THE QUEEN'S AWARDS FOR ENTERPRISE: INNOVATION 2011

Research & Development

White Paper

WHP 282

July 2014

The Application of Sampling Theory to Television Frame Rate Requirements

K.C. Noland

BRITISH BROADCASTING CORPORATION

BBC Research & Development
White Paper WHP 282

The Application of Sampling Theory to Television Frame Rate Requirements

Katy C. Noland

Abstract

Digital television pictures are sampled in three dimensions: two spatial dimensions, and time. This paper investigates the effect of treating video capture as a traditional temporal sampling problem, such that the frame rate is double the highest frequency in the video, or, conversely, the video signal is temporally band-limited to below half the frame rate. A significant contribution of this work is to find the fastest motion that is of interest, from which a maximum temporal frequency and hence a minimum frame rate can be calculated. To find the fastest motion of interest, a model of the human spatio-temporal contrast sensitivity function is used. For each spatial frequency, the velocity at which humans are able to resolve moving detail as well as the detail on a static object of the highest possible spatial frequency in a particular spatial format is found. This subjective matching procedure can be interpreted as finding the minimum frame rate that does justice to a specified spatial resolution, assuming that classical sampling theory is adhered to. A model of human eye tracking is then included, to take account of the effect that humans are able to resolve detail on a moving object more easily when our eyes are following the object than when our eyes are static. Incorporating the eye tracking model results in minimum frame rate requirements that are many times higher than those in use today. However, this does not take account of all the effects of eye movements: following a moving object can also reduce the visibility of aliasing artefacts, and hence the paper concludes with a discussion of why a degree of aliasing can be permitted, and hence traditional sampling theory does not necessarily need to be applied in television sampling.

Additional key words: high frame rate, ultra-high definition, UHD, resolution, aliasing, strobing, motion blur, human visual system, HVS

White Papers are distributed freely on request.

Authorisation of the General Manager, Research & Development is required for publication.

BBC Research & Development
White Paper WHP 282

The Application of Sampling Theory to Television Frame Rate Requirements

Katy C. Noland

1 Introduction

An increase in frame rate will improve the temporal resolution of television pictures, reducing motion blur or strobing effects. It is not completely clear, however, how high a frame rate would be required for these motion artefacts to be completely imperceptible. This paper contributes to our understanding of motion perception in television, by investigating the application of traditional sampling theory [Nyquist, 1928] [Shannon, 1948] to the time domain in video capture. This is one step towards a complete understanding of the visibility of motion artefacts: it allows the visibility of motion blur to be investigated, but does not address its counterpart, a strobing effect caused by temporal undersampling. In reality traditional sampling theory is not fully applicable to video, since the motion of the human eye can reduce the visibility of the strobing effect caused by aliasing. The analysis presented nonetheless supplies a useful basis for further discussion and research.

The work comes at a time when ultra-high definition (UHD) formats with a spatial resolution of 3840×2160 (UHD-1) or 7680×4320 (UHD-2) pixels have been proposed, and consumer displays that support UHD-1 are already available, but the appropriate frame rate to match these spatial resolutions remains a topic of lively debate. Recommendation ITU-R BT.2020 [ITU-R, 2012a] defines the parameter values for UHDTV, and includes all frame rates from the High Definition (HD) television parameter specification [ITU-R, 2002], plus one higher frame rate of 120 frames per second (fps). Additional frame rates of 100 and 119.88 fps may be subsequently included to address problems of compatibility with existing systems. These higher values are an acknowledgement that HD frame rates may not be sufficient for UHD.

The work presented here complements subjective tests [ITU-R, 2012b][Kuroki et al., 2007] [Driesnack, 2013] that seek to inform discussion on higher frame rates, by taking a theoretical approach to the same problem: finding the most appropriate frame rate for a given spatial resolution. Assuming that traditional sampling theory is applied, such that all frequencies above half the sampling frequency are removed by an anti-alias filter prior to sampling, the lowest frame rate that would cause motion blur to be no more visible than the worst spatial blur is calculated. This is done using existing data on the visibility of moving spatial patterns at a range of frequencies and velocities: a model of the human contrast sensitivity function. The contrast sensitivity function alone does not take account of all eye movements, so a model of human eye tracking is subsequently incorporated into the calculations.

In section 2 an introduction to eye movements and the subjective characteristics of video motion that are affected by the choice of frame rate is given, and then previous approaches to determining an appropriate frame rate are discussed in section 3. The contrast sensitivity function model is introduced in section 4, followed in section 5 by a demonstration, using purely objective calculations, of how the frame rate must be increased in proportion to the spatial resolution in order to maintain dynamic resolution. Section 6 is the core of the paper, where a method of matching the frame rate to the spatial resolution using the human contrast sensitivity function model is proposed and then combined with an eye tracking model. The results of the frame rate calculations are presented in section 7, followed by a discussion of the assumptions made and the implications of the results in section 8. Section 9 concludes the paper.

2 Subjective Characteristics Influenced by the Frame Rate

The human eye is an integral part of any television system, and hence its performance should also be taken into account when choosing a frame rate. In this section the different kinds of eye movement are introduced, followed by a description of the subjective characteristics of television pictures that are affected by the frame rate.

2.1 Eye Movements

Human eye movements can be categorised into three types [Daly, 2001, pp. 185–186]:

retinal drift, a continuous slow and random motion of the eye around a fixation point at speeds of about 0.8–0.15 degrees per second that is usually not perceived or under conscious control;

smooth pursuit, or tracking of a moving object, at speeds up to about 80 degrees per second, normally under the control of the viewer; and

saccades, very fast movements of 160–300 degrees per second that occur when jumping between fixation points, during which sensitivity is significantly reduced.

One difficulty in setting up subjective experiments to measure quality of motion is taking account of how viewers' eyes move. Experiments can either allow free eye movements, which would be representative of a natural viewing scenario, or control for eye movements in order to separate the effects of movement in the scene and movements of the eye.

2.2 Motion Artefacts

The most important subjective temporal effects in television pictures are flicker, strobing and motion blur. The critical flicker frequency is the rate at which refreshing of the whole screen is just visible, on a display that includes some blanking between frames. Since humans are known to be more sensitive to flicker in their peripheral vision [Barten, 1999, pp. 115–117], it is related to the percentage of the field of vision occupied by the screen, which can be determined by the screen size and viewing distance. It also strongly depends on the brightness of the flicker and surroundings, as well as the duty ratio (on-time as proportion of refresh rate) of the flicker [Roberts, 2009, p. 291].

Strobing is the result of temporal aliasing, where the frame rate is insufficient to represent the motion in the scene and so objects do not move smoothly. It can appear as either juddery motion or multiple imaging. The effect is reduced when a moving object is tracked by the eye, reducing the speed of the object's image on the retina. However, only one linear velocity can be tracked at a time, so any objects that either rotate or move at a different speed, or in a different direction, will still suffer from strobing if the frame rate is too low. This occurs even for nominally stationary objects such as a static background, whose image will move across the retina if the eye is tracking another moving object. Strobing is most severe when both the camera and display apertures are short, and hence individual frames are sharp. Any increase in the camera or display aperture time will reduce strobing, but increase motion blur.

Motion blur is the result of finite integration times at the camera or display. In the camera, just as when taking a photograph, if an object is moving fast and the camera shutter is open for a finite time, the object will be smeared across the image. Reducing the shutter time to give a shorter aperture helps to sharpen the image, since high temporal frequencies would then be less attenuated by temporal integration, but this increases the potential for temporal aliasing. The effects at the display are similar: a short display aperture time such as those achieved by cathode ray tubes (CRTs) results in very sharp images and smooth motion for tracked objects if the camera aperture time was also short, but distortion of untracked motion due to temporal aliasing. Longer display aperture times, on the other hand, such as those of conventional liquid crystal displays (LCDs) which approach 100 % of the frame period, will cause the images to be smeared across the

camera aperture	display aperture	tracked motion	untracked motion
short	**short**	smooth and sharp	strobed and sharp
short	**long**	smooth and blurred by retinal slip	strobed and blurred by retinal slip
long	**short**	smooth and blurred by camera integration	mainly smooth with potential for some strobing, and blurred by camera integration
long	**long**	smooth, and blurred by camera integration and retinal slip	mainly smooth with potential for some strobing, and blurred by camera integration and retinal slip

Table 1: Description of motion for long and short camera and display aperture times, when the frame rate is insufficient. N.B. Camera integration causes motion blur only when the scene is moving, and retinal slip occurs only when the eye is moving.

retina when an object is tracked, since the displayed image will stay constant for a finite period of time whilst the eye is moving smoothly. This effect is sometimes called “retinal slip”. The effects of different camera and display apertures are summarised for tracked and untracked motion in table 1. Adjusting the camera and display aperture time allows strobing and motion blur to be balanced, but the only way to simultaneously produce smooth motion and sharp pictures for both tracked and untracked objects is to use a high frame rate.

In this paper it is assumed that traditional Nyquist-Shannon capture [Nyquist, 1928] [Shannon, 1948] is possible with an optical anti-alias filter at half the sampling frequency followed by impulsive sampling, and similarly that the video signal is viewed via an impulsive display followed by an optical reconstruction filter. Such optical filters currently do not exist, but the conditions can be approximated by using temporal oversampling at the camera and display. Analysis of alternative capture and display methods is left to future work.

3 Previous Approaches to Estimating the Required Frame Rate

Previous research into the frame rates required for television all suggests that an increase in frame rate leads to perceptible, and sometimes large, improvements in subjective quality. This section gives an overview of the most relevant earlier studies.

3.1 Results Summarised in ITU-R BT.2246

The recent ITU-R Report, *The present state of ultra high definition television* [ITU-R, 2012b], contains an overview of subjective experiments carried out by Japanese Broadcaster NHK on required frame rates for television. Their results are presented, but details of the experiments are generally not given. The results have been used to inform standardisation of the parameters for UHD television [ITU-R, 2012a].

The results show that the critical flicker frequency increases with the width of the field of view and decreases with display duty ratio. Around 80 fps were needed to eliminate flicker with 100 degrees field of view—this would be the field of view for UHD-2 viewed from 0.75 times the screen height (0.75 H)—using the worst-case duty ratio for each subject. This frame rate is lower than those proposed for smooth motion portrayal (discussed next), so flicker is less important than strobing in the high frame rate debate. This is especially true given that many LCDs have a duty ratio of close to 100%, which essentially eliminates visible flicker [Emoto and Sugawara, 2012].

Motion blur was shown to increase with long aperture times either in the camera or display, and to be more obvious for fast moving objects. To reduce motion blur to an acceptable level for objects moving at between 8 and 32 degrees per second, it was found that the aperture length should be less than about 3 ms. This is equivalent to 30% shutter at 100 fps, 45% shutter at 150 fps or 100% shutter at 333 fps. In separate experiments, strobing was shown to be worse with a shorter aperture time for fast-moving material (25% aperture was worse than 50% at 240 fps), but better with higher frame rates (120 fps was rated much better than 60 fps, and additional perceptible improvements were measured at 240 fps, all with a constant 1/240 s aperture). In combination these results indicate that around 50% aperture at 150 fps or 100% aperture at 300 fps may be suitable parameters for reducing both strobing and motion blur to acceptable levels. The higher frame rate of 300 fps would have the benefit of integer conversion from both 50 and 60 fps.

Dynamic visual acuity, measured by the speed at which an object's orientation can be determined, was shown to increase as the field of view widens. This means that higher spatial and temporal resolution may be needed for large screens than otherwise expected. The effects of a wide field of view and in general differences between foveal and peripheral vision are yet to be fully understood.

The experiments presented so far all aimed to measure very specific aspects of video quality. However, the most comprehensive measure of quality is overall subjective ratings, that encompass all perceptible artefacts. The report includes results showing that 120 fps offers a big improvement in overall subjective quality compared to 60 fps, and 240 fps is better still. This confirms earlier results that focussed on motion blur and strobing.

3.2 Further Literature

A number of further articles in the literature also address the question of video frame rate. Most aim to either create a completely transparent system, or to balance severe artefacts in very low bit rate applications.

Kuroki et al. [2007] report a series of subjective tests that examine the impact of frame rate on both visual acuity and perceived quality. They started by measuring the threshold of perceivable horizontal resolution for square wave gratings mounted on a mechanical rig moving at different horizontal velocities, and showed that even without any head or eye stabilisation, the velocity at which subjects are no longer able to resolve a pattern is much lower for higher spatial frequencies. This data is a form of contrast sensitivity threshold, and is equivalent to one contour line in the contrast sensitivity model that will be described in section 4. The authors then repeated the experiment using recordings of the moving grating. They recorded the gratings in both 24 frames per second progressive and 30 frames per second interlaced formats on 35 mm film and HD cameras, and found that although spatial visual acuity is only slightly lower for stationary patterns in the recordings, it drops rapidly to less than 1/8th of the equivalent values for the real grating as soon as there is any movement. This is clear evidence that the frame rates tested are insufficient for a transparent system. The authors then demonstrate that motion blur occurs as a result of both camera integration and retinal slip.

Their next experiments, reported in the same article, asked subjects to rate blurriness and jerkiness for some real material shot at 1000 fps, with frames then averaged to synthesise a range of frame rates from 62 to 500 fps. The spatial resolution of the sequences is not given, but the sample images suggest several different formats. The results show a dramatic improvement in both blur and jerkiness at 125 fps over 62 fps, then flattening off above 250 fps. Jerkiness was rated worse with short shutter times. These experiments independently verify the results presented in ITU-R BT.2246 [ITU-R, 2012b], that a great improvement is possible by changing to a frame rate in the region of 100 to 150 fps, and that further quality improvements are possible by using a frame rate of over 200 fps. Findings from subjective experiments conducted by the European Broadcasting

Union's Broadcast Technology Futures group appear to support this conclusion [Driesnack, 2013].

Watson [2013] takes an analytical approach to the problem. He presents a frequency domain analysis of a simplified video capture and display system, and classifies artefacts according to whether they appear in the "window of visibility", which he defines as a threshold on the contrast sensitivity function measured by Robson [1966]. He proposes this kind of analysis as a method of measuring the visibility of aliasing artefacts for a given system and signal. In section 6 of this paper the same question is approached from a different angle, starting by finding the range of signals we would like to be able to represent, then calculating the system parameters that would be required to achieve this with no aliasing.

Tonge [1986] discusses frame rate requirements under a number of different assumptions, and suggests that a frame rate of up to 2000 fps would be needed for television pictures to appear exactly equivalent to continuous motion, based on measured limits of dynamic resolution. He notes that with eye tracking the rate can be reduced to the critical flicker frequency, and even suggests that with motion-compensated frame insertion the transmitted frame rate could be less than 10 fps to represent motion vectors without aliasing, based on the acceleration capabilities of the moving eye. This, however, would rely on knowing the motion accurately: information that would not be contained in a signal at only 10 fps, and so would have to be sent in addition to the pictures. It also does not take account of revealed backgrounds or shot changes.

A number of articles [Apteker et al., 1995] [Huynh-Thu and Ghanbari, 2008] [McCarthy et al., 2004] [Ou et al., 2011] assess the importance of frame rate in the context of very low bit rates and resolutions for streaming applications (up to 400 pixels horizontal resolution at between 5 and 30 frames per second), with most finding that very low frames rates are more acceptable than very severe compression. However, this paper is concerned with much higher resolution video, so the conclusions from these low bit rate studies are not applicable.

Some earlier work from BBC Research and Development provides evidence that the balance between spatial and temporal resolution is tipped in favour of the spatial dimensions with HD television. Stone [1986] performed a detailed investigation into parameters for HD television, and demonstrated significant perceived improvements when increasing the frame rate from 50 to 80 fps. He reports that although reducing the camera shutter time to 50 % also improved the quality of motion, this approach caused severe problems when converting to other frame rates due to the strong temporal aliases in the signal. Tanton and Stone [1989] compared a range of interlaced formats using a still image as test material, and showed that, even with no motion in the source, significant benefits could be achieved using a 75 Hz field rate in preference to 50 Hz. The experiments were conducted using a CRT display, so the result may in part be due to improved flicker performance at 75 Hz, which may not apply to hold-type displays.

Armstrong et al. [2008] and Salmon et al. [2011] give excellent descriptions of the arguments for higher frame rates, and report observations of significant quality improvements over conventional rates in material captured at 300 fps and converted to 100 fps for display. This paper adds to their arguments from an analytical angle: the frame rate needed to balance the visibility of temporal and spatial artefacts for a given spatial resolution is predicted using a model of the human contrast sensitivity function, assuming that classical Nyquist-Shannon sampling is applied.

4 Contrast Sensitivity Function Model

The human spatio-temporal contrast sensitivity function (CSF) represents the visibility of pure frequencies, i.e. sinusoidal gratings, at a range of spatial and temporal frequencies. The temporal frequency τ is the rate at which a point in space changes its brightness, ususally as a result of an object moving past that point. It can be calculated from the object's spatial frequency ρ and velocity v, using the relationship

$$\tau = \rho v \tag{1}$$

[Tonge, 1986]. A moving or stationary flashing object can also create a temporal frequency, in which case the flash frequency should also be incorporated into equation 1, but this occurs relatively rarely in real scenes. Since moving objects occur more often in nature than flashing patterns, the spatio-temporal CSF is often plotted on axes of spatial frequency and velocity, rather than spatial frequency and temporal frequency. Kelly [1979] suggests that analysis in terms of velocity is also a better fit to the functionality of the human visual system.

In this paper an adapted version of Kelly's model of the human contrast sensitivity function is used [Kelly, 1979]. The original Kelly model is derived from human CSF measurements collected using greyscale rolling sinusoidal gratings as stimuli, with a (presumably average) retinal luminance of 300 Trolands[1]. Subjects were presented with each grating in turn and asked to adjust the contrast until the grating was just visible, giving the contrast visibility threshold for that stimulus. Retinal stabilisation was used to exclude the effects of any involuntary eye movements from the experimental results. A mathematical model was then fitted to the data. Kelly showed that the form of the contrast sensitivity function remains the same for different velocities, only the peak spatial frequency position and maximum height vary. The work also showed that data from experiments without retinal stabilisation can be interpreted as the stabilised results at a constant velocity of approximately 0.15 degrees per second, which is known to be approximately the retinal drift velocity.

Daly modified Kelly's model by adding three constants, c_0, c_1 and c_2, for fine-tuning the peak sensitivity, maximum spatial frequency cutoff and critical fusion frequency [Daly, 2001]. The modification allows the model to take account of brighter images and to match data from non-stabilised experiments that show higher overall sensitivity. The contrast sensitivity at retinal velocity v_d, measured in degrees per second, and spatial frequency ρ_d, measured in cycles per degree, is modelled as:

$$\mathrm{CSF}(\rho_d, v_d) = k \cdot c_0 \cdot c_1 \cdot c_2 \cdot v_d \cdot (c_1 2\pi\rho_d)^2 \exp\left(-\frac{c_1 4\pi\rho_d}{\rho_{d\,\max}}\right) \tag{2}$$

where

$$k = s_1 + s_2 \cdot \left|\log\left(\frac{c_2 v_d}{3}\right)\right|^3 \tag{3}$$

and

$$\rho_{d\,\max} = \frac{p_1}{c_2 v_d + 2} \tag{4}$$

with $s_1 = 6.1$, $s_2 = 7.3$ and $p_1 = 45.9$ (reproduced from [Daly, 2001] and [Laird et al., 2006]).

