High Frame Rates and Human Vision: A View Through the Window of Visibility

Andrew B. Watson

SMPTE Mot. Imag. J 2013, 122:18-32.
doi: 10.5594/j18266

TECHNICAL PAPER

High Frame Rates and Human Vision: A View Through the Window of Visibility

By Andrew B. Watson

There is new interest in higher frame rates for digital cinema. The frame rate and subsequent processing have a large impact on the presence of artifacts in the final presentation. The structure of these artifacts is revealed by transforming the sequence of frames into the spatiotemporal frequency domain. The visibility of these artifacts can be determined through application of a tool called the window of visibility. This is a simplified representation of human visual sensitivity to spatial and temporal frequencies. This paper describes the capture and display of movies in signal processing terms. It shows how the several steps in the process can by represented in space and time and in spatial and temporal frequency. It then introduces the window of visibility as a simplified version of the human visual contrast sensitivity function. It shows how the tool can be used to compute the lowest artifact-free frame rate under simplified circumstances. It also shows how it can be used to visualize the impact of various steps in processing the image sequence. Finally, the paper describes the influence of luminance, eccentricity, color, and eye movements on the size and shape of the window of visibility. In conclusion, the tool provides a useful aid to understanding the role of human visual sensitivity in the selection of frame rates and frame processing. The paper also provides an interactive demonstration that allows exploration of the trade-offs in movie capture and display.

INTRODUCTION

All movies consist of a sequence of still images, presented rapidly enough that, to the human observer, they convey the impression of smooth motion. Historically, movies have been captured from life at a frame rate of 24 Hz. Recently, proposals and actions have been made to increase the frame rate to higher values of 48, 60, or even 120 Hz.[1] Ultimately, the value of such proposals depends on the response of the human visual system to image sequences. The purpose of this paper is to review the movie capture, processing, and display process in the context of human visual sensitivity and thus provide a scientific underpinning for both aesthetic and economic decisions in this area.

The fundamental theory of fidelity in time-sampled stroboscopic motion displays was provided by Watson, Ahumada, and Farrell.[2] They pointed out that the effect of sampling in the time domain was to produce characteristic spectral artifacts in the frequency domain. They also noted that human visual sensitivity could be characterized by a distinct region in spatiotemporal frequency (the window of visibility) and that as long as the artifacts were outside that window, the stroboscopic display would appear smooth.

While the essentials of that theory remain sound, current interest in higher frame rates has encouraged an update and expansion of the window of visibility analysis. In particular, this paper adopts a more accurate shape for the window and a more detailed consideration of the various steps in the movie capture and display process.

MOVIES IN THE FREQUENCY DOMAIN

It is natural to think of a movie, or a segment of a movie, as a rectangular cuboid volume occupying dimensions of space (width and height, measured in degrees of visual angle) and time (measured in seconds). A point within this volume is a luminance at one point in space and time (neglecting three dimensions, or 3D, and color for simplicity). It is also possible, however, to consider the Fourier transform of this movie segment, which is also a rectangular cuboid volume but now occupying dimensions of spatial frequency (horizontal and vertical, measured in cycles/degree) and temporal frequency (measured in Hz). A point within this volume corresponds to a modulation of luminance that is sinusoidal in both space and time (a spatiotemporal sinusoid). While it is less familiar, the Fourier representation has two advantages: it allows both a simple depiction of the effects of the capture and display process and a simple representation of the limits of human visual spatial and temporal sensitivity.

MOVIE CAPTURE AND DISPLAY PROCESS

Fundamental Equations

This section describes the individual steps in the process of capture and display of moving imagery. It also shows how those steps can be viewed in the frequency domain. First, it introduces a mathematical description that provides the justification for this analysis. The complete process can be described by the equation

$$l(x,t) = \Big(\Big(\Big(\Big(\big(m(x,t) * o(x)\big) * e(t)\big) s(t)\Big) * a(t)\Big) g(t)\Big) * f(t) * h(t) \quad (1)$$

where * indicates convolution and $l(x,t)$ is the final rendered luminance image on the display. Each of the remaining functions in this expression represents one step in the capture and display process, and the details of each function are provided in the fol-

lowing sections. For simplicity, we consider here only luminance, one spatial dimension x, and the time dimension *t*. These ideas can be extended to two spatial dimensions, color, and stereo. Also we consider only temporal sampling, although spatial sampling is also involved.

While **Eq. 1** may appear complex, it is actually a simple sequence of convolutions and multiplications. Representing the process in this way enables a simple transformation to the frequency domain. This is because in the Fourier transform, multiplication and convolution are exchanged. Thus, the Fourier transform $L(u,w)$ of $l(x,t)$ is given by

$$L(u,w) = \Big(\Big(\big(\big(M(u,w)O(u)E(w)\big) * S(w)\big)A(w)\Big) * G(w)\Big)F(w)H(w) \quad (2)$$

where the uppercase function names indicate individual Fourier transforms and u and w represent spatial and temporal frequency in cycles/degree and Hz, respectively. Each convolution in **Eq. 1** has been replaced with a multiplication in **Eq. 2**, and each multiplication has been replaced with a convolution. Additional simplicity is provided because the component functions in **Eq. 1** often have simple Fourier transforms, as shown later.

Motion

To illustrate the movie capture and display process, consider a simple scene: a vertical line moving from left to right at a constant speed of *r* degrees/sec. Because the line does not vary over its vertical extent, we can ignore the vertical dimension and just consider the horizontal and temporal dimensions. Motion of the line is defined by the motion function $m(x,t)$, an impulse whose spatial location is a linear function of time:

$$m(x,t) = \delta(x - rt) \quad (3)$$

In a two-dimensional plot with axes of space (in degrees) and time (in seconds), this forms a line impulse with a slope of r^{-1} (**Fig. 1 (a)**). This example sets r = 2 degrees/sec.

The Fourier transform of this function is the spatiotemporal frequency spectrum of the moving line:

$$M(u,w) = \delta(w + ur) = \delta\left(u + \frac{w}{r}\right) \quad (4)$$

It is also an oblique line impulse, in this case with a slope of $-r$ **(Fig. 1 (b)**). The magnitude of the temporal frequency associated with each spatial frequency increases in proportion to the speed ($w = -r\,u$). The lines representing the space-time and space-time frequency images of a moving line are orthogonal.

This and succeeding figures show in the two rightmost panels the space-time image and spatiotemporal spectrum that result from all of the preceding steps. Because motion is the start of the process, the motion distribution and spectrum are reproduced in **Fig. 1 (c) and (d)**.

Blur

Despite beginning with an abstract line of zero width, a real line must have some width, if only as a result of the camera optics. We characterize this by the blur function $o(x)$ and represent it here as a Gaussian with a width b in degrees:

$$o(x) = \frac{1}{b}\exp\left(-\pi\frac{x^2}{b^2}\right) \quad (5)$$

The example **(Fig. 2 (a))** assumes b = 1/60 degree (1 arcmin), close to the limit of human visual acuity.[3] The Fourier transform of the blur function **(Fig. 2 (b))** is also a Gaussian, with a width of b^{-1} cycles/degree:

$$O(u) = \exp\left(-\pi u^2 b^2\right) \quad (6)$$

The blur function is convolved with the motion function in the space domain to yield a blurred moving line **(Fig. 2 (c))**. To obtain the spectrum of the blurred moving line, multiply $M(u,w)$ by $O(u)$ to obtain **Fig. 2 (d)**. The blur in the space domain attenuates the higher spatial frequencies in the frequency domain.

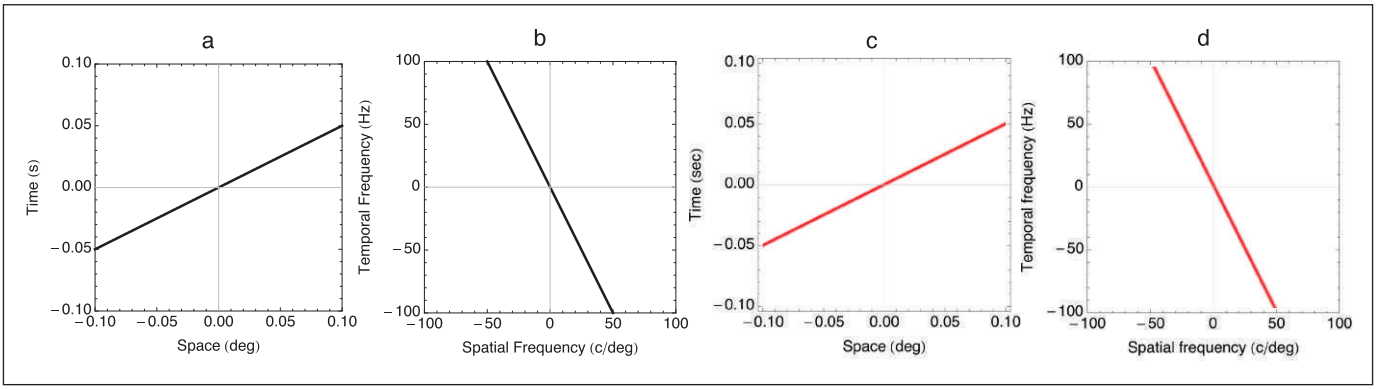


Figure 1. Image and spectrum of a moving line.

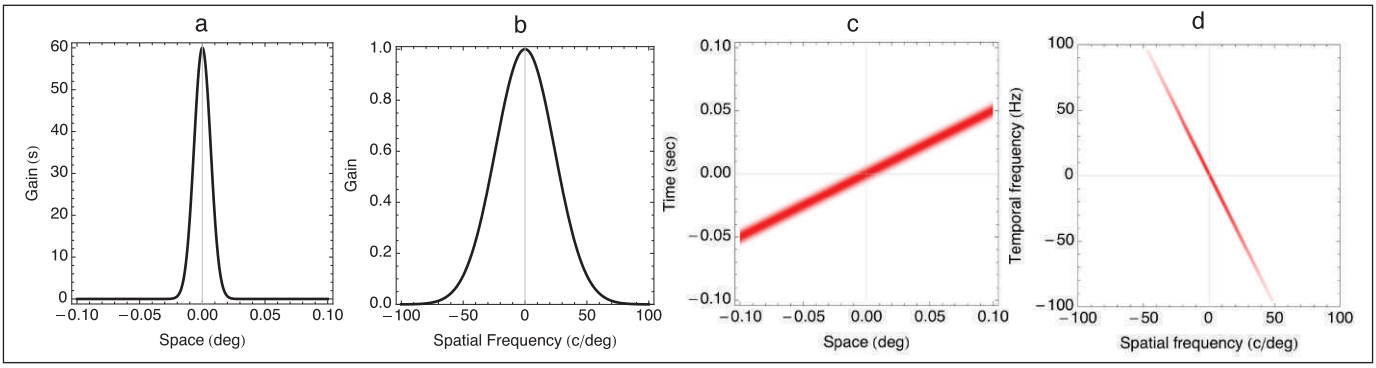


Figure 2. The blur function.

