

Without low spatial frequencies, high resolution vision would be detrimental to motion perception

Cong Shi

School of Microelectronics and Communication Engineering, Chongqing University, Chongqing, China
Schepens Eye Research Institute of Mass Eye and Ear, Department of Ophthalmology, Harvard Medical School, Boston, MA, USA



Shrinivas Pundlik

Schepens Eye Research Institute of Mass Eye and Ear, Department of Ophthalmology, Harvard Medical School, Boston, MA, USA



Gang Luo

Schepens Eye Research Institute of Mass Eye and Ear, Department of Ophthalmology, Harvard Medical School, Boston, MA, USA



A normally sighted person can see a grating of 30 cycles per degree or higher, but spatial frequencies needed for motion perception are much lower than that. It is unknown for natural images with a wide spectrum how all the visible spatial frequencies contribute to motion speed perception. In this work, we studied the effect of spatial frequency content on motion speed estimation for sequences of natural and stochastic pixel images by simulating different visual conditions, including normal vision, low vision (low-pass filtering), and complementary vision (high-pass filtering at the same cutoff frequencies of the corresponding low-vision conditions) conditions. Speed was computed using a biological motion energy-based computational model. In natural sequences, there was no difference in speed estimation error between normal vision and low vision conditions, but it was significantly higher for complementary vision conditions (containing only high-frequency components) at higher speeds. In stochastic sequences that had a flat frequency distribution, the error in normal vision condition was significantly larger compared with low vision conditions at high speeds. On the contrary, such a detrimental effect on speed estimation accuracy was not found for low spatial frequencies. The simulation results were consistent with the motion direction detection task performed by human observers viewing stochastic sequences. Together, these results (i) reiterate the importance of low frequencies in motion perception, and (ii) indicate that high frequencies may be detrimental for speed estimation when low frequency content is weak or not present.

Introduction

Visual motion perception plays an important role in a wide range of tasks in an organism's life, such as navigation and detection of obstacles. Motion perception in humans (or in primates in general) is not just restricted to the detection or registration of moving patterns, but also includes the ability to estimate speed. Primate visual cortex is known to house circuitry necessary for speed perception. Neurons in the MT region, as well as some complex V1 neurons in primates are specially tuned for speed perception (Perrone & Thiele, 2001; Perrone & Thiele, 2002; Priebe, Cassanello, & Lisberger, 2003). A particularly intriguing aspect of these speed tuned MT neurons is that they are known to be somewhat invariant to the spatial frequency of the stimuli, which is especially important given the wide band of spatiotemporal frequencies received by our visual system (Simoncelli & Heeger, 2001). However, despite this invariance of the speed tuned MT neurons to changes in the spatial frequency spectrum of the stimuli, motion perception is not the same across the entire visible spatial frequency spectrum.

Our ultimate goal is to explore motion perception in people with low vision in various situations. One interesting aspect of this problem is to study motion perception in different situations, specifically looking at the interaction of motion speed and spatial frequency content of the scene. We implemented a computational model that is broadly based on some of the existing

Citation: Shi, C., Pundlik, S., & Luo, G. (2020). Without low spatial frequencies, high resolution vision would be detrimental to motion perception. *Journal of Vision*, 20(8):29, 1–13, <https://doi.org/10.1167/jov.20.8.29>.



theories of motion perception in primate visual cortex. Our aims in this work are to simulate different visual conditions stratified by spatial frequency bands and examine motion perception at different motion speeds in different kinds of stimuli. The simulation findings are then compared with human subject data that we report in support of the computational model along with previous studies regarding motion perception in humans.

Previous studies have explored the role of spatial frequencies in motion perception via motion psychophysics (Gilden, Bertenthal, & Othman, 1990; Hess & Aen-Stockdale, 2008; Ramachandran, Ginsburg, & Anstis, 1983; Shioiri, Ito, Sakurai, & Yaguchi, 2002; Wichmann & Henning, 1998; Yang & Stevenson, 1997), and a general conclusion emerging from these studies is that low spatial frequencies play a key role in motion perception. Studies measuring human performance at different visual acuities (VAs) provide further evidence of the same conclusion. Normally sighted humans were able to understand the scene content in the presence of motion information even when their vision was blurred (i.e. they lacked high spatial frequency visual cues; Pan & Bingham, 2013; Saunders, Bex, Rose, & Woods, 2014). Low vision only degraded the capability to perceive very slow motions ($< 2^\circ/\text{s}$), and there was no significant difference in perception of faster motion in individuals with small to moderate VA loss compared with normally sighted subjects (Lappin, Tadin, Nyquist, & Corn, 2009). Thus, based on the previous studies, it is clear that low spatial frequencies are critical for motion perception, which may not be affected by the loss of high-resolution vision.

However, the range of speeds and spatial frequencies used in these studies was limited, excluding extremities of frequency and speed values. In addition, the testing generally involved a single spatial frequency presented in terms of an artificial stimulus at a given moment, or using spatially band-pass filtered random dot kinematogram stimuli (Cleary & Braddick, 1990). A range of spatial frequencies are commonly present in complex real-world stimuli, and the interaction between the various spatial frequency bands and motion perception performance is not completely clear for such natural images. As natural images would be more relevant to our visual experience than random dot stimuli commonly used in laboratory experiments, we intend to explore some of the pertinent questions concerning the relationship between speed and the presence and/or absence of various spatial frequency bands in natural visual stimuli.

In this study, we used an implementation of an algorithm based on the well-known biological motion energy model (Adelson & Bergen, 1985; Etienne-Cummings, Van der Spiegel, & Mueller, 1999; Grzywacz

& Yuille, 1990; Ogata & Sato, 1991), to quantitatively derive motion speed across different spatiotemporal frequency bands in natural and stochastic (binary noise patterns) visual stimuli. Such a model used multiple channels of spatial and temporal narrow-band filters (Bex & Dakin, 2002) to capture the motion information (for reviews about earlier motion energy models, see [Burr & Thompson, 2011; Nishida, 2011]). Using the computational model, we quantitatively analyzed the interaction of different spatial frequency channels in visual motion speed estimation by simulating different VA levels, such as normal vision (20/20), low vision (20/50 and 20/200), and complementary vision conditions containing only high spatial frequency components. Although the complementary vision conditions do not exist in the context of daily visual experience, they allowed us to explore independently the effect of spatial frequencies on speed perception. Furthermore, given that spatial frequency distribution differs greatly between natural and stochastic visual stimuli, independently looking at high and low spatial frequency bands in these two different kinds of stimuli allow us the opportunity to examine causal relationship between spatial frequency and speed estimation. We replicated earlier findings about the importance of low spatial frequencies in speed perception, while showing that, in certain situations, high spatial frequencies can actually be detrimental for speed estimation.

Methods

An overview of our methods for this work is shown in Figure 1. First, we generated motion sequences by translating different natural and stochastic images with known speed (Figure 2). Next, we filtered these sequences to simulate effects of different visual conditions: low vision or loss of VAs at different levels using low pass filters with different cutoff frequencies and complementary vision conditions (hypothetical) using high pass filters (Figure 3). Spatiotemporally white noise was added to the sequences before the vision condition filtering to simulate external (physical world) noise. Finally, we applied Shi & Luo's (Shi & Luo, 2018) implementation of Grzywacz and Yuille's motion energy model (see Figure 1) to estimate the speed of motion in these sequences from their spatiotemporal frequency components called motion energy. We examined the relationship between spatial frequencies and speed estimation accuracy of the computational model under different simulated vision conditions and at different speeds. We further tested motion direction discrimination ability of human observers for different speeds.