The constants c_0, c_1 and c_2 were tuned to fit data from an experiment by Laird et al. [2006] in which human sensitivity to moving Gabor patterns was measured, without retinal stabilisation but with fixation on the centre of the screen. This means that they take account of involuntary retinal drift, but not smooth pursuit (tracking). The fitted constants are given as $c_0 = 0.6329$, $c_1 = 0.8404$ and $c_2 = 0.7986$. These are the values used for the calculations in section 6. A plot of the modelled contrast sensitivity function is shown in figure 1.

5 Objective Matching of Spatial and Temporal Resolution

In this section it is shown that from a purely objective perspective the frame rate must be increased in proportion to the spatial resolution, if dynamic resolution is to be maintained. Standard sampling theory is used to determine the frame rate required to maintain a given static horizontal resolution when an object moves.

The working involves a number of changes of unit. Subscript notation is used to distinguish between them, with r referring to units with angles measured in radians, d to units with angles

[1] 1 Troland is the retinal illuminance produced when a surface of luminance 1 candela per square metre (1 nit) is viewed through a pupil of area $1\,\mathrm{mm}^2$.

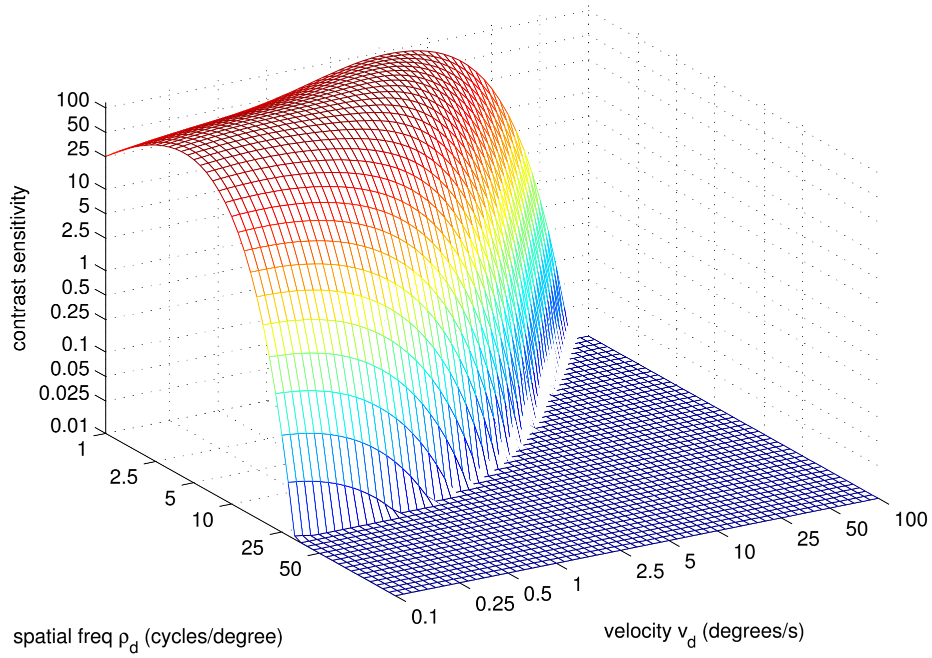


Figure 1: Kelly-Daly model of the human contrast sensitivity function, as described by Laird et al. [2006].

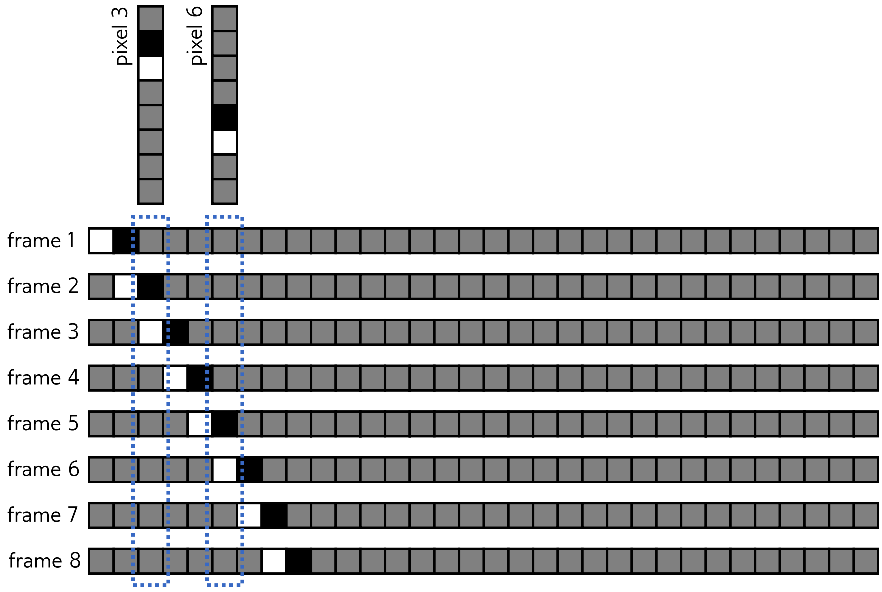
measured in degrees, w to units with distances measured in pictures (i.e. number of picture widths) and p to units with distances measured in pixels.

If the number of horizontal pixels in a video sequence is h, sampling theory tells us that horizontal frequencies up to $h/2$ cycles per picture can be represented. If the scene is sampled temporally, as all conventional video recordings are, then the signal is also subject to aliasing constraints in the time domain. Figure 2 illustrates the principles using a single period of a sinusoid at the critical horizontal frequency, moving at a constant velocity across the screen. If the amplitude at one physical pixel on the screen is plotted over time, it should be possible to see the shape of the object as it moves through the selected pixel, as is the case for figure 2(a). If the object is moving too fast for the frame rate, as in figure 2(b), it will jump several pixels between one frame and the next, so a plot of the amplitude at a single pixel location will not show the object accurately. The object will appear and disappear at different locations across the screen rather than moving smoothly. This is temporal aliasing.

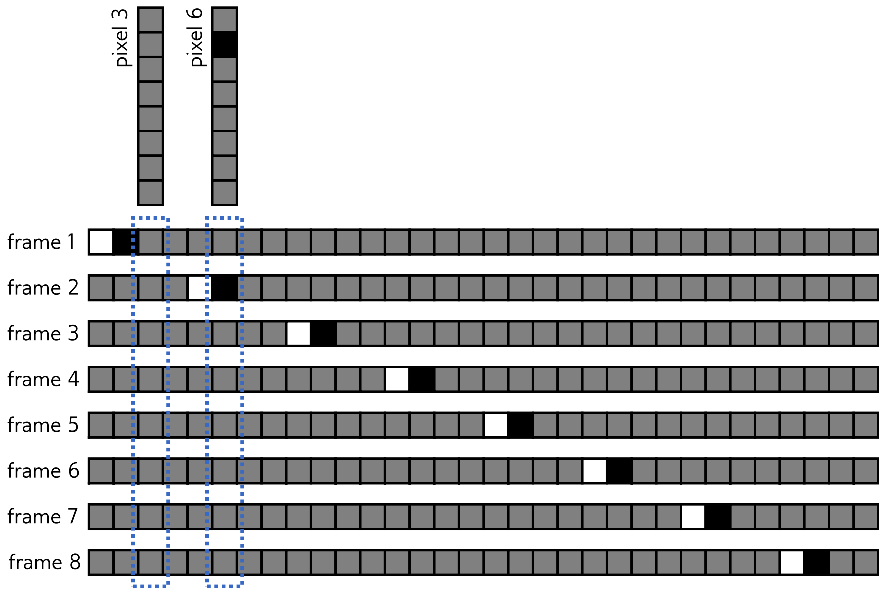
Aliasing constraints are the same in time as in space, so the critical temporal frequency at any particular pixel is the half the frame frequency. The brightness at one pixel position can be changing at the critical frequency as a result of a sinusoidal object moving through the pixel at a horizontal velocity of half its period per frame. The object in figure 2(a) is moving at the critical velocity.

These relationships, derived from standard sampling theory, can be used to calculate how many frames per second are required to maintain a given static spatial resolution for different velocities, under the assumption of no aliasing being present. If the following variables are defined:

$$
\begin{aligned}
h &= \text{number of horizontal pixels} \\
v_w &= \text{horizontal object velocity in pictures per second} \\
v_p &= \text{horizontal object velocity in pixels per second} \\
\rho_p &= \text{static horizontal frequency in cycles per pixel} \\
\mathrm{F} &= \text{required frame rate to avoid aliasing}
\end{aligned}
$$



(a) Frame rate is just high enough to capture the motion without aliasing.



(b) Frame rate is too low to capture the motion without aliasing.

Figure 2: Illustrations of a single period of a sampled sinusoid travelling along a picture line. The pixel values over time at two fixed pixel positions are also shown as the object passes through them. In 2(a) the frame rate is just high enough, and so the whole object is seen passing through any given pixel position. In 2(b) the frame rate is too low, so the moving object is not properly represented at any given pixel position.

the relationship can be written down:

$$v_p = hv_w \tag{5}$$

From equation 1 and the Nyquist-Shannon theorem:

$$\mathrm{F} = 2\rho_p v_p \tag{6}$$

For the maximum spatial frequency that can be represented by a pixel format, $\rho_p = 1/2$, so

$$\mathrm{F} = hv_w \tag{7}$$

This frame rate would be able to represent, without motion blur or temporal aliasing, an object of the highest possible spatial resolution moving at a given velocity v_w. It can be calculated using the number of horizontal pixels and the highest object velocity. The results are presented in section 7.1, and equation 6 is also used in section 6.3 where a method of specifying the highest desired velocity is proposed.

6 Application of Human Perception Measurements to Match the Frame Rate to a given Spatial Resolution

This section forms the core of the paper: a method of finding the required frame rate to subjectively match a given horizontal resolution, using the Kelly-Daly model of the human spatio-temporal contrast sensitivity function [Laird et al., 2006]. After explaining some necessary unit conversions (sections 6.1 and 6.2), this is done first for a fixed gaze (section 6.3), then a model of eye tracking is introduced (section 6.4) to take account of increased sensitivity to motion blur for tracked objects.

