A MISSING FACTOR IN OBJECTIVE VIDEO QUALITY MODELS: A STUDY OF COLOR

Margaret H. Pinson

National Telecommunications and Information Administration

ABSTRACT

This paper examines color from multiple perspectives. A theory is proposed that the current color spaces are poorly designed for video quality and video coding purposes. The implication is that objective video quality metrics could be improved by the development of a field-specific color edge detection algorithm or color distance measurement. A prototype subjective database is provided that could be used to train such an algorithm.

1. INTRODUCTION

The most widely known objective video quality metric is peak signal to noise ratio (PSNR). While many formulae exist, PSNR is traditionally calculated on the luma plane of the YCbCr color space. This characteristic is not unique—many objective video quality metrics depend primarily or entirely upon the luma plane.

The Video Quality Experts Group (VQEG) validation reports provide high quality analyses of objective video quality metrics on subjective video quality databases. Within these analyses, the Pearson correlation between the best objective metric and subjective data rarely exceeds 0.9. This indicates a missing factor, an element that is obvious to subjects yet missed by the metrics.

I believe the missing element is color, and the culprit a flawed representation of the human visual system. My supposition is that improved formulae for calculating very large color distances would increase the performance of all objective video quality metrics.

This paper explores color perception as the missing element in objective video quality metrics. We will explore challenges in the development of a perceptual color space. Information about the visual system is gathered from multiple fields, including linguistics and art. A subjective dataset is presented that estimates color distances. This dataset is made available to encourage the development of a perceptual color edge detector.

2. RGB AND YCBCR

Let us begin with a simple demonstration of the problem. Computers use the RGB color space, while video codecs use the YCbCr color space. The consequences are in some sense trivial, because the transformation requires a simple matrix computation. Given RGB pixels from [0..1], we obtain YCbCr values in the range [0..255] as follows [1]:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} R \\ G \\ B \end{bmatrix} \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112.000 \\ 112.000 & -93.786 & -18.214 \end{bmatrix}$$

where both color spaces describe perfect cubes.

The RGB color space represents a wide range of colors by mixing red (R), green (G) and blue (B) light. Fig. 1 shows the red, green and blue axes, each calculated with a neutral value for the other two planes (128 assuming RGB pixels from [0..255]). These axes are not perceptually uniform.

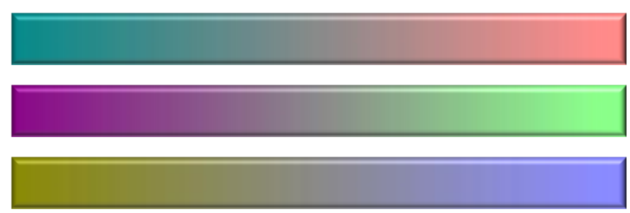

Figure 1. The red (top), green (middle) and blue (bottom) axes of the RGB color space with the other planes set to a neutral value.

YCbCr claims to be a perceptual color space, meaning perceptually uniform. There are obvious empirical problems with this claim. Broadcast video formats routinely subsample Cb and Cr by two horizontally (4:2:2), and this is known to have a negligible impact on quality. Many non-broadcast video applications subsample by two both horizontally and vertically (4:2:0), and this is known to have a very small quality impact. By contrast, subsampling Y (the luma plane) by four is the difference between standard definition television and HDTV—a dramatic difference indeed.

Objective video metrics provide additional empirical evidence. For example, Stephen Wolf and I trained the video quality metric with variable frame delay (VQM_VFD) on 11,255 subjectively rated video clips. This metric uses the luma plane only.

To better understand YCbCr's claim of perceptual uniformity, let us visually examine the color space. Fig. 2 shows the Y, Cb and Cr axes, each calculated with a neutral value for the other two planes (i.e., Y = 128, Cb = Cr = 0). The luma axis seems to have a reasonable claim of perceptual uniformity. The usefulness of this luma plane is also evident in the MPEG committee having chosen the luma plane as a core structural element for all modern video codecs.

Distances along the Cb and Cr axes do not appear to be perceptually uniform. Only the middle of each axis shows color gradations. Shades of green appear at one end of both axes. While the Euclidean distance between Cb = ±128 and Cr = ±128 is large, the perceptual distance can be larger (e.g., purple to mint green) or smaller (e.g., mint green to yellow green). This perceptual non-uniformity is demonstrated further by Fig. 3, which shows the range of Cb and Cr values for Y = 16, 128 and 240. Notice that gray is barely visible, purple spans a narrow range, and magenta (hot pink) spans a large range.

