

Spatiotemporal flicker detector model of motion silencing

Lark Kwon Choi¹, Alan C Bovik¹, Lawrence K Cormack²

¹Department of Electrical and Computer Engineering, The University of Texas at Austin, UT Austin WNCG, 1616 Guadalupe St, UTA 7.518, Austin, TX 78701, USA; ²Department of Psychology, The University of Texas at Austin, 108 E Dean Keeton Stop A8000, Austin, TX 78712-1043, USA; e-mail: larkkwonchoi@utexas.edu
Received 7 April 2014, in revised form 12 October 2014

Abstract. Motion can impair the perception of other visual changes. Suchow and Alvarez (2011a, *Current Biology*, **21**, 140–143) recently demonstrated a striking ‘motion silencing’ illusion, in which the salient changes among a group of objects’ luminances (or colors, etc) appear to cease in the presence of large, coherent object motion. To understand why the visual system might be insensitive to changes in object luminances (‘flicker’) in the presence of object motion, we constructed similar stimuli and did a systematic spectral analysis of them. We conducted human psychophysical experiments to examine motion silencing as a function of stimulus velocity, flicker frequency, and spacing; and we created a simple filter-based model as a working hypothesis of motion silencing. From the results, we found that the threshold of silencing occurs when the log frequency of object replacement is roughly one quarter of the log flicker frequency (the mean slope is approximately 0.27). The dependence of silencing on object spacing may be explained as a phenomenon of temporal sampling of the stimuli by the visual system. Our proposed model successfully captures the psychophysical data over a wide range of velocities and flicker frequencies.

Keywords: motion silencing, illusion, flicker detector, spatiotemporal model, primary visual cortex

1 Introduction

A recent study has demonstrated that objects changing in color, luminance, size, or shape appear to stop changing when they move rapidly in collective motion (Suchow & Alvarez, 2011a). One hundred small dots were randomly arranged in a ring-shaped pattern around a central fixation mark (see the illusion at <http://visionlab.harvard.edu/silencing/>). Each dot changed continuously over time in color, luminance, size, or shape. The changes were easily noticeable when the dots were stationary, but the changes were undetectable when the dots were suddenly sent into continuous rotational motion. This motion-induced failure to detect change, known as silencing, not only suggests the tight coupling of motion and object appearance but also reveals that motion can disrupt the perception of salient changes in visual objects. The mechanisms underlying this phenomenon remain unknown, although we expect that it might be explainable using known mechanisms. Here, we aim to model the effects of the illusion using well-known spatiotemporal filter-based models.

The brief window hypothesis states that change detection relies on the success of local detectors, which fail when a fast-moving object affords them only a brief glance. Suchow and Alvarez (2011a) suggested this hypothesis as an explanation for silencing, but they later argued that it was insufficient, suggesting suppression or misattribution as alternative hypotheses (Suchow & Alvarez, 2011b). The misattribution hypothesis posits that, when there is an actual motion signal, the dynamic signal from the flicker is misattributed to the motion signal, and hence no flicker is perceived (Choi, Bovik, & Cormack, 2012). In addition, a combination of motion and crowding was suggested to explain silencing (Turi & Burr, 2013), whereby silencing depends on both the spacing between the target and the nearby object (‘flanker’) and eccentricity, with critical spacing about half eccentricity in

agreement with Bouma's Law (Bouma, 1970). Peirce (2013) tested whether the awareness of motion signals can be silenced by coherent changes in color or size and whether coherence is necessary for silencing. His results suggest that neither motion nor coherent changes are necessary for silencing. Although these studies have contributed greatly to our understanding of why silencing occurs, the topic is still controversial because the underlying neural systems have not yet been revealed. Here, we propose an explanation that does not require crowding.

To determine why the visual system silences changes of objects in the presence of rapid motion, we constructed similar visual stimuli to induce the conditions under which silencing occurs. Observations and systematic spatiotemporal spectral analyses of the presentation data led us to develop a simple filter-based hypothesis and a spatiotemporal flicker detector model to explain the motion silencing phenomenon. We conducted a series of human psychophysical experiments to understand whether our filter-based hypothesis could explain the motion silencing illusion.

2 Methods

2.1 Observers

Three University of Texas students served as naive observers. One researcher (an author) also served as an observer. All observers were between the ages of 20 and 35 years. They had normal or corrected-to-normal vision when an optometrist checked their vision with an acuity test (Snellen Test) and a color perception test (Ishihara Test) during subject screening. This project was approved by the Institutional Review Board at The University of Texas at Austin.

2.2 Stimuli

We first placed the changing objects in a regular array with uniform radial and angular spacing. Three concentric circles with 24 equally spaced dots per circle were shown in each video frame. During our observations we hypothesized that the degree of silencing could be impacted by the spacing between dots, so we next created separate videos, each containing one ring made of 24, 18, 12, or 10 dots, as shown on http://live.ece.utexas.edu/research/motion_silencing/index.html, and in figure 1. We have also placed these data on the public repository at <https://osf.io/izrby/>.

A cluster of 24, 18, 12, and 10 dots was distributed in a circle at equally spaced intervals (which corresponded to $\pi/12$, $\pi/9$, $\pi/6$, and $\pi/5$ radians, respectively) around a red central fixation mark at a distance of 6.42 deg (242 pixels) from the center. Each dot was 0.92 deg (35 pixels) in diameter. The initial luminance of each dot was chosen randomly from a uniformly distributed eight-bit grayscale. The luminance changed sinusoidally in a gray background and ranged from 'black' at the weakest intensity to 'white' at the strongest. The luminance of 'black' dots, 'white' dots, and the gray background was 0.20, 255.9, and 45.27 cd m⁻², respectively.

The flicker frequency (the number of oscillations in luminance per second) of each dot was 1/12, 1/6, 1/4, 1/3, or 1/2 Hz, and the flicker frequency did not change from the start to the end of a trial. The ring of dots continuously rotated in either a clockwise or a counter-clockwise direction; the direction was alternated. The velocity (v) was the instantaneous tangential speed of a dot as it followed its circular trajectory on the screen, which was expressed in terms of deg s⁻¹. The range of velocity was 0–29.1 deg s⁻¹, and the initial velocity did not change during each trial.

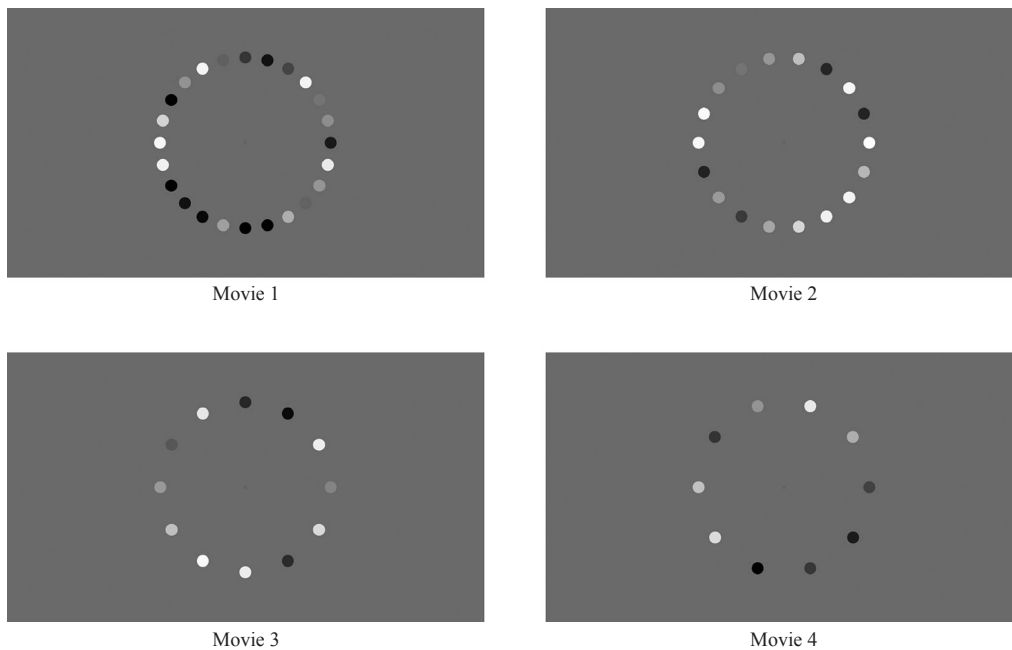


Figure 1. Examples of stimuli used in the human experiments. A cluster of 24, 18, 12, or 10 dots was arranged in a ring around a central fixation mark. The distance from the center to each dot was 6.42 deg, and each dot was 0.92 deg in diameter. The luminance of each dot was chosen randomly from black to white and changed sinusoidally. The flicker frequency was 1/12, 1/6, 1/4, 1/3, or 1/2 Hz and did not change during each trial. The ring of dots continuously rotated. The rotational velocity varied between trials and was expressed in terms of deg s^{-1} .