Equation 6 allows the frame rate to be calculated that can represent, without aliasing, an object of a given spatial frequency moving a given velocity. The pixel format determines the highest possible spatial frequency, but gives no indication of the highest velocity we might want to represent. In television, especially live recordings, it is difficult to predict what the fastest motion will be. In this section an approach grounded in visual perception research is taken: the calculated frame rate can represent all moving detail that is as easy to resolve as detail in a static object of the highest representable spatial frequency.

The Kelly-Daly contrast sensitivity model describes how easy it is to resolve any particular static frequency, and allows the set of points to which humans are equally sensitive in the spatial-frequency–velocity space to be found. This set of points is equivalent to one contour line in the CSF, as illustrated by the black line in figure 4. Equation 6 is then used to find the required frame rate for each point on the contour line, and the largest of these values is taken as the overall required frame rate.

The contrast sensitivity model does not contain any information about the visibility of strobing artefacts, and hence the analysis does not allow for any temporal aliasing in the signal, it only takes account of sensitivity to motion blur. If some strobing is found to be acceptable, visible motion blur can be eliminated at a lower frame rate than those indicated here.

6.1 Expressing Cycles per Degree as Cycles per Picture

The Kelly-Daly model specifies contrast sensitivity function values at a range of spatial frequencies, ρ_d, measured in cycles per degree, and a range of velocities, v_d, measured in degrees per second. In order to use the CSF to specify a frame rate suitable for a particular pixel format, it is necessary to relate that pixel format, or rather the maximum spatial frequency it can represent, to the corresponding spatial frequency at the eye. Hence the first step is to convert units of the CSF from cycles per degree to cycles per picture. It is similarly necessary to convert the CSF velocity units from degrees per second to the more convenient pictures per second, which will be explained in section 6.2.

To do the conversions requires some trigonometry based on the viewing distance, A_w, and screen width, W_w, both measured in number of picture widths (i.e. $W_w = 1$). These relationships are illustrated in figure 3. For a fixed flat screen and fixed viewer, the distance between the viewer and the centre of the screen will be smaller than the distance between the viewer and the screen edges. Hence for constant values of ρ_d and v_d at the eye, the spatial frequency at the screen will be warped downwards towards the screen edges, and the velocity at the screen will be warped upwards towards the edges.

Let the angle θ_r radians be the angle subtended at the viewer between the centre of the screen and the fixation point, and the angle ϕ_r radians be the angle subtended at the viewer by one period

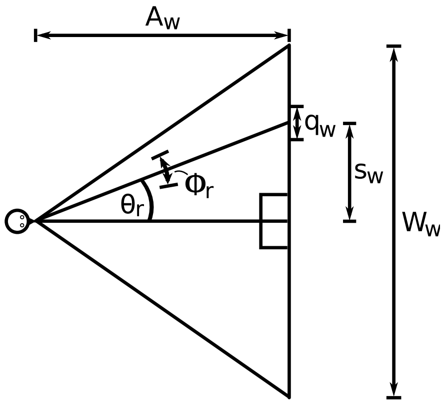


Figure 3: Illustration of the parameters used to calculate the required frame rate in section 6.

of the spatial pattern. Length s_w is the distance between the centre of the screen and the fixation point, and length q_w is the length of the projection of ϕ_r onto the screen.

From figure 3:

$$s_w = A_w \tan\theta_r \tag{8}$$

The length q_w (one period of the image frequency at the screen) is related to the angle ϕ_r (one period of the image at the eye) by the rate at which distance s_w changes with respect to θ_r, $\frac{\mathrm{d}s_w}{\mathrm{d}\theta_r}$.

$$\frac{\mathrm{d}s_w}{\mathrm{d}\theta_r} = \frac{A_w}{\cos^2\theta_r} \tag{9}$$

$$q_w = \phi_r \frac{\mathrm{d}s_w}{\mathrm{d}\theta_r} = \frac{\phi_r A_w}{\cos^2\theta_r} \tag{10}$$

Now let $\rho_r = \frac{1}{\phi_r}$ be the number of cycles in the image per radian subtended at the eye, and $\rho_w = \frac{1}{q_w}$ the number of image cycles per picture. From equation 10:

$$\rho_w = \frac{\rho_r \cos^2\theta_r}{A_w} \tag{11}$$

Equation 11 provides the basis for converting between image cycles per radian at the eye ρ_r and image cycles per picture ρ_w, it only remains to convert the angle units to degrees to match the CSF data, and substitute information about the viewing distance A_w, which affects the level of spatial detail at the screen that can be resolved. Equation 12 performs the conversion between image cycles per degree at the eye and image cycles per radian at the eye:

$$\rho_r = \rho_d \times \frac{360}{2\pi} \tag{12}$$

The viewing distance is left as a parameter that is specified in number n of picture heights H_w, and a 16:9 aspect ratio is assumed for calculating the screen height from the width.

$$A_w = nH_w = \frac{9nW_w}{16} \tag{13}$$

It may also be preferable to specify the fixation point in terms of distance along the screen, rather than θ_r. Distance s_w is therefore expressed as a proportion x of the screen width, so $s_w = xW_w$. From equation 8:

$$\theta_r = \arctan\frac{s_w}{A_w} \tag{14}$$

$$\theta_r = \arctan\left(\frac{16xW_w}{9nW_w}\right) \tag{15}$$

$$\theta_r = \arctan\left(\frac{16x}{9n}\right) \tag{16}$$

Values for ρ_r, A_w and θ_r, and $W_w = 1$ can then be substituted into equation 11 to give:

$$\rho_w = \frac{320\rho_d}{n\pi} \times \cos^2\left[\arctan\left(\frac{16x}{9n}\right)\right] \tag{17}$$

Equation 17 is used to convert the CSF spatial frequency units from cycles per degree to cycles per picture.

6.2 Expressing Degrees per Second as Pictures per Second

The same principle can be used to express the velocity, given as v_d degrees per second at the eye, as the more convenient v_w pictures per second. Let v_r be the velocity in radians per second at the eye, so $v_r = v_d \times \frac{2\pi}{360}$, and let t be the time in seconds.

$$v_w = \frac{\mathrm{d}s_w}{\mathrm{d}t} = \frac{\mathrm{d}s_w}{\mathrm{d}\theta_r}\frac{\mathrm{d}\theta_r}{\mathrm{d}t} \tag{18}$$

Substitute equation 9:

$$v_w = \frac{A_w v_r}{\cos^2\theta_r} \tag{19}$$

Then substitute for A_w, v_r and θ_r:

$$v_w = \frac{nW_w v_d \pi}{320\cos^2\left[\arctan\left(\frac{16x}{9n}\right)\right]} \tag{20}$$

Equation 20 is used to convert the CSF velocity units from degrees per second to pictures per second. Figure 4 shows the contrast sensitivity model of figure 1, plotted on axes of cycles per picture against pictures per second. A viewing distance of 1.5 H was used, and it was assumed that the viewer is looking at the centre of the screen, i.e. $n = 1.5$ and $\theta_r = 0$. Larger values of θ_r would lead to finer spatial resolution, and hence a matching higher frame rate would be calculated to match spatial detail at the screen edges. However, the recommended viewing distance for a spatial format is intended to put the pixel structure at limit of spatial resolution, so if a viewer sits at the recommended viewing distance the additional spatial resolution at the edges would not be perceived. Hence $\theta_r = 0$ gives the most representative value of perceived spatial resolution when the viewer is at the recommended viewing distance.

6.3 Determining the Frame Rate to Match a given Horizontal Resolution

The aim of this work is to determine a frame rate that is somehow perceptually equivalent to a particular horizontal resolution. An upper bound is set on the velocity of a moving object for every spatial frequency, based on the ability of the human visual system to resolve that spatial frequency as the object moves. If the spatial detail on the moving object can be resolved as well or better than a still object of the highest possible spatial frequency, the frame rate should support that motion. Assuming that no temporal aliasing can be tolerated (which is not necessarily the case—see discussion in section 8.1), the required frame rate can be calculated using equation 6. With reference to figure 4, this means that the frame rate must be high enough to represent every frequency on or to the left of the contour line, marked in black, that passes through the highest spatial frequency at zero velocity. This contour line can be thought of as a line of equal blur: any objects with finer spatial detail moving at the same speed, or the same object moving any faster,

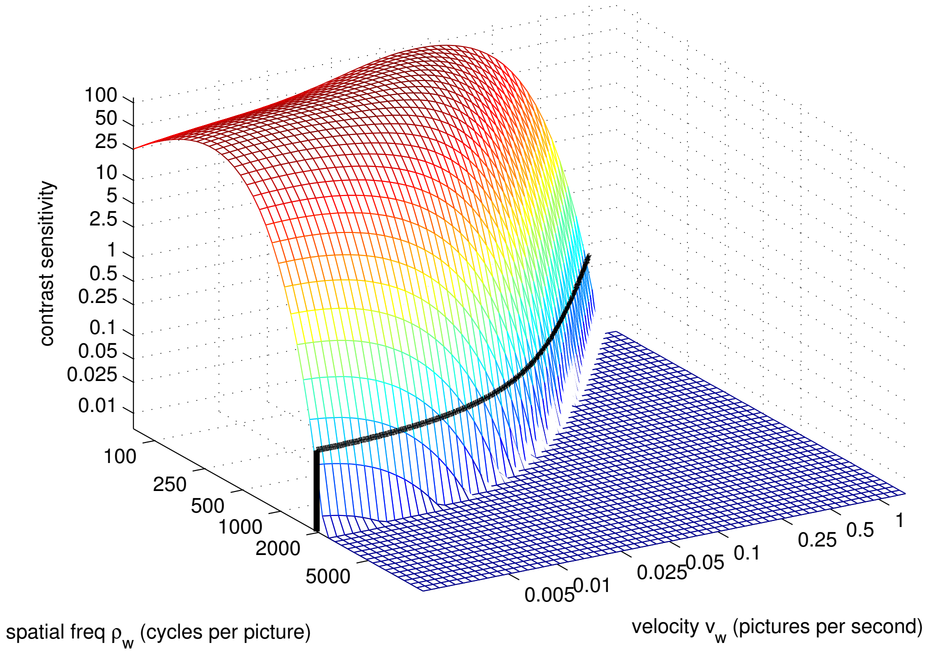


Figure 4: Model of the human contrast sensitivity function with units converted to spatial frequency ρ_w in cycles per picture and velocity v_w in pictures per second. In this case a viewing distance of 1.5 H was used, and it was assumed that the viewer is looking at the centre of the screen. The black line shows the threshold of spatio-temporal frequencies to be represented to perceptually match a spatial resolution of 3840 horizontal pixels.

will be more blurred than one on or to the left of the line. The line shown in figure 4 is for a spatial resolution of 3840 pixels, which corresponds to a maximum spatial frequency of 1920 cycles per picture, or just under 30 cycles per degree at a distance of 1.5 H.

The peak of the contrast sensitivity function is not at the origin, so for the rare cases where the highest spatial frequency falls below the peak, the frame rate required to represent only the points with lower spatial frequencies is calculated, even though some higher frequencies would be easier to resolve. This only occurs for the shortest viewing distance of 0.75 H with the lowest resolutions of 720 and 1080 horizontal pixels, so is not important for realistic viewing conditions; in all other cases the highest spatial frequency falls above the peak in the contrast sensitivity function.

Having determined the spatial frequencies and velocities that should be representable, the required frame rate for each one is calculated using equation 6. These are shown in figure 5 for some selected spatial resolutions at a viewing distance of 1.5 H. The peak of each curve is taken as the overall frame rate to match its corresponding spatial resolution.

6.4 Incorporating a Model of Eye Tracking

The frame rate requirements calculated using the method described in section 6.3 are relevant for a fixed gaze on the centre of the screen, but under normal viewing conditions, the eye is free to track objects as they move. When the eye is exactly tracking an object, the relative velocity of the object to the retina is reduced. Daly describes, based on experimental data, how in reality we do not track motion exactly, but rather follow at about 82 % of the object speed, catching up with saccades when the lag becomes too great [Daly, 2001, pp.185–187]. He finds a maximum eye tracking velocity of around 80 degrees per second. This model is shown in figure 6.

In order to incorporate knowledge of eye tracking into the contrast sensitivity model, it is

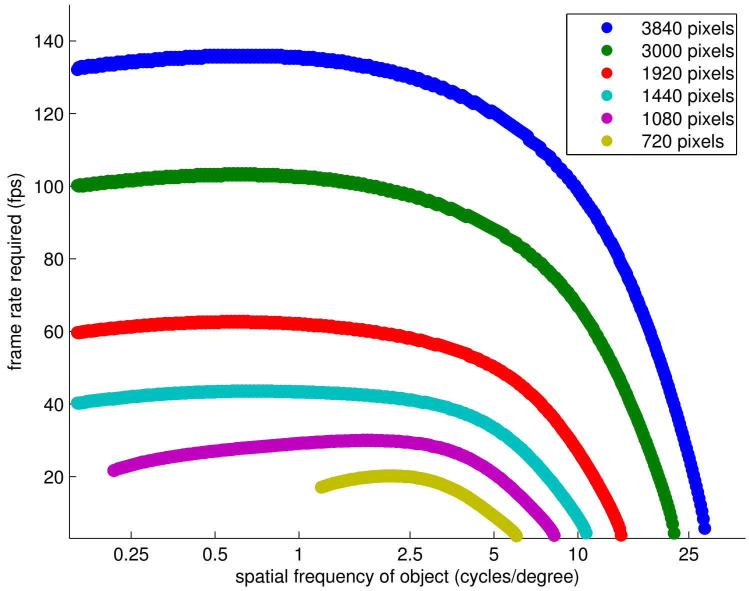


Figure 5: Minimum frame rate required for each point along the “equal blur” contours, for selected spatial resolutions at a viewing distance of 1.5 H. The contour marked in black in figure 4 is for 3840 horizontal pixels, and corresponds to the dark blue (uppermost) curve here. The peak of each curve is taken as the required frame rate to match that spatial resolution.

necessary to scale the CSF velocities. An object moving at v_d degrees per second, if tracked at $0.82v_d$ degrees per second, presents a velocity of $(1-0.82)v_d$ degrees per second at the retina—this is a result of the lag between the object and the eye. Hence, to convert the velocity of the image moving across the retina, for which the contrast sensitivity function of figure 1 is valid, to the absolute velocity of a tracked object, which is needed to determine the required frame rate by means of equation 6, all velocities in the contrast sensitivity function are scaled by $1/(1-0.82) \approx 5.6$. An upper limit of 80 degrees per second was also imposed on the tracking velocity. The corresponding maximum trackable object velocity is $(80/0.82) = 97$ degrees per second. Figure 7 shows a comparison of the original and modified contrast sensitivity functions. A discussion of the sensitivity of the results to the eye tracking measurements is given in section 8.4.

7 Results

In this section the results of the frame rate calculations described in sections 5 and 6 are presented. The required frame rates for various spatial frequencies at a range of velocities using only objective criteria are shown first (section 7.1, method described in section 5), then the results of using the Kelly-Daly contrast sensitivity function model to determine the maximum desired velocity are presented (section 7.2, method described in section 6). Daly's model of eye tracking is then incorporated into the calculations, to give frame rates that should eliminate blur from tracked motion (section 7.3, method described in section 6.4). The limitations of the approach will be discussed in section 8, the most significant being the assumption that no strobing artefacts can be tolerated, which means that the frame rate values should not be interpreted as recommendations for real television systems, but simply as approximations of the frame rate required to remove motion blur in the absence of strobing.

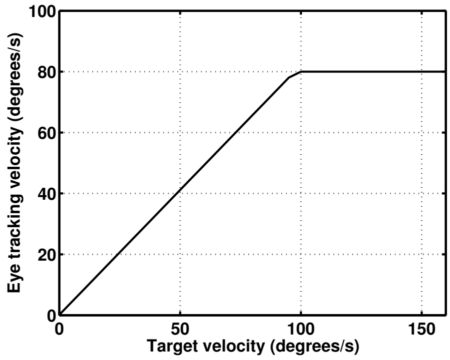
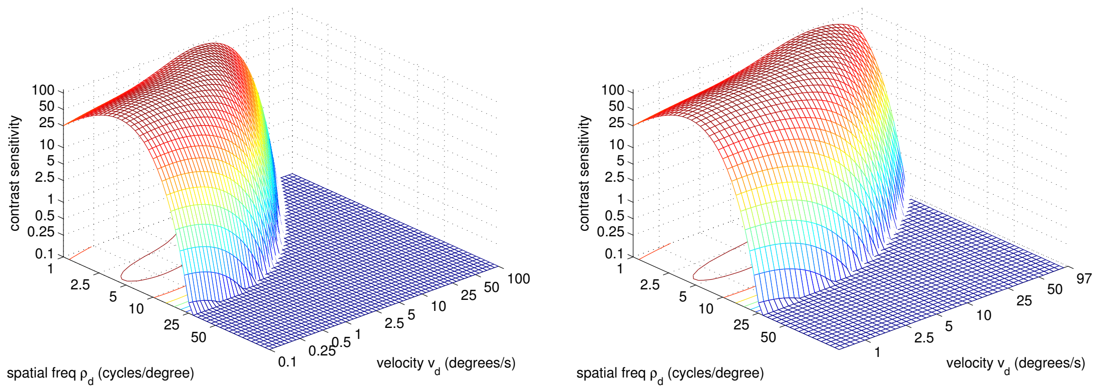


Figure 6: Model of eye tracking from Daly [2001, pp.185–187].



(a) Without eye tracking

(b) With eye tracking

Figure 7: Comparison of contrast sensitivity function without and with modification to include a model of eye tracking.

7.1 Frame Rate Calculated from Sampling Theory Only

The frame rates required to represent objects with different spatial frequencies moving at a range of velocities are shown in figure 8. These frame rates are calculated using only equation 6, and hence they do not take account of the human visual system. The plot shows that the frame rate required to represent a moving object without aliasing increases with the object speed and spatial frequency. The different colours indicate the spatial frequencies and velocities that can be represented with frame rates of 50, 100, 300 and higher than 300 fps. Finer spatial detail, made possible by formats with greater spatial resolution, requires a proportionally higher frame rate for a given velocity in pictures per second. The maximum velocities for objects containing the highest spatial frequency in a given pixel format at 50 fps seem quite slow, but it should be noted that in a real system the higher spatial frequencies would be attenuated in moving objects due to the finite camera shutter time. The object, with added motion blur, can therefore move more quickly without aliasing than these calculations indicate.

It is valid to compare the calculations for different spatial formats if a given real object velocity remains constant in pictures per second as the number of horizontal pixels changes. However, the velocity in pictures per second would not remain the same if, for example, the camera operator tends to zoom out of a scene when the target format is known to have a high spatial resolution, in effect using the extra pixels for a larger scene area rather than for extra resolution. This point is discussed further in section 8.5.