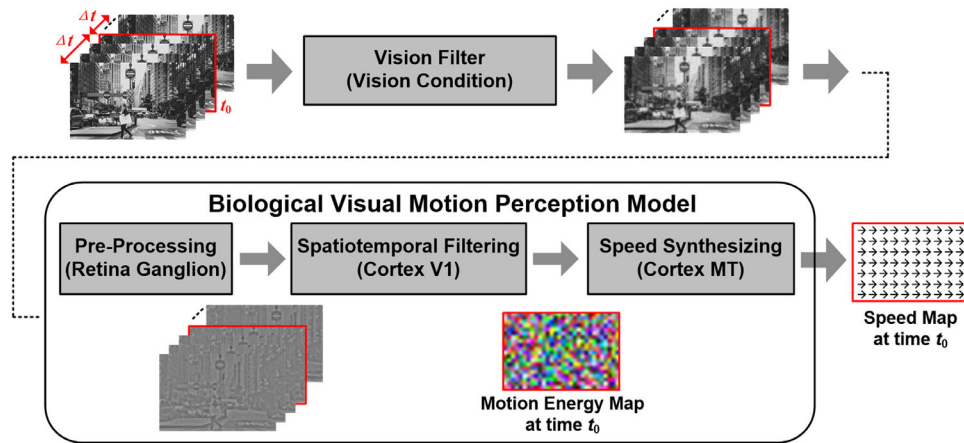


Figure 1. An overview of our method for motion speed estimation. Motion sequences were generated by translating images with different speeds. We filtered these sequences to simulate effects of different visual conditions and then we applied the biological motion perception model to these sequences. Motion estimation at time t_0 requires the frames within a time window of $2\Delta t$ centered at t_0 , where $2\Delta t$ is the length of temporal filters used in the biological model. The preprocessing stage filtered out very high frequency noise and spatial DC components. Then the spatiotemporal filtering stage sampled multiple spatiotemporal frequency components to generate motion energy maps as the spatiotemporal filtering results. Finally, the speed synthesizing stage used the motion energy maps to infer the real motion speed, which was then compared with the known value to obtain speed estimation error at different speed values as well as for different simulated vision conditions.

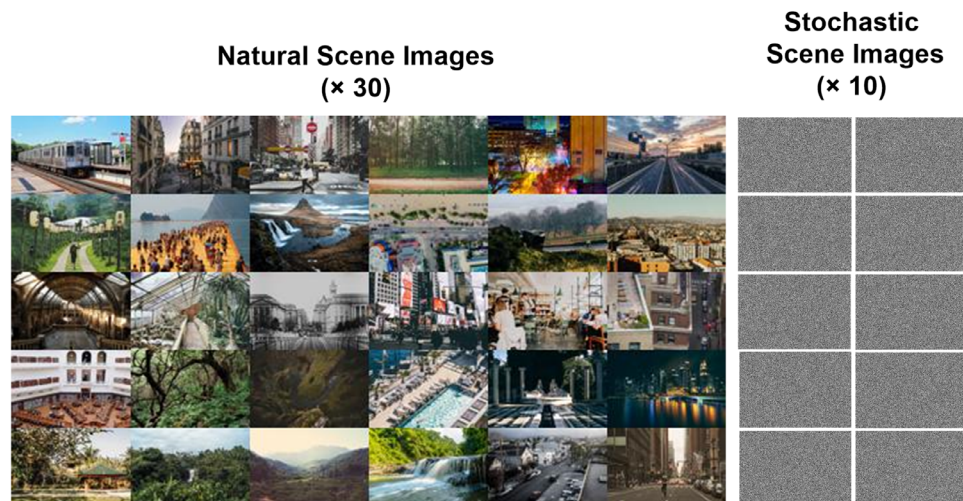


Figure 2. The high-resolution daily life natural images ($n = 30$) downloaded from the internet and artificially generated binary stochastic images ($n = 10$) were used for generating the motion sequences.

Generation of motion sequences

To generate motion sequences of natural scenes, we used the Google image search engine with keywords like natural scenes, urban scenes, rural scenes, street blocks, buildings, railways, beaches, and so on, and randomly picked 30 high resolution images out of the search results (see Figure 2). These natural images were then down-sampled (with anti-aliasing processing) to 900×600 grayscale pixels to ensure

that they contained sufficient high spatial frequency components for simulating the 20/20 normal vision. To simulate the inherently continuous physical world, we set a high view-field resolution of $120 \text{ pixel}/^\circ$ on these images so that each such 900×600 image corresponded to a field-of-view of $900/120 \times 600/120 = 7.5^\circ \times 5^\circ$ in the physical world. To compare with low spatial frequency dominated natural images, we also generated 10 additional binary stochastic images of the same resolution in which each pixel was randomly

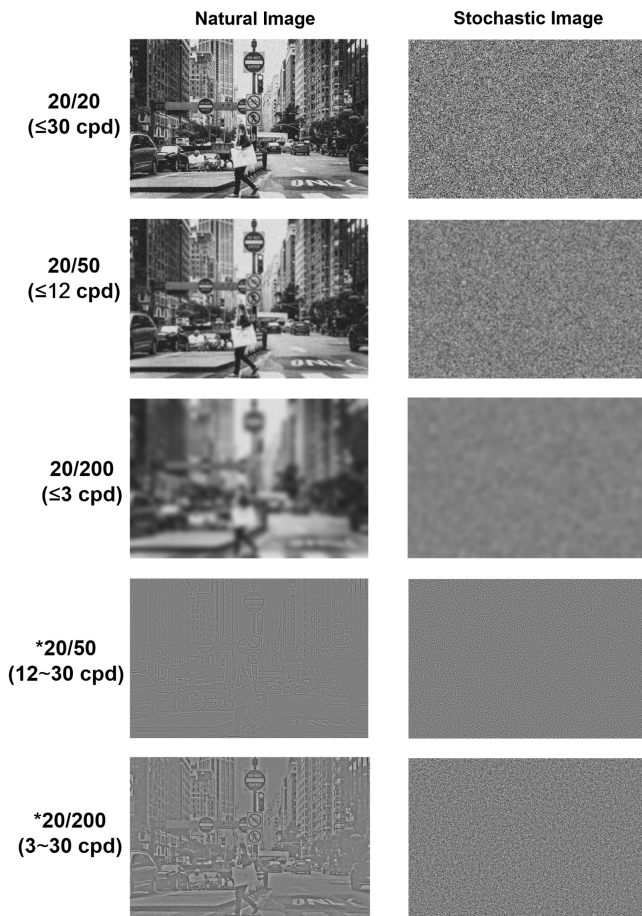


Figure 3. Simulated vision conditions. A frame from a natural sequence (left column) and a stochastic sequence (right column) filtered to simulate the five different vision conditions (each row). The visual acuity levels and the corresponding cutoff frequency is indicated for each row (* indicates complementary vision conditions). The low vision sequences only contained lower spatial frequencies (blurring effect), whereas the complementary conditions, which do not exist in real world, contained only high spatial frequency components (preserving edges).

assigned as either 0 (dark) or 1 (white) with equal probabilities. The spatial frequency spectrums of these stochastic images were nearly flat from low to high frequencies.

We generated motion sequences from the above 40 images by horizontally shifting them in a cyclic manner (the part shifted out on one side was to be shifted in on the other side). For each image, we generated 5 motion sequences with different speeds: 0.1 °/s, 1 °/s, 5 °/s, 15 °/s, and 30 °/s. The time duration of each sequence was set to $2\Delta t = 0.2s$, exactly covering the length of temporal filters used in the biological motion perception model, as will be introduced later. Therefore, the estimated speed for a sequence was actually for the central time instant $t_0 = 0.1$ seconds (see Figure 1).

Because the speed did not change over time and the image contents did not change despite the shift, the estimated motion at any frame was representative of the motion for the entire sequence. Thus, the need for using longer motion duration time for motion estimation was obviated.

Simulation of visual conditions

By filtering the motion sequences in different manner, we simulated 5 visual conditions: VA of 20/20 (normal vision), VA of 20/50 (moderate vision loss), VA of 20/200 (severe vision loss), a complementary VA of 20/50, and a complementary VA of 20/200. The underlying assumption in simulating these visual conditions was that low vision conditions correspond to loss of high spatial frequency components in the scene. On the other hand, the complementary vision conditions simulated a situation where only high spatial frequency components were present. Although this is not realistic, it helps us to investigate the causal relationship between spatial frequency and motion speed estimation. In terms of spatial frequency, a VA of 20/20 (normal vision) simulated the perceivable spatial frequencies up to 30 cycle/° (cpd), whereas the 20/50 and 20/200 low vision conditions simulated perceivable frequencies no higher than 12 and 3 cpd, respectively. The normal and low vision conditions were realized by applying low-pass spatial filters (called vision filters in this paper) with cutoff frequencies of 30, 12, and 3 cpd, respectively, to every frame of the original motion sequences. The complementary vision conditions that simulated only high spatial frequency components were realized by subtracting their corresponding low vision sequences from the normal vision sequence frame by frame. For example, the complementary VA of 20/50 simulated a condition where frequencies within between 12 and 30 cpd were present. These imaginary complementary vision conditions that simulate the presence of only high spatial frequencies do not exist in the real world. They were used as hypothetical references for studying the effects of low and high spatial frequencies on motion perception. The frames were filtered first and then displaced to generate the motion sequences corresponding to different visual conditions.

Moreover, to simulate external noise in the physical world, we added spatiotemporally white Gaussian noises with a standard deviation of 10% of the dynamic range of image brightness to all sequences before the vision condition filtering. Figure 3 shows the first frames of one natural motion sequence and one stochastic motion sequence after being filtered to simulate the 5 visual conditions, all with the simulated 10% external noise.

Biological visual motion perception model

More detailed description regarding the computational model is provided in the Supplementary Material. We implemented the widely accepted computational motion perception model (Adelson & Bergen, 1985; Grzywacz & Yuille, 1990) with the following modifications (see Figure 1). (1) The 2D spatial filters were decomposed to faster 2-stage cascaded 1D filtering (Etienne-Cummings, Van der Spiegel, & Mueller, 1999; Shi & Luo, 2018). (2) In the pre-processing stage, we used a DoG filter to embrace a wider spatial frequency band from 0.5 to 36 cpd to facilitate successive processing. (3) In the spatiotemporal filtering stage to extract motion energy, the sampling density was higher: $0.6 \text{ cpd} \times \{1, 2, \dots, 50\}$ for spatial frequencies, and $5 \text{ Hz} \times \{0, \pm 1, \pm 2, \dots, \pm 10\}$ for temporal frequencies. (4) In the speed synthesizing stage, the speed candidates were sampled from -40 to 40 %/s, with a step of 0.01 %/s from -0.2 to 0.2 %/s, and a step of 0.1 %/s elsewhere. In addition, because all spatial locations of a motion sequence moved with the same speed, the speed estimates of the model were sampled at spatial locations at an interval of $1/6^\circ$ along both dimensions (excluding locations near the boundary lying within the radius of the spatial filters). The average estimated speed for these sampled locations was used as the perceived speed for the entire sequence.

Evaluation of the simulation and statistical analysis

We presented the results of our simulation in terms of motion energy distributions under different visual conditions for natural and stochastic motion sequences. For a given speed and vision condition, the distributions were obtained by accumulating the motion energy for each sampled spatial frequency (with the temporal frequencies integrated out) for all the probing spatial image locations and then averaged over all the 30 natural sequences. Speed estimation errors were calculated with respect to the ground truth. We performed repeated measures analysis of variance (ANOVA) to determine the within-subject effects of different speeds and visual conditions on speed estimation error for a given type of image sequence, and between subject effects of sequence type (each image sequence can be considered to be a subject). The relative speed estimation errors were inverted for natural sequences to ensure normality of the data; the data for stochastic sequences were normal. Normality was tested using Shapiro-Wilks test and data for three natural images and one stochastic image were considered outliers (outside 99% confidence interval). Thus, the statistical analysis was done using

27 natural and 9 stochastic sequences. Wherever required, the ANOVA results were corrected for sphericity and the corrected results are reported. The nonparametric testing method was used for comparing the speed estimation error for natural and stochastic sequences (2 sample Kolmogorov-Smirnov test). Statistical analysis was performed using IBM-SPSS.

Human subject evaluation

We wanted to verify the findings of the computational model in humans by testing the hypothesis that low spatial frequency components are critical for motion perception. For this purpose, we generated stochastic sequences from binary images in which each pixel was randomly assigned as either 0 (dark) or 1 (white) with equal probabilities (same as described previously for the simulation). The motion sequence was generated by shifting the image frame horizontally in a cyclic manner. Stochastic stimuli moving at two different speeds were used for testing: 26 %/s and 50 %/s, representing relatively slower and faster speed values. The stimuli were presented in a circular patch (13° size) surrounded by black background for a duration of 1.5 seconds for a given trial on a computer screen. The direction of motion in the stimulus for a given trial was randomly either from left to right or vice versa. The human observer viewed the stimuli monocularly from a distance of 33 cm and responded by pressing arrow keys (left or right arrow) corresponding to the perceived direction of the motion stimulus. The observer either viewed the screen directly (normal vision condition [NV]) or while wearing a $+7\text{D}$ blur lens in front of the test eye (low vision [LV]). The experiment was carried out in 4 conditions: 2 different stimuli speeds 26 and 50 %/s \times 2 visual conditions (NV and LV). There were 32 trials per condition, 16 in each direction, with the direction of motion changing randomly between successive trials.

A total of 8 normally sighted observers, with near VA of 20/20 or better, participated in the study. The study followed the tenets of the Declaration of Helsinki and informed consent was obtained from all the study participants. The protocol was approved by the institutional review board at the Massachusetts Eye and Ear. When wearing blur lens (simulating the LV condition), the monocular VA of the participants dropped to 20/200 or worse. The order of the experimental conditions was balanced across subjects for vision and speed. Motion direction detection accuracy in each condition was obtained from the subject responses. The motion direction detection accuracy was compared between speed and visual conditions via pairwise testing (using nonparametric Wilcoxon signed rank test), as well as using a

mixed effect model (vision, speed, and vision-speed interaction as fixed effects and subject wise random effects).

Task equivalence between model and human judgment

Because speed estimation error was the performance metric for the computational model, whereas motion direction discrimination was the task in our human subject evaluation, the question of direct comparison between the model and behavior needs to be addressed. The model computes motion energy, with the speed estimation as its final step. Therefore, how accurately the model predicts speed is based on the “perception of motion” by the model. To show a direct link with the human judgment data we obtained in our experiments, we used the intermediate results of the speed computation to model the motion direction judgment probabilistically.

The procedure for making a motion direction judgment based on the computational model is as follows. An intermediate result of the computational model is the speed value computed in each of the 735 local blocks in the images. In addition to speed, each block also outputs the direction of motion. However, not all blocks estimate the same direction of motion and the same speed value. Naively averaging the direction of all the blocks to obtain the final motion direction would not represent the correct overall motion direction judgment, especially when the variability (error) across the image is large. Therefore, we average directions for the blocks that estimate speed within 75% to 125% of the mean estimated speed for the entire sequence. The decisions of the rest of the blocks are randomized 50% in either direction. Thus, we arrive at a motion direction judgment and its probability for a given sequence. The overall probability of correct direction estimation is arrived at by averaging the results for 10 sequences of the stochastic pixel images described above for 2 motion speeds (5 °/s and 30 °/s) for two visual conditions (NV: 20/20 and LV: 20/200).

Results

The raw motion energy distributions over spatial frequencies in natural sequences for different speeds in normal vision condition are highly skewed toward low spatial frequencies (Figure 4a). The normalized cumulative motion energy distributions of Figure 4a have about 90% of the total motion energy concentrated below 12 cpd frequencies for 0.1 °/s speed curve and this amount increases for other speed curves, going up to

99.9% for 30 °/s speed. Due to the higher concentration of motion energy at lower spatial frequencies in natural sequences, there is also a relatively large overlap between the motion energy distributions for NV and the LV conditions (85% for VA 20/50 and 45% for 20/200), but relatively smaller overlap for the complementary vision conditions (38% for VA *20/200 and 3% for *20/50; see Figure 4b). A larger amount of overlap indicates more similarity in the motion energy distributions. Because the cutoff frequencies for vision condition simulating filters are 3 and 12 cpd (the LV conditions are low pass cases, whereas the complementary conditions are high pass cases around these cutoff frequencies), larger overlap of NV and LV conditions is the expected outcome.

Compared with natural sequences, the motion energy distributions for stochastic images in NV condition are no longer concentrated near low spatial frequency bands but instead appear flat across the entire frequency spectrum (see Figure 4c). The motion energies are also of significantly smaller magnitude (reduced by about 89%). If we compute the normalized cumulative distribution for 0.1 °/s speed curve, then 40% of the motion energy is concentrated below 12 cpd in stochastic sequences compared with about 90% for natural sequences for the same spatial frequency band. Furthermore, contrary to the natural sequences, there is a relatively larger overlap between NV and complementary vision conditions (86% for *20/200 and 12% for *20/50) than the LV condition (58% for 20/50 and 6% for 20/200; see Figure 4d). Again, this is expected because there is a relatively small amount of motion energy present in the low spatial frequency band in the stochastic sequences to begin with.

There is a discernable effect of speed on the motion energy distributions, as higher speeds lead to a higher concentration of motion energy in the lower spatial frequency region in both natural and stochastic sequences (see Figure 4e). Because 3 cpd was the lowest cutoff frequency for simulation of LV conditions, the fraction of motion energy at or below 3 cpd was used as a way to quantify the effect of speed on motion energy distributions. Predictably for the 20/200 vision condition with 3 cpd cutoff frequency, the motion energy fraction at 3 cpd is already at 99% at 0.1 °/s. In natural sequences for NV and 20/50 condition, the fraction of motion energy \leq 3 cpd at 0.1 °/s speed is at 51% and 68%, respectively. As the speed increases, this amount increases close to about 100% at 30 °/s. The same effect is also seen in stochastic sequences. However, the motion energy fraction \leq 3 cpd for 0.1 °/s is a lot lower in stochastic sequences compared with natural sequences (5% and 18% for 20/20 and 20/50 conditions for stochastic, whereas 51% and 68% for the same in natural sequences, respectively), before increasing to close to 100% at 30 °/s. There is also a noticeable interaction of visual conditions and speed

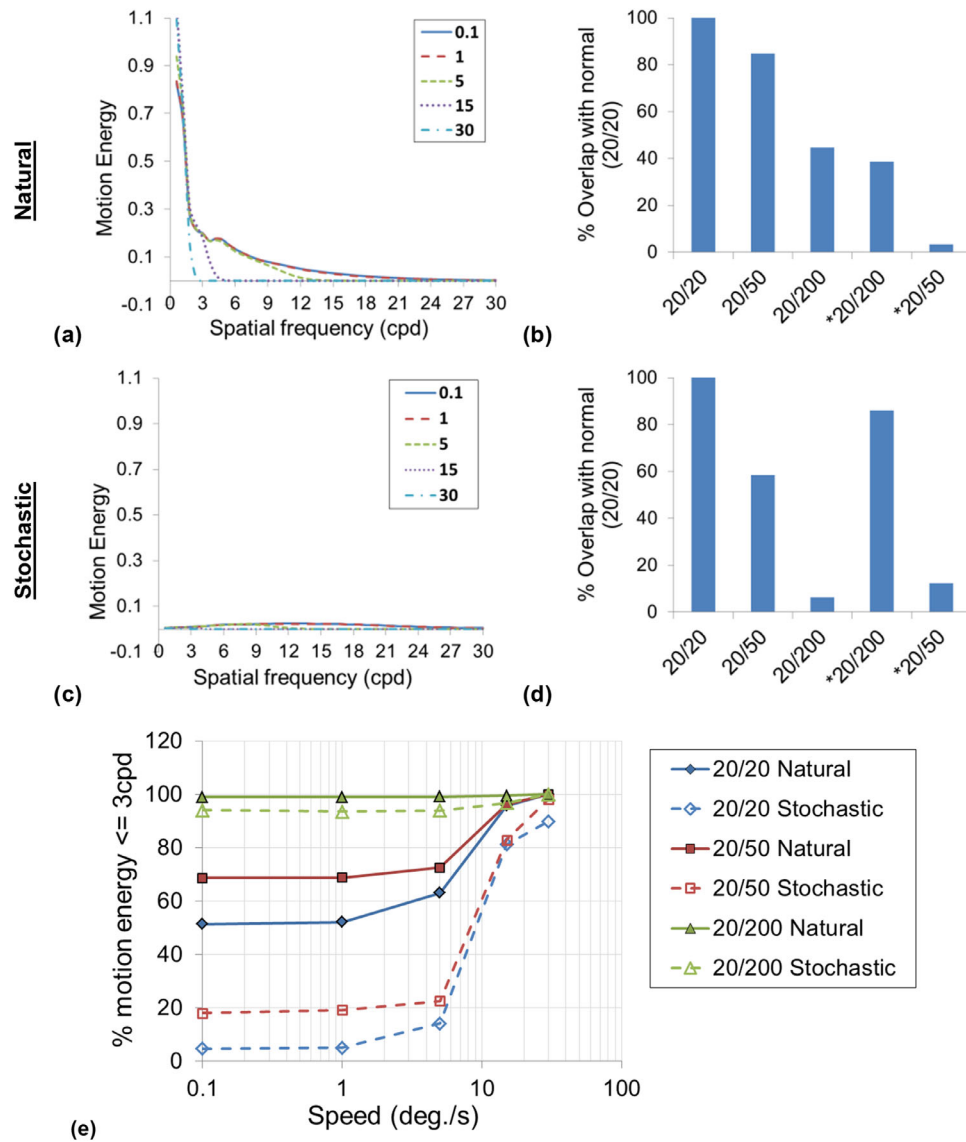


Figure 4. Motion energy distribution over spatial frequency in normal vision condition for different speeds in natural (a) and stochastic (c) sequences. Percent overlap of motion energy distributions for various visual conditions (low-vision and complementary vision conditions) with normal vision (20/20) in natural (b) and stochastic (d) sequences. A larger overlap means higher degree of similarity between the distributions. Median speed (5 °/s) curve was used for computing the overlap. Effect of speed on the motion energy distributions for normal and low vision conditions in natural and stochastic sequences (e). The fraction of motion energy below 3 cpd frequency indicates its concentration in low spatial frequency region, which increases with increasing speed.

on the motion energy distribution within natural and stochastic sequences: lower visual acuity leads to a less steep increase in motion energy fraction ≤ 3 cpd for higher speeds. This is again expected, because more motion energy is concentrated in lower spatial frequency regions for LV conditions to begin with.

Speed estimation error did not change significantly in normal vision and the two LV conditions in natural sequences ($F(1, 2.745) = 2.745$, $p = 0.099$; Figure 5a). There was a significant effect of speed as the error increased with speed in all visual conditions

($F(1, 1.161) = 34.92$, $p < 0.001$). The interaction between speed and visual conditions was significant ($F(1, 1.308) = 4.39$, $p = 0.034$). In contrast to natural sequences, speed estimation error was significantly larger in 20/20 case compared with other LV conditions in stochastic sequences ($F(1, 1.403) = 849.2$, $p < 0.001$ see Figure 5b). The error also significantly increased with speed in stochastic sequences ($F(1, 1.701) = 1029.34$, $p < 0.001$). The interaction of speed with the visual conditions was significant ($F(1, 2.091) = 428.37$, $p < 0.001$), as the error in the

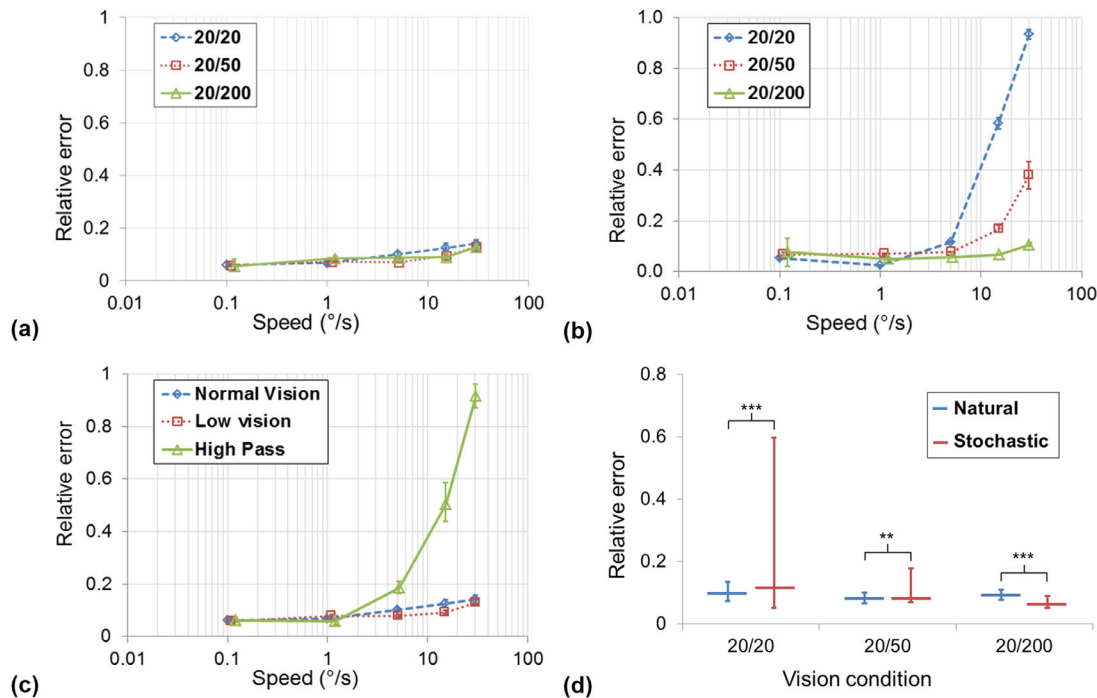


Figure 5. Speed estimation error results. Comparison of relative error (ratio of absolute error value and the ground truth speed) in normal vision (20/20) and low vision conditions (20/50 and 20/200) over the tested speed range is shown for natural (a) and stochastic (b) sequences. Comparison of relative speed estimation error among normal vision, low vision, and complementary vision conditions (high pass) in natural sequences are shown in (c). Low vision condition in this case is the mean of 20/50 and 20/200 conditions, whereas the high pass condition is the mean of *20/50 and *20/200 conditions. In plots a through c, error bars denote standard error of mean. In addition, in these plots, please note the logarithmic scale for the horizontal axis. Comparison of the relative error for natural and stochastic sequences is shown in (d). The pooled relative error distributions across all speeds for 20/20, 20/50, and 20/200 conditions are shown, with horizontal lines representing the median and the error bars showing the 25th and 75th percentile (***: $p < 0.001$, **: $0.01 < p \leq 0.001$).

20/20 condition was significantly higher than 20/50 and 20/200 vision conditions at higher speeds. Speed estimation was impaired in complementary vision conditions as the error was significantly larger in complementary conditions compared with the NV or LV conditions ($F(1, 1.301) = 26.56, p < 0.001$; see Figure 5c). The error increased significantly with speed in the complementary vision conditions ($F(1, 2.032) = 106.13, p < 0.001$) and the interaction between speed and visual conditions was significant ($F(1, 1.889) = 21.39, p < 0.001$) as the error between complementary and LV conditions was larger at higher speeds. When comparing the two kinds of sequences in normal and LV conditions, the speeds estimation error was significantly different in natural and stochastic sequences ($F = 9.23, p = 0.005$; see Figure 5d). There was a significant differences in error distribution (pooled across all speeds) in all three vision conditions between natural and stochastic sequences (20/20: $Z = 2.32, p < 0.001$; 20/50: $Z = 1.94, p = 0.001$; and 20/200: $Z = 3.01, p < 0.001$), even as the medians did not differ. This was due to the highly skewed nature of

the error distribution, which necessitated use of the nonparametric statistical approach in this particular case.

The data collected from the human observers for the motion direction detection task involving stochastic stimuli (Figure 6a) showed a significant effect of speed: overall, the correct response rate was lower for 50 %/s speed compared with 26 %/s in both conditions ($t = -8.03, df = 21, p < 0.001$). Importantly, a significant interaction was seen between speed and visual conditions, with the correct response rate improving from an average of 68% to 92% from NV to LV condition for 50 %/s speed ($t = 4.3, df = 21, p < 0.001$). For 26 %/s speed, all subjects detected the motion direction with 100% accuracy in both vision conditions. Similar results were seen for an equivalent direction discrimination task simulated with the computational model (Figure 6b). The probability of correct direction estimation was lower (at 62%) for NV condition at the higher speed of 30 %/s than the lower speed of 5 %/s, but was the same (100%) for both speeds in the LV condition.