2.3 Apparatus

The experiments were programmed using MATLAB and the Psychophysics Toolbox (Brainard, 1997). The Psychophysics Toolbox interfaced with an NVIDIA GeForce G 105M graphics card in a Windows computer. The study was conducted using a liquid crystal display (LCD) (Hewlett-Packard, Palo Alto, CA). The spatial resolution was 1366×768 with a pixel density of 96 ppi (38 pixels cm^{-1}). The LCD refreshed at 60 Hz and was illuminated by a 200 Hz backlight. Measurements made with a V-lambda corrected fast photocell (United Detector Technologies PIN-10AP) confirmed that the display correctly and consistently rendered single frame stimuli, and was additive over frames with no interaction (so that a 3 frame stimulus, for example, was just a repeated longer version of a 1 frame stimulus with a 60 Hz temporal frequency response). Despite the known problems with early LCD displays, we have had good experiences with more recent models, and we are confident that our monitor correctly rendered our stimuli. The viewing distance was approximately 57 cm.

2.4 Design and procedure

In each trial the observer was shown a stimulus in the center of the screen, as shown in figure 1. Observers indicated whether the moving dots were flickering (ie were changing luminances) or not while holding their gazes on a red central fixation mark.

We ran four dot spacings ($\pi/12$, $\pi/9$, $\pi/6$, and $\pi/5$ radians of angular separation) and five flicker frequencies (1/12, 1/6, 1/4, 1/3, and 1/2 Hz) in a fully crossed design yielding 20 experimental conditions. Within each condition the stimulus rotational velocity was varied over five values to yield a psychometric function. The particular velocities were chosen per observer and condition to yield a good psychometric function based on pilot data

(‘good’ meaning spanning the ascending part of the function without unduly sampling the tails, as in figure 2). For each stimulus velocity per condition 30 trials were executed, 15 of which presented flickering dots and the remaining 15 of which did not. The order of these 150 trials was randomized. The 20 conditions (four spacings $\times$ five frequencies) were blocked and were also presented in a random order.

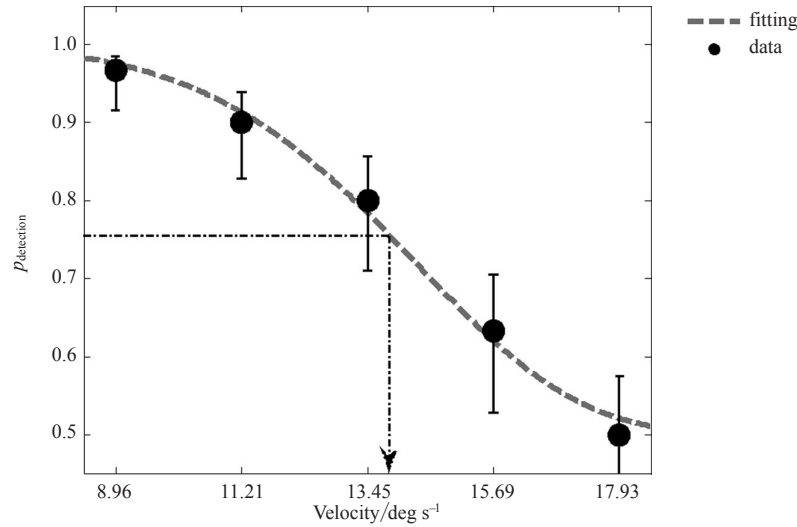


Figure 2. Psychometric data fitted to the Weibull function. The x-axis is rotational velocity (in deg s⁻¹), and the y-axis is a probability of detection ($p_{\text{detection}}$). The velocity corresponding to a probability of detection equal to 0.75 was chosen as the threshold of velocity for silencing, v_{th} . Error bars using the binomial distribution with 68% confidence intervals are shown.

For each trial a ‘yes/no’ procedure was used, in which ‘yes’ corresponded to “I think the dots are flickering” (because half the trials were catch trials, chance performance was 50%). Each stimulus was presented until an observer made his or her judgment (average response time was approximately 3–5 s), and the following trial was automatically initiated after recording the response. To minimize the effects of observer fatigue and eyestrain, observers were allowed to rest for as long as needed after finishing a condition (150 trials). After completing the first set of all 20 conditions, each observer repeated them twice more, for a total of three runs per condition ($20 \times 150 \times 3$, or a total of 9000 trials per observer). A short training exercise—which was similar to a real test but had different flicker frequency, velocity, and dot spacing of the visual stimuli—was conducted prior to data collection.

We found the threshold of velocity for silencing, v_{th} , by fitting the acquired psychometric data to the Weibull function (Weibull, 1951; Wichmann & Hill, 2001). The Weibull function used in our experiments is:

$$W(x, \alpha, \beta, q) = \frac{1}{2} + \frac{1 - \exp(-x/\alpha)^\beta}{2}, \quad (1)$$

where α is the (log) offset parameter and β is the (log) slope parameter. On the basis of 150 trials at each spacing and frequency condition, we selected the velocity corresponding to a probability of detection equal to 0.75 at determining whether the dots were actually flickering or not as v_{th} . We found that the observer could detect the dot flicker easily in a low-velocity display, but detection of the dot flicker was difficult in a high-velocity display. Figure 2 shows an example fit of the psychometric data with the Weibull function. The error bars using the binomial distribution with 68% confidence intervals are also shown.

3 Results

The results of the human psychophysical experiment show that motion silencing occurs when objects move fast enough within a given flicker frequency and object spacing. Our results also reveal that the threshold of silencing occurs when the log frequency of object replacement is roughly one quarter of the log flicker frequency (the mean slope is approximately 0.27) for all spacing.

This experiment measured the threshold of velocity for silencing (v_{th}) as a function of flicker frequency ($f_{flicker}$), given dot spacing Δx . We analyzed the average v_{th} (in deg s^{-1}) of four observers over all trials because each observer's result was similar to the result of all other observers. As shown in figure 3a, when $f_{flicker}$ increased, v_{th} also increased for all Δx , and at a given $f_{flicker}$, silencing occurs when the motion velocity is larger than the value of the curve. For example, when the flicker frequency is 1/4 Hz, and dot spacing is $\pi/12$ radians (the third circle marker in figure 3a), the salient change of luminance stops when velocity is roughly 14.57 deg s^{-1} . Furthermore, v_{th} increases when Δx increases for all $f_{flicker}$. These results show that flicker frequency, velocity, and object spacing impact motion silencing.