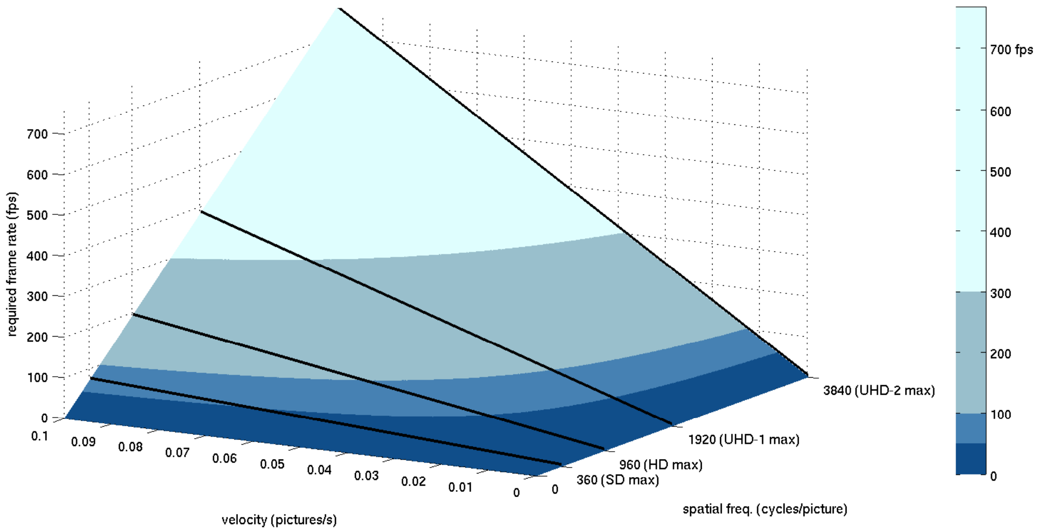


Figure 8: Frame rates required to represent a range of horizontal frequencies moving at different velocities, without aliasing. The four thick lines indicate the maximum spatial frequency (half the number of horizontal pixels) for some common spatial formats.

Figure 8 gives the frame rates needed for a range of velocities, but does not provide an upper limit on the desired velocity. For this the contrast sensitivity function model is used.

7.2 Frame Rate Calculated using the Contrast Sensitivity Function Model to Limit the Representable Velocity

Figure 9 shows the frame rate required to perceptually match a range of horizontal resolutions by limiting the maximum object velocity using the method described in section 6, for viewing distances of 3 H, 1.5 H, and 0.75 H, with no eye tracking. It was assumed that the viewer is looking at the centre of the screen, i.e. $\theta_r = 0$. The contrast sensitivity function data on which the model is based is limited in both temporal and spatial frequency, and hence the required frame rate for horizontal resolutions above 2048 pixels viewed at 3 H and above 4096 pixels viewed at 1.5 H could not be calculated: extrapolating the mathematical model is not valid since there is no evidence that the same model shape would apply beyond the data range tested. Recommended viewing distances are defined such that the width of a pixel is at the limit of human spatial resolution [ITU-R, 2012c, p. 3], and hence sitting further from the screen to obtain a finer pixel structure would offer no perceived improvement in spatial detail. The calculations have therefore been limited to viewing distances at or closer than the recommended viewing distances.

For 720 horizontal pixels viewed at 3 H the required frame rate is a little below 50 fps, assuming that there is no aliasing and no eye tracking. The frame rate then increases approximately linearly with the number of horizontal pixels up to around 140 fps for HD with 1920 horizontal pixels. Given that the viewer is much closer to the screen for UHD-1 and UHD-2, with recommended viewing distances of 1.5 H and 0.75 H respectively also designed to put the pixel structure at the limit of spatial resolution for the human eye, the data suggests that with no eye tracking 140 fps would also be suitable for these formats.

A frame rate of 140 fps fits with the literature discussed in section 3, that suggests that a great improvement in quality can be obtained by increasing frame rates to a value in the region of 100 to 150 fps. It should be noted that in those experiments, there is likely to have been some degree of strobing in the signals, and viewers were in general permitted to move their eyes freely. Extending the contrast sensitivity function model to take account of eye tracking will lead to a

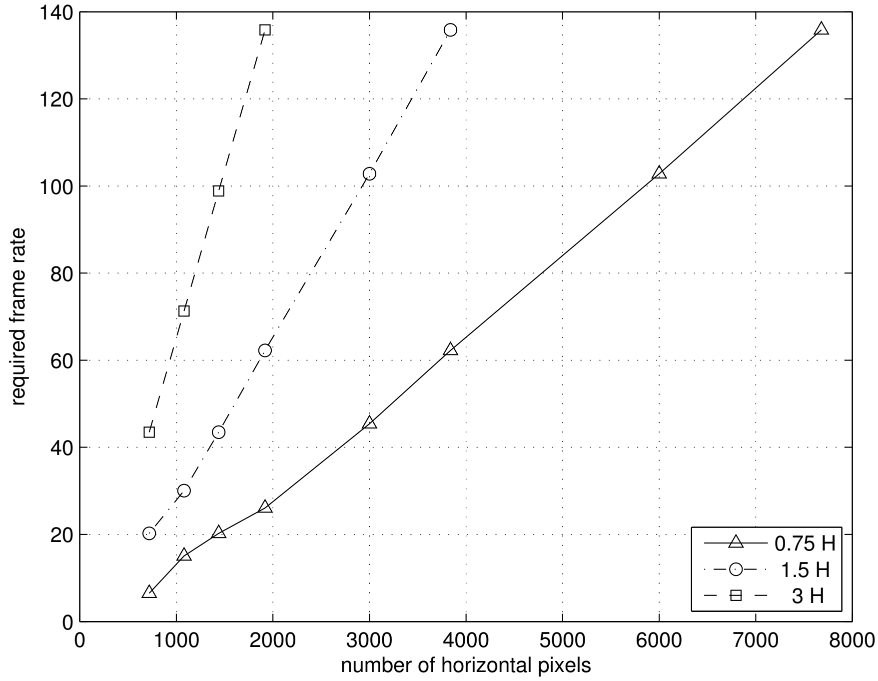


Figure 9: Frame rate required to perceptually match various common horizontal resolution values, with viewing distances of 0.75 H, 1.5 H and 3 H and a fixed gaze, assuming no strobing can be tolerated.

need for yet higher frame rates, that could account for the further quality improvements detected experimentally using a frame rate of over 200 fps.

7.3 Frame Rate Calculated using the Contrast Sensitivity Function Model with Incorporated Eye Tracking

Figure 10 shows the calculated frame rates according to the contrast sensitivity model with incorporated eye tracking. The frame rates shown are simply scaled versions of those shown in figure 9, limited by the maximum eye tracking velocity of 80 degrees per second. The figures demonstrate just how important an effect eye tracking is in the discussion of motion portrayal and frame rates. Frame rates of up to 700 fps are now coming into consideration.

8 Discussion

The results presented in this paper are an indication of the kind of frame rates that might be needed to eliminate motion blur if standard sampling theory is adhered to, such that all aliasing in the traditional sense is removed by a low pass anti-alias filter. This means that the frame rate values primarily relate to the capability of the human visual system to perceive motion blur caused by removing high temporal frequencies from the signal. The frame rates suggested to eliminate blur for tracked motion are clearly higher than is currently practicable to implement, and there are many other considerations to take into account, discussed in this section, which mean that the values calculated are only one factor amongst many in the wider discussion of television frame rates. They are not the result of a complete analysis of all the complexities of television systems and human vision.

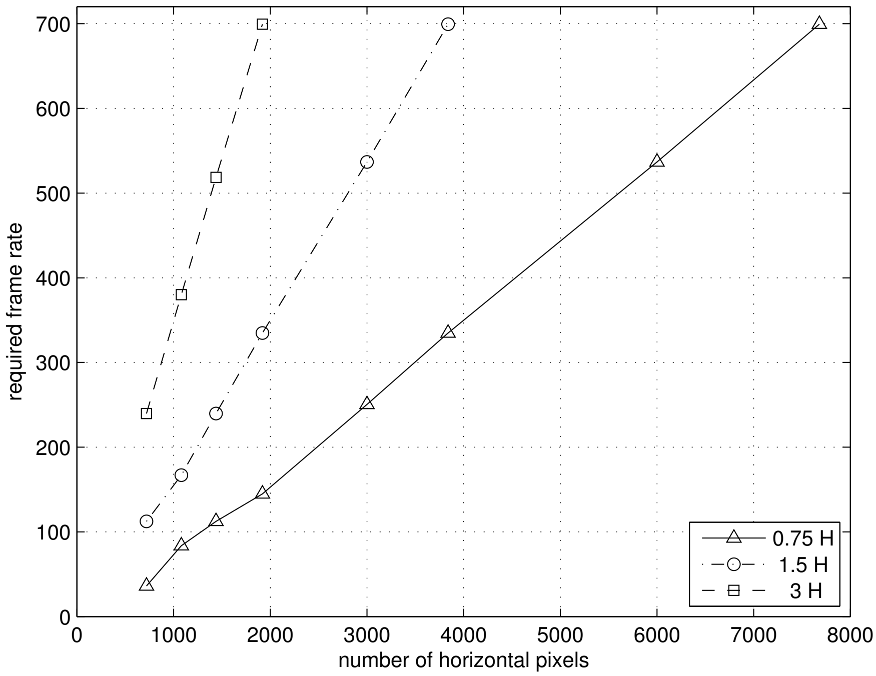


Figure 10: Frame rate required to perceptually match various common horizontal resolution values, with a viewing distance of 0.75 H, 1.5 H and 3 H, during eye tracking, assuming no strobing can be tolerated.

8.1 The Importance of Strobing Perception

The most important factor missing from the analysis presented is an investigation into what happens when traditional sampling theory is violated, and some aliasing is permitted in the signal. Television has always operated under these conditions [Drewery, 1995], and the reason it is not more objectionable is related to eye tracking. The moving eye is an important part of video signal reconstruction, and is not taken into account by traditional sampling theory, which requires no aliasing only from the perspective of a fixed position on the screen. As described in section 2, when the eye moves to follow an object, the relative velocity between the eye and the object is reduced. This can be interpreted as the human visual system doing a form of motion-compensated interpolation by tracking the object with the eyes: parts of the signal that would traditionally be called aliases are interpreted as part of the baseband as a result of the eye movement. This means that, for tracked motion, aliasing in the signal is not objectionable, and at low frame rates it is even desirable when the alternative is severe motion blur.

