SMPTE Meeting Presentation

How Independent are HDR, WCG, and HFR in Human Visual Perception and the Creative Process?

Sean T. McCarthy, Ph.D.

ARRIS Group

Written for presentation at the

SMPTE 2015 Annual Technical Conference & Exhibition

Abstract. *UHD, HDR, WCG, HFR are bound to be powerful creative tools with which to engage the viewer. Such acronyms (and would-be logos) could also prove to be influential marketing aids. But how justified would standards bodies, content creators, and distributors be in thinking of each feature as independent?*

This paper will provide principles of applied vision science to quantify the extent of interdependence of luminance, field–of-view, color perception, and temporal sensitivity. This paper will also identify situations in which luminance, color, and frame rate should perhaps be considered in concert rather than as independent creative dials.

Keywords. Ultra HD, UHD, vision science, visual perception, high dynamic range, HDR, color gamut, WCG, high frame rate, HFR, creative intent, content distribution

Introduction

Ultra HD TV is more than just higher resolution. It can be thought of as a step towards enabling digital media to seem almost as real as the real world. When the International Telecommunication Union (ITU-R) set out to define the parameters[1,2] of Ultra HD, a motivation was to provide viewers with an experience that virtually covers all of the human visual field[3]. Covering the visual field means going beyond HDTV resolutions[4,5], which is where most conversations about Ultra HD began and ended, at least until recently. More and more, Ultra HD is also about expanded color spaces that can match real world colors; enhanced luminance and contrast that can convey the natural sense of light in a scene; and higher frame rates that would make motion feel real even for high action sports. Each of these components – resolution, field-of-view, color space, luminance, contrast, and frame rate – are tools that creatives can use to help viewers connect more intimately with stories, characters, and emotions.

We as engineers can easily think of resolution, color space, luminance, and frame rate as independent technological features. Yet, the human vision system makes no such stark distinctions. Perception of motion is influenced by display size, viewing distance, frame rate, refresh rate, luminance, and the current adaptation state of the photoreceptors in the retina[3,5-7]. Similarly, perceived hue is not uniquely defined by the spectral composition of the light coming from a display: It also depends on luminance, adaptation, and the composition of the scene[8-13]. Every aspect of physical stimuli influences the whole of a viewer's perception, awareness, and experience.

The aim of this paper is to provide concepts from vison science in the context of Ultra HD with specific attention to potential perceptual interdependences of the emerging technologies of High Dynamic Range (HDR), Wide Color Gamut (WCG), High Frame Rate (HFR), and Wide Field-of-View (WFOV). A main motivation is to help further conversations around creative intent and engineering design so that the development of technical standards and the ongoing commercialization processes for Ultra HD result in great consumer experiences.

In particular, this paper explores the following:

- The interaction of field-of-view and frame rate on smooth high-acuity motion tracking
- The interaction of luminance and screen size on flicker perception
- The interaction of luminance, perceived contrast, and color appearance
- The impact of speed of visual adaptation on scene changes, program changes, and commercials

Smooth High-Acuity Motion Tracking, Frame Rate, and Display Size

Several empirical subjective studies[3,5] have shown that frame rates higher than those used previously in HD production can significantly affect, and often improve, a viewer's experience, particularly for high-action and sports content having a lot of motion. On the other hand, the amount of raw spatial detail in a frame – the sharpness of the image – can be significantly limited by motion blur[14-16] created in the camera and/or in the display.

Display size, and consequently field-of-view, can also have a large impact on the viewer's experience. An object in motion will move a larger distance between frames on a larger display than on a smaller display when all other things such as viewing distance and frame rate are the same. At some point, object motion will be so great that a viewer's eye cannot keep up and the viewer will perceive a series of still images (stroboscopic effect, jerkiness) rather than smooth motion.

Ideally, the creative process would be free to choose frame rate, spatial sharpness, smooth motion, and stroboscopic motion independently to deliver the artistic intent; but, when is a person actually able to perceive high-spatial detail for moving features? What are the viewing conditions in which smooth motion could become unintended stroboscopic motion, or *vice versa*?

This section aims to provide visual performance data to aid in being more quantitative about the perception of smooth motion and the impact of frame rate, display size, and field of view.

Retinal Topography & Visual Acuity

Retinal topography is illustrated in Figure 1. The retina is made of layers of specialized neurons, each having a unique critical role in vision[17,18]. The light-gathering layer is the photoreceptor layer. It is constructed of different kinds of light-sensitive cells: long-wavelength- (L, "red'), medium-wavelength- (M, "green"), and short-wavelength- (S, "blue") sensitive cones, and achromatic rods. Cone photoreceptors dominate the central parts of the photoreceptor layer and are primarily responsible for high-acuity daylight (photopic) viewing. Rod photoreceptors dominate the periphery and are responsible for dark-adapted (scotopic) vision.

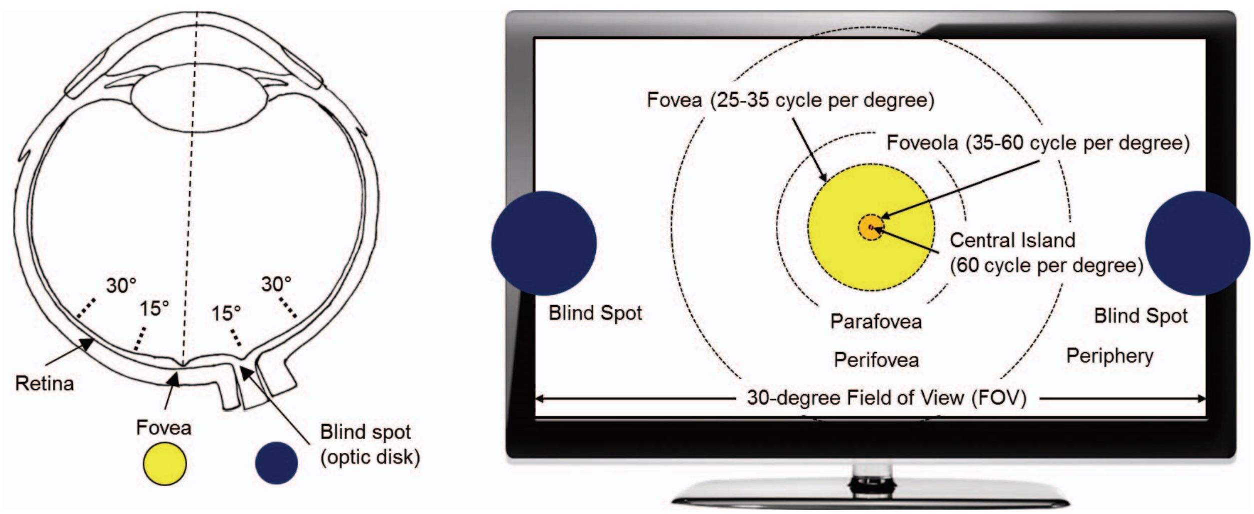


Figure 1. Retinal topography projected onto a display subtending 30 degrees of visual field horizontally. Smaller Ultra HD displays and tablets can also be expected to subtend approximately 30 degrees in normal viewing conditions. Note that the left- and right-eye blind spots flank the borders of the HDTV display when a viewer gazes at the center of the display.