Expose

When the camera captures the moving image, it does so with a shutter (mechanical or electronic) that remains open for some fraction d of the frame duration (in the movie industry, this fraction is usually characterized by the "shutter angle" ranging from 0° to 360°). The action of opening the shutter can be represented by convolution with an exposure function $e(t)$, which we treat as a pulse of width $d\ w_s^{-1}$, where w_s is the frame rate and w_s^{-1} is the frame duration:

$$e(t) = \frac{w_s}{d}\Pi\left(\frac{t\,w_s}{d}\right) \quad (7)$$

The example **(Fig. 3 (a))** sets w_s = 48 Hz and d = 1 (a shutter angle of 360°). The formula uses Bracewell's notation Π for the Heaviside pi function, a pulse of height and width 1, centered at 0.[4] The Fourier transform $E(w)$ is a sinc function **(Fig. 3 (b))**, with a first zero at the inverse of the pulse width, or $d^{-1}\ w_s$ (in the example, 48 Hz):

$$E(w) = \mathrm{sinc}\left(\frac{d\pi w}{w_s}\right) \quad (8)$$

If the line is moving, it will suffer from motion blur for the extent of the exposure function. This blur is imposed by convolving the line **(Fig. 2 (c))** with the exposure function in the space domain **(Fig. 3 (a))**, resulting in a more blurred moving line **(Fig. 3 (c))**. In the frequency domain, multiply the previous spectrum **(Fig. 2 (d))** by $E(w)$ **(Fig. 3 (b))** to obtain **Fig. 3 (d)**. Note the additional narrowing of the spectrum. Though $O(u)$ and $E(w)$ multiply across different dimensions, they both attenuate high spatial frequencies for an object in motion.

Because the shutter is applied before sampling, and because it partially bandlimits the temporal frequency extent of the spectrum, it limits subsequent aliasing and consequently has a powerful influence over the intrusion of visual artifacts. Furthermore, although it causes spatial blur, it does so primarily for features in motion, because the band limits are in the temporal, not spatial, domain. Additionally, as shown later, the eye is less sensitive to blur when features are in motion.

Sample

The next step is to capture the image at discrete moments in time by multiplying by the sampling function $s(t)$. This is a sequence of impulses separated by time intervals of 1 frame, or w_s^{-1} seconds. This is represented using Bracewell's[4] Shah function Ш (also known as the Dirac comb):

$$s(t) = \frac{1}{w_s}\text{Ш}(t\,w_s) = \frac{1}{w_s}\sum_{k=-\infty}^{\infty}\delta\left(t - \frac{k}{w_s}\right) \quad (9)$$

Figure 4 (a) shows the example of w_s = 48 Hz. The Fourier transform S(w) is also a sequence of impulses **(Fig. 4 (b))**, but they are separated by w_s Hz:

$$S(w) = \text{Ш}\left(\frac{w}{w_s}\right) \quad (10)$$

Multiplying the blurred moving line **(Fig. 3 (c))** by the sampling function yields the result shown in **Fig. 4 (c)**. Each of the horizontal slices is a static spatial image, captured on film, on tape, or in digital form. In the frequency domain, we convolve **Fig. 3 (d)** with S(w), which causes a replication of the spectrum of the moving line

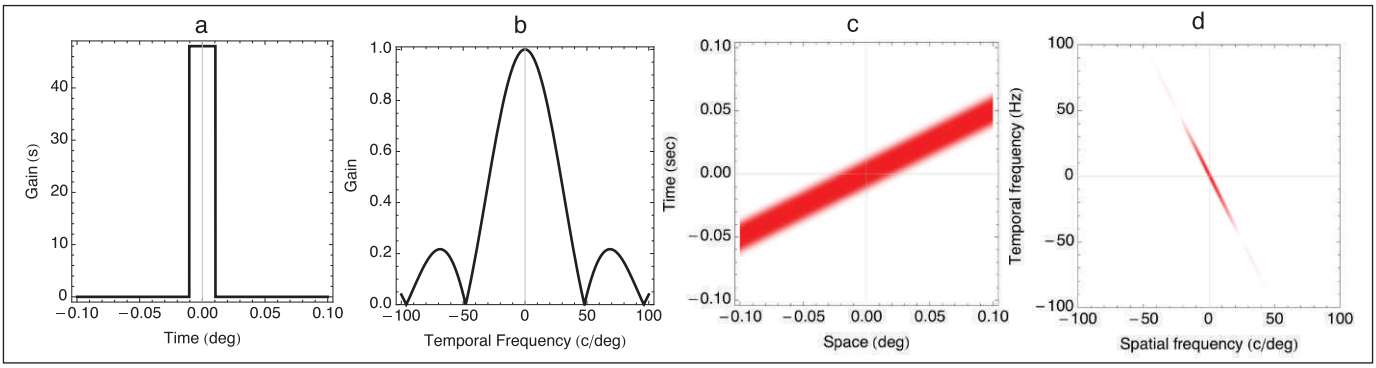


Figure 3. The expose function.

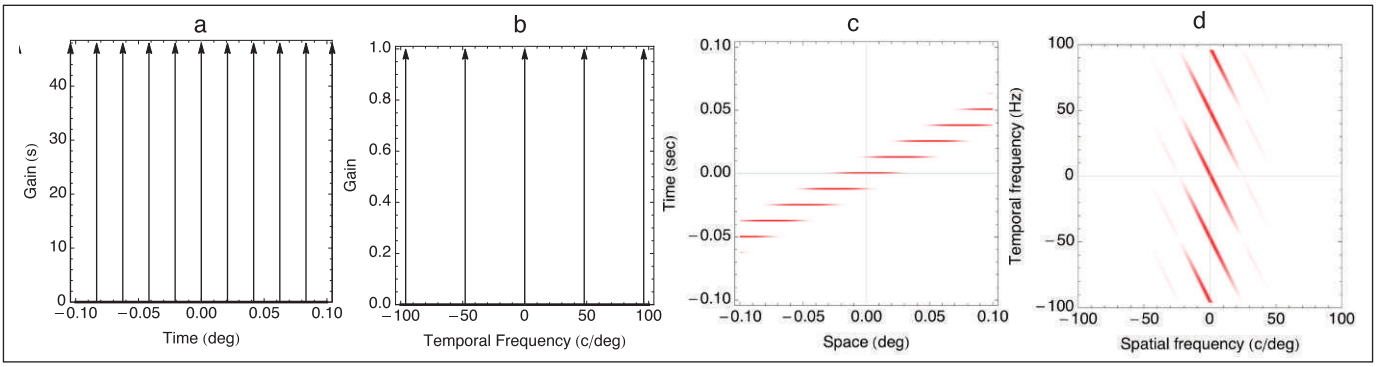


Figure 4. The sample function.

Fig. 4 (d). This replication of the spectrum is the essential reason for temporal sampling artifacts. As shown later, if these replicas lie in the visible region of the spatiotemporal frequency domain, they manifest as visible artifacts.

Filter

The remaining steps comprise the display process. First, consider a possible reduction in the display frame rate relative to the capture frame rate. Specifically, consider reductions by an integer factor k, for example, a reduction from 48 Hz, by a factor k = 2, to 24 Hz. Prior to such downsampling, it is typical to apply a prefilter to reduce aliasing. For illustration, assume the prefilter is an average of k adjacent frames. In that case, the prefiltering function $a(t)$ can be represented as a pulse of width k frames:

$$a(t) = \frac{w_s}{k}\Pi\left(\frac{tw_s}{k}\right) \quad (11)$$

Figure 5 (a) shows the example of k = 2. The Fourier transform of this pulse is a sinc function **(Fig. 5 (b))**, with a first zero at the inverse of the pulse width, or $k^{-1}\, w_s$ (in the example, 24 Hz):

$$A(w) = \mathrm{sinc}\left(\frac{k\pi w}{w_s}\right) \quad (12)$$

Convolving the samples **(Fig. 4 (c))** by the filter yields the filtered samples **(Fig. 5 (c))**. In the frequency domain, we multiply the spectrum of the samples **(Fig. 4 (d))** by the filter transform to yield the filtered spectrum **(Fig. 5 (d))**. The filtering has removed most spectral artifacts but also has produced additional narrowing of the spectrum (blurring).

Downsample

Filtering is followed by a second sampling operation, using the downsampling function $g(t)$, which is a sequence of pulses at intervals of k w_s^{-1} **(Fig. 6 (a))**:

$$g(t) = \frac{k}{w_s}\text{Ш}\left(\frac{tw_s}{k}\right) \quad (13)$$

The Fourier transform is also a sequence of pulses but at intervals of $k^{-1}\, w_s$ **(Fig. 6 (b))**:

$$G(w) = \text{Ш}\left(\frac{kw}{w_s}\right) \quad (14)$$

Multiplication of the filtered line **(Fig. 5 (c))** by the downsampling function produces the new samples **(Fig. 6 (c))**. Note the slight broadening of each sample in the space dimension because of the averaging of adjacent frames of an object in motion. This is an additional form of motion blur, similar to that in the spectrum after filtering **(Fig. 5 (d))**. The new spectrum **(Fig. 6 (d))** is obtained by convolution of the spectrum in **Fig. 5 (d)** with the downsampling spectrum **(Fig. 6 (b))**.

Flicker

In traditional movie projection, each frame is presented multiple times (usually 2 or 3 fields) to avoid flicker. We implement this with the flicker function $f(t)$, which consists of a sequence of n impulses, at intervals of $k\, n^{-1}\, w_s^{-1}$:

$$f(t) = \frac{1}{n}\sum_{k=1}^{n}\delta\left(t - \frac{2k-n-1}{2nw_s}\right) \quad (15)$$

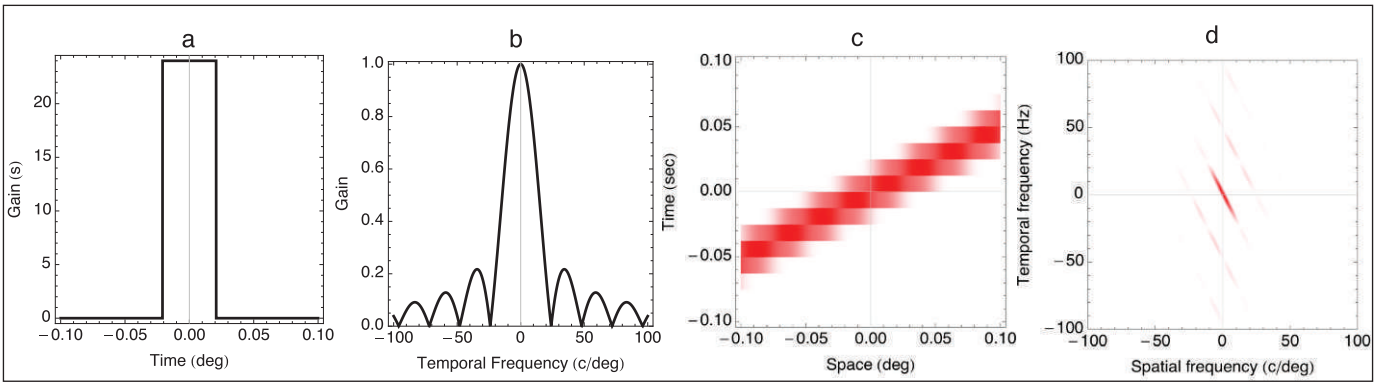


Figure 5. The filter function.

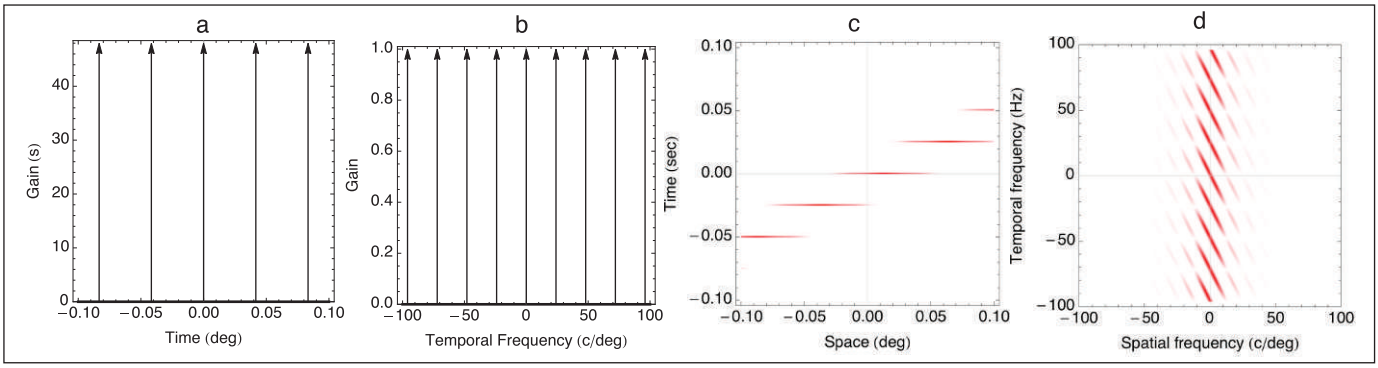


Figure 6. The downsample function.

The Fourier transform $F(w)$ depends on n, but in general it tends to attenuate higher frequencies and has a first zero at $k^{-1}\ w_s$:

$$F(w) = \frac{1}{n}\sum_{k=1}^{n}\exp\left(-i\pi w\frac{2k-n-1}{nw_s}\right) \quad (16)$$

An example for n = 3 is shown in **Fig. 7 (a) and (b)**. The first zero of the spectrum is at 24 Hz. The new samples **(Fig. 6 (c))** are convolved with $f(t)$ to produce the multiple fields **(Fig. 7 (c))**. In the frequency domain, multiply the spectrum **(Fig. 6 (d))** by $F(w)$ to obtain **Fig. 7 (d)**. This attenuates much of the spectral replicas.

Hold

The final step in the process is that each field is displayed for a fraction p of the field interval. This is represented by the hold function h(t), which is represented here by a pulse of width $p\ k\ n^{-1}\ w_s^{-1}$:

$$h(t) = \frac{nw_s}{p}\Pi\left(\frac{t\,nw_s}{p}\right) \quad (17)$$

This reflects the actual display duration of each field and is illustrated in **Fig. 8 (a)** for the case of p = 3/4. The Fourier transform $H(w)$ is another sinc function,

$$H(w) = \mathrm{sinc}\left(\frac{p\pi w}{nw_s}\right) \quad (18)$$

with a first zero at $p^{-1}\ k^{-1}\ n\ w_s$, which in this example is at 96 Hz **(Fig. 8 (b))**.

To obtain the final luminance distribution on the display $l(x,t)$ **(Fig. 8 (c))** we convolve the fields **(Fig. 7 (c))** with the hold function **(Fig. 8 (a))**. To obtain the final display luminance spectrum $L(u,w)$ **(Fig. 8 (d))**, we multiply the spectrum **(Fig. 7 (d))** by $H(w)$ to obtain **Fig. 8 (b)**.

The value of p depends on the technology employed. For liquid crystal displays, for example, the value is near 1. For cathode ray tube (CRT) televisions, it is much briefer. For film projection, it is typically 0.5, while for digital cinema projection it is controllable and may be near to 1.[5] Longer durations produce a brighter image, and attenuate some sampling artifacts, but they produce more motion blur when observers track moving edges.[6]

In the example, the uncorrupted original signal is depicted in spectral terms in **Fig. 1 (d)**. In comparing this to the final spectrum in **Fig. 8 (d)**, the latter exhibits artifacts embodied in several spectral replicas. Several of the preceding steps have ameliorated these replicas, but some persist. The visibility of these fragments determines the quality of the rendered movie. To determine that visibility, we turn to a consideration of human spatial and temporal contrast sensitivity in the frequency domain.

THE WINDOW OF VISIBILITY

The previous section described the sampling and display process for movies and showed how the processing artifacts could be depicted in the frequency domain. This section covers human visual sensitivity in the frequency domain to predict sensitivity to these artifacts.

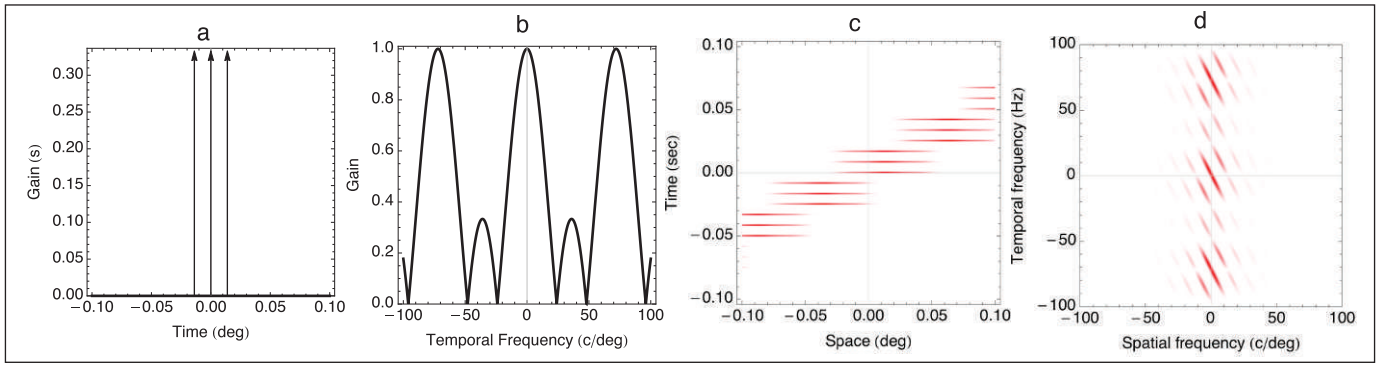


Figure 7. The flicker function.

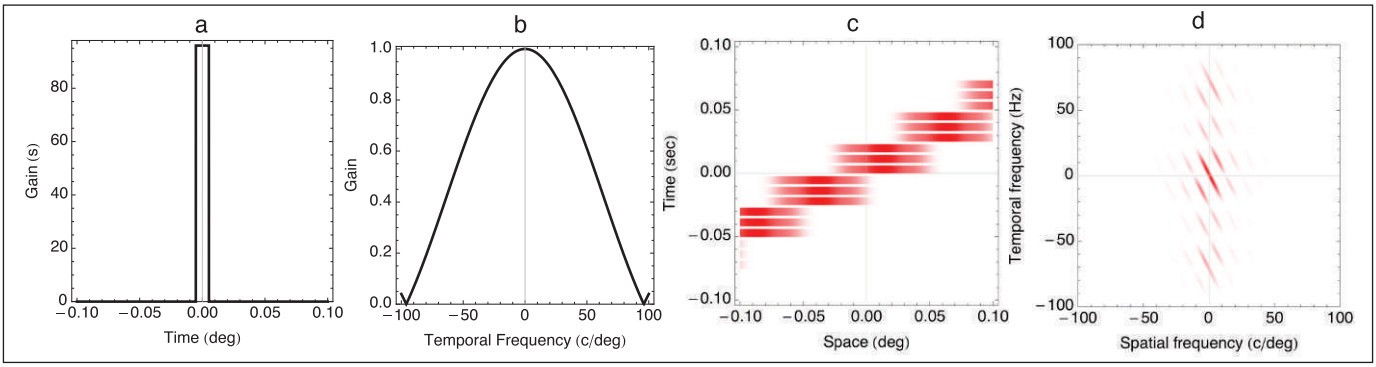


Figure 8. The hold function.

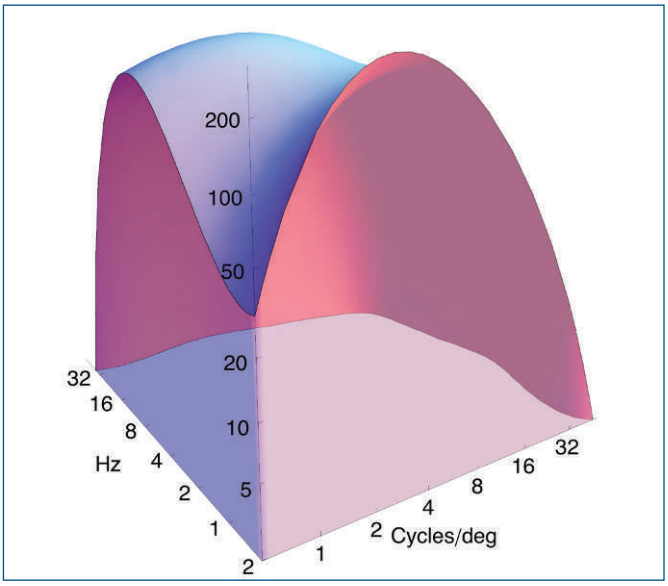


Figure 9. An STCSF.

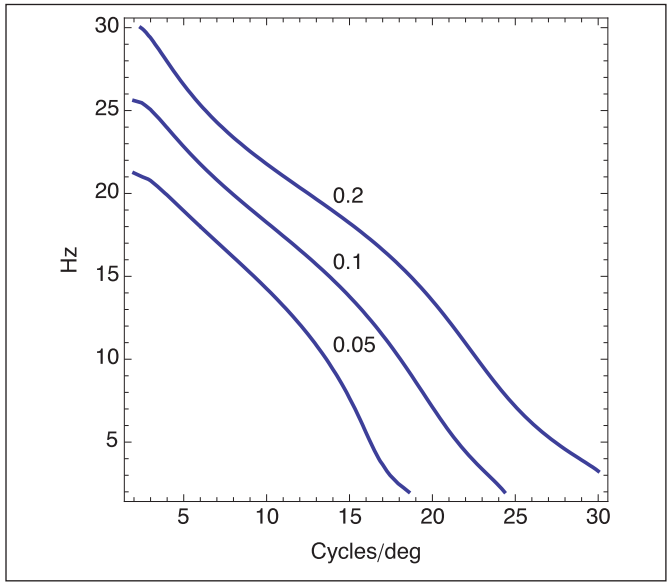


Figure 10. Isosensitivity curves of the STCSF.

Schade[7] introduced the idea of using the modulation transfer function to characterize human visual sensitivity to spatial detail. In the work of van Nes and Bouman,[8] this was expressed as the spatial contrast sensitivity function (SCSF) that specified the minimum contrast required to detect a sinusoidal luminance grating of a specified spatial frequency in cycles/degree. (Contrast is defined as the luminance variation divided by the average luminance.) In a similar fashion, de Lange[9] introduced the notion of the temporal contrast sensitivity function (TCSF), specifying the minimum visible contrast of a temporal sinusoid. Robson,[10] using targets that were sinusoidal in both space and time, collected data to specify a more general spatiotemporal contrast sensitivity function (STCSF). **Figure 9** shows a version of the STCSF created by fitting Robson's data with a fourth-order polynomial surface.

This surface, plotted in logarithmic coordinates, shows that for Robson's conditions (20 candela per square meter, or cd/m^2; 2.5 degree square; indefinite duration), contrast is visible within a range of spatial and temporal frequencies, up to a spatial acuity limit just above 32 cycles/degree, and a temporal acuity limit just above 32 Hz. Let us refer to these acuity limits as u_0 and w_0 and define the window of visibility as the region of spatiotemporal frequency that is visible to the human observer. It is given graphically by the floor of the surface plot in **Fig. 9**. However, that plot is in logarithmic coordinates. A more useful picture of the window is given by a contour plot in linear coordinates, as shown in **Fig. 10**.

The three curves indicate isosensitivity contours for contrast thresholds of 0.05, 0.1, and 0.2. The outermost limit of the window would be at a contrast threshold of 1, but the data of Robson do not permit us to extrapolate to that point. (However, the rapid decline of sensitivity at higher frequencies means that the outer limit will not be far beyond the 0.2 threshold.) The curves suggest that to a first approximation, the window of visibility has the shape of a diamond, as shown diagrammatically in **Fig. 11**.

As a simplification, regard everything within the window as visible and everything outside the window as invisible. The boundaries of the window are given by the acuity limits of u_0 cycles/degree and w_0 Hz, as shown by the green diamond in **Fig. 11**. This simplified

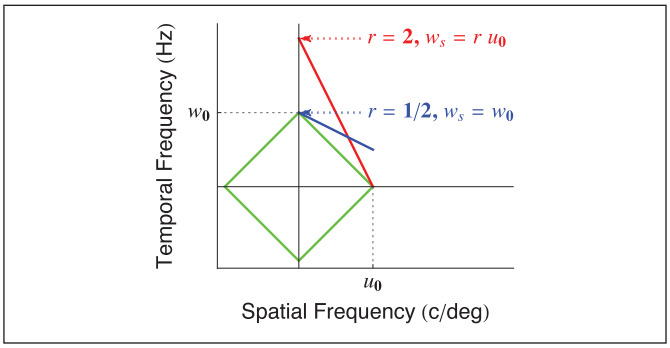


Figure 11. Diagrammatic representation of the window of visibility. Red and blue lines are spectral replicas at the critical frame rate.

representation allows us to predict, in the simplest case, when temporal sampling artifacts become visible.

Consider **Fig. 4 (d)**. This shows the spectral replicas that result for a particular speed of motion r and frame rate w_s. If the replicas fall outside of the window of visibility, they will not be seen, and the motion will appear smooth (neglecting, for the moment, the subsequent rendering and display steps that may alter artifact visibility). The slope of the replicas is $-r$. If we consider just the first replica and ask when it will enter the window visibility, the answer is given by the red and blue lines in **Fig. 11**, which illustrate fragments of the first spectral replica in two speed regimes. For slow speeds (blue), the replica will abut the window at its upper apex when $w_s = w_0$. For higher speeds (red), the replica will abut the window at its right apex and will intersect the vertical axis at $w_s = r\,u_0$. The boundary between slow and fast speeds occurs when the slope of the replica equals the slope of the window, that is, when $r = w_0/u_0$. This is called the critical speed r_0 degrees/sec. To summarize, the critical frame rate w_c at which motion will appear smooth is given by

$$\begin{aligned} w_c &= w_0 \quad r \le r_0 \\ &= r u_0 \quad r > r_0 \end{aligned} \qquad (19)$$

In these two speed regimes, the replicas enter the window in different regions. For speeds below the critical speed, the replicas enter in the region of low spatial and high temporal frequency. This is typically described as "flicker." For speeds greater than the critical speed, replicas enter the window at high spatial, low temporal frequencies. These artifacts are typically experienced as multiple images.

An earlier report on the window of visibility[2] tested this general theory by measuring the critical sampling rate for a moving line for two observers. This was done for several motion speeds ranging up to 16 degrees/sec. The data are reproduced in **Fig. 12**. They show a steady increase in critical frame rate as speed increases, extending for one observer to more than 200 Hz. **Equation 19** has been fit into these data by fixing w_0 at the value of the critical rate for the slowest speed and estimating u_0. The values of the two parameters are shown in each panel.

This simplified prediction shows that the critical frame rate can be well above the temporal acuity limit. As discussed later, there are ways to ameliorate this requirement, but in general the critical frame rate is higher than the human temporal resolution limit w_0.

EXAMPLES

This section provides a few examples that illustrate some effects of steps in the capture and display process. Each figure shows the spatiotemporal spectrum of the result (similar to **Fig. 8 (d)**) and, for reference, the window of visibility as a green diamond, assuming u_0 = 30 cycles/degree and w_0 = 30 Hz. The examples that follow vary the parameters r, w_s, b, d, n, and p. Where not stated otherwise, the parameters are fixed at values typical of movie capture and display (r = 4 degrees/sec, w_s = 24 Hz, b = 1/60 degree, d = 0.25, k = 1, p = 0.75, n = 3).

It is not possible to show all possible combinations of possibilities and parameters, but for those interested in exploring this further, an interactive demonstration is introduced later.

Speed

Recall that the spectrum has a slope of –r and that the replicas are spaced at intervals of w_s. **Figure 13** shows the effect of speeds of 0, 0.25, 1, and 4 degrees/sec, captured at w_s = 24 Hz. As speed increases, more of the artifact enters the window of visibility and visual quality declines.

Blur

For the blur example, the speed is set to 4 degrees/sec and the blur is varied over values of 1, 4, and 16 arcmin. This is a spatial blur imposed on the imagery before temporal sampling. **Figure 14** shows that with very strong preblurring, it is possible to remove most spectral replicas, but this comes at the cost of some visible

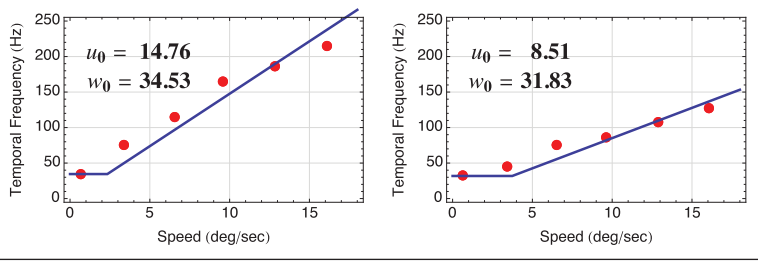


Figure 12. Critical frame rate for a moving line for two observers.

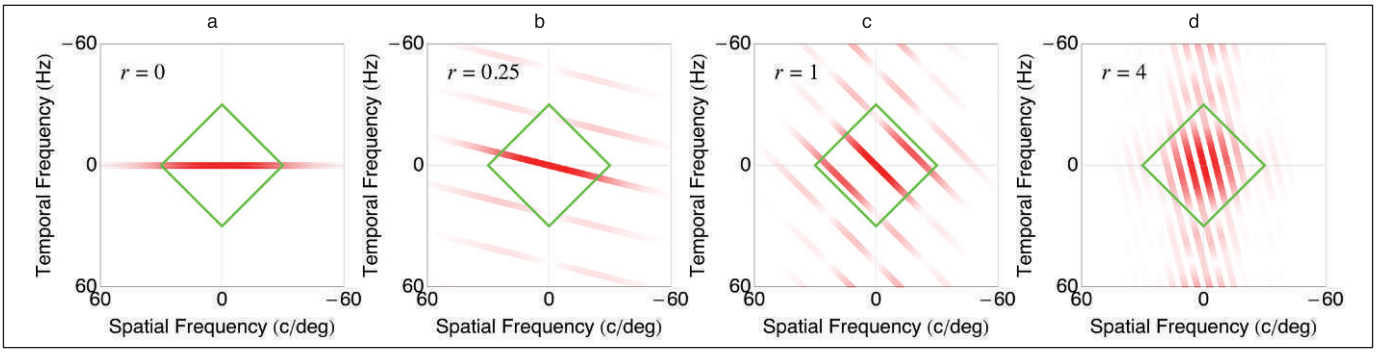


Figure 13. Effect of speed on the spectrum of a sampled moving line.

blur (compare **Fig. 14 (a) and (d)**). This example also illustrates that temporal sampling artifacts are most severe for high spatial frequencies. The use of preblurring to avoid motion artifacts in computer graphics animations relies on this idea.[11]

Shutter Angle

Shutter angle controls the amount of motion blur that is imposed on moving imagery before it is time sampled. Because it removes high spatial frequencies, and their corresponding high temporal frequencies, it can reduce sampling artifacts. **Figure 15** shows the spectrum for the moving line captured with shutter angles of 90° and 360° (d = 0.25 and 1). The larger value attenuates the spectral replicas but also results in some spatial blur. As noted previously, computer graphic animations lack an actual shutter and may contain very high spatial frequencies, so they may require preblurring to simulate the beneficial effects of motion blur.[11]

Flicker

The spectrum of unsampled stationary line (r = 0 degrees/sec) is shown in **Fig. 16 (a)**. The energy all lies on the axis at 0 Hz. The fading of the line at high spatial frequencies is due to the spatial blur of the line, specified by the parameter b. When sampled at w_s = 24 Hz, the typical rate for movie capture, the spectral replicas, at intervals of 24 Hz, intrude into the window and lead to visible flicker (**Fig. 16 (b)**). If each frame is then exposed three times (n = 3), using the flicker function, the replicas are removed and the sampled and unsampled spectra are identical **(Fig. 16 (c))**. However, when the speed is 4 degrees/sec **(Fig. 16 (d))**, the tripling of each frame does not remove the replicas and artifacts are visible **(Fig. 9 (e))**. These examples set the hold parameter to p = 0.25, because a larger value also attenuates flicker, as noted later.

Hold

The hold parameter p is the fraction of the display field duration that is illuminated. **Figure 17 (a)** shows the result when r = 0 and n = 1 (no duplicate display fields) and p = 0.1, showing visible artifacts. Increasing p to 1 removes the artifacts **(Fig. 17 (b))**. This is not surprising, because the line now appears continuously on the display. In **Fig. 17 (c)**, the speed is increased to 4 degrees/sec and p is returned to a value of 0.1, with substantial visible artifacts. Increasing p to 1 attenuates the replicas at high spatial frequencies but does little to remove the artifacts from within the window of visibility.

Under standard film projection conditions (n = 3), changing p from 0.1 to 1 has little effect on the visible artifacts. It has a proportional effect on screen brightness, for which reason it is desirable to have a value close to the maximum of 1. However, a virtue of small values is that they minimize the motion blur that results when the eye tracks a moving target.[6]

Frame Rate

The simplest way to remove sampling artifacts is to increase the frame rate. **Figure 18 (a)** shows the unsampled spectrum, and

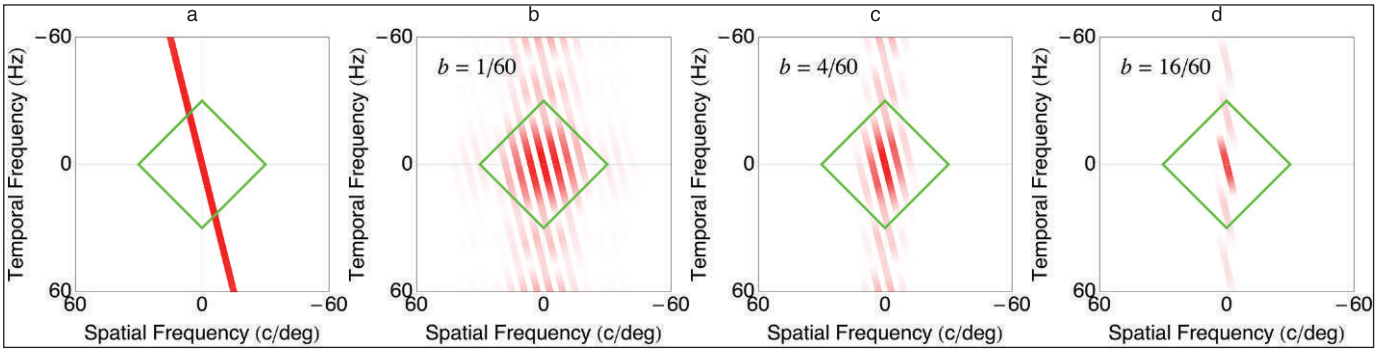


Figure 14. Effect of blur on the spectrum of a sampled moving line.

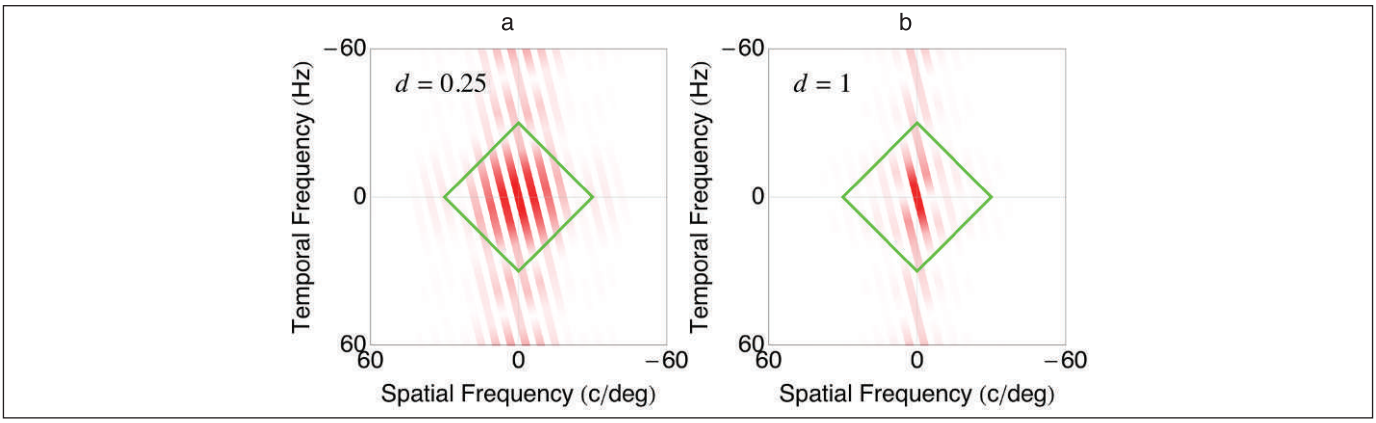


Figure 15. Effect of shutter angle on the spectrum of a sampled moving line.

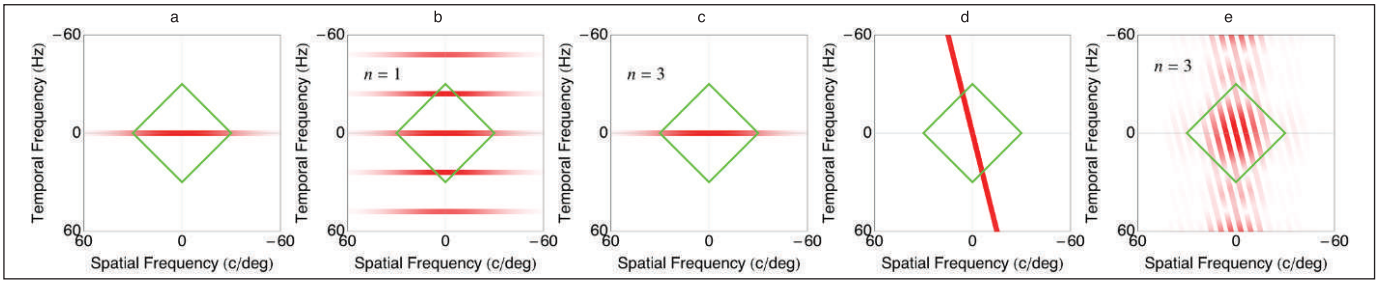


Figure 16. Effect of flicker on the spectrum of a sampled line.

Fig. 18 (b) shows the sampled spectrum (n = 1, d = 1/4, p = 1/4) with a multiplicity of visible artifacts. These can be somewhat ameliorated through flicker, exposure, and hold, as shown in **Fig. 18 (c)** (n = 3, d = 1, p = 1). However, they are moved entirely outside the window when the frame rate is increased to 120 Hz, as shown in **Fig. 18 (d)**. This is the critical frame rate as specified by **Eq. 3**. This has the additional advantage of preserving the sharpness that is otherwise lost through the various forms of motion blur.

FACTORS AFFECTING THE WINDOW OF VISIBILITY

The preceding section showed how the visibility of sampling and rendering artifacts could be explained by their position relative to the window of visibility. However, the window is not of fixed size; rather, it and the underlying contrast sensitivity function depend upon a variety of conditions. To predict the visibility of frame rate artifacts, it is important to understand the size of the window under the conditions in question.

Luminance

Temporal contrast sensitivity depends upon display luminance.[9,12] The temporal acuity limit w_0, also known as the critical flicker frequency (CFF), increases linearly with luminance, as shown in **Fig. 19**.[13,14] The red and blue data are from Tyler and Hamer[13] for a foveal target of 0.5 degree diameter and a 5.7 degree target at an eccentricity of 35 degree temporal, respectively. On very bright backgrounds, it may attain values as high as 90 Hz. Their retinal illuminance values have been transformed to luminance using the

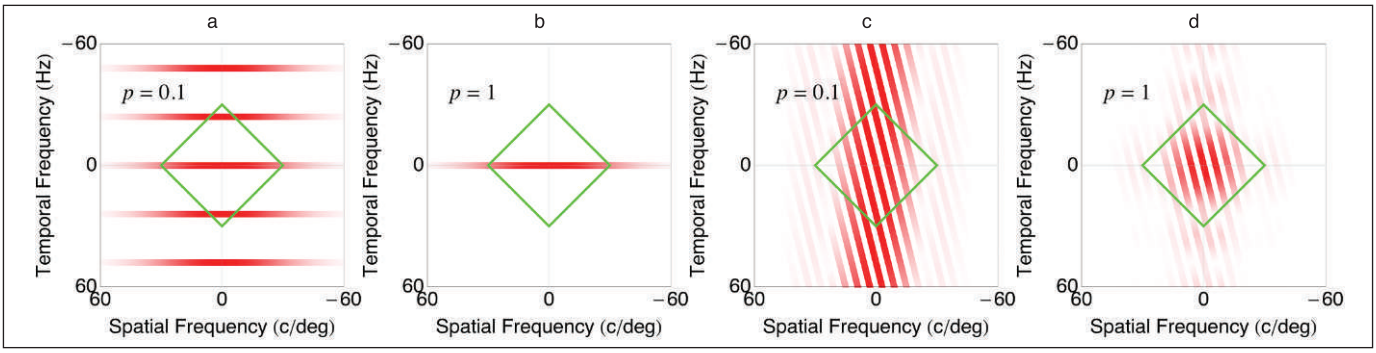


Figure 17. Effect of hold on the spectrum of a sampled line.

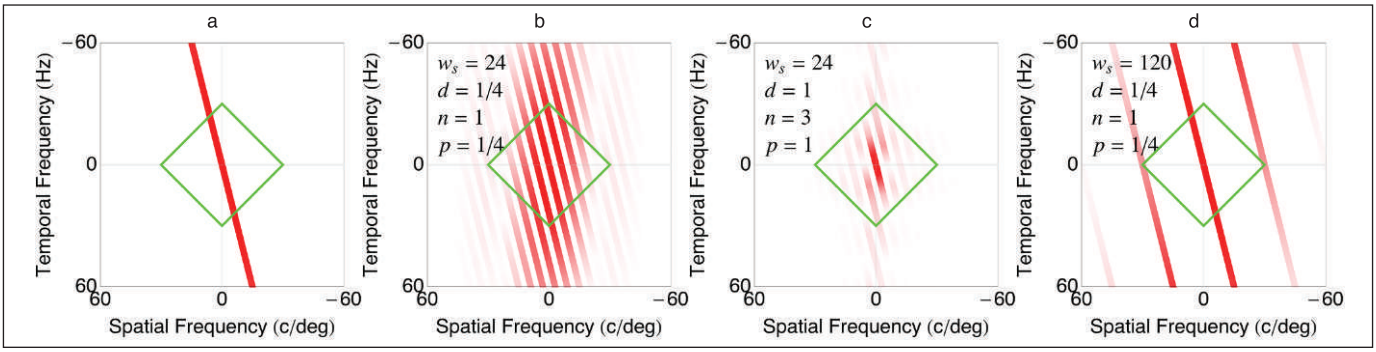


Figure 18. Effect of frame rate on the spectrum of a sampled moving line.

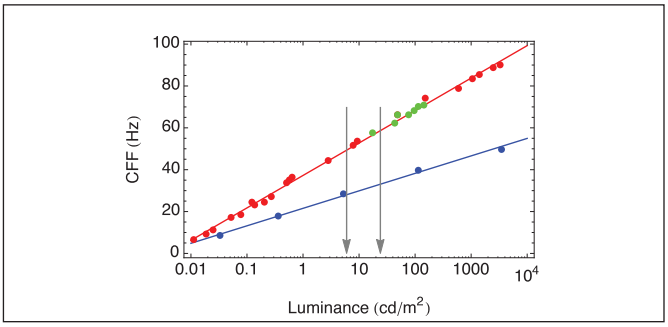


Figure 19. Temporal acuity limit (CFF) as a function of luminance.

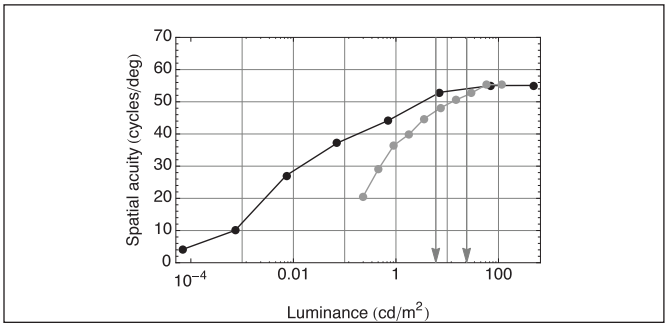


Figure 20. Spatial acuity as a function of background luminance.

formula of Watson and Yellott,[14] assuming a viewing distance of 2 screen heights and monocular viewing.

The green points are data from Farrell et al.,[15] as plotted by Barten,[16] for CFF of a CRT display with a width of 30 degree. Because the green data points are for actual judgments of fusion of a large, centrally viewed display, their agreement with the red curve suggests this is the curve of practical relevance. The arrows indicate upper and lower limits of current standards for digital cinema screen brightness (12 and 48 cd/m^2, divided by two to represent average rather than peak brightness). For the higher standard, the CFF is around 60 Hz.

The spatial acuity limit u_0 also depends on display luminance, as shown in **Fig. 20**. The black points are grating acuities from figure 8 of van Nes and Bouman[8]; the gratings were 4.5 x 8.25 degree, at 525 nm, and were viewed monocularly through a 2mm artificial pupil. The gray points are monocular decimal letter acuities measured by Rabin,[17] inverted and scaled by 38 to match the grating acuities at the highest luminances. Most of the increase in acuity occurs at very low (scotopic) levels, and acuity saturates at luminances of around 100 cd/m^2. It is likely that with a natural pupil grating acuity would have declined more rapidly, because lower luminances would result in a larger pupil and lower optical quality due to aberrations. This suggestion is reinforced by letter acuities, measured with a natural pupil, that fall more rapidly with declining luminance. The gray arrows are reproduced from **Fig. 19** and again show digital cinema standards. At the higher standard, acuity is a little more than 50 cycles/degree.

Color

Both spatial and temporal acuity limits are substantially lower for color than for luminance modulation. Mullen[18] showed that the SCSF for isoluminant chromatic gratings was shifted by as much as a factor of 3 to lower spatial frequencies. Likewise Varner[19] showed similar shifts for the TCSF. Both of these results suggest that window of visibility is much smaller for color signals and that frame rate artifacts due to luminance are more visible than those for color. This result is of more practical relevance to coding and compression than to capture, unless luminance and color can be captured at different frame rates.

Distance from Fixation

Perhaps the most powerful influence on the size of the window of visibility is position within the visual field. Visual spatial resolution is ultimately limited by the density of retinal ganglion cells (RGCs) that send their signals up the optic nerve to the brain. **Figure 21** shows a surface plot of the Nyquist limit imposed by RGCs over the central 30 degrees of the visual field, as well as a slice through this surface along the temporal meridian.[20] From a high above 80 cycles/degree, the curve falls rapidly by about a factor of 4 in the space of 4 degrees. The significance of this result is that the spatial extent of the window shrinks rapidly as we move away from the point of fixation. What about the temporal extent? Although there is some change in temporal behavior with eccentricity (**Fig. 19**), there is little or no systematic reduction in the temporal acuity.[21] Thus, while the aspect ratio w^0/u_0 of the window and equivalently

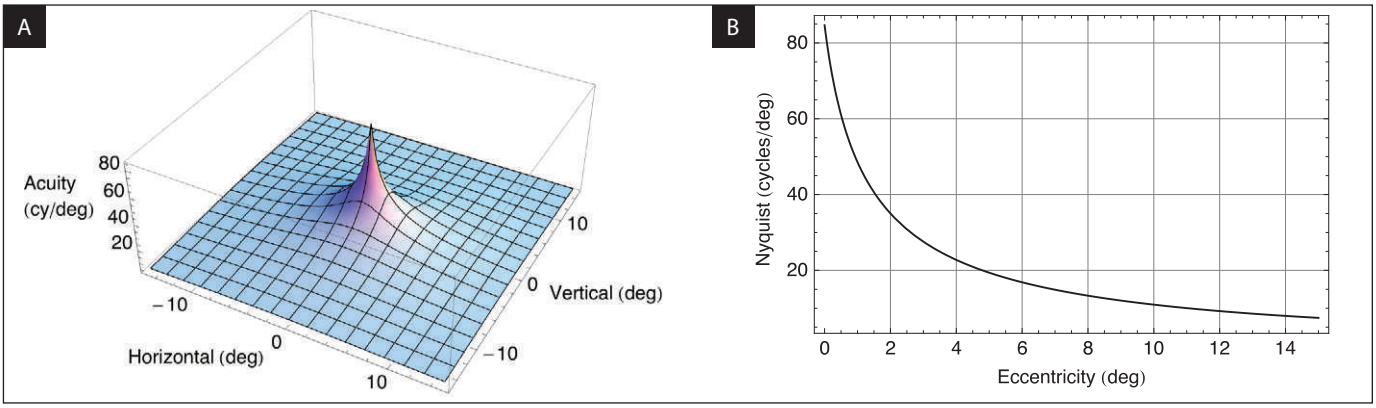


Figure 21. Spatial acuity as a function of retinal location.

the critical speed r_0 is approximately 1 degree/sec at the fovea, in the periphery it is much larger. This result implies that sampling artifacts due to motion in the periphery are likely to be less visible than those in the fovea and that they are more likely to manifest as flicker rather than as multiple images or other spatial artifacts.

Eye Movements

The descriptions of both the rendered movie and the human visual sensitivity have been expressed in retinal spatial coordinates. In simple terms, they assume that the eye remains stationary and fixated on the center of the display. In practice, the eye is always in motion. In this context, the most significant eye movements are saccades and smooth pursuit movements. The latter occur when the eyes track the motion of a feature, in which case the retinal speed of that feature is reduced nearly to 0. Saccades are very rapid eye movements that move fixation from one feature to another. We can depict the influence of eye movements by representing the window of visibility in display coordinates, transforming it geometrically in the appropriate way. In the coordinates of spatial and temporal frequency, motion is represented by a shearing transform in the temporal frequency domain. The temporal frequency associated with each spatial frequency u is increased by $-r\,u$; that is, (u,w) is transformed to $(u,w - r\,u)$. When the window of visibility consists of a diamond shape, with vertices at $(u_0,0)$, $(0,w_0)$, $(-u_0,0)$, and $(0,-w_0)$, the transformed result is a parallelogram with coordinates of $(u_0,-r)$, $(0,w^0)$, $(-u_0,r)$, and $(0,-w_0)$. Several cases are illustrated in **Fig. 22**.

Figure 22 (a) illustrates the case of motion at 2 degrees/sec, sampled at 30 Hz, with the eye stationary (d = 1/4, n = 1, h = 1/4, b = 1/60). Several replicas lie well within the window, yielding a highly visible artifact. When the eye tracks the motion (**Fig. 22 (b)**), the window is transformed in such a way that the replicas are largely excluded and the artifacts are diminished. **Figure 22 (c)** shows a case in which the motion is 0, the frame rate is 30 Hz, and the eye is stationary. No replica intrudes within the window. But if the eye moves at 2 degrees/sec, the replicas fall within the window and artifacts are visible. Thus, we see that eye movements can both reduce and enhance the visibility of artifacts. In general, smooth pursuit eye movements enhance the quality of the features that are tracked but may reduce the quality of other features that are not tracked. But features that are not tracked are likely to be displaced from fixation, and as noted previously, the window there is smaller in the spatial frequency dimension.

DISCUSSION

It is widely understood that effective depiction of smooth motion in stroboscopic displays such as movies and video depends upon the response of the human visual system. However, vague concepts such as "persistence of vision" and "apparent motion" do not lead readily to quantitative predictions or prescriptions. In contrast, the window of visibility provides a clear and simple explanation for the qualitative presence or absence of visible artifacts and may, in certain circumstances, provide quantitative prescriptions.

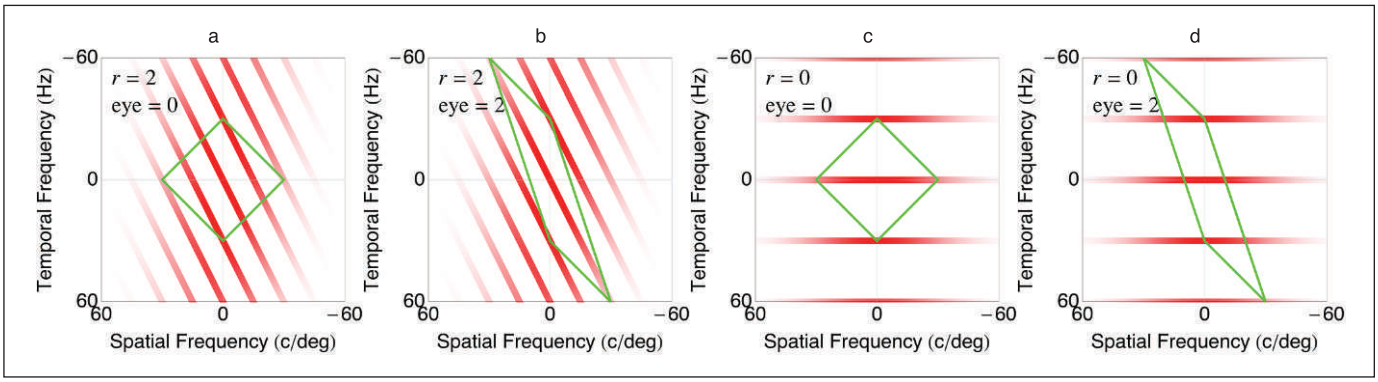


Figure 22. Effect of eye movements on the spectrum of a sampled moving or stationary line.

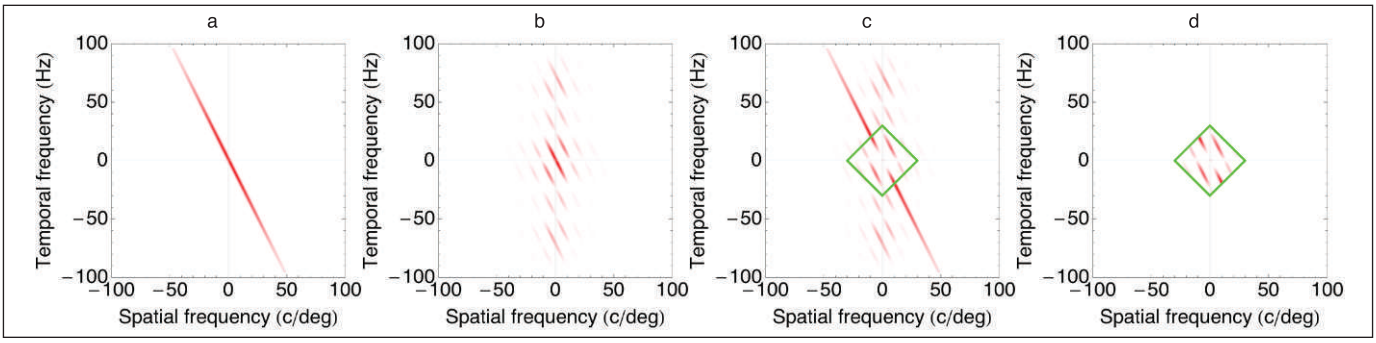


Figure 23. Error spectrum and the window of visibility.

This paper has shown that a segment of movie can be represented in the spatiotemporal frequency domain, and compared to the window of visibility, to allow understanding of the visibility of sampling and rendering artifacts. The illustrations used a simple movie: that of a vertical line moving horizontally at a fixed speed across the screen. While this movie may win no awards, it provides an example of image movement that is as simple as possible. In a real movie, many lines move in many directions at a variety of speeds, but the principles remain the same as those of a single line, because the complex movie may be regarded as the sum of many simple ones and the Fourier transform is linear.

Because this paper's treatment of this problem emphasizes simplicity, it is useful to consider what a less simple but more comprehensive treatment would look like. First, it would begin with a set of representative segments of real movie footage, containing varying amounts and kinds of motion. To explore the sampling and rendering artifacts, it would be best if the original "reference" footage was an exact copy of the wavefront reaching the human eye. Because this is not possible, each segment should be as high in quality as possible in terms of spatial and temporal resolution. Each segment would be subjected to a particular sequence of capture, sampling, and rendering operations, sometimes called a hypothetical reference circuit (HRC), to yield a set of "processed" segments. Each segment would then be transformed into the frequency domain (now 3D, with two spatial and one temporal dimensions). The difference between these spectra could then be computed, and rather than being gated by the window of visibility, it would be weighted by the full STCSF, and pooled over all dimensions, to compute an overall metric of artifact visibility. Further embellishments to this calculation might include masking and various nonlinearities. Finally, some average of the results for the several segments might serve as a global measure of the quality of the HRC. In general form, this calculation is similar to that employed in objective metrics for quality of compressed digital video.[22,23]

In the application of the window of visibility, we have been mainly concerned with artifact additions (the intrusion of replicas within the window) rather than subtractions, such as the attenuation of the spectrum due to blurring. The more complete treatment described previously would account for both additions and subtractions. But subtractions can also be appreciated with a simple analysis, by considering the difference between the reference and the processed signals, relative to the window of visibility. This is illustrated in **Fig. 23**, which reproduces in **Fig. 23 (a) and (b)** the reference and final spectra from **Figs. 1 (d) and 8 (d)**, respectively. These are the spectra before sampling or rendering and the final result. **Fig. 23 (c)** shows the difference between these two spectra, along with the window. **Figure 23 (d)** removes fragments of the error spectrum that lie outside the window. This is the spectrum of the visible error, and it includes both additions (replicas) and subtractions (blur).

Many functions used earlier for illustration can be replaced by other functions. For example, the blur function may not be a Gaussian, and the exposure, filter, and hold functions need not be rectangular pulses. The flicker function could also have a different form. Modifying these functions may provide useful benefits in reducing the visibility of sampling artifacts or otherwise enhancing quality. Whatever their precise form, however, they must play essentially the same role assigned to them in the discussion of the capture and display process.

Aliasing and Nyquist Sampling

In standard signal processing theory, it is well known that after periodic sampling of a signal at frequency w, signal components above the Nyquist frequency of $w/2$ are aliased into the spectrum of the signal. To guard against aliasing, the signal must be prefiltered before sampling to remove components above the Nyquist frequency. That ensures that the spectral replicas do not overlap the original spectrum. A postfilter is then typically used to remove the replicas if the signal must be returned to a continuous analog form.

Because temporal frequency is the product of spatial frequency and speed, and because there is no practical limit to the speed at which features can travel in the optical signal, there are effectively no bounds to its temporal frequency spectrum. Thus, no temporal sampling frequency (frame rate) will ever be sufficient to capture optical reality. However, the window of visibility defines the region of spatiotemporal frequency that we would like to preserve from that signal, because it is the region to which we are sensitive. (Think of it as a postfilter that reconstructs a continuous signal from the stroboscopic samples.) Accordingly, a prefilter designed to preserve just the window, followed by sampling at $2w_0$, would preserve all visible information in the optical signal.

Unfortunately, such a filter is not causal, is not separable in space and time, and employs both positive and negative weights; as such,

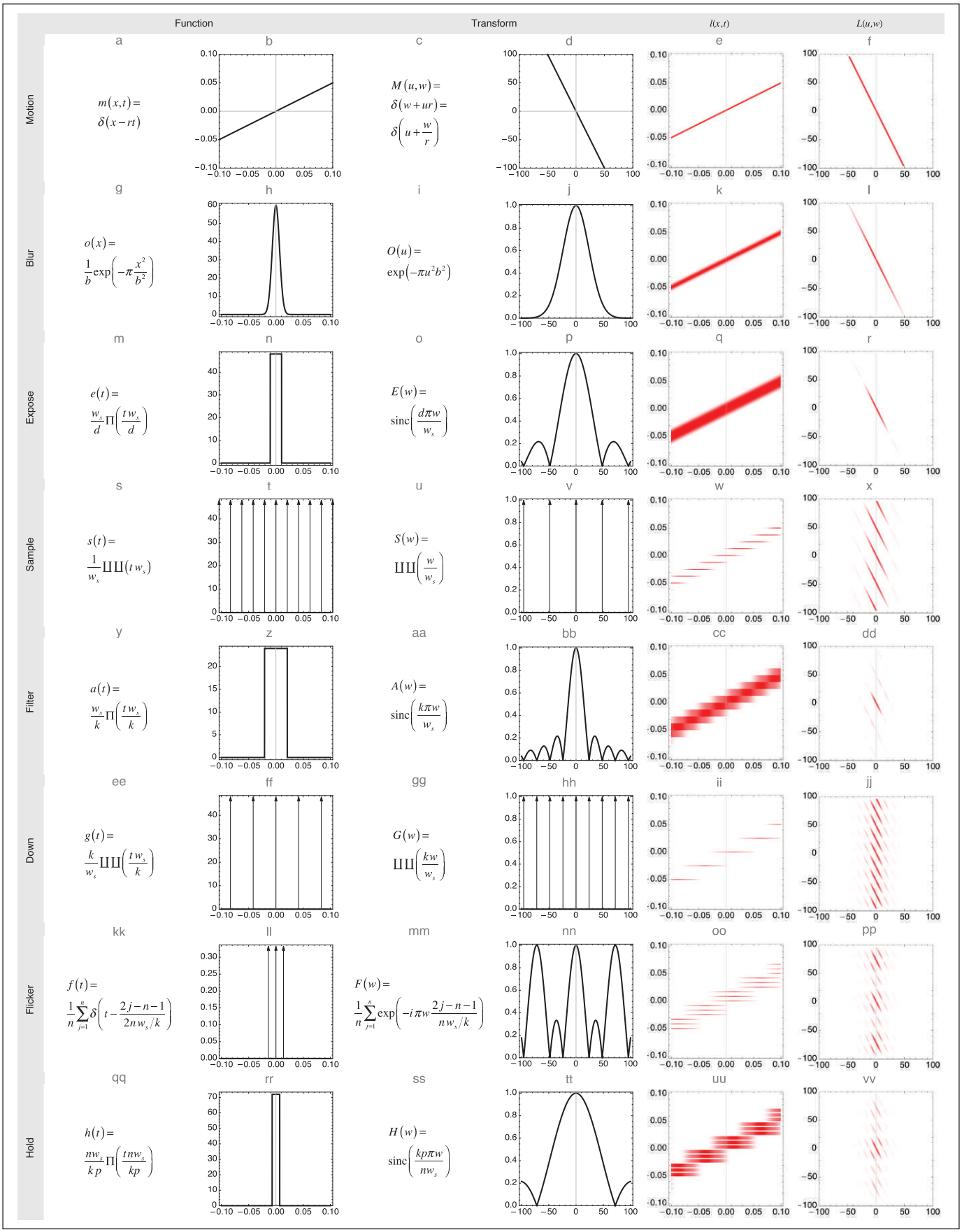


Figure 24. The process of capture and display of moving imagery. Parameters shown are $r = 2$ degrees/sec, $b = 1/60$ degree, $w_s = 48$ Hz, $d = 1$, $k = 2$, $n = 3$, $p = 0.75$.

it is difficult to construct from analog methods. Simpler solutions that prefilter only in the time domain, through analog control of the shutter (the expose function $e(t)$), have been developed.[24] More precise control of the passband could be arranged through analog or digital methods.

The formula provided earlier for the critical frame rate (**Eq. 19**) assumes that a prefilter has not been applied, in which case the Nyquist frequency depends upon the highest temporal frequency in the source, which in turn depends upon the speed. If an effective prefilter can be applied, to eliminate frequencies above w_0, then a capture frame rate of $2w_0$ will always be sufficient.

Fidelity vs. Preference

The window of visibility provides a tool for identifying when visible artifacts are present that will allow the observer to distinguish smooth and sampled motion. It cannot, however, explain whether the observer will prefer sampled or smooth. It is well known that many observers express a preference for a lower frame rate when viewing "movies," while they may not object to higher frame rates when viewing "video." In an age of digital presentation, and a great heterogeneity of displays, including smartphone, tablet, laptop, home TV, home theater, and commercial theater, it is unclear where the boundary lies between movie and video. User preferences are a complex result of cultural, artistic, and technical influences, and they change with time. I do not attempt to explain them here. Nevertheless, the window of visibility identifies the nature of the artifacts as well (flicker, multiple images, blur, strobing, etc.) and may help to explain what viewers desire when they seek the "film look."

Demonstration

Because the illustrations in the earlier figures can show only a few of the many possible combinations of sampling and display parameters, as well as visual parameters, this paper provides an interactive demonstration of the window of visibility. A static view of this demonstration is provided in **Fig. 16**. A working version of the demonstration is available on the SMPTE website at http://tinyurl.com/Watson-Demo. (Note: The demonstration requires the use of the free CDF player browser plug-in. Please download the player and then reload this page. The demo file can also be downloaded and run as a standalone application, provided that the CDF player is installed. Flicker difficulties have been encountered with Firefox browsers on some computers. Try using Safari or Chrome as alternatives.) Questions should be addressed to andrew.b.watson@nasa.gov.[25]

Summary Diagram

For reference, and to allow a survey of the complete sequence, **Fig. 24** reproduces all steps in the movie capture and display process, shown separately in **Figs. 1 to 8**. The component functions are illustrated in mathematical and graphical form in the first four columns. The rightmost two columns show the cumulative result of the process. To reduce clutter, the axes labels given in **Fig. 1** were omitted. In the first row, the vertical axis is time (function) or temporal frequency (transform), while the horizontal axis is space (function) or spatial frequency (transform). In the remaining rows, columns 2 and 4 always depict gain by the vertical axis, and the horizontal axis always depicts time or temporal frequency, except for row 2 where it shows either space or spatial frequency. In columns 5 and 6, the vertical axis is always time (column 5) or temporal frequency (column 6), and the horizontal axis is always space (column 5) or spatial frequency (column 6). Columns 5 and 6 show luminance distributions or magnitude spectra. The magnitude is shown by the red intensity.

CONCLUSION

The window of visibility is a simplified representation of the region of spatial and temporal frequencies that are visible to the human eye. By transforming the movie signal into the spatiotemporal frequency domain, and by comparing it to window of visibility, it is possible to obtain a simple visualization of the visibility of temporal sampling and rendering artifacts. This visualization may help guide the selection of frame rates and processing steps to ensure highest-quality movies.

ACKNOWLEDGMENT

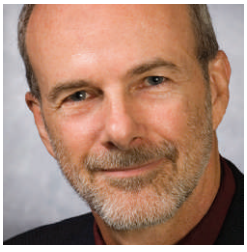
Thanks to Paul Hearty and Dave Stump for inviting me to the 2012 Society of Motion Picture and Television Engineers Symposium on High Frame Rates that inspired this paper.

REFERENCES

1. W. Aylsworth, "High Frame Rate Distribution," *SMPTE Mot. Imag. J.*, 121(6):66-68, 2012, http://journal.smpte.org/content/121/6/66.short.
2. A. B. Watson, A. J. Ahumada, Jr., and J. Farrell, "Window of Visibility: Psychophysical Theory of Fidelity in Time-Sampled Visual Motion Displays," *J. Opt. Soc. Am. A*, 3(3):300-307, 1986, http://www.opticsinfobase.org/viewmedia.cfm?uri=josaa-3-3-300&seq=0.
3. A. B. Watson and A. J. Ahumada, "Blur Clarified: A Review and Synthesis of Blur Discrimination," *J. Vis.*, 11(5):1-23, 2011, http://journalofvision.org/11/5/10/.
4. R. N. Bracewell, *The Fourier Transform and Its Applications*, 3rd ed., McGraw-Hill: Boston, 2000.
5. G. Hewlett and G. Pettitt, "DLP Cinema™ Projection: A Hybrid Frame-Rate Technique for Flicker-Free Performance," J. SID, 9(3):221-226, 2001, http://dx.doi.org/10.1889/1.1828795.
6. A. B. Watson, "Display Motion Blur: Comparison of Measurement Methods," *J. SID*, 18(2):179-190, 2010.
7. O. H. Schade, Sr., "Optical and Photoelectric Analog of the Eye," *J. Opt. Soc. Am.*, 46(9):721-739, 1956.

8. F. L. van Nes and M. A. Bouman, "Spatial Modulation Transfer in the Human Eye," *J. Opt. Soc. Am. A*, 57:401-406, 1967.
9. H. de Lange, "Research into the Dynamic Nature of the Human Fovea–Cortex Systems with Intermittent and Modulated Light. I. Attenuation Characteristics with White and Colored Light," *J. Opt. Soc. Am. A*, 48:777-784, 1958.
10. J. G. Robson, "Spatial and Temporal Contrast Sensitivity Functions of the Visual System," *J. Opt. Soc. Am.*, 56:1141-1142, 1966.
11. M. Potmesil and I. Chakravarty, "Modeling Motion Blur in Computer Generated Images," *Comput. Graph.*, 17(3):389-399, 1983.
12. A. B. Watson, "Temporal Sensitivity," in K. Boff, L. Kaufman, and J. Thomas (Eds.), *Handbook of Perception and Human Performance*, Wiley: New York, 1986.
13. C. W. Tyler and R. D. Hamer, "Analysis of Visual Modulation Sensitivity. IV. Validity of the Ferry-Porter Law," *J. Opt. Soc. Am. A*, 7(4):743-758, 1990, http://www.ncbi.nlm.nih.gov/pubmed/2338596.
14. A. B. Watson and J. I. Yellott, "A Unified Formula for Light-Adapted Pupil Size," *J. Vis.*, 12(10), 2012, http://journalofvision.org/12/10/12/.
15. J. E. Farrell, B. L. Benson, and C. R. Haynie, "Predicting Flicker Thresholds for Video Display Terminals," *Proc. SID*, 28(4):449-453, 1987.
16. P. G. J. Barten, *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality*, SPIE Optical Engineering Press: Bellingham, WA, 1999.
17. J. Rabin, "Luminance Effects on Visual Acuity and Small Letter Contrast Sensitivity," *Optom. Vis. Sci.*, 71:685-688, 1994.
18. K. T. Mullen, "The Contrast Sensitivity of Human Colour Vision to Red–Green and Blue–Yellow Chromatic Gratings," Journal of Physiology, 359:381-400, 1985.
19. D. Varner, "Temporal Sensitivities Related to Color Theory," *J. Opt. Soc. Am. A*, 1:474-481, 1984.
20. N. Drasdo, C. L. Millican, C. R. Katholi, and C. A. Curcio, "The Length of Henle Fibers in the Human Retina and a Model of Ganglion Receptive Field Density in the Visual Field," Vis. Res., 47(22):2901-2911, 2007, http://www.ncbi.nlm.nih.gov/pubmed/17320143.
21. J. J. Koenderink, M. A. Bouman, A. E. Bueno de Mesquita, and S. Slappendel, "Perimetry of Contrast Detection Thresholds of Moving Spatial Sine Wave Patterns. II. The Far Peripheral Visual Field (Eccentricity 0°-50°)," *J. Opt. Soc. Am.*, 68:850-854, 1978.
22. A. B. Watson and J. Malo, "Video Quality Measures Based on the Standard Spatial Observer," presented at the International Conference on Image Processing, Rochester, NY, 2002.
23. A. B. Watson, J. Hu, and J. F. McGowan, III, "Digital Video Quality Metric Based on Human Vision," *J. Electron. Imag.*, 10(1):20-29, 2001, http://vision.arc.nasa.gov/publications/jei2000_2wc.doc.pdf.
24. E. Renz-Whitmore, "Beyond FPS: Shutter Angle & Shape Drive Our Experience of Motion Onscreen," Pro Video Coalition, http://provideocoalition.com/pvcexclusive/story/beyond_fps_shutter_angle_shape_drive_our_experience_of_motion_onscreen, 2012.
25. A. B. Watson, "Vision Science: Demonstration of the Window of Visibility," http://visionscience.com/cdf/watson-windowofvisibility-demo.html.

A contribution received from Andrew B. Watson, NASA Ames Research Center, Moffett Field, CA, USA.



Andrew B. Watson is the senior scientist for vision research at NASA Ames Research Center in California. He is the author of more than 100 papers and 6 patents on topics in vision science and imaging technology. In 2001, he founded the *Journal of Vision* (http://journalofvision.org) where he serves as editor-in-chief. Watson is a Fellow of the Optical Society of America, of the Association for Research in Vision and Ophthalmology, and of the Society for Information Display. He is vice chair for vision science and human factors of the International Committee on Display Measurement. In 2007, he received the Otto Schade Award from the Society for Information Display, and in 2008 the Special Recognition Award from the Association for Research in Vision and Ophthalmology. In 2011, he received the Presidential Rank Award from the President of the United States.