However, allowing aliasing by impulsive sampling without an anti-alias filter is not a universal solution, because only one object at a time can be tracked. Aliases in the signal are revealed in background objects, objects moving at different speeds, and rotating objects, and can appear as strobing, multiple imaging or judder. Table 2 summarises the visible motion artefacts for tracked and untracked motion under the two kinds of sampling discussed so far. It is possible to achieve a balance of blur and strobing between these two extremes using an anti-alias filter that attenuates potential aliases to a greater or lesser degree. In order to find the right balance, tests of the human visual system are needed to establish the visibility of strobing artefacts at different frame rates and for different types of motion. The results may vary for different types of content—for example, strobing may not be problematic in a drama production where the background is de-focussed and the foreground contains little motion, but it may be highly objectionable during a camera pan in a football match when the strobed motion of the crowd in the background fills the screen.

	Nyquist-Shannon sampling	Impulsive sampling (no anti-alias)
Tracked motion	No strobing, may need **700 fps** to eliminate blur	No blur, low sensitivity to strobing, may only need to meet the critical flicker frequency of **80 fps**
Untracked motion	No strobing with fixed gaze, may need **140 fps** to eliminate blur	No blur, high sensitivity to strobing, required frame rate **unknown**

Table 2: Comparison of blur and strobing artefacts with traditional Nyquist-Shannon sampling (no strobing) and impulsive sampling (no motion blur). A short display aperture time is assumed.

When discussing untracked motion the calculations presented in this paper have assumed that the gaze is fixed. An additional, subtle point is that even when traditional sampling theory is adhered to, eye movements can cause a strobing effect to be perceived in untracked motion. If an untracked object is moving in the opposite direction to the eye, the relative velocity between the object and the eye will be greater than the absolute velocity of the object. The positions of the object from one frame to the next will hence be further apart on the retina than if the eye were still, and hence a higher sampling frequency than otherwise expected would be needed to completely avoid the perception of aliasing.

8.2 Camera and Display Effects

The calculations presented have assumed Nyquist-Shannon sampling, with an infinitely short shutter and no spatial or temporal aliasing in the signal, followed by a display with an infinitely short aperture and an ideal reconstruction filter. In section 8.1 the possibility of balancing ideal anti-alias filtering with a degree of strobing was also suggested. In reality the non-band-limited image of a scene will be passed through a camera aperture that integrates over space and time, and through a display that is likely to have a sample-and-hold response in both space and time. Both of these processes amount to gentle low pass filtering, so will reduce aliasing to some degree but not entirely, as well as attenuating higher passband frequencies. A balance of attenuation and strobing can be achieved by varying the aperture time at the camera or display, or by using a more sophisticated oversampling approach. Analysis of these alternative capture and display processes is left to future work.

8.3 Limitations of the Contrast Sensitivity Function Model

The data used to define the contrast sensitivity function model were collected under restricted experimental conditions, which means that the model cannot represent the human sensitivity to all video patterns in a television signal. The data provides information about the just visible threshold of contrast sensitivity, but it does not describe how the contrast sensitivity function might vary for greater levels of contrast. A set of equal brightness surfaces similar to the equal loudness curves developed for audio (reproduced in [Watkinson, 2004, p. 182]), that describe how the visual contrast sensitivity varies with brightness, would allow greater accuracy. It is also likely that the light level of the surroundings will have an effect on the shape of the contrast sensitivity function, given that it is known to affect perceived lightness [Hunt, 1987, p. 40], and that visibility of noise has been found to vary with surrounding luminance [Roberts, 2009, pp. 294–295].

The model also does not take into account any changes in the contrast sensitivity that may occur in peripheral vision. It is generally accepted that, for television brightnesses and flicker frequencies, humans are more sensitive to flicker at the edges of our field of view, although this is a complex subject with measured flicker sensitivity varying depending on flicker brightness, surround brightness, flicker frequency, stimulus size, and colour [Roberts, 2009, p. 291], [Swanson et al., 1987]. Since flicker and motion are both fundamentally temporal percepts, it is likely that our motion sensitivity will also be affected by an object's location within the field of view. Sensitivity in peripheral vision will become increasingly important as very large screens become more common.

The CSF data is based on sensitivity to monochrome stimuli, and assumes that sensitivity to vertical and horizontal motion are equivalent. These two simplifications are likely to be acceptable, given that humans have greater sensitivity to brightness variation than to colour [Hunt, 1987, p. 382], and that differences between horizontal and vertical resolution are likely to consist of only a small difference in eye tracking capability, based on the physiology of the human eye [Daly, 2001, p. 197].

8.4 Limitations of the Eye Tracking Model

The model of eye tracking used [Daly, 2001, p. 186] is also based on experimental data, but in this case using only five subjects. Therefore there may be some inaccuracy in the mean eye tracking velocity of 82 % of the object velocity, and in the maximum eye velocity of 80 degrees per second. The frame rates calculated for tracked motion are highly sensitive to variations in the 82 % figure, since it appears in the denominator of the frame rate scaling factor, $1/(1-0.82)$, but this sensitivity is tempered by lower sensitivity in the maximum trackable velocity of $(80/0.82)$ degrees per second, which puts an upper limit on the frame rate. Nonetheless, the figure of 700 fps should be regarded as approximate.

8.5 Discussion of the Effects of Viewing Distance and Camera Zoom

The observation from figures 9 and 10 that the required frame rate does not increase as the viewing distance decreases is somewhat counter-intuitive. It certainly does not mean that a higher frame rate will produce no visible improvements for close viewing. Rather, it shows that the spatial artefacts are more visible close-up, so the frame rate need not be as high to perceptually match the lower spatial resolution at the eye. A higher frame rate would almost certainly bring perceptible improvements to the video when viewed from a short distance, but would provide temporal resolution that is perceptually better than the spatial resolution.

The reason for the apparent paradox is that the calculations are agnostic to the video content, they simply allow everything that is as easy to see as the highest possible spatial frequency to be represented. Of interest to the viewer, however, is the detail visible on a particular object at different distances and in different formats. The difficulty arises because, in stepping back to view the screen from a greater distance, a particular important piece of video information, such as the detail required to just recognise a face, does not retain the same spatial frequency at the eye. The spatial frequency at the eye increases, so the detail becomes harder to see, hence it may be intuitive to assume that a lower frame rate is required. Indeed, any motion in the scene will be slower in units of degrees per second at the eye when viewed from a greater distance. The calculations in this paper do not take account of this variation in spatial frequency and movement for particular content as the viewing distance changes. The three lines in figures 9 and 10 relating to different viewing distances show only the frame rate needed to match the highest possible spatial frequency of a particular pixel format, regardless of how often that frequency occurs.

Similarly, the results at any one of the three viewing distances are agnostic to any change in zoom level between different resolutions, which may affect the extent to which the spectrum is filled up to the maximum spatial resolution, depending on the fractal dimension of the scene. If the camera is zoomed out for high spatial resolutions, there may be more high spatial frequencies in the signal, so a matching higher frame rate may be desirable. However, this is counterbalanced by the fact that any motion in the same scene will be lower in units of pictures per second, making a higher frame rate less important: a footballer running at a constant speed will have a higher velocity in pictures per second when the camera is zoomed in than when the camera is zoomed out. Regardless of the effects of zooming, it is valid to compare the frame rate values at a particular viewing distance for the case where material is recorded in a UHD format and viewed on a range of different displays: a likely use case for UHD in the coming years.

8.6 Depth of Field

One possible solution that minimises motion blur in tracked motion at the same time as minimising strobing for untracked motion at a moderate frame rate would be to develop an adaptive system, that imitates a short camera aperture for tracked objects and a long aperture for untracked objects. This kind of approach has a number of problems, one of them being that it is not possible to predict which objects a viewer will track.

However, a relative of this technique has been used for many years in drama and movie productions. A lens with shallow depth of field is used together with a short camera aperture, which means that foreground objects are in focus and do not suffer from motion blur, whilst background scenery is defocussed (blurred) and so does not suffer from strobing. This technique is only suitable for programmes where it is not necessary to be able to resolve the detail in the background. Whilst it works for drama, and can even assist the storytelling by guiding the viewer towards the foreground, it would not be appropriate for sport where it is desirable to be able to focus on objects at widely varying distances from the camera. Hence different solutions may be desirable for different programme genres.

8.7 Assumptions in the Approach to Calculating the Frame Rate

It is also important to discuss the overall approach taken to determining an appropriate frame rate. The basis of the work is that the parts of the video spectrum that are easiest to see are also the most important. This ignores any unevenness in how the signal is distributed, treating all in-band frequencies equally. It may however be the case that, for example, very high temporal resolution is rarely needed, in which case it could be argued to favour spatial over temporal resolution. Alternatively, the sampling format could be adapted to the programme genre, using high frame rates only when they are really required.

A fundamental assumption of the work is that when two artefacts are equally perceptible, they are also equally objectionable. This is an assumption that is widely used, and has led to development of highly successful perceptual video compression techniques [Watkinson, 2004, pp. 8–9] as well as being implicit in the ITU five-grade impairment categories [ITU-R, 2009, p. 18] that include "perceptible" and "annoying" as different points on the same quality scale.

It has also been assumed that there is no trade-off between spatial and temporal resolution, i.e. that temporal artefacts cannot be masked by good spatial resolution, and vice versa. Informal viewing suggests that this assumption is valid, but a complex series of subjective tests would be needed to confirm it conclusively.

Equations 17 and 20 additionally assume a 16:9 aspect ratio, but this is used only in the definition of the viewing distance, and is easy to remove from the calculations if desired.

8.8 Other Considerations

Further considerations for determining an appropriate frame rate include noise levels in the signal, which have not been taken into account, and may vary depending on frame rate and spatial resolution. Although with a high frame rate, or short frame period, there will be less light reaching the camera sensor, it is possible that humans have a higher noise tolerance at high frame rates due to an integrating effect in the human visual system. This would need to be confirmed with subjective tests, and the point found at which the detrimental effects of increased noise become more important than the benefits obtained by higher frame rates.