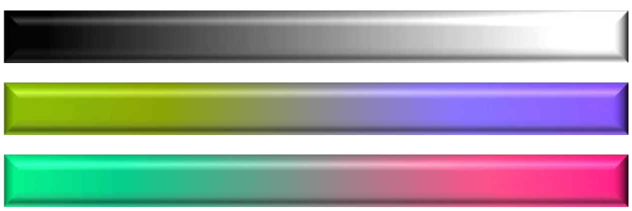

Figure 2. The Y (top), Cb (middle) and Cr (bottom) axes of the YCbCr color space, with other planes set to a neutral value.

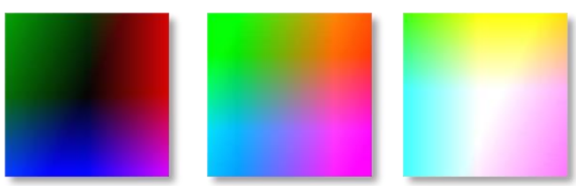

Figure 3. (Cb,Cr) planes of the YCbCr color space for Y = 16 (left), 128 (middle) and 240 (right) are not perceptually uniform.

There are other color spaces available with perhaps greater claim to perceptual uniformity. Alas, all are flawed [2]. Rather than delving into the strengths and weaknesses of each, let us explore the challenges involved in developing a perceptually uniform color space.

3. COLOR IS NOT EUCLIDEAN

Human perception of color describes a space that is not Euclidean [2]. This non-uniformity is reflected in the ellipses used within papers to describe uniform threshold color differences within International Commission on Illumination (CIE) color spaces [3]. There are irregularities, which can be seen by paging online through the *Atlas of the Munsell Color System* from 1913 [4]. The Munsell color space has three dimensions with luma as one axis, saturation increasing as the distance from the central axis increases, and hue configured radially around the luma axis. Unlike the perfect cube of RGB or YCbCr, Munsell spans an irregular blob.

Judd [5] identifies two problems with creating an ideal color space. First, twice as many hue differences as saturation differences can be accommodated in Euclidean space. For example, the color distance circumference of the Munsell color space appears to be 720˚ instead of 360˚. Second, surrounding colors impact color perception. The tendency is to shift the perceived color toward the surrounding color's complementary color.

Consequently, Euclidean distance is a poor choice of statistic—as can be seen by the failures within objective video quality metrics, where Euclidean geometry is typically assumed. Another consequence is the multitude of color spaces found in literature, each flawed and to some degree arbitrary [2]. These color spaces were typically designed to meet the needs of one particular community, such as artists or manufacturers (e.g., to specify dye lot tolerances). Often, distances along one dimension cannot be meaningfully compared to distances along another dimension.

Kuehni [3] identifies subjective datasets that characterize human perception of color differences and a variety of formulae to estimate color differences with some degree of accuracy. The problem from a video quality standpoint is that these models and the related research focus on small color differences (e.g., just noticeable differences, color matching, and color ordering). Within the color theory community, a "large" color difference is still nonetheless quite small from a video perspective—for example, spanning 5% to 20% of the distance between black and white.

4. COLOR IN THOUGHT AND LANGUAGE

Research summarized and simplified in *Color Categories in Thought and Language* [6] describes the work of linguists, visual psychologists, and anthropologists to understand human perception of color. This section provides a brief overview that is intended to bridge the gap between color differences that can be noticed and color differences that are judged important.[1]

[1] This section draws primarily upon articles by Bill Wooten, David Miller, Greville Corbett, Ian Davies, Kimberly Jameson, Roy D'Andrade and Robert Boynton.

International studies of color terms across dozens of languages indicate two hypotheses:

- There is a limited set of basic color terms.
- There is a constrained order in which these colors appear in languages.

Put another way, if a language has *N* basic color terms, we can predict what those colors will mean (e.g., as defined by a focal point and a set of color chips). Fig. 4 shows the initial proposal for basic color term entry into language. These eleven basic color terms were theorized to have a neurological basis.

A language with two color terms will contain two composite colors:

- White/warm: a composite of white, yellow & red
- Black/cool: a composite of black, blue, & green

The foci are not necessarily at black and white, as would be presumed from Fig. 4.

$$\begin{matrix} White \\ \& \\ Black \end{matrix} \rightarrow Red \rightarrow \begin{vmatrix} Green \\ Yellow \end{vmatrix} \rightarrow Blue \rightarrow Brown \rightarrow \begin{vmatrix} Purple \\ Pink \\ Orange \\ Gray \end{vmatrix}$$

Figure 4. Initial theory for the order that colors enter language. Where colors are listed vertically in brackets, the colors might enter a language in any order (e.g., green then yellow; or yellow then green).

Later refinements indicate two processes at work. The first process divides composite color categories into the six primary basics: black, white, red, yellow, