust NC, Mante V, Simoncelli EP, Movshon JA (2006) How MT cells analyze the motion of visual patterns. Nat Neurosci 9:1421–1431. CrossRef Medline

Simoncelli EP, Heeger DJ (1998) A model of neuronal responses in visual area MT. Vis Res 38:743–761. CrossRef Medline

Simoncini C, Perrinet LU, Montagnini A, Mamassian P, Masson GS (2012) More is not always better: adaptive gain control explains dissociation between perception and action. Nat Neurosci 15:1596–1603. CrossRef Medline

Smith AT, Edgar GK (1991) Perceived speed and direction of complex gratings and plaids. J Opt Soc Am A 8:1161–1171. CrossRef Medline

Solomon SS, Tailby C, Gharaei S, Camp AJ, Bourne JA, Solomon SG (2011) Visual motion integration by neurons in the middle temporal area of a New World monkey, the marmoset. J Physiol 589:5741–5758. CrossRef Medline

Sotiropoulos G, Seitz AR, Seriès P (2014) Contrast dependency and prior expectations in human speed perception. Vis Res 97:16–23. CrossRef Medline

Stocker AA, Simoncelli EP (2006) Noise characteristics and prior expectations in human visual speed perception. Nat Neurosci 9:578–585. Medline

Stocker AA, Simoncelli EP (2009) Visual motion aftereffects arise from a cascade of two isomorphic adaptation mechanisms. J Vis 9(9):9:1–14. CrossRef Medline

Stocker A, Majaj N, Tailby C, Movshon J, and Simoncelli E (2009) Decoding velocity from population responses in area MT of the macaque. Paper presented at Vision Sciences Society Meeting, Naples, FL, May.

Stone LS, Thompson P (1992) Human speed perception is contrast dependent. Vis Res 32:1535–1549. CrossRef Medline

Thompson P (1982) Perceived rate of movement depends on contrast. Vis Res 22:377–380. CrossRef Medline

Vintch B, Gardner JL (2014) Cortical correlates of human motion perception biases. J Neurosci 34:2592–2604. CrossRef Medline

Weiss Y, Simoncelli EP, Adelson EH (2002) Motion illusions as optimal percepts. Nat Neurosci 5:598–604. CrossRef Medline

Wei XX, Stocker A (2012) Bayesian inference with efficient neural population codes. In: Lecture notes in computer science, artificial neural networks and machine learning, volume 7552, pp 523–530. Lausanne, Switzerland: ICANN 2012.

Yang J, Lee J, Lisberger SG (2012) The interaction of Bayesian priors and sensory data and its neural circuit implementation in visually guided movement. J Neurosci 32:17632–17645. CrossRef Medline

Yuille A, Bülthoff H (1996) Bayesian decision theory and psychophysics. In: Perception as Bayesian inference (Knill D, Richards W, eds), pp 123–161. New York: Cambridge UP.

Zeki SM (1974) Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. J Physiol 236:549–573. CrossRef Medline