The topography of the retina is more nuanced than a blunt demarcation between high-acuity center and scotopic periphery[19,20]. Figure 1 illustrates concentric zones of retinal topography and notes the maximum resolvable spatial frequency for each zone. The tiniest center of the retina contains only L & M cones (no S “blue” cones or rods) and is the center of our visual field. Polyak called this the central island[21]. The central island corresponds to only approximately 0.2 degrees (12 min of arc) of the visual field, yet it has the highest density of photoreceptors and is thus responsible for our most acute vision. The foveola, which contains the central island, spans approximately 1.2 degrees of visual angle. Being free of rods and blood vessels, the foveola also supports very high cone density and high-acuity vision. Encompassing the foveola is the fovea, which spans approximately 6 degrees of visual field. The fovea is composed of a mix of cones and rods, with cones becoming scarcer and rods becoming denser farther from the center. Moving outward from the fovea, the cone density continues to decline, as does photopic (daylight & color) acuity.

Cone photoreceptors in the central island of the foveola are arranged in a close-packed hexagonal array. The spacing of cones is approximately 30 seconds of arc of the visual field (~2 cones per minute of arc), which translates to ~120 cone photoreceptors per degree. According to Nyquist sampling theory[22], the maximum theoretically spatial frequency that can be encoded by the retina would thus be about 60 cycles/degree; i.e., ½ the spatial sampling frequency. Subject experiments have shown that humans are able to achieve this theoretical “simple” acuity limit in highly controlled test conditions[23] using interference patterns. In actual experience, the ability to resolve small details extends to lower and, perhaps surprisingly, to higher acuity levels depending on the visual task. Reading acuity – so called 20/20 vision as defined by Snellen[24] – relates to the ability to recognize symbols and their orientation. 20/20 acuity corresponds to a spatial frequency of 30 cycles/degree[25].

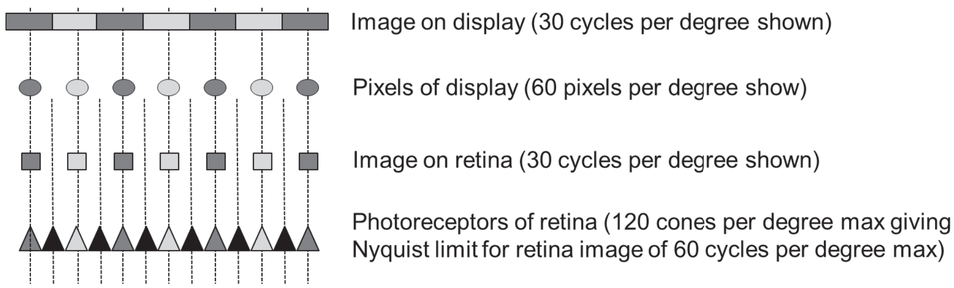
Hyperacuity[26, 27] is the ability to notice details seemingly beyond the Nyquist limit such as the misalignment of line segments (called vernier acuity) in which the misalignment is smaller than the diameter of a photoreceptor. Hyperacuity does not actually violate the Nyquist limit, rather it is a result of the visual brain processing the collective activity of many photoreceptors. Hyperacuity and Snellen acuity relate to the brain’s interpretation of light stimuli whereas the Nyquist limit relates to signal processing constraints on data.

Field-of-View & Visual Angle

High-definition television was defined for viewing distances of 3-times picture height[28], which corresponds to a horizontal field-of-view of approximately 30 degrees. For an HD display having 1920x1080 pixels, the horizontal pixel density is thus approximately 60 pixels per visual degree as imaged on the retina, matching the theoretical Nyquist resolution limit of the retina’s central island.

For an Ultra HD display having 3840x2160 pixels, the proposed optimal field-of-view corresponds to 60 degrees[3,29] (1.6-times picture height), and thus also 60 pixels per visual degree as imaged on the retina. For smaller Ultra HD TVs and tablets viewed at a 30-degree field of view, the result would be approximately 120 pixels per visual degree.

It is worth noting that the retina's Nyquist limit is not the only Nyquist limit to be considered[30]. Whereas the density of photoreceptors sets a Nyquist limit for the maximum spatial frequency of the retinal image that can be sampled by the visual system, the pixel density of the display sets a Nyquist limit for the maximum spatial frequency of the image content rendered by the display (see Figure 2). For 60 display pixels per visual degree, the displayed image can contain spatial frequencies up to 30 cycles per degree; i.e., ½ the display pixel density. As noted in Figure 2, the Nyquist limit for an image rendered on a display is approximately equal to the Nyquist limit of a retinal image sampled by the photoreceptors of the central island for a "4k" (3840x2160) display subtending 30 degrees, or an "8k" (7680x4320) display subtending 60 degrees of visual field.



Horizontal Resolution (pixels)	Viewing Angle (degrees)	Pixels Density (pixels per degree)	Nyquist Limit for Displayed Image (cycles per degree)	Corresponding Acuity Level
1920	30	64	32	Snellen
3840	30	128	64	Simple
	60	64	32	Snellen
7680	30	256	128	Hyperacuity
	60	128	64	Simple
	120	64	32	Snellen

Figure 2. Illustration of the influence of resolution, field-of-view, and Nyquist limits for displayed images adapted from McCarthy (2012). The amount of spatial frequency information that can be displayed depends on the sampling frequency (pixels per degree) of the display, which depends on resolution and field-of-view. The amount of spatial detail that can enter the visual system depends on the maximum sampling frequency of the photoreceptors of the retina, which is maximum at ~120 cones per degree in the central island. The corresponding acuity level that can be supported for various combinations of display resolution and field-of-view is shown in the right-most column: Snellen corresponds to "reading acuity"; simple corresponds to the retina's maximum Nyquist limit; and hyperacuity corresponds to the ability to perceive even smaller details of extended features. Venier acuity is an example of hyperacuity.

Each eye has a blind spot approximately 15 degrees to each side of the fovea. (The blind spot is the photoreceptor-free optic disk created by the optic nerve as it passes through the photoreceptor layer on the way to the brain.) Together, the right and left blind spots span about 30 degrees.

It is interesting to note that the blind spots tend to flank the edges of an HDTV screen at normal viewing distances. The blind spots would also align with the outer edges of tablets, computer screens, and smaller Ultra HD displays where the screen subtends approximately 30 degrees of visual angle. In such viewing conditions, one can look around the screen by simply moving one's eyes with little or no head motion.

For larger Ultra HD displays (Figure 3) where the screen would subtend the recommended 60 degrees of visual angle[3,29], then the blind spots would be inside the boundaries of the screen. Looking around the screen would thus often involve head motion as well as eyeball rotation.

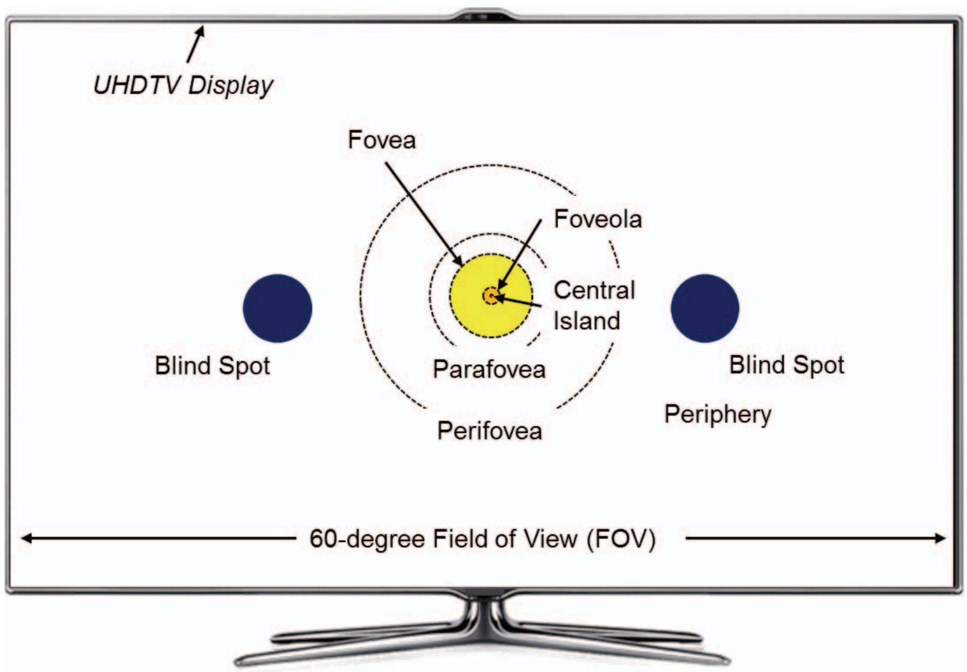


Figure 3. Retinal topography projected onto an Ultra HD display subtending 60 degrees of visual field horizontally. Note the left- and right-eye blind spots are well within the boundaries of the Ultra HD display, which would thus tend to evoke coordinated eye and head motion to view the entire Ultra HD image.

Eye Movements

People move their eyes within the eye socket in various ways[31, 32] that, for the purposes of this paper, can be classified into 3 broad categories:

- **Saccades** are high speed movements (up to 1,000 degrees of visual field per second[33]) in which the center of vision is shifted from one feature of interest to another in jerky succession. We use saccades to inspect the world. Research[33] has shown that gaze shifts of less than 20-25 degrees are typically accomplished using only eye motion (no head motion). Larger gaze shifts result in larger saccades followed by compensating head motion that tends to reorientate the eyes to a head-centered position.

- **Smooth pursuit** is a special kind of eye motion used to track moving objects. People cannot generally move their eyes smoothly within their eye sockets without tracking an object of interest. The maximum speed of smooth pursuit is approximately 40 degrees of visual field per second[34]. This upper limit is important to understanding the impact of frame rate, as discussed below.
- **Catch-up saccades and head movement** are used when an object of interest moves faster than the eye can track smoothly (more that approximately 40 degrees per second). The visual system combines smooth pursuit with directed high-speed catch-up saccades with or without coordinated head movement to minimize retinal slip (the motion of the retinal image relative to the fovea).

The upper angular velocity limit of smooth pursuit relates to the minimum frame rate needed to keep a feature within the high-acuity center of vision without head motion. Subjective tests[35] indicate that movement of the retinal image relative to the retina at a rate greater than 2-4 degrees per second (threshold retinal slip velocity) results in significant deterioration of visual acuity. On a per frame basis for video displayed at 60 frames per second, the threshold retinal slip would be 0.03 to 0.07 degrees per frame (2-4 minutes of arc), or about 15-30% the diameter of the central island.

Table 1. Relationship between frame rate, retinal topology, and spatial acuity at the limit of smooth pursuit (40 degrees per second).

Frame Rate (fps)	Degrees per Frame	Spatial Acuity → Maximum			
		Foveola		Central Island	
		diameter	radius	diameter	radius
24	1.7	X			
30	1.3	X			
60	0.7		X		
120	0.3		X	X	
240	0.2			X	
300	0.1				X

Table 1. provides information about high spatial visual acuity at the limit of smooth pursuit. For various frame rates, a moving feature would move a certain amount (degrees per frame) that can be compared to the size of the specialized regions of the retina. For example, a feature moving at 40 degrees per second and captured at 60 frames per second would move approximately 0.7 degrees per frame, which is approximately the radius of the foveola. The same feature captured at 120 frames per second would move approximately 0.3 degrees per frame, which is slightly larger than the diameter of the central island that is responsible for maximum spatial acuity. At 300 frames per second, the feature would move only about the radius of the central island. Use of higher frames rates could help reduce retinal slip and might thus assist a viewer in maintaining maximum visual acuity even for features moving at the limiting angular velocity of smooth pursuit.

In comparison, a feature displayed at 60 frames per second and moving horizontally at 40 degrees per second moves approximately 0.67 degrees per frame, which is about equal to the radius of the foveola (~1.2 degrees diameter). During smooth pursuit, the eyes would move between frames in anticipation of the position of the feature in the next frame in the attempt to keep retinal slip below the threshold required for high-acuity vision. For high motion content, particularly content in which the direction of motion changes often, higher frame rates would reduce how far a feature would move between frames relative the size of the high-acuity regions of the retina (see Table 1.). Thus, higher frame rates could be expected to progressively increase the possibility for the viewer to perceive a high level of spatial detail and smooth natural motion simultaneously.

Although obvious, it is perhaps worth noting that a feature moving completely from one edge of a display to the other edge of a display at the limiting angular velocity of smooth pursuit would take twice as long (~ 1.5 seconds) for a display subtending 60 degrees compared to one subtending 30 degrees and would, thus, provide a different viewer experience.

Note also that content that was created having features that move at the limit of smooth pursuit on a display subtending 30 degrees of visual field would result in motion beyond the limit of smooth pursuit when rendered on a larger display subtending 60 degrees. The viewer experience in the two scenarios could be very different even at high frame rates. The higher angular motion on the wider field-of-view display would tend to evoke catch-up saccades, head motion, and a reduced spatial visual acuity compared to the 30-degree field-of-view situation.

Luminance, Screen Size, and Flicker Perception

The emergence of high dynamic range (HDR) technologies[36-41] provides numerous new creative options. For example, use of the greater absolute luminance available in HDR could be used to affect a viewer's sensation of flicker and temporal change intentionally. In other cases, increased luminance could affect a viewer's experience in a manner that was not intended.

Sensitivity to flicker and temporal variations depends on luminance, field-of-view, frame rate, display refresh rate, and the adaptation state of the photoreceptors.

Figure 4A illustrates the impact of luminance and screen size, respectively, on flicker. The critical flicker fusion (CFF) rate is the temporal frequency above which flicker is perceived as a steady light. CFF may be predicted from luminance with the Ferry-Porter Law[17], which states that the CFF increases in proportion to the logarithm of luminance. The increased sensitivity to flicker with luminance is related to the speed at which the visual system responds to changes in light intensity. As the visual system adapts to brighter stimuli, the speed of response increases and flicker becomes more noticeable.

As illustrated in Figure 4B, flicker sensitivity also depends on the location of the stimulus on the retina and the size of the stimulus; i.e., the field-of-view. The CFF for flicker in the center of vision is lower than in the peripheral at all luminance levels. And the CFF for large stimuli (such as from a UHD display or cinema screen) is higher than for smaller stimuli at all luminance levels. The phenomenon is called the Granit-Harper law.[17]

Given that TV manufacturers have increased frame rates in newer models to 120 and even 240 Hz, there's no reason to expect flicker perception to be caused by HDR display systems. However, the stroboscopic effect of content captured at the 24 frame-per-second (fps) rate used with films and episodic series could be even more noticeable to viewers for 60-degree field-of-view displays (large Ultra HD displays) compared to 30-degree field-of-view (smaller Ultra HD displays, HDTVs, and tablets). The use of frame interpolation technologies in display would tend to reduce stroboscopic effects, but might not be consistent with the original creative intent.

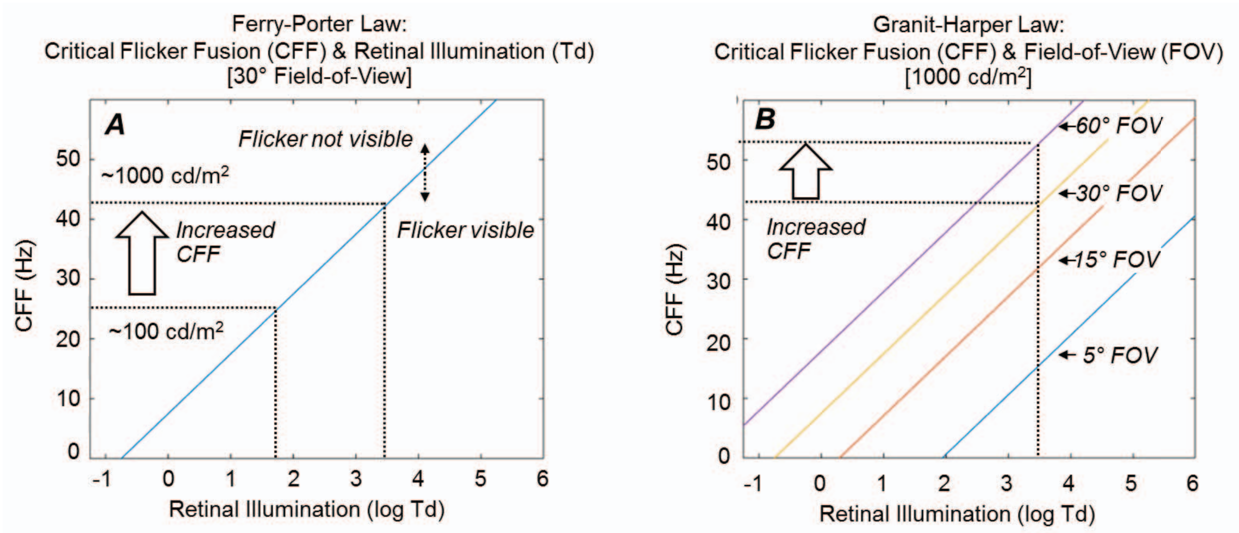


Figure 4. Illustration of the Ferry-Porter law, *A*, and Granit-Harper law, *B*. Brighter HDR displays may be expected to make non-smooth motion and flicker more noticeable. The Ferry-Porter law states that the Critical Flicker Fusion (CFF) rate increases with the log of retinal illumination. The Granit-Harper law states the CFF increases with the size of the retinal image. Thus, larger displays subtending wider field-of-view may be expected to make non-smooth motion and flicker more noticeable. (The horizontal axis is in units of log Trolands (log Td). A Troland is a measure of the illumination of the retina and is equal to luminance (cd/m^2) multiplied by the area of the pupil of the eye. See Table 2.)

Luminance and Contrast and Color Perception

The perception of color and relative lightness is complicated and, as yet, not fully understood. Color cannot be described completely as a set of coordinates on a CIE chromaticity diagram[42], nor simply as a combination of monochromatic stimuli[8]. The development and testing of color appearance theories and models is an area of significant ongoing research.

The wider luminance and color range made possible by HDR and WCG arm the creative process with very powerful tools; yet, as with all things related to perception, an alteration of one variable such a luminance can have spill-overs effects on other variables such as perceived contrast or color appearance.

Adaptiaton to sustained luminance, for example, affects the relationship between perceived brightness and variations in luminance. The Stevens Effect[8,43,44] of adaptation describes, in part, the phenomenon that "…the brightness of any constant luminance decreases with increasing adaptation level."[44] As illustrated in Figure 5, perceived brightness at any adaptation level can be described by a Stevens power law:

$$B = k(L - L_0)^{\beta}$$

where *B* is perceived brightness, β is the power exponent (slope in log-log), *k* is scale factor (intercept in log-log), L_0 is the absolute luminance threshold, and *L* is luminance of the stimulus.

The Stevens effect of adaptation on perceived brightness relates to changes in the Stevens power law parameters as adaptation level changes. Changes in the log-log intercept, k, describe the decreased brightness of constant luminance with increasing adaptation. Changes in the absolute threshold, L_0, reveal a strong dependence of the absolute threshold for stimulus luminance on perceived brightness with increasing adaptation. Changes in the log-log slope, β, characterize the stronger dependence of brightness on stimulus luminance at higher adaptation levels than at lower adaptation levels.

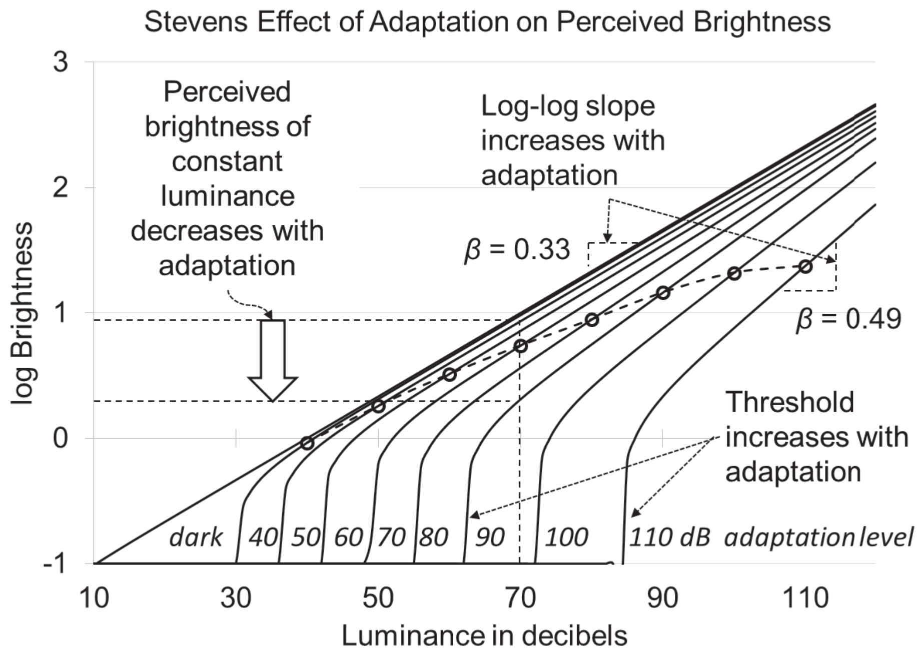


Figure 5. Illustration of the Stevens effect of adaptation on perceived brightness adapted from Stevens & Stevens (1963). Each solid curve follows a Stevens power law and describes the perceived brightness (vertical axis) of a brief test stimulus when the eye has been light adapted to a sustained luminance as noted in the figure. The luminance of the test stimulus is indicated on the horizontal axis in terms of decibels relative to 10^{-10} Lambert (L) (1 L = $10^4/\pi$ cd/m^2) (90 dB corresponds to ~318 cd/m^2) The dashed line is the terminal equilibrium function, which is the brightness perceived at any level of luminance when the eye is fully adapted to that level. As noted in Stevens & Stevens (1963), perceived brightness would relax over time from the level indicated by corresponding power law to the level indicated by the terminal equilibrium function.

It should be noted that, in the strictest sense, the Stevens effect relates only to the perception of a brief stimulus (2 seconds in the original research) in a light-adapted eye relative the perception of a brief stimulus in a dark adapted eye. As such, the Stevens effect does not account for simultaneous contrast or other compositional and temporal context, any of which can have significant impact on perceived brightness and can be expected to have major influences on a viewer's experience of HDR content. Nonetheless, Stevens effect and Stevens power law could provide useful starting points for exploring the creative impact of minimum, maximum, and average luminance levels.

The Hunt effect[8] is very similar to the Stevens effect, but it relates to the perception of colorfulness rather than contrast. Colors appear more colorful (more vivid and intense) as overall luminance increases. A color viewed in dim illumination would be perceived as a match to a much less saturated color under brighter illumination. As with the Stevens effect, the Hunt effect could provide new creative flexibility, particularly for HDR content.

Luminance also affects the perception of the hue (Figure 6). The Bezold-Brücke[8,45] effect describes the perception of two stimuli having the same wavelength but different luminance as different hues. As luminance increases, light stimuli (such as pixels on a display) appear bluer for wavelengths below ~500 nm and yellower for wavelengths above ~500 nm. It is thought that the Bezold- Brücke effect is a result of non-linearity in higher-level processing centers of the visual system after the photoreceptors.

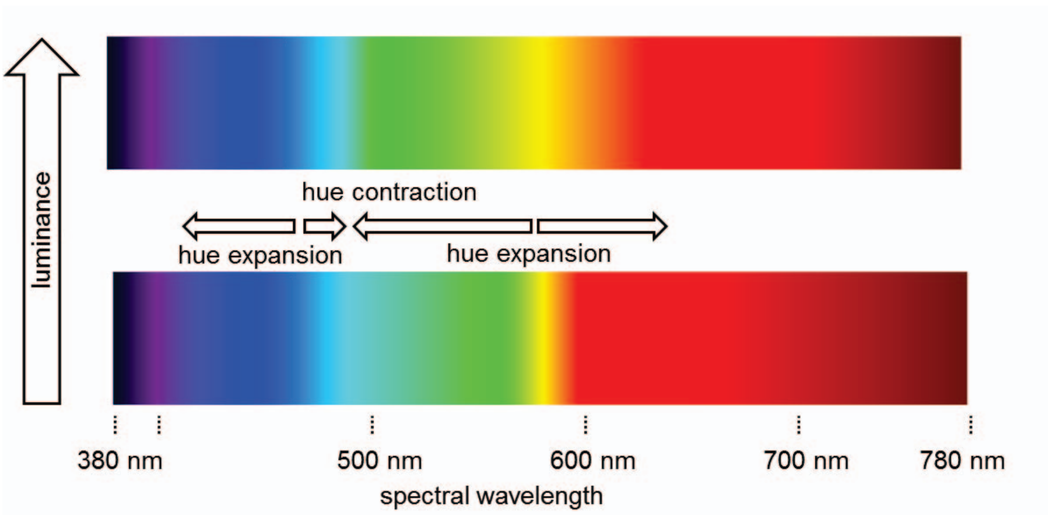


Figure 6. Illustration of the Bezold-Brücke effect in which luminance affects perception of hue. As illustrated, a 10-fold increase in luminance causes light stimuli to appear noticeably bluer below ~500 nm or yellower above ~500 nm.

Light and dark adaptation

HDR creates an opportunity to use wide swings in luminance levels for creative effects in ways that have not been possible before. The luminance ranges proposed[36-41] are such that they can affect the adaptation state of the visual system. A dark-adapted viewer's experience can be very different from a light-adapted viewer's experience. In viewing situations in which dramatic content is mixed with advertising, it will not always be possible to know ahead of time what a viewer's adaptation state would be, particularly when HDR and non-HDR content is presented sequentially.

The human visual system adapts to 9 log units of light intensity[17]. Several physiological processes contribute to the human visual system's ability to adapt to this enormous range of light intensities, including: changes in pupil diameter; progressive shift from rod-dominated to cone-dominated vision; changes in the gain of the photoreceptor responses; and change in the concentration of light-sensitive photopigments in the photoreceptors.

Changes in pupil diameter

Changes in the size of the pupil regulate the area through which light enters the eye by a factor of approximately 16. Between fully dark-adapted and light-adapted conditions, the pupil diameter changes in a graded manner. Exposure to bright light can reduce the area of the pupil, and consequently reduce retinal illumination, by approximately a log unit in a half second.

Pupil size depends also on the size of the stimulus (field-of-view) in addition to luminance. Table 2 provides results from the Watson-Yellot[46] model that includes all of the known factors that influence pupil size. In this calculation, an image displayed at an average luminance of 10 cd/m^2 may be expected to have a diameter of ~4.2 mm. At 10,000 cd/m^2, the pupil diameter can be expected to be ~2.0 mm, which would correspond to an approximately 4.5-fold reduction in pupil area.

Table 2. Luminance, pupil size and retinal illuminance.
Calculated using the Watson-Yellot model for a visual stimulus of 30-degrees.

Luminance (cd/m^2)	0.001	0.01	0.1	1	10	100	1000	10000
Pupil Diameter (mm)	7.3	7.1	6.5	5.6	4.2	3.1	2.3	2.0
Pupil Area (mm^2)	42.2	39.4	33.6	24.2	14.1	7.4	4.3	3.1
Retinal Illuminance (troland)	0.042	0.39	3.4	24	141	736	4309	31286

Vision scientists use the term “troland” (Td), equal to luminance multiplied by pupil area, as a measure of retinal illuminance. Table 1 provides troland values for various luminance levels calculated using the Watson-Yellot model of pupil diameter for a visual stimulus of 30-degrees. (Note that between 10 and 1,000 cd/m^2, luminance changes by a factor of 100 but retinal illuminance changes only by a factor of approximately 30.)

Although changes in pupil size are only a small part of the overall adaptation range, reduced pupil size also reduces glare, increases depth of field, decreases some kinds of optical aberrations, and increases diffraction blur, any of which could be considered in crafting the artistic intent.

Bleaching Adaptation

It is well known that rods are responsible for night vision and cones for daylight and color perception, but it is more accurate to think of visual adaptation in three categories that better reflect the gradual shift from rod-dominated vision to cone-dominated vision as light conditions brighten:

- Scotopic (below 0.001 Cd/m^2) dominated by rods
- Mesopic (0.001 to 10 Cd/m^2) mix of rods and cones
- Photopic (above 10 Cd/m^2) dominated by cones

In darkened cinema theaters and home environments, mesopic-level adaptation might be a significant consideration. In bright home and mobile environments, photopic-level adaptation would be more typical.

With the growing commercialization of brighter HDR displays and content, a phenomenon known as “bleaching adaptation”[44] could become significant and impact a viewer’s experience. Bleaching adaptation is a reduction in the concentration of the light-sensitive biological pigment inside photoreceptors (rhodopsin in rods and cone-opsin in cones). When a photopigment absorbs a photon, the now-activated pigment initiates a chain of reactions that ultimately creates the neural signal that a photon was absorbed. Before the photopigment can be used again, it needs to be recycled, which takes time. Over a period of steady illumination, the photopigment concentration will reach a new lower steady-state level when the rate at which photopigments absorb photons is balanced by the rate at which they are replenished through the retina’s recycling program. Consequently, the optical density of the photoreceptors also decreases; i.e., photoreceptors absorb less light because there are fewer photopigments available to catch photons.

Figure 7 provides a more quantitative view of bleaching adaptation. The curve illustrates the concentration of excitable photopigment (relative to the concentration in a dark adapted photoreceptor). According to a 1st-order Rushton model[47,48], the concentration would be halved by sustained retinal illumination of 4.3 log trolands (approximately 6,000 cd/m^2 according to the Watson-Yellot model, see Table 2). In comparison, sustained luminance of 100 cd/m^2 (approximately 3 log troland) – the luminance level of standard dynamic range (SDR) content[49] - would result in only an approximately 5% decrease in excitable photopigment. Note that 100 cd/m^2 SDR level is at the beginning of the knee of the bleaching curve. Thus, in moving from SDR to HDR content, viewers could experience significantly more bleaching adaptation, which could result in afterimages and reduced visual sensitivity particularly following prolonged scenes with high average luminance.

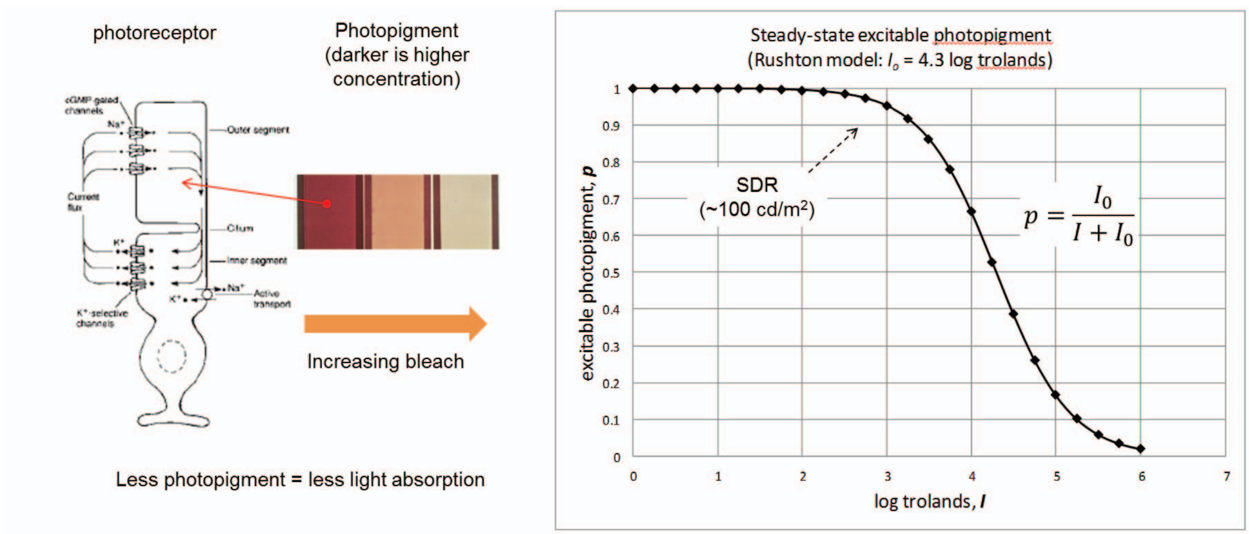


Figure 7. Illustration of the effect of luminance on photoexcitable photopigment in retinal photoreceptors.

Speed of Response and Adaptation

Rod-driven vision differs from cone-driven vision not only in terms of absolute sensitivity but also in terms of the speed at which light variations are signaled. The peak of the response of a dark-adapted rod to a flash of light occurs at approximately 120 msec[506]. The peak of a light-adapted rod response occurs at approximately 75 msec[506]. The peak of cones response occurs at approximately 20 msec[517]. The change in the time-scale of neural responses is part of the reason that sensitivity to flicker and motion increases with luminance.

Photoreceptors adapt to moderate non-bleaching step changes in increased illumination on the time scale of seconds[528,539]. During this light adaptation process, the absolute sensitivity of photoreceptors decreases and the kinetics of responses increase in speed.

The time course of adaptation to decreased illumination (dark adaptation) depends on the intensity and duration of preceding stimuli[17]. If the preceding exposure was low enough that no significant photopigment bleaching occurred, then dark adaptation may also be measured on the time scale of seconds but slightly slower than light adaptation. For beaching adaptation, dark adaptation may take longer, on the order of tens of seconds to minutes. Dark adaptation also has two distinct phases: one is driven by recovery of cone sensitivity; the other even slower phase is driven by rod sensitivity.

In bright home and mobile viewing environments, both light and dark adaptation to modulations in illumination may be expected to proceed on a time scale measured in seconds. In dark home and theater environments, rapid changes going back and forth from mesopic-level to photopic-level luminance might result in slower dark adaptation.

Speed of Adaptation, Scene Transitions, and Commercials

As detailed in the preceding discussion, the speed of light and dark adaptation depends on the level of retinal illuminance, the duration of illumination, the kinetics of the changes in pupil size, the rate of change of photoreceptor sensitivity, the rate of change of excitable photopigment, and the state of overall visual adaptation (scotopic, mesopic, or photopic).

In some situations, content creators could craft scene transitions to leverage adaptation for artistic effect. In other situations, the impact of rapid local or global luminance changes in could have unwanted impact on a viewer's experience, as could conceivable be the case when HDR and non-HDR content is presented sequentially to viewers.

Conclusion

Ultra HD is bringing a welcome transformation in the television viewing experience to the benefit of producers, distributors and consumers. Wider field-of-view, better portrayal of real world light and color, and smooth natural motion are now options that content creators may use to connect with the viewer in a more immediate and immersive way.

In this paper, our goal was to provide quantitative information about how various Ultra HD parameters could interact based on known principles of vision science. Over time, content creators will be creating practical rules-of-thumb as they produce more and more Ultra HD assets. We hope that some of the background on vision science presented in this paper might be useful in that creative process. Specifically, the following key points could prove germane:

- Visual acuity may be thought of, and leveraged, in a more nuanced manner than simply stating whether or not individual pixels can be seen. Visual acuity depends of the visual task in question. Acuity ranges from the ability to recognize symbols, such as when reading, to the ability to notice spatial details that are smaller than the size of a photoreceptor. The latter is called hyperacuity.

- The visual system is capable of high spatial acuity for moving features provided the velocity of the slip of the image on the retina is less than 2-4 degrees per second. Increasing captured frame rate progressively shrinks the frame-to-frame displacement of a feature moving at the limit smooth pursuit. At 60 frames per second, the frame-to-frame displacement would be about the same as the radius of the foveola. At 300 frames per second, the displacement would be about the size of the radius of the central island, which is the region of maximum spatial acuity, and thus high spatial acuity and smooth motion might be achieved simultaneously.
- High-speed motion crafted for smaller Ultra HD TVs and tablets (those subtend ~30 degrees field-of-view) could create a very difference experience when viewed on a larger device (those that subtend ~60 degrees). On the smaller display, motion can be followed smoothly using only eye rotation without head movement. On the larger display, the angular velocity could exceed the velocity limit of smooth pursuit and would tend to evoke coordinated head and eye movement.
- Flicker and sensitivity to temporal change increases with both luminance and field-of-view. Thus, HDR content rendered on large Ultra HD display could evoke a sense of non-constant light either deliberately as part of the creative intent or inadvertently.
- To the extent that flicker and stroboscopic effects are unwanted, the move toward brighter and larger Ultra HD display could increase the desire of creatives to use higher frame rate captures.
- Luminance affects the perception of contrast, colorfulness, and hue. HDR luminance levels by themselves could conceivably be leveraged to create noticeable changes in contrast and color. To the extent that color and lightness constancy is desired, extra care might be needed.
- The visual system adapts to changes in luminance. Moderate changes in luminance occur on the order of seconds. Recovery from bright and/or sustained retinal illumination can take minutes as the photoreceptors replenish light-sensitive photopigments. The time course of adaptation could be leveraged for creative impact. In situations in which HDR and non-HDR content is rendered sequentially -- such as non-HDR commercials within an HDR program, for example – the slower recovery from bright stimuli could have unintended effects.
- Changes in pupil size can occur in less than a second. Although pupil size plays a relatively small role in light adaptation, light-induced reduction of pupil size also reduces glare, increases depth-of-field, reduces certain optical aberrations, and increases diffraction blur, all of which could be leveraged for creative effect.

Ultra HD is clearly an exciting new direction for television. Creatives are actively exploring the new ways the various Ultra HD technologies could be used to connect viewers with story, characters, and emotions. Likewise, standards bodies and equipment vendors are developing specifications and best practices to aid in delivering the intended artistic effect to viewers. We hope that the information we presented in this paper can help inform efforts to bring Ultra HD to its full potential.

Acknowledgements

Thank you to Dawn Danaher, Melissa Malley, Kevin Wirick, and David L. Long for helpful suggestions and insights that improved this manuscript.

References

1) ITU-R BT.2020. “Parameter values for ultra-high definition television systems for production and international programme exchange,” http://www.itu.int/rec/R-REC-BT.2020
2) SMPTE ST 2036-1-2013, “Ultra high definition television – Image parameter values for program production,” http://standards.smpte.org
3) ITU-R Report BT.2246-4 “The present state of ultra-high definition television,” http://www.itu.int/pub/R-REP-BT.2246
4) ITU-R BT.709-5, “Parameter values for the HDTV standards for production and international programme exchange,” http://www.itu.int/rec/R-REC-BT.709
5) T. Yamashita, K. Masaoka, K. Ohmura, M. Emoto, Y. Nahida, and M. Sugawara. “”Super Hi-Vision” video parameters for next-generation television.” SMPTE Motion Imaging Journal, May/June 2012
6) P. Hanhart, P. Korshunov, T. Ebrahimi, Y. Thomas, and H. Hoffman. (2015). “Subjective Quality Evaluation of High Dynamic Range Video and Display for Future TV” SMPTE Motion Imaging Journal, May/June 2015
7) S.T. McCarthy. “Perception: How UHD relates to the human visual sense,” presented at SMPTE 2013 Symposium – Next generation imaging format: More, faster, and better pixels, October, 2013.
8) M. D. Fairchild. *Color Appearance Models*, 3rd Edition. John Wiley & Sons. 2013.
9) G. Wyszecki and W.S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae,* 2nd Edition, John Wiley & Sons. 2000.
10) M. H. Kim, T. Weyrich, and J. Kautz. "Modeling human color perception under extended luminance levels.” SIGGRAPH 2009, Vol 28 ,Issue 3. 2009
11) E.H. Adelsen “Checkershadow illusion.” http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html
12) R.B. Lotto and D. Purves. “The effects of color on brightness,” Nature Neuroscience, Vol. 2, No. 11, November, 1999.
13) R.B. Lotto. University College London www.lottolab.org
14) P. Routhier and E. Perez-Pellitero “Temporal vs. Spatial Resolution: Comparative Tests for Broadcast Sports.” SMPTE Motion Imaging Journal, May/June 2015
15) J. Defilippis.”Does video need to go faster? Can higher frame rates eliminate image blur?” TVTechnology, September 15, 2015.
16) J. Defilippis. “Adventures in high frame rate, part 2.” presented at HPA Tech Retreat, February, 2015. http://hollywoodpostalliance.org/wp-content/uploads/2015/02/JDeFilippis_W545_HPA_2015.pdf
17) L.A. Levin, S.F.E. Nilsson, J. Ver Hoeve, S. Wu, P.L. Kaufman, and A. Alm, *Alder’s Physiology of the Eye*, 11th edition. Saunders – Elsevier, 2011.
18) H. Kolb, “Simple anatomy of the retina,” Webvision: The organization of the retina and visual system, http://webvision.med.utah.edu