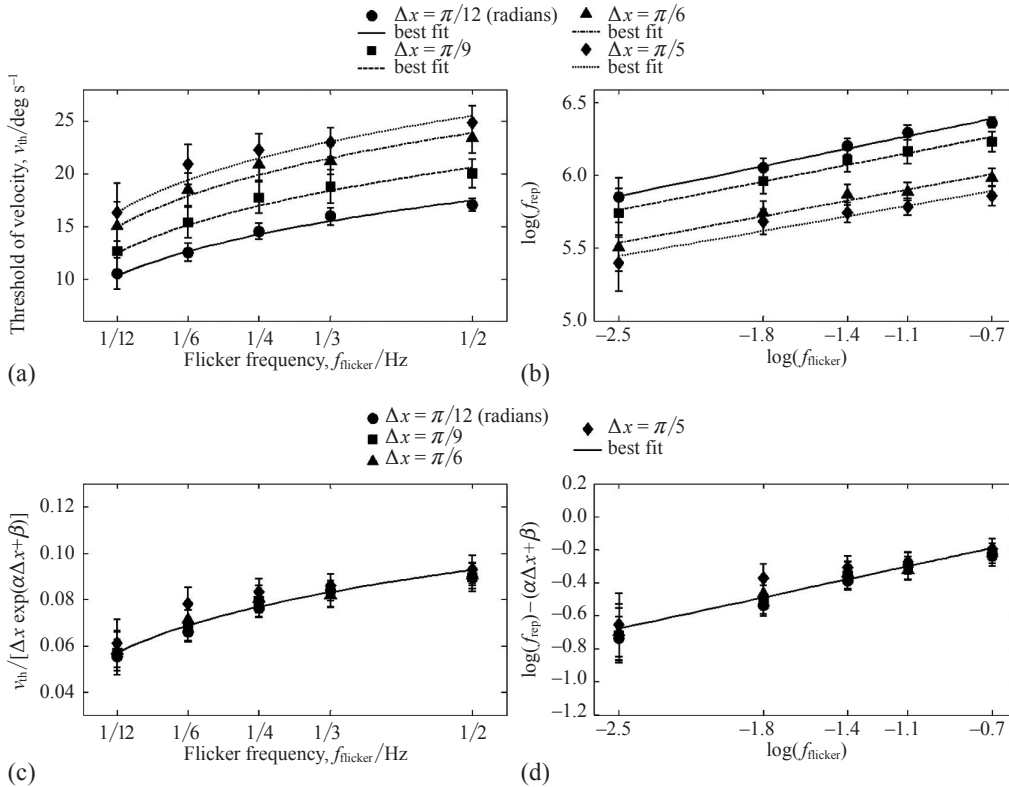


Figure 3. Results of human psychophysical experiments when motion silencing occurs. Results are averaged over all observers, and error bars are shown with 95% confidence intervals. (a) Each marker indicates the obtained threshold of velocity (v_{th}) with respect to flicker frequency ($f_{flicker}$) for different dot spacings— $\Delta x = \pi/12$ ($\bullet$), $\pi/9$ ($\blacksquare$), $\pi/6$ ($\blacktriangle$), and $\pi/5$ ($\blacklozenge$) radians—and each line represents the best fit of corresponding markers. (b) Each marker denotes the log frequency of dot replacement [$\log(f_{rep})$] as a function of the log flicker frequency [$\log(f_{flicker})$], and the line is the best fit (least squares) of corresponding markers. Frequency of dot replacement (f_{rep}) was computed from the obtained v_{th} given Δx (see the appendix). (c) The data of figure 3a replotted after dividing by $\Delta x \exp(\alpha \Delta x + \beta)$ with $\alpha = -1.4622$ and $\beta = 6.9720$. The curve, $(f_{flicker})^a$ with $a = 0.2725$, is the best least squares fit of corresponding markers. (d) The data of figure 3b replotted after subtracting the offset of $\log(f_{rep})$, $(\alpha \Delta x + \beta)$, from $\log(f_{rep})$. The line is the best fit of corresponding markers.

We next investigated whether a lawful relationship exists among stimuli velocity, flicker frequency, and spacing when motion silencing occurs. In the circular trajectory of dots on the screen (eg figure 1), when motion is sufficiently rapid, one dot replaces another dot so quickly that humans do not perceive the change. The time for dot replacement (Δt , the reciprocal amount of time it takes for one dot to move to the position formerly occupied by its neighbor) when silencing occurs can be calculated by dividing Δx by v_{th} , and the frequency of dot replacement (f_{rep}) is the inverse of Δt . Thus, figure 3b is obtained from figure 3a. As can be seen in figure 3b, f_{rep} is roughly proportional to $f_{flicker}$ on a log–log scale by a factor of approximately one quarter (the mean of the experiments is 0.2725), while the intercept of $\log(f_{rep})$ is linearly proportional to the dot spacing when silencing occurs.

An interesting scaling behavior is evident in the human experimental data. To see this, we have replotted the same data in figures 3c and 3d. If we divide the threshold of velocity (v_{th}) by a function of object spacing—that is, if we plot $v_{th} / [\Delta x \exp(\alpha \Delta x + \beta)]$ —then all the data closely cohere to a single curve, as shown in figure 3c. In a similar way, if we plot $\log(f_{rep}) - (\alpha \Delta x + \beta)$, then all data nearly coincide with a single line, as shown in figure 3d, where $\alpha \Delta x + \beta$ is an offset of $\log(f_{rep})$, and α and β are calculated by a least squares fitting of the human experimental data. For a detailed quantitative explanation see the appendix. The underlying scaling behavior shows that there exists a strong inseparable relationship between the threshold of velocity, flicker frequency, and dot spacing when motion silencing occurs.

The results of the human psychophysical experiments in figure 3a imply spectral constraints on motion silencing over a wide range of velocities and flicker frequencies, given dot spacing. The fits act as threshold contours, where silencing occurs in the region above each contour for any given condition.

4 Model

4.1 Framework

As the human visual system processes visual information efficiently by deconstructing it into spatial and temporal frequency channels (Daugman, 1985; Maffei & Fiorentini, 1973; Pollen, Lee, & Taylor, 1971), we constructed a simple filter-based model using an established spatiotemporal energy model of simple cells in the primary visual cortex (Adelson & Bergen, 1985; DeAngelis, Ohzawa, & Freeman, 1993, 1995; Emerson, Bergen, & Adelson, 1992; Watson, Ahumada, & Farrell, 1986) as a working hypothesis of motion silencing. It is believed that there are three broad temporal frequency channels in human vision: one that is low pass, one peaking at around 8 Hz, and another that peaks at a slightly higher temporal frequency. We first measured the local frequency and motion of the objects, and executed a two-dimensional (2-D) discrete Fourier transform (DFT) (Oppenheim, Schaffer, & Buck, 1999) for the spectral analysis. We applied a simple filter to the spectral signatures of the stimuli, and then determined whether the visual stimuli generated silencing or not based on the filter output.

4.1.1 The stimulus in space–time. To separately access and quantify the local flicker frequency and collective motion of the objects, we constructed a space–time 2-D diagram (indexed array) from the continuous stimuli and represented it as an image. This method aided in measuring and visually understanding the interplay between flicker frequency and object motion.

We used a circle that passes through the center of each dot in the visual stimuli to create the 2-D diagram shown in figure 4a. The luminance along the extent of each circle was ‘straightened’ into a horizontal row or vector starting at a fixed angle (0°) and continuing in a clockwise direction. These circular traces through the changing dots constituted the rows of the space–time diagrams displayed in figure 4b. Uniformly sampling the luminance over time

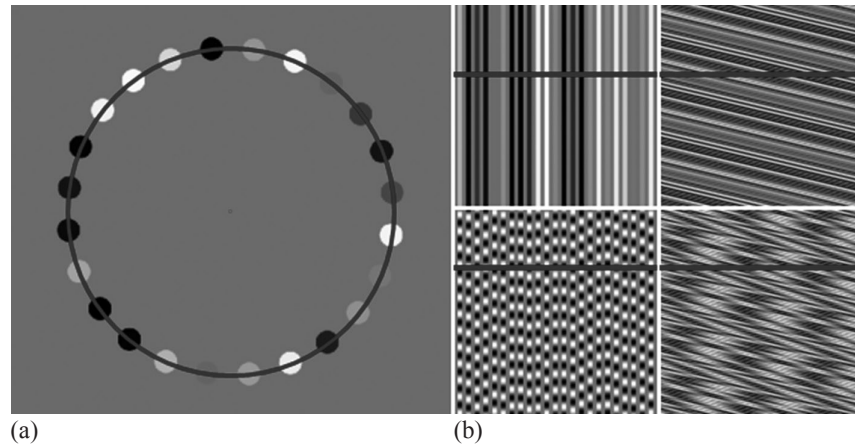


Figure 4. A space–time diagram of the visual stimulus. (a) A circle overlaid on dot objects. In the experiments the ring of dots rotated around its center at a specified rotational velocity (v), at the same time as the dot luminance oscillated at a specified flicker frequency (f_{flicker}). (b) Some space–time diagrams displayed as images under various v and f_{flicker} [from top left to bottom right: $(v, f_{\text{flicker}}) = (0, 0)$, $(6.7, 0)$, $(0, 0.5)$, and $(6.7, 0.5)$, respectively]. v and f_{flicker} were expressed in terms of deg s^{-1} and Hz, respectively.

at 60 Hz generated additional rows, the stack of which constituted the space–time diagram. Thus, vertical columns of the space–time diagram contained temporal luminance variances at fixed spatial positions (on a circle) in the video.

Different velocities and flicker frequencies produced distinct patterns in the space–time diagram. For example, when there was no motion and no luminance change, the first row remained unchanged over time as shown at the top left of figure 4b. When the ring rotated faster without luminance change, the representation became tilted towards the horizontal (eg the top-right panel in figure 4b), which indicates that motion corresponds to orientation in space–time. The bottom-left panel of figure 4b shows modulations when only luminance changed, while the bottom-right panel of figure 4b shows modulations when both motion and flicker frequency were altered. Although the representation of the space–time diagram was specific to our experimental presentation, we believe that the conclusions we reached are relevant to more general space–time diagram tessellations and patterns.

4.1.2 Spectral domain analysis. Each space–time diagram that measured the local flicker frequency and collective motion of the objects was subjected to a 2-D DFT, and the distributions of the spectral signatures were analyzed. The distributions of the spectral signatures constituted a consistent change pattern according to the combination of velocity and flicker frequency of objects, as can be seen in figure 5. The high-energy spectral signatures which are represented by red straight lines moved away from the center when there was no motion ($v = 0$) and flicker frequency increased ($f_{\text{flicker}} = 0, 0.5$, and 1 Hz from left to right), as shown in figures 5a–5c. In addition, the orientation of high-energy spectral signatures increased with respect to the horizontal when velocity increased ($v = 0, 3.36$, and 16.81 deg s^{-1} from left to right) and $f_{\text{flicker}} = 0$, as depicted in figures 5d–5f. These consistent change patterns of the spectral signatures reflect how each velocity and flicker frequency component is deconstructed into spatial and temporal frequency channels.

When both velocity and flicker frequency were induced simultaneously, the separated high-energy spectral signatures due to a given flicker frequency ($f_{\text{flicker}} = 1/4$ Hz) became narrower as they oriented when velocity increased ($v = 0, 3.36$, and 16.81 deg s^{-1} from left to right), as shown in figures 5g–5i. This systematic movement of the spectral signatures

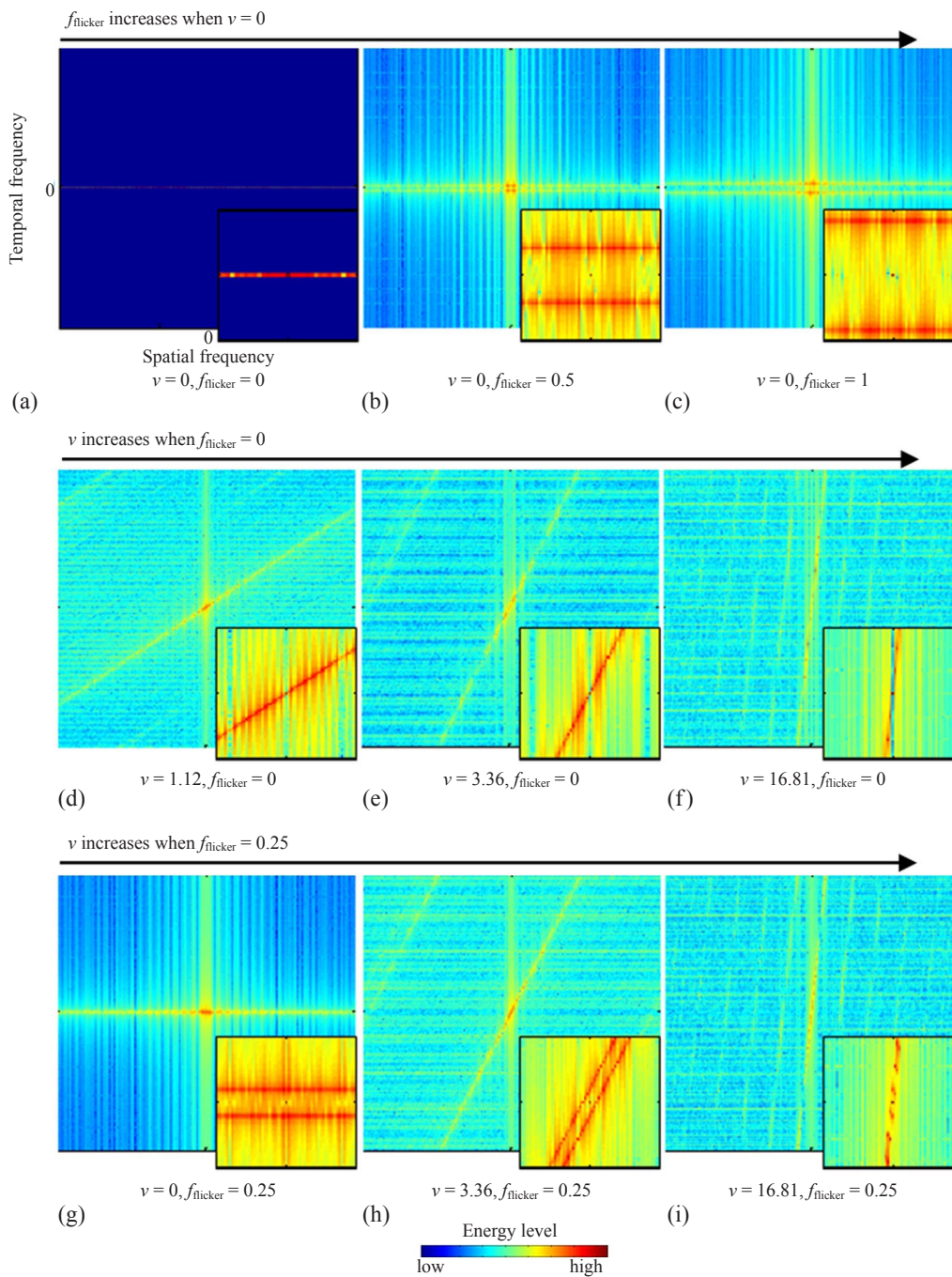


Figure 5. [In color online, see <http://dx.doi.org/10.1068/p7772>] The discrete Fourier transform (DFT) magnitudes of space-time diagrams (centered and logarithmically compressed to reveal the DFT structure) at different combinations of velocity (v) and flicker frequency (f_{flicker}). The unit of v and f_{flicker} is deg s^{-1} and Hz, respectively. In each panel the x-axis is spatial frequency (w_s), and the y-axis is temporal frequency (w_t). The center is a zero spatiotemporal frequency [direct current—that is, $(w_s, w_t) = (0, 0)$]. The arrows on top of each row indicate the changes in stimuli. The energy levels of the spectral signatures are rendered from cool (low energy) to hot (high energy), and the red straight lines denote high-energy spectral signatures. Subfigures located in the bottom right of each panel represent the magnified spectral signatures of the center area.

implies that the tightly coupled motion and changes of an object can be probed in a spatiotemporal frequency domain. Further, the movement of the spectral signatures might serve as a psychometric factor to detect motion silencing.

4.1.3 Simple spatiotemporal filter. The space–time diagrams of the visual stimuli and subsequent spectral analyses may cast against spatiotemporal mechanisms by which neurons in the central visual pathways process images within a localized region of space and time. Therefore, to detect motion silencing, we used a motion energy model of the simple cell receptive fields (Adelson & Bergen 1985; DeAngelis et al., 1993, 1995; Emerson et al., 1992; Watson et al., 1986) to design a spatiotemporal filter that we will hereafter term ‘simple filter’.

A space–time plot was used to characterize the time-varying dynamics of the simple cell receptive fields. Some simple cells have space–time separable receptive fields, but a majority of simple cells exhibit marked space–time inseparability (DeAngelis et al., 1995; McLean & Palmer, 1989). Separable receptive fields can be well approximated by the product of a spatial impulse response and a temporal impulse response, while inseparable receptive fields can be constructed directly from a combination of separable ones (Watson & Ahumada, 1983).

To construct spatiotemporal impulse responses of inseparable receptive fields in the visual system, we implemented a spatial profile using a Gabor function, which is a widely used approximation of simple cell receptive field responses (Bovik, Clark, & Geisler, 1990; Daugman, 1985).

A Gabor function is the product of a sinusoid with a Gaussian envelope:

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \cos(2\pi f_s x - \phi), \quad (2)$$

where σ_x is a receptive field size, f_s is a spatial frequency, and ϕ is a spatial phase. The temporal impulse response is based on linear filters of the form (Adelson & Bergen, 1985):

$$F(t) = (kt)^n \exp(-kt) \left[\frac{1}{n!} - \frac{(kt)^2}{(n+2)!} \right]. \quad (3)$$

We needed a filter with narrow low-response and wide high-response regions in the frequency domain to capture changes in the spectral signatures being used to detect motion silencing. Therefore, we used an even Gabor function with $\sigma_x = 0.06$ and $f_s = 2.2$, and used $n = 3$ for the fast temporal response and $n = 5$ for the slow temporal response with $k = 100$. Figures 6a and 6b show the spatial and the temporal impulse response, respectively.

We combined one spatial (Gabor) and two temporal (fast and slow response) filters into separable spatiotemporal filters, then constructed a simple filter to model inseparable receptive fields by adding those two filters. Figures 6c–6d and 6e–6f show the spatiotemporal profile of the model spatiotemporal filter in the space–time domain and in the spatiotemporal frequency domain, respectively, with close-ups shown on the right.

We applied some simple preprocessing steps to refine the 2-D DFT of the space–time diagram and to emphasize high-energy signatures arising from the moving dots. First, we detected the locations of the pairs of peak values along each column of the matrix, and smoothed the DFT with a small Gaussian filter. We then identified the pairs of peak values along each column using the location that we detected and zeroed all values further than a distance $L = 10$ pixels from each peak. This results in a distinctive isolated signature without discretization artifacts. The flicker detector computes the output of the simple filter as the summed product of the filter and the 2-D DFT represented by high-energy spectral signatures after preprocessing.

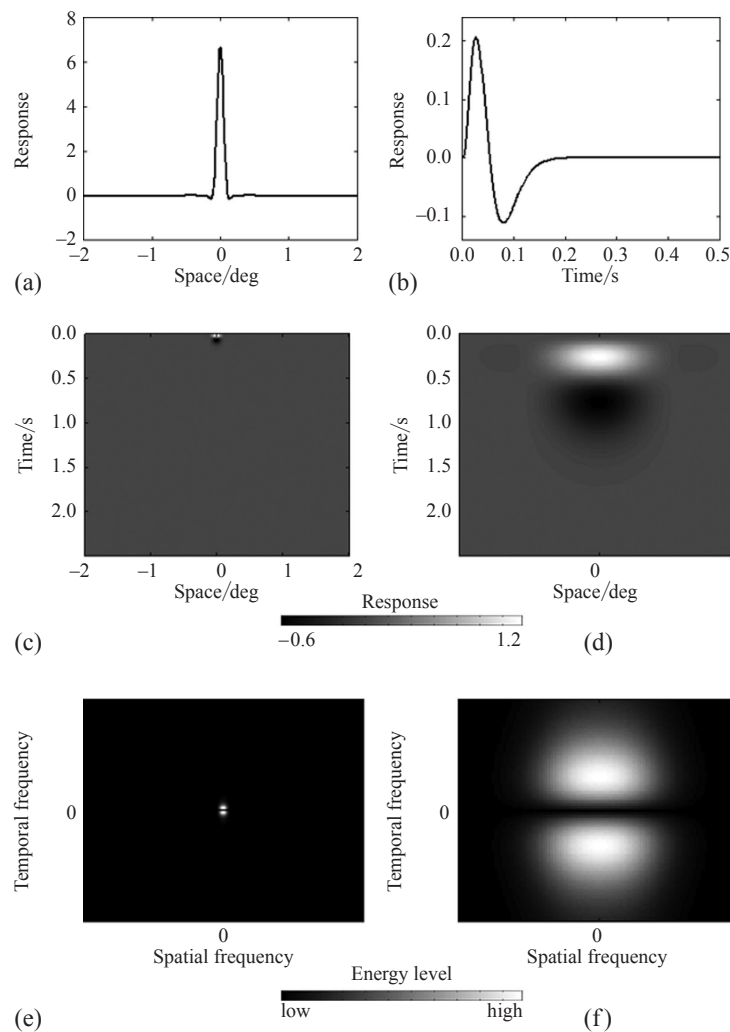


Figure 6. The proposed spatiotemporal filter to detect motion silencing. (a) The spatial impulse response was constructed by using one even Gabor function. (b) The temporal impulse response was implemented by summing two temporal linear filters: fast response and slow response. (c) The spatiotemporal profile in a space–time domain. (d) Magnification of the spatiotemporal profile near 0 deg and 0 s in a space–time domain. Response was rendered from -0.6 to 1.2 with grayscale. (e) Spatiotemporal profile in a frequency domain. (f) Magnification of spatiotemporal profile near direct current. Energy levels are shown from black (low) to white (high).

4.2 Silencing effect

On the basis of systematic spatiotemporal spectral analyses of the visual stimuli and that the proposed filter has a spatiotemporal profile not unlike that of cortical simple cells (Adelson & Bergen, 1985; Carandini et al., 2005; De Valois, Cottaris, Mahon, Elfar, & Wilson, 2000; Hubel & Wiesel, 1962, 1968; Priebe & Ferster, 2008), we pose a simple filter-based hypothesis that, following Occam's Razor, might probe the movement of spectral signatures and detect silencing.

The simple filter shown in figure 7 is a plausible model of the silencing effect: if there was no motion and high flicker (eg figure 7c), a large response would occur but this would be

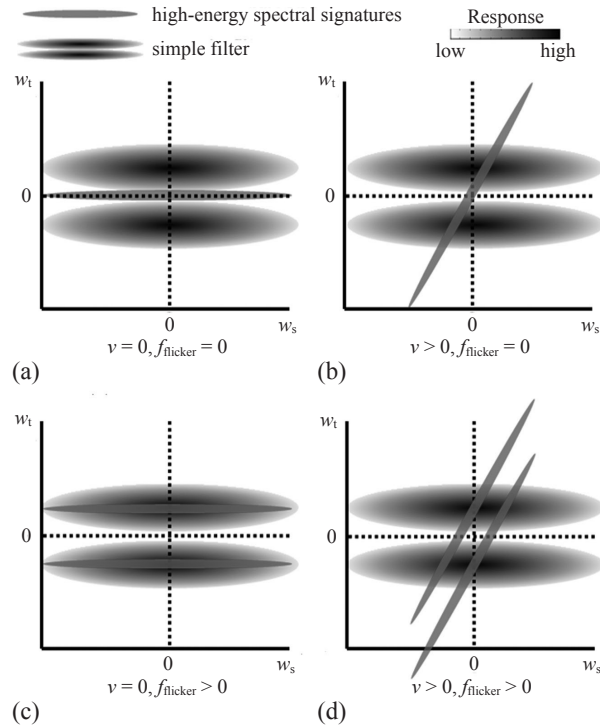


Figure 7. A simple filter-based hypothesis to probe the movement of spectral signatures and to detect motion silencing. (a)–(d) All possible cases of moving and flickering visual stimuli. In each panel the x -axis and the y -axis indicate spatial (w_s) and temporal frequency (w_t), respectively. The flicker frequency (f_{flicker}) moves the signatures towards or away from a zero spatiotemporal frequency [$(w_s, w_t) = (0, 0)$], and motion velocity (v) affects its orientation. A simple filter interacts with different instances of the spectral signatures of flickering objects in motion. The response profile of the filter is displayed from white (low) to black (high). Maximum filter output occurs when (c) $v = 0, f_{\text{flicker}} > 0$ decreases as v increases.

canceled as object motion increased (eg figure 7d). Ostensibly, silencing would occur when the filter response became subthreshold.