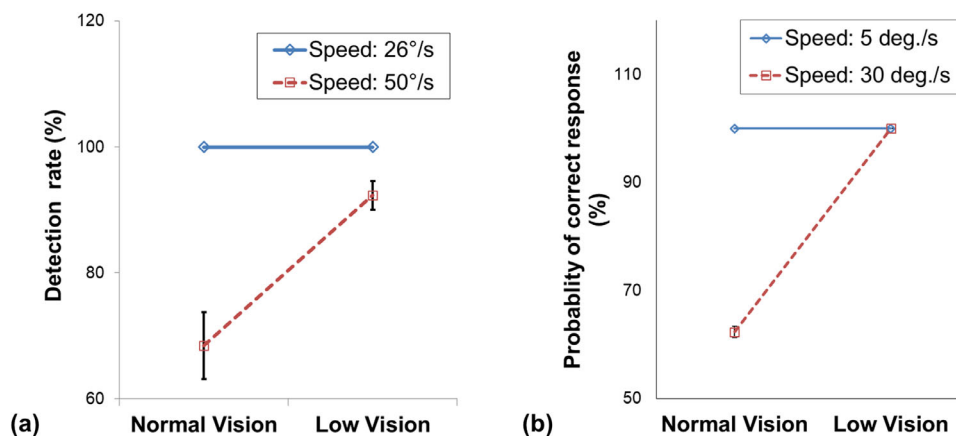


Figure 6. Direction discrimination task in stochastic sequences. **(a)** Correct response rate for human subjects ($n = 8$) are plotted for two speeds and viewing conditions (NV: normal vision, LV: low vision induced with +7D blur lens). Error bars show standard error of mean. A significant interaction between stimulus speed and vision condition can be seen. For the 26 °/s stimulus speed, the detection rate was 100% in both the vision conditions. On the other hand, for the 50 °/s speed the detection rate increased significantly from the NV to LV condition (avg. \pm sem.: NV = $68.4 \pm 5.3\%$, LV = $92.3 \pm 2.24\%$, $p < 0.001$). **(b)** Analogous motion direction discrimination task simulation for two speeds (5 °/s and 30 °/s) and two simulated visual conditions. The probability of direction discrimination is estimated based on the computed speed and direction at multiple locations in the sequences. The probability of correct direction response is averaged over 10 stochastic sequences.

Discussion

We used an analytical model based on motion energy along with data from human observers to examine the relationship between motion perception and spatial frequency. By using a computational model, we were able to address two issues that have not been addressed in previous psychophysics studies measuring motion perception thresholds: (i) the model performed direct speed estimation, providing a more objective and quantitative way of determining speed perception sensitivity, and (ii) by simulating different visual conditions, including complementary vision conditions with only high spatial frequencies that do not exist in natural world, we were able to independently explore the causal relationships between speed estimation accuracy and different spatial frequency bands in natural images. We further verified some of the main findings from the computational model simulation with human subject experimentation. Overall, our results show the dominant role played by low spatial frequency components in speed perception that largely agree with a wide variety of previous research in motion psychophysics and primate neurobiology. At the same time, we demonstrate the detrimental effects of high spatial frequencies in speed estimation in certain special cases.

The speed estimation results (see Figure 5) in natural images showed that the error was the same in NV and LV conditions, but was significantly larger in complementary vision conditions. Because

LV conditions predominantly contained low spatial frequency components of the scene compared with the NV condition, and the complementary conditions did not contain low frequencies, this suggests that low frequencies are critical in speed estimation. Because large errors will occur (especially for high speeds) if only high frequencies are present in the stimuli, and high frequencies did not seem to help improve speed estimation accuracy when low frequencies are present, we argue that the high spatial frequencies may actually have detrimental effects on speed estimation in most daily activities, where most motion speeds are not low. High frequencies make positive contributions only when we observe slow motion (e.g. watching for subtle motion clues in surveillance videos). The detrimental effect would not happen to those (high frequency) feature tracking-based computer vision algorithms.

Our human subject experimentation supports the importance of low spatial frequency components in perception of motion. When viewing stochastic sequences normally (NV condition), the observers performed far worse for higher speed stimuli than for the lower speed stimuli. However, when viewing with a blur lens (LV condition), the subjects performed significantly better for higher speed stimuli than they did in the NV condition. The explanation from the motion energy perspective is as follows. When viewing the stochastic sequences with NV, the eyes received relatively weaker motion energy associated with low spatial frequencies as compared with high frequencies. Through the blurred lens, which smoothed the

stimuli, the interference from the motion energy content associated with the high spatial frequencies was suppressed. This helped improve the direction detection performance. The experiment demonstrated that seeing more details does not necessarily help with motion perception when the low spatial frequency content is not predominant in the scene. The qualitative consistency between model judgment simulation and human judgment may support the validity of using the computational model to understand motion perception in humans in certain situations.

There is a strong support for the importance of low spatial frequencies and/or relative irrelevance of high frequency components in motion perception in the literature (Gilden, Bertenthal, & Othman, 1990; Hess & Aen-Stockdale, 2008; Lappin, Tadin, Nyquist, & Corn, 2009; Morgan, 1992; Pan & Bingham, 2013; Ramachandran, Ginsburg, & Anstis, 1983; Saunders, Bex, Rose, & Woods, 2014; Shioiri, Ito, Sakurai, & Yaguchi, 2002; Tadin, Nyquist, Lusk, Corn, & Lappin, 2012; Wichmann & Henning, 1998; Yang & Stevenson, 1997). However, testing with complementary vision conditions was not extensively explored previously, except in the case of Smith et al., where testing was also done with a high pass version of random dot kinematogram (RDK; Smith, Snowden, & Milne, 1994). Interestingly, they found no difference in global motion perception between normal and high pass RDKs and concluded that global motion perception was not dependent on low spatial frequencies, which *prima facie* seems contrary to our results. This inconsistency can be explained with two observations. First, it is possible that human observers in the study by Smith et al. (Smith, Snowden, & Milne, 1994) were using feature tracking to estimate global motion (as is their conclusion), unlike our model that is based on motion energy. It has been suggested that the feature tracking mechanism remains unaffected by spatial filtering and that it could simultaneously co-exist with motion energy-based motion perception mechanism (Smith & Ledgeway, 2001). Our second observation relates to the experimental methodology, as the speeds of the test stimuli in Smith et al. (Smith, Snowden, & Milne, 1994) ranged from 2.5 °/s to 5.5 °/s, which were on the lower side of the range used in our simulations. We can see that at these speeds, our simulation shows relatively small speed estimation errors between complementary vision (high pass cases) and normal vision conditions (see Figure 5c). It is only for higher speeds, where we can see an appreciable difference in the estimation errors between the LV and complementary vision conditions.

Thus, separately evaluating the LV (low spatial frequency) and complementary vision conditions (high frequency bands) for a wide range of speeds further verified the known results regarding the importance of low spatial frequencies in motion perception, showing validity of our motion perception model. At the same

time, it helped explain apparent inconsistencies in the previous work.

Implicit within the speed estimation error results presented in Figure 5 is the observation that accurate estimation at higher speeds requires lower spatial frequencies, which is clearly seen in the biasing of motion energy curves toward low frequency bands at higher speeds in Figure 4. This is consistent with the findings previously reported in different forms and contexts in a number of previous studies showing increasing speeds led to a shifting of the psychophysical response curves toward low spatial frequencies (Chen, Bedell, & Frishman, 1998; Chung, Levi, & Bedell, 1996; Levi, 1996; McKee, Silverman, & Nakayama, 1986; Mechler & Victor, 2000; Smith & Edgar, 1990). Although the actual ranges tested in these studies differed and were not as wide ranging in terms of spatial frequencies or speeds used in these experiments, the overall trend still holds. Serving as a proxy for the psychophysical motion thresholds, the motion energy distributions generated by our computational model were able to replicate the observed relationship between speed and spatial frequency content of the stimulus.

Although the results in Figure 5 were reported in terms of absolute error values, it should be noted that the model underestimated the speed at higher spatial frequencies. This is consistent with the earlier finding reported by Smith & Edgar (1990).