It has also been assumed that the scene is illuminated by a continuous light source. Most television productions will make use of artificial light that is modulated, some at twice the local mains frequency (tungsten lights), and others at much higher frequencies (LEDs or other types of high-frequency lighting). When the frame period does not contain a whole number of periods of the lighting pattern, both the brightness and colour of the captured frames can vary. Even if the changes are too fast to be seen, they will have detrimental effects on compression algorithms that rely on consecutive frames being similar.

In a practical system there will also be a need to convert between frame rates, for example to deliver legacy content through a new UHD transmission chain. The extent of temporal aliasing in existing television signals makes simple frame rate conversion to non-integer multiples of the original rate difficult, resulting in a strong preference for any new frame rate to be an integer multiple of existing ones. Conversion difficulties should, however, be eased for signals captured at higher frames rates, since they will not be so severely undersampled, hence interpolation will be more effective.

The overall data rate is clearly also an important consideration, and if higher frame rates are introduced compression systems will become ever more important. HEVC (high efficiency video coding) promises improved coding efficiency [Sullivan et al., 2012], and it is likely that increasing the frame rate will not require a proportional increase in bit rate, since consecutive frames have greater similarity if they are closer in time, and a lower proportion of intra-coded frames would be needed to maintain a given maximum time to start decoding a stream. Initial tests by the European Broadcasting Union's Broadcast Technology Futures group suggest an average bit rate increase of only 8 % when upgrading from 60 to 120 fps, using the reference HEVC implementation [Gabriellini, 2014]. Nonetheless, within a limited bit budget the value of a higher frame rate must also be judged in comparison to the value of other potential enhancements that require extra data, such as higher dynamic range or a wider colour gamut.

9 Conclusion

High frame rate television is one factor under discussion as a potential key component for any public service broadcast of ultra-high definition television. A review of earlier work into television frame rates revealed a consensus that subjective quality improvements can be made with a higher frame rate, but estimates of the precise frame rate required vary considerably. One reason for the variation is that in subjective tests, the most reliable way of judging perceived quality, it is difficult to separate capture and display artefacts, and to control for viewer's eye movements which can have a significant effect on motion perception.

In this paper, to complement subjective test results, a theoretical study was presented that applied traditional sampling theory to the temporal sampling of video signals. It was assumed that the presence of aliases in the signal can cause strobing effects in untracked motion. Traditional sampling theory requires that all aliases are removed, and hence if traditional sampling theory is adhered to there should be no strobing effects visible.

Sampling theory alone is not able to place a limit on the frame rate required, because it is not possible to know in advance the maximum object velocity that could occur in live television. A model of human perception was used to place an upper limit on velocity derived from human sensitivity to moving detail: the frame rate was set such that any moving detail that humans can see as clearly as the finest static detail in a particular spatial format was maintained. This effectively matches the anti-alias filter at the capture stage to the reconstruction filter that is the human eye, capturing the parts of the video signal that can be perceived and rejecting everything else.

The first stage of the analysis assumed a fixed gaze, and found that a frame rate of 140 fps may be needed to prevent any strobing effects in untracked motion at the same time as maintaining detail in moving objects. However, whilst a stationary eye can be modelled as a conventional reconstruction filter, a tracking eye is able to reduce strobing effects in the object it is tracking, but not in other parts of the scene, with sensitivity to motion blur also increasing for tracked objects. This means that eye movements need to be taken into account when considering normal viewing conditions: any system that relies solely on models of a stationary eye is likely to fail. An additional model of eye tracking was incorporated into the analysis to find the frame rate that would be needed to prevent all aliasing and maintain moving detail in both tracked and untracked objects when the eye is free to move—approximately 700 fps.

This work has assumed that all aliasing in the signal is undesirable. However, we know that the human visual system is able to tolerate some strobing effects, particularly in tracked motion. A relaxation of the requirement for zero strobing would allow the frame rate to be reduced, but to understand how far it can be reduced it will be necessary to quantify the trade-off between motion blur and strobing. This is an important area for further study.

In addition to the analytical work, this paper presented a discussion of related practical considerations that would need to be taken into account before recommending any specific parameters. Any new production grammar that may emerge for UHD, such as a tendency for wide shots in preference to close-ups, would need to be understood, as well as any differences in requirements for different programme genres. Technical issues such as lighting interference, ease of format conversion and bit rate restrictions would also need to be considered.

This work is a significant step towards understanding the capabilities of the human visual system in resolving moving detail. The frame rates presented are not intended to be practical recommendations, but can be used to steer subjective test design and interpret subjective test results, and will be valuable for informing discussions on frame rate standardisation.

10 Acknowledgements

Particular thanks to Graham Thomas, Andrew Cotton, Tim Borer and Matthias Mauch for their contributions to this work.

References

R. T. Apteker, J. A. Fisher, V. S. Kisimov, and H. Neishlos. Video acceptability and frame rate. *IEEE MultiMedia*, 2(3):32–40, Autumn/Fall 1995.

M. Armstrong, D. Flynn, M. Hammond, S. Jolly, and R. Salmon. High frame-rate television. Technical Report WHP 169, BBC Research White Paper, September 2008.

Peter G. J. Barten. *Contrast Sensitivity of the Human Eye and its Effect on Image Quality*. SPIE Optical Engineering Press, 1999.

S. Daly. Engineering observations from spatiovelocity and spatiotemporal visual models. In Christian J. van den Branden Lambrecht, editor, *Vision Models and Applications to Image and Video Processing*. Kluwer Academic Publishers, 2001.

J. O. Drewery. TV-the grand illusion. In *IEE Colloquium on Motion Reproduction in Television*, pages 1/1–1/7, May 1995.

D. Driesnack. Another important "H". *EBU Tech-I Magazine*, 18:16, December 2013.

M. Emoto and M. Sugawara. Critical fusion frequency for bright and wide field-of-view image display. *Display Technology, Journal of*, 8(7):424–429, July 2012.

A. Gabriellini. HFR and video compression. *EBU Tech-I Magazine*, 19:7, March 2014.

R. W. G. Hunt. *The Reproduction of Colour in Photography, Printing & Television.* The Bath Press, 4th edition, 1987.

Quan Huynh-Thu and M. Ghanbari. Temporal aspect of perceived quality in mobile video broadcasting. *IEEE Transactions on Broadcasting*, 54(3):641–651, September 2008.

ITU-R. Parameter values for the HDTV standards for production and international programme exchange. Recommendation ITU-R BT.709-5, April 2002.

ITU-R. Methodology for the subjective assessment of the quality of television pictures. Recommendation ITU-R BT.500-12, 2009.

ITU-R. Parameter values for ultra-high definition television systems for production and international programme exchange. Recommendation ITU-R BT.2020, August 2012a.

ITU-R. The present state of ultra high definition television. Report ITU-R BT.2246-1, August 2012b.

ITU-R. General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays. Recommendation ITU-R BT.2022, August 2012c.

D. H. Kelly. Motion and vision. II. Stabilized spatio-temporal threshold surface. *J. Opt. Soc. Am.*, 69(10), October 1979.

Y. Kuroki, T. Nishi, S. Kobayashi, H. Oyaizu, and S. Yoshimura. A psychophysical study of improvements in motion-image quality by using high frame rates. *Journal of the Society for Information Display*, 15(1):61–68, January 2007.

J. Laird, M. Rosen, J. Pelz, E. Montag, and S. Daly. Spatio-velocity CSF as a function of retinal velocity using unstabilized stimuli. In *Proc. SPIE-IS&T Electronic Imaging*, volume 6057. The International Society for Optical Engineering (SPIE), 2006.

John D. McCarthy, M. Angela Sasse, and Dimitrios Miras. Sharp or smooth? Comparing the effects of quantization vs. frame rate for streamed video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*, pages 535–542. ACM, 2004.

H. Nyquist. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644, April 1928.

Yen-Fu Ou, Zhan Ma, Tao Liu, and Yao Wang. Perceptual quality assessment of video considering both frame rate and quantization artifacts. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(3):286–298, March 2011. ISSN 1051-8215.

A. Roberts. *Circles of Confusion.* EBU, 1st edition, 2009. ISBN 978-2-8399-0574-9.

J. G. Robson. Spatial and temporal contrast-sensitivity functions of the visual system. *J. Opt. Soc. Am.*, 56(8):1141–1142, 1966.

R. A. Salmon, M. G. Armstrong, and S. J. E. Jolly. Higher frame rates for more immersive video and television. BBC R&D White Paper 209, October 2011.

C. E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3, 4):379–423, 623–656, 1948.

M. A. Stone. A variable-standards camera and its use to investigate the parameters of an HDTV studio standard. BBC Research Department Report RD 1986/14, December 1986.

G.J. Sullivan, J. Ohm, Woo-Jin Han, and T. Wiegand. Overview of the High Efficiency Video Coding (HEVC) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12):1649–1668, Dec 2012.

William H. Swanson, Takehiro Ueno, Vivianne C. Smith, and Joel Pokorny. Temporal modulation sensitivity and pulse-detection thresholds for chromatic and luminance perturbations. *J. Opt.Soc. Am.*, 4(10):1992–2005, October 1987.

N. E. Tanton and M. A. Stone. HDTV displays: subjective effects of scanning standards and domestic picture sizes. BBC Research Department Report 1989/09, January 1989.

G. J. Tonge. Time-sampled motion portrayal. In *Proceedings of the Second International Conference on Image Processing and its Applications*, London, June 1986.

John Watkinson. *The MPEG Handbook*. Focal Press, 2nd edition, 2004.

Andrew B. Watson. High frame rates and human vision: A view through the window of visibility. *SMPTE Motion Imaging Journal*, pages 18–32, March 2013.