4.3 Performance

To test whether a simple filter-based hypothesis works or not, we compared the model results against human psychophysical results. A wide range of visual stimuli including human experiments were generated at four dot spacings ($\Delta x = \pi/12, \pi/9, \pi/6$, and $\pi/5$ radians), six flicker frequencies ($f_{\text{flicker}} = 1/12, 2/12, 3/12, 4/12, 5/12$, and $1/2$ Hz), and twelve velocities ($v = 0, 6.7, 9.0, 11.2, 13.4, 15.7, 17.9, 20.2, 22.4, 24.7, 26.9$, and 29.1 deg s⁻¹). Each stimulus was subjected to the simple filter, and the filter output was computed and compared against the human responses.

The simple filter responses to the visual stimuli over a wide range of v and f_{flicker} when $\Delta x = \pi/6$ radians are depicted in figure 8a. The output of the filter is visually displayed as transitioning from white (high response) to dark gray (low response). The high-energy spectral signatures of the visual stimuli when the stimuli's velocity was zero for a fixed flicker frequency ($f_{\text{flicker}} > 0$) maximally overlapped with the passband (high response) of the simple filter, while as the velocity increases, the high-energy spectral signatures overlapped the stopband (low response) of the filter, as illustrated in figures 7c and 7d. Hence, the output of the filter took maximum value when the velocity was zero for a fixed flicker frequency ($f_{\text{flicker}} > 0$), and decreased as the velocity increased.

We have found that the filter's equireponse contour at a given dot spacing is strongly correlated with the human study results representing the threshold of velocity when motion silencing occurs. The filter's equireponse contour when $\Delta x = \pi/6$ radians, as represented by the thick dashed line (red online) in figure 8a, was superimposed on the human study results for comparison. The corresponding filter's equireponse of $\log(f_{\text{rep}})$ against $\log(f_{\text{flicker}})$ was also plotted with a thick dashed line (red online) in figure 8b. As can be seen in figures 8a and 8b, for all tested flicker frequencies the human study results and the model results were strongly correlated. For other dot spacings (eg $\Delta x = \pi/12$, $\pi/9$, and $\pi/5$ radians) the correlations were high, similar to when $\Delta x = \pi/6$ radians. A method of predicting the threshold of the simple filter output as a function of dot spacing when silencing occurs is described in the appendix.

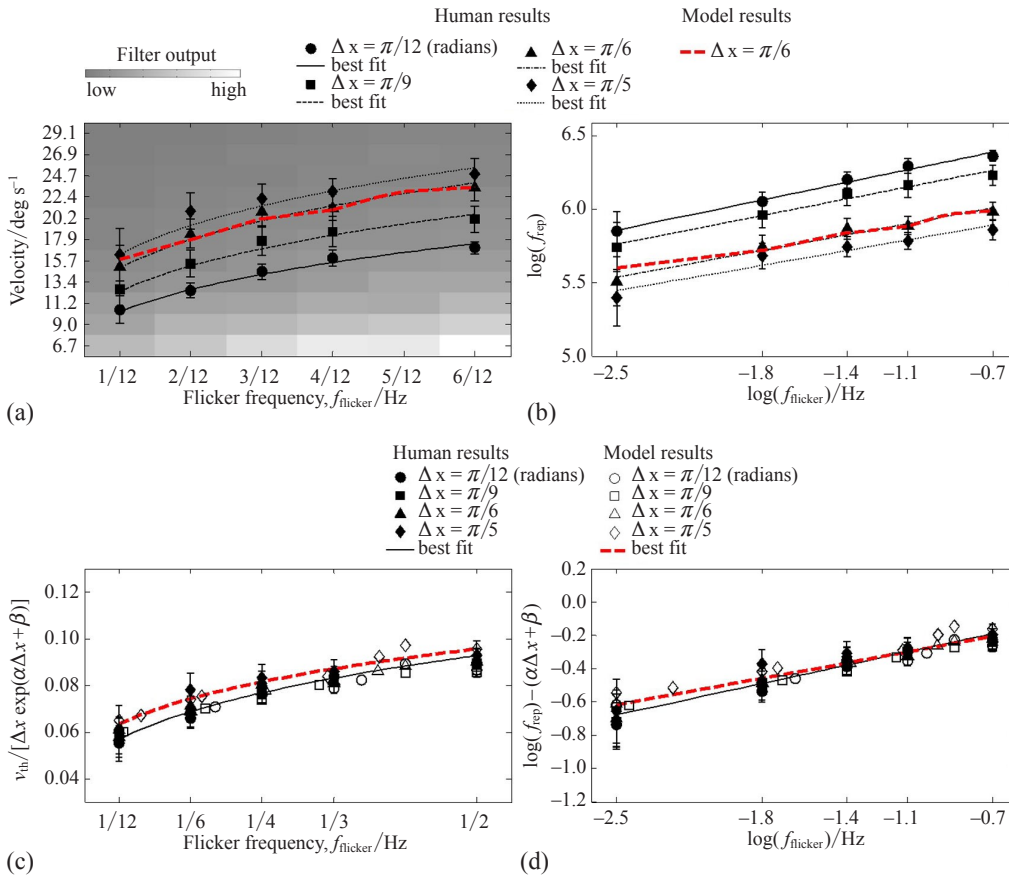


Figure 8. [In color online.] Results of the proposed simple filter-based hypothesis and corresponding flicker detector model. For comparison, the model results are superimposed on the human study results at different dot spacings [represented by Δx , the human results are shown as $\Delta x = \pi/12$ (●), $\pi/9$ (■), $\pi/6$ (▲), and $\pi/5$ (◆) radians, while the model results are shown as $\Delta x = \pi/12$ (○), $\pi/9$ (□), $\pi/6$ (△), and $\pi/5$ (◇) radians]. (a) Velocity (v) plotted against flicker frequency (f_{flicker}). The gray background pattern indicates simple filter outputs for the tested v and f_{flicker} . (b) $\log(f_{\text{rep}})$ plotted against $\log(f_{\text{flicker}})$. In figures 8a and 8b the red dashed line, obtained for equal filter outputs, represents a threshold contour for motion silencing when the dot spacing $\Delta x = \pi/6$ radians. (c) $\hat{v}_{\text{th}}/[\Delta x \exp(\alpha \Delta x + \beta)]$ plotted against f_{flicker} with $\alpha = -1.4622$, and $\beta = 6.9720$, where $\hat{v}_{\text{th}}$ is the predicted v_{th} . (d) $\log(f_{\text{rep}}) - (\alpha \Delta x + \beta)$ plotted against $\log(f_{\text{flicker}})$, where f_{rep} is the predicted f_{rep} . In figures 8c and 8d the red dashed line is the best fit (least squares) of corresponding markers of all model results. There is a strong correspondence between the data and the model.

To evaluate the performance of the proposed model against the human results via prediction accuracy of scaling behavior between the threshold of velocity, flicker frequency, and dot spacing when motion silencing occurs, all predicted model results were superimposed on the human results. Figures 8c and 8d display $\hat{v}_{th}/[\Delta x \exp(\alpha \Delta x + \beta)]$ as a function of $f_{flicker}$ and $\log(\hat{f}_{rep}) - (\alpha \Delta x + \beta)$ as a function of $\log(f_{flicker})$, respectively, where $\hat{v}_{th}$ and $\hat{f}_{rep}$ are the predicted v_{th} and f_{rep} , respectively. We found $\hat{v}_{th}$ from the threshold equivalence contour (eg figure 8a when $\Delta x = \pi/6$ radians), and then $\hat{f}_{rep}$ was obtained from $\hat{v}_{th}$. All model data points fall on a single curve (figure 8c) and a line (figure 8d) close to the human results. This means that the proposed simple filter-based hypothesis and flicker detector model work well to distinguish motion silencing.

Although all dots are physically flickering for all regions, the upper side of the threshold contour in figure 8a indicates a motion silencing (perceptually no-flickering) region, while the lower side represents a perceptually flickering region. We define a silencing map as a diagram including the threshold contour for motion silencing. It can predict when silencing occurs over all ranges of velocity and flicker frequency studied. We recognize, of course, that the occurrence of silencing is a gradual process as rotation speed increases. The threshold contour, therefore, simply represents the location where $p_{detection} = 0.75$ (see figure 2).

Furthermore, it is possible to estimate the degree of silencing for an arbitrary stimulus in the same range studied as a ratio of v to v_{th} . We call this prediction model the motion silencing index (see the appendix).

5 Discussion

We have proposed a physiologically plausible flicker mechanism to model the Suchow–Alvarez motion silencing illusion. This model aims to predict the effect based on an analysis of the spectral signatures of the stimuli using the designed simple filter. Our proposed flicker detector model describes the quantitative relationship between velocity and flicker frequency with regard to silencing, wherein a threshold contour on the designed space–time filter responses divides regions of perceived and unperceived flicker over all velocity and flicker frequency ranges studied. The threshold contour works as a spectral constraint to determine motion silencing. We suggest that the bandpass responses of cortical neurons may be implicated in whether changes in moving objects are perceived.

We also conducted a human psychophysical study, which enabled us to quantitatively model a recently discovered lawful relationship between motion and object change for silencing. The model implies that the log frequency of the dot replacement is roughly one quarter (the mean slope is approximately 0.27) of the log flicker frequency where silencing occurs. Although the proposed flicker detector model accurately predicts motion silencing over a wide range of velocities and flicker frequencies, it may not detect silencing in more extreme cases than the range studied.

The filter's equiresponse contour representing the threshold of motion silencing (eg the red dashed line in figure 8a) was observed to rise as Δx increased. This dependence of silencing on object spacing could be a phenomenon of temporal sampling of the stimuli by the visual system. For a given rotational velocity of the visual stimuli, since the luminance is uniformly sampled over time at 60 Hz, the product of the flicker frequency and the spatial frequency of dots determines the temporal change of luminance along the ring. At the same flicker frequency, when the dot spacing is large along the ring, the temporal frequency of luminance change is low, while when the dot spacing is small, the temporal frequency of change is large. Following the 'strobe-in-head' theory (Simpson, Shahani, & Manahilov, 2005) or temporal subsampling (Purves, Paydarfar, & Andrews, 1996; VanRullen & Koch, 2003), it is possible that, when the stimulus temporal frequency is low enough, the strobe-in-head is fast enough to adequately sample motion and luminance changes, so that the changing luminances of the dots can be seen.

On the other hand, at higher stimulus temporal frequencies undersampling by the strobe-in-head may cause reduced visibility of the luminance changes (silencing). A method of predicting the threshold of the proposed simple filter output when silencing occurs as a function of dot spacing was derived using the human experimental data (see the appendix).

The quantitative relationship between motion silencing and dot spacing follows a consistent model that agrees with human experimental data, possibly revealing a strong relationship between the threshold of velocity, flicker frequency, and dot spacing when motion silencing occurs. The human psychophysical data points, $v_{th}/[\Delta x \exp(\alpha \Delta x + \beta)]$, closely follow a single curve, $(f_{flicker})^a$ with $a = 0.2725$, as shown in figure 3c. This result implies that the thresholds of velocity, flicker frequency, and dot spacing are not separable when motion silencing occurs. Since the proposed flicker detector model measures the spatiotemporal spectral characteristics of time-varying visual stimuli using a nonseparable space–time filter, the hypothesis finds additional support. The corresponding flicker detector model may capture the underlying scaling behavior of motion silencing, as indicated by figures 8c and 8d. The high correlation between the human results and the model results suggest that our model may be a starting point for better understanding neural processing of motion silencing in the visual pathway.

With regard to the underlying scaling behavior of the human experimental data—that is, $v_{th}/[\Delta x \exp(\alpha \Delta x + \beta)] \approx (f_{flicker})^a$ —if we view v_{th} as a function $F(\omega)$ of temporal frequency, ω , and $\Delta x \exp(\alpha \Delta x + \beta)$ as a function $F(f)$ of spatial frequency, f , the scaling behavior has the form:

$$\frac{F(\omega)}{F(f)} \sim (f_{flicker})^a. \quad (4)$$

Thus, we assert that the ratio of temporal frequency to spatial frequency of a time-varying stimulus is proportional to the flicker frequency of the stimuli with an exponent a when motion silencing occurs. This ratio of temporal frequency (or function of) to spatial frequency (or function of) appears to relate to Dong and Atick's (1995) model of natural video statistics. In addition, the value of α is approximately -1.46 , which is not far from the exponent value of -1.2 observed by Tolhurst, Tadmor, and Chao (1992) on natural image spectra.

The gradual change of the filter output represented by transitioning from white (high response) to dark gray (low response) in figure 8a is closely related to the graded effect presented by Suchow and Alvarez (2011a). They reported that the faster the ring rotated, the slower the dots seemed to change, and that the fastest rotation (0.33 Hz) in their experiments produced nearly complete silencing. In figure 8a, at any arbitrary flicker frequency, as the velocity increases, the output of the simple filter response gradually decreases. This implies that human sensitivity to change in moving objects gradually decreases as the velocity increases, and then may become completely lost. Suchow and Alvarez (2011a) showed that silencing can occur at one fixed flicker frequency by abruptly and alternately changing stationary and moving phases of dots. This can drastically affect the temporal characteristics of the display, introducing a range of temporal frequencies not present in the rotating display (Burr, 2011). We captured the graded effect using the responses of the designed filter over a wider range of velocity and flicker frequency conditions as shown in figure 8a and empirically determined the points of nearly complete silencing as indicated by the threshold contours in figure 8a.

There are a number of other observed phenomena that may be relevant to the silencing illusion. For example, Anstis and Ho (2014) found that the apparent speed of rotating dots increases with dot density. Earlier, Thompson (1982) observed that apparent speed is affected by contrast, with lower contrast stimuli being seen as faster. It is plausible that the silencing effect is determined as a function of apparent, rather than real speed. Therefore, it would be

of interest to perform a study that extends the one here, whereby the dot density is increased and/or contrast manipulated, although the problem of estimating the apparent velocity would also have to be solved in order to quantitatively explore the relationship between apparent motion and silencing.

The stimulus used here can be characterized as a moving envelope (the dots) having a temporally modulated carrier (the flicker in the dots). It would be of interest to understand the relevance between the temporal modulation and silencing. Although flicker is one way that temporal modulation can occur, the temporal modulation of the stimuli in our experiments combines multiple factors including motion velocity, flicker frequency, and dot spacing. A deep examination of the separate and interactive effects of these and other controlling factors on temporal contrast modulation with respect to silencing would be a valuable future study, which we envision would require a carefully parsed stimulus setup to isolate the effects of each factor.