Simulating both LV and complementary vision conditions and testing on different stimuli (natural versus stochastic) also allowed us to answer questions about the interaction between low and high spatial frequency components, which are available at the same time to the human eye in the real world. Specifically, given the importance of low spatial frequencies in speed perception, does it matter if high frequencies are present or not as long as low frequencies are present? Based on the speed estimation results for natural sequences (see Figure 5a,c), one may conclude that as long as low spatial frequencies are present, speed estimation is equally accurate with and without high spatial frequencies. However, speed estimation in stochastic sequences (see Figure 5b) serves as a counter to this claim, as error is significantly higher in NV conditions compared with LV conditions. In addition, the error is higher in 20/50 vision conditions at higher speeds compared with 20/200 vision condition. This indicates that in stochastic sequences, presence of high spatial frequencies is detrimental to speed estimation.

A possible explanation for this effect can be provided based on the signal-to-noise ratio (SNR) in the input stimuli, while noting the difference between the motion energy distributions of natural and stochastic sequences (see Figure 4). Unlike in natural sequences where motion energy pertaining to low spatial frequencies was dominant, stochastic sequences had a relatively flat motion energy distribution. This is the consequence

of spatial frequency distribution in natural imagery following power-law (van der Schaff & van Hateren, 1996), being rich in low frequency content compared with stochastic sequences. Speed estimation at higher speeds is highly reliant on the motion energy extracted from low spatial frequency bands. For stochastic sequences with added noise, the SNR in the low spatial frequency band was much lower compared with the natural sequences, which led to higher speed estimation errors in NV condition. When the sequences were low pass filtered to simulate 20/50 and 20/200 vision conditions, the SNR improved and allowed more accurate speed estimation compared with NV conditions. This finding is in line with Van Doorn's study, which found that a high SNR is required for accurate speed estimation in the case of fast visual motion (Van Doorn & Koenderink, 1982).

Thus, we can conclude that sufficient low spatial frequency content is required for accurate speed estimation, otherwise detrimental effects of high spatial frequencies may become evident. Natural images have more low frequency components than higher ones, and vision loss usually starts with impairment of high-resolution visual function. Therefore, the detrimental effect of high spatial frequencies is not typically seen, unless specific artificial stimuli are used. Probably because more low spatial frequency components may help minimize motion perception error, primate visual systems seem to have more neurons tuned to low frequencies than those tuned to high frequencies (Perrone & Thiele, 2001).

We further predict that there is a threshold spatial frequency separating signal (consisting of motion energy corresponding to the frequencies below the threshold) from the noise (motion energy of frequencies above the threshold) for any given speed that determines the required SNR for speed perception in that scene. Slower speeds push this signal-noise separating threshold toward higher spatial frequencies, which means a larger frequency band becomes signal, thus improving the SNR. On the other hand, high speeds push the threshold to low spatial frequency region thereby increasing the reliance of speed perception on low frequencies. Changes to the spatial frequency spectrum on either side of this separating threshold accordingly affect the SNR and in turn affect the speed perception. Further psychophysics studies are needed to confirm this prediction in human observers.

Our findings can be explained at a more fundamental level of V1 and MT neurons using the speed perception model proposed by Simoncelli & Heeger (Simoncelli & Heeger, 2001) that postulates that oriented spatial receptive field (SRF) of speed tuned MT neurons can be derived by combining SRFs of different directionally tuned V1 neurons (not speed tuned). In spatiotemporal space, a speed tuned neuron has an oriented SRF that shows speed invariance over a range of spatial and

temporal frequencies (Perrone & Thiele, 2001; Priebe, Cassanello, & Lisberger, 2003). On the other hand, a neuron that is not speed tuned (V1 simple cells) will have an SRF that is parallel to the axes of the spatiotemporal space. Larger slope of the oriented MT neuron SRFs indicate tuning for higher speeds and thus would involve combining different V1 neurons as opposed to those tuned for slow speed. In the absence of low frequencies and in the presence of high spatial frequencies (such as in the case of our stochastic sequences), there will be a relatively larger error in the estimation of the slope (speed) at higher speeds (Figure 7). Because there is a limit on our perception of spatial and temporal frequencies, higher temporal frequencies are not available to compensate for the lack of low spatial frequency for estimating higher speeds (temporal limit [Nakayama, 1990]).

One high-level task potentially relevant to this work is driving behavior in people with reduced VA. By exploring the relationship between speed and spatial frequency content of the scene using a motion perception model (and supported by human subject experiments), we could determine the importance of different spatial frequency bands for motion estimation. Particularly, this work tried to address questions related to whether reduced VA affects perception of motion in natural stimuli. Using a computational model allows us to isolate separate high and low frequency bands and study the speed estimation for different speed values.

The speed perception model and the simulations of LV conditions used for evaluation in this work have some limitations. First, extremely low VAs (below 20/200) were not simulated, as they required a very high spatiotemporal resolutions and large filter sizes in order to satisfy the Nyquist rate. At very low spatial frequencies, we expect motion sensitivity to be impaired for very slow speeds (Yang & Stevenson, 1997), which means that people with very low VAs will not be able to perceive slow motion speeds. Due to the limits of the Nyquist sampling rate, we were not able to simulate ultra-low VA cases to try and reproduce this result. In addition, the assumption that LV conditions correspond to the inability in perceiving higher spatial frequencies is somewhat simplistic, and, in the real world, LV corresponds to a variety of conditions where the perception of a very wide range of spatial frequencies can be impaired in a nonuniform manner (Chung & Legge, 2016). Thus, our simulation of LV conditions is only an approximate representation of real-world cases. The model presented here is an implementation of a spatiotemporal motion energy based model and the speed estimation error may not generalize to other kinds of motion estimation models (such as Bowns, 2011; Bowns, 2018). Speed processing in humans is not completely understood and this work explores only certain specific aspects of speed processing and

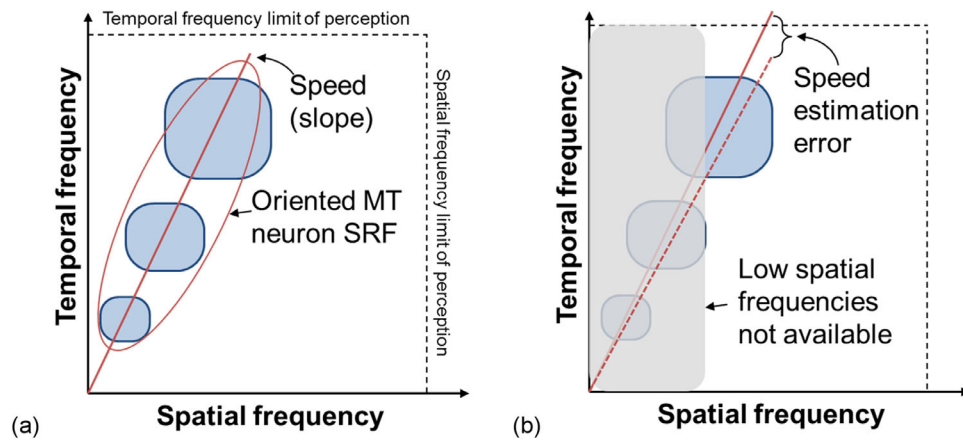


Figure 7. An explanation of why speed perception at higher speeds is erroneous in the absence of low spatial frequencies. (a) According to the model proposed by Simoncelli and Heeger (Simoncelli & Heeger, 2001), directionally sensitive V1 neurons that are not speed tuned (shown as blue blobs) are pooled together to arrive at an estimate of speed given by the slope of the oriented ellipse (red curve) in the spatiotemporal space. The speed tuned MT neuron has oriented spectral receptive field (SRF), with higher slope corresponding to higher speed. (b) Due to the limits on perception, when the band containing low spatial frequencies are not available (for example, as is the case in the complementary vision conditions), results in an error in speed estimation that increases with increasing motion speed of the stimuli.

its relationship to spatial frequency content, using a computational model of speed estimation. Further human subject studies are warranted.