19) H. Kolb, R. Nelson, E. Fernandez, and B. Jones, “Part XIII:Facts and figures concerning the human retina,” Webvision: The organization of the retina and visual system, http://webvision.med.utah.edu
20) C.A Curicio, K.R. Slaon, R.E. Kalina, and A.E. Hendrickson, “Human photoreceptor topography,” J. Comp. Neurol., 292:497-523, 1990
21) S.L. Polyak. *The Retina*, University of Chicago Press: Chicago, IL, 1941
22) E.W. Weisstein. "Nyquist Frequency." From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/NyquistFrequency.html
23) D.G. Green, “Regional variations in the visual acuity for interference fringes on the retina,” J. Physiol. 207:351-356, 1970
24) H. Snellen. *Probebuchstaben Zur Bestimmung Der Sehschaerfe (trans. Sample letters for determination of visual acuity)*, PW van de Weijer: Utrechy, the Netherlands, 1862
25) M. Kalloniatis and C. Luu, “Visual acuity,” Webvision: The organization of the retina and visual system, http://webvision.med.utah.edu
26) G. Westheimer, “Editorial: Visual acuity and hyperacuity,” Investig. Ophthalmol., 14(8):570-572, 1975.
27) G. Westheimer, “Hyperacuity,” Scholarpedia, 6(8):9973, 2011.
28) ITU-R Report BT.801-4 “The present state of high-definition television,” http://www.itu.int/pub/R-REP-BT.801
29) ITU-R BT.1845 “Guidelines on metrics to be used when tailoring television programmes to broadcasting applications at various image quality levels, display sizes and aspect ratios.” www.itu.int/rec/R-REC-BT.1845-0-200810-S
30) S.T. McCarthy. “Quantitative evaluation of human visual perception for multiple screens and multiple CODECs,” presented at SMPTE Annual Technical Conference and Exposition, Hollywood, CA. 2012
31) J.-J. Orban de Xivry and P. Lefevre. “Saccades and pursuit: two outcomes of a single sensorimotor process,” J Physiol 584.1, 2007.
32) J.M. Henderson. “Human gaze control during real-world scene perception,” TRENDS in Cognitive Science, Vol. 7 No. 11, 2003.
33) E.G. Freedman. “Coordination of the eyes ad head during visual orienting.” Exp Brain Res, 190(4):369-387, 2008.
34) Millodot, M. *Dictionary of Optometric and Visual Science*, 7th edition. Elsevier Health Sciences. 2009
35) C.C.A.M Gielen, S.F. Gabel, and J. Duysens. “Retinal slip during active head motion and stimulus motion,” Exp Brain Res, 155:211-219, 2004.
36) S. Miller, S. Daly, and M. Nezamabadi. “Perceptual signal coding for more efficient usage of bit codes,” SMPTE Motion Imaging Journal 122, 52:59, May 2013.
37) T. Borer, “Non-linear Opto-Electrical Transfer Functions for High Dynamic Range Television,” BBC R&D White Paper. WHP 283, July 2014.
38) SMPTE ST-2084 “High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays,” http://standards.smpte.org
39) S. Miller, S. Daly, and M. Nezamabadi. “Perceptual signal coding for more efficient usage of bit codes”, SMPTE Motion Imaging Journal, May/June 2013
40) C. Poynton, J. Stessen, and R. Nijland. “Deploying wide color gamut and high dynamic range in HD and UHD.” SMPTE Motion Imaging Journal, April 2015

41) S. Farrell, T. Kunkel, and S. Daly. "A cinematic luminance range by the people, for the people: Viewer preferences on luminance limits for large-screen environment." SMPTE Motion Imaging Journal, July/August 2015.
42) CIE. *Commission international de l'Eclairage proceedings*. Cambridge University Press, 1931.
43) J.C. Stevens and S.S. Stevens. "Brightness function: Effects of adaptation." J. Opt Soc. Am., 53, 375-387. 1963.
44) J.C. Stevens and L.E. Marks. "Stevens' power law in vision: Exponents, intercepts, and thresholds." Proceedings of the Fifteenth Annual Meeting of the International Society for Psychophysics, 87-92, 1999.
45) R.W. Pridmore. "Bezold–Brucke hue-shift as functions of luminance level, luminance ratio, interstimulus interval and adapting white for aperture and object colors," Vision Research 39, 3873-3891, 1999.
46) A.B. Watson, and J.I. Yellot. "A unified formula for light-adapted pupil size," J. Vis 12(10):12, 2012.
47) W.A.H. Rushton and G.H. Henry. "Bleaching and regeneration of cone pigments in man," Vision Research 8, 617-63, 1968.
48) A.C.C. Coolen and D. van Norren "Kinetics of human cone photopigments explained with a Rushton-Henry model." Biol. Cybernetics Vol. 58, No. 2. 1988
49) ITU-R BT.1886 "Reference electro-optical transfer function for flat panel displays used in HDTV studio production," http://www.itu.int/rec/R-REC-BT.1886
50) C. Friedburg, M.M. Thomas, and T.D. Lamb. "Time course of the flash response of dark- and light-adapted human rod photoreceptors derived from the electroretinogram," J. Physiol. 534.1, pp. 217-242. 2001.
51) J.H. Van Hateren and T.D. Lamb. "The photocurrent of human cones is fast and monophasic," BMC Neuroscience, 7:34, 2005.
52) L.-H. Cao, D.-G. Luo D.-G., and K.-W. Yau. "Light responses of primate and other mammalian cones," PNAS. Vol. 11 No. 7, 2007.
53) J.I. Korenbrot. "Speed, adaptation, and stability of the response to light in cone photoreceptors: The functional role of Ca-dependent modulation of ligand sensitivity in cGMP-gated ion channels," JGP vol. 139, No. 1, 31-56. 2011.