A consistent physiological and computational model that detects motion silencing might be useful to probe related motion perception effects, such as distortion visibility in compressed video, such as is commonly seen online. As motion silencing is related to the simultaneous perception of tightly coupled motion and object appearance, the proposed flicker detector model could be applied to address practical problems such as developing perceptual video quality algorithms that can predict ‘silenced’ distortions.

Acknowledgments. This work was supported by Intel and Cisco Corporations under the VAWN program, and by the National Science Foundation under Grants IIS-0917175 and IIS-1116656.

References

- Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, **2**, 284–299.
- Anstis, S., & Ho, A. (2014). Apparent speed of a rotating disk varies with texture density. *Journal of Vision*, **14**(10), 1333. (Abstract)
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, **226**, 177–178.
- Bovik, A. C., Clark, M., & Geisler, W. S. (1990). Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**, 55–73.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10**, 433–436.
- Burr, D. (2011). Visual perception: More than meets the eye. *Current Biology*, **21**, R159–R161.
- Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., Gallant, J. L., & Rust, N. C. (2005). Do we know what the early visual system does? *The Journal of Neuroscience*, **25**, 10577–10597.
- Choi, L. K., Bovik, A. C., & Cormack, L. K. (2012). A flicker detector model of the motion silencing illusion. *Journal of Vision*, **12**(9), 777. (Abstract)
- Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, **2**, 1160–1169.
- DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat’s striate cortex. I. General characteristics and postnatal development. *Journal of Neurophysiology*, **69**, 1091–1117.
- DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1995). Receptive-field dynamics in the central visual pathways. *Trends in Neurosciences*, **18**, 451–458.
- De Valois, R. L., Cottaris, N. P., Mahon, L. E., Elfar, S. D., & Wilson, J. A. (2000). Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity. *Vision Research*, **40**, 3685–3702.
- Dong, D. W., & Atick, J. J. (1995). Statistics of natural time-varying images. *Network: Computation in Neural Systems*, **6**, 345–358.
- Emerson, R. C., Bergen, J. R., & Adelson, E. H. (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research*, **32**, 203–218.

-
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, **160**, 106–154.
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, **195**, 215–243.
- Maffei, L., & Fiorentini, A. (1973). The visual cortex as a spatial frequency analyser. *Vision Research*, **13**, 1255–1267.
- McLean, J., & Palmer, L. A. (1989). Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat. *Vision Research*, **29**, 675–679.
- Oppenheim, A. V., Schaffer, R. W., & Buck, J. R. (1999). *Discrete-time signal processing* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Peirce, J. W. (2013). Is it just motion that silences awareness of other visual change? *Journal of Vision*, **13**(7):17, 1–10.
- Pollen, D. A., Lee, J. R., & Taylor, J. H. (1971). How does the striate cortex begin the reconstruction of the visual world? *Science*, **173**(3991), 74–77.
- Priebe, N. J., & Ferster, D. (2008). Inhibition, spike threshold, and stimulus selectivity in primary visual cortex. *Neuron*, **57**, 482–497.
- Purves, D., Paydarfar, J. A., & Andrews, T. J. (1996). The wagon wheel illusion in movies and reality. *Proceedings of the National Academy of Sciences of the USA*, **93**, 3693–3697.
- Simpson W. A., Shahani U., & Manahilov V. (2005). Illusory percepts of moving patterns due to discrete temporal sampling. *Neuroscience Letters*, **375**, 23–27.
- Suchow, J. W., & Alvarez, G. A. (2011a). Motion silences awareness of visual change. *Current Biology*, **21**, 140–143.
- Suchow, J. W., & Alvarez, G. A. (2011b). Which kinds of motion silence awareness of visual change? *Journal of Vision*, **11**(11), 734. (Abstract)
- Thompson, P. (1982). Perceived rate of movement depends on contrast. *Vision Research*, **22**, 377–380.
- Tolhurst, D. J., Tadmor, Y., & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic & Physiological Optics*, **12**, 229–232.
- Turi, M., & Burr, D. (2013). The “motion silencing” illusion results from global motion and crowding. *Journal of Vision*, **13**(5):14, 1–7.
- VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences*, **7**, 207–213.
- Watson, A. B., & Ahumada, A. J. Jr. (1983). A look at motion in the frequency domain. In J. K. Tsotsos (Ed.), *Motion: Representation and Perception* (pp. 1–10). Baltimore, MD: ACM.
- Watson, A. B., Ahumada A. J. Jr., & Farrell, J. E. (1986). Window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays. *Journal of the Optical Society of America A*, **3**, 300–307.
- Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, **18**, 293–297.
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Attention, Perception, & Psychophysics*, **63**, 1293–1313.

Appendix

Quantitative formulation of human study results

We calculated the time for dot replacement (Δt , the reciprocal amount of time it took for one dot to move to the position formerly occupied by its neighbor) when motion silencing occurs, by dividing each angular dot spacing (Δx) by the measured threshold of velocity for silencing (v_{th}). The frequency of dot replacement (f_{rep}) was computed by inverting Δt :

$$\Delta t = \frac{\Delta x}{v_{th}}; \quad (A1)$$

$$f_{rep} = \frac{1}{\Delta t}. \quad (A2)$$

As shown in figure 8b, $\log(f_{\text{rep}})$ was linearly proportional to $\log(f_{\text{flicker}})$ for all dot spacing when motion silencing occurs. In addition, as the lines indicating $\log(f_{\text{rep}})$ were almost parallel, and the intercept increased as Δx increased in a similar ratio, we could approximate the relationship between f_{rep} and f_{flicker} as follows:

$$\begin{aligned}\log(f_{\text{rep}}) &= a \log(f_{\text{flicker}}) + b; \\ \log(f_{\text{rep}}) &= \log(f_{\text{flicker}})^a + b; \\ f_{\text{rep}} &= \exp(b)(f_{\text{flicker}})^a.\end{aligned}\tag{A3}$$

a was found to be approximately 0.2725 from the human experimental data, and b could be calculated from a least squares fit.

$$b = \alpha \Delta x + \beta,\tag{A4}$$

where $\alpha = -1.4622$ and $\beta = 6.9720$.

By using equations (A1), (A2), and (A3), we could formulate v_{th} quantitatively as follows:

$$v_{\text{th}} = \Delta x \exp(b)(f_{\text{flicker}})^a.\tag{A5}$$

The threshold of the velocity for silencing is a function of dot spacing and flicker frequency.

By using equations (A4) and (A5), and by dividing v_{th} by $\Delta x \exp(b)$, we obtained the underlying scaling behavior on the human experimental data as follows:

$$\frac{v_{\text{th}}}{\Delta x \exp(\alpha \Delta x + \beta)} = (f_{\text{flicker}})^a.\tag{A6}$$

Motion silencing index

The motion silencing index (MSI) represents the degree of silencing for an arbitrary stimulus as a ratio of v to v_{th} as follows:

$$\text{MSI} = \frac{v}{v_{\text{th}}} = \frac{v}{\Delta x \exp(b)(f_{\text{flicker}})^a}.\tag{A7}$$

When there is no motion ($v = 0$), the MSI is zero. This means that there is no silencing, and humans perceive luminance changes well. As v increases, the MSI also increases. When v approaches v_{th} , the filter output approaches the threshold in the silencing map, as shown in figure 8, and motion silencing starts to occur. The degree of silencing increases as the MSI increases.

Dependence of the filter threshold on dot spacing

Predicting the threshold of the simple filter output, y , as a function of dot spacing when silencing occurs was accomplished by a least squares fit of the observed thresholds against dot spacing as follows:

$$\text{threshold, } y = \frac{\gamma}{\Delta x},\tag{A8}$$

where $\gamma = 1.086 \times 10^{12}$. The observed threshold was 7.24, 5.35, 3.62, and $3.15 (\times 10^{10})$ when dot spacing was $\pi/12$, $\pi/9$, $\pi/6$, and $\pi/5$ radians, respectively, based on human experiments.