Keywords: motion perception, spatial frequency, biologically inspired speed perception model, simulation of vision conditions

Acknowledgments

Supported by the National Institutes of Health (NIH) under Grant R01 AG041794.

Commercial relationships: none.

Corresponding authors: Shrinivas Pundlik; Cong Shi.
Email: shrinivas_pundlik@meei.harvard.edu;
shicong@cqu.edu.cn.

Addresses: Schepens Eye Research Institute of Mass Eye & Ear, 20 Staniford Street, Boston MA 02114, and School of Microelectronics and Communication Engineering, Chongqing University, Chongqing, China.

References

- Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, 2, 284–299.
- Bex, P. J., & Dakin, S. C. (2002). Comparison of the spatial-frequency selectivity of local and global motion detectors. *Journal of the Optical Society of America. A, Optics Image Science and Vision*, 19, 670–677.
- Bowns, L. (2011). Taking the energy out of spatiotemporal energy models of human motion processing: The component level feature model. *Vision Research*, 51, 2425–2430.
- Bowns, L. (2018). Motion estimation: A biologically inspired model. *Vision Research*, 150, 44–53.
- Burr, D., & Thompson, P. (2011). Motion psychophysics: 1985–2010. *Vision Research*, 51, 1431–1456.
- Chen, Y., Bedell, H. E., & Frishman, L. J. (1998). The precision of velocity discrimination across spatial frequency. *Perception & Psychophysics*, 60, 1329–1336.
- Chung, S., & Legge, G. (2016). Comparing the shape of contrast sensitivity functions for normal and low vision. *Investigative Ophthalmology & Visual Science*, 57, 198–207.
- Chung, S. T. L., Levi, D. M., & Bedell, H. E. (1996). Vernier in motion: What accounts for threshold elevation? *Vision Research*, 36, 2395–2410.
- Cleary, R., & Braddick, O. J. (1990). Direction discrimination for band-pass filtered random dot kinematograms. *Vision Research*, 30, 303–316.
- Etienne-Cummings, R., Van der Spiegel, J., & Mueller, P. (1999). Hardware implementation of a visual-motion pixel using oriented spatiotemporal neural filters. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 46, 1121–1136.

- Gilden, D. L., Bertenthal, B. I., & Othman, S. (1990). Image statistics and the perception of apparent motion. *Journal of Experimental Psychology*, *16*, 693–705.
- Grzywacz, N. M., & Yuille, A. (1990). A model for the estimate of local image velocity by cells in the visual cortex. *Proceedings of the Royal Society of London B: Biological Sciences*, *239*, 129–161.
- Hess, R. F., & Aen-Stockdale, C. (2008). Global motion processing: The effect of spatial scale and eccentricity. *Journal of Vision*, *8*, 1–11.
- Lappin, J. S., Tadin, D., Nyquist, J. B., & Corn, A. L. (2009). Spatial and temporal limits of motion perception across variations in speed, eccentricity, and low vision. *Journal of Vision*, *9*, 1–14.
- Levi, D. M. (1996). Pattern perception at high velocities. *Current Biology*, *6*, 1020–1024.
- McKee, S., Silverman, G., & Nakayama, K. (1986). Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision Research*, *26*, 609–619.
- Mechler, F., & Victor, J. D. (2000). Comparison of thresholds for high-speed drifting vernier and a matched temporal phase-discrimination task. *Vision Research*, *40*, 1839–1855.
- Morgan, M. J. (1992). Spatial filtering precedes motion detection. *Nature*, *355*, 344–346.
- Nakayama, K. (1990). “Properties of early motion processing: Implications for the sensing of ego motion,” in *The Perception and Control of Self Motion*, eds. R. Warren, & A.H. Wertheim. (Hillsdale NJ: Lawrence Erlbaum), 69–80.
- Nishida, S. (2011). Advancement of motion psychophysics: Review 2001-2010. *Journal of Vision*, *11*, 11.
- Ogata, M., & Sato, T. (1991). Motion perception model with interaction between spatial frequency channels. *Systems and Computers in Japan*, *22*, 30–39.
- Pan, J. S., & Bingham, G. P. (2013). With an eye to low vision: Optic flow enables perception despite image blur. *Optometry and Vision Science*, *90*, 1119–1127.
- Perrone, J. A., & Thiele, A. (2001). Speed skills: measuring the visual speed analyzing properties of primate MT neurons. *Nature Neuroscience*, *4*, 526–532.
- Perrone, J. A., & Thiele, A. (2002). A model of speed tuning in MT neurons. *Vision Research*, *42*, 1035–1051.
- Priebe, N. J., Cassanello, C., & Lisberger, S. G. (2003). The neural representation of speed in macaque area MT/V5. *The Journal of Neuroscience*, *23*, 2650–2661.
- Ramachandran, V. S., Ginsburg, A. P., & Anstis, S. M. (1983). Low spatial-frequencies dominate apparent motion. *Perception*, *12*, 457–461.
- Saunders, D. R., Bex, P. J., Rose, D. J., & Woods, R. L. (2014). Measuring information acquisition from sensory input using automated scoring of natural-language descriptions. *PLoS One*, *9*, e93251.
- Shi, C., & Luo, G. (2018). A compact VLSI system for bio-inspired visual motion estimation. *IEEE Transactions on Circuits and Systems for Video Technology*, *28*, 1021–1036.
- Shioiri, S., Ito, S., Sakurai, K., & Yaguchi, H. (2002). Detection of relative and uniform motion. *Journal of the Optical Society of America. A, Optics Image Science and Vision*, *19*, 2169–2179.
- Simoncelli, E. P., & Heeger, D. J. (2001). Representing retinal image speed in visual cortex. *Nature Neuroscience*, *4*, 461–462.
- Smith, A., & Ledgeway, T. (2001). Motion detection in human vision: A unifying approach based on energy and features. *Proceedings of the Royal Society of London B: Biological Sciences*, *268*, 1889–1899.
- Smith, A. T., & Edgar, G. K. (1990). The influence of spatial frequency on perceived temporal frequency and perceived speed. *Vision Research*, *30*, 1467–1474.
- Smith, A. T., Snowden, R. J., & Milne, A. B. (1994). Is global motion really based on spatial integration of local motion signals? *Vision Research*, *34*, 2425–2430.
- Tadin, D., Nyquist, J. B., Lusk, K. E., Corn, A. L., & Lappin, J.S. (2012). Peripheral vision of youths with low vision: Motion perception, crowding, and visual search. *Investigative Ophthalmology & Visual Science*, *53*, 5860–5868.
- van der Schaff, A., & van Hateren, J. H. (1996). Modelling the power spectra of natural images: statistics and information. *Vision Research*, *36*, 2759–2770.
- Van Doorn, A., & Koenderink, J. (1982). Spatial properties of the visual detectability of moving spatial white noise. *Experimental Brain Research*, *45*, 189–195.
- Wichmann, F. A., & Henning, G. B. (1998). No role for motion blur in either motion detection or motion-based image segmentation. *Journal of the Optical Society of America - A, Optics Image Science and Vision*, *15*, 297–306.
- Yang, J., & Stevenson, S. B. (1997). Effects of spatial frequency, duration, and contrast on discriminating motion directions. *Journal of the Optical Society of America - A, Optics Image Science and Vision*, *14*, 2041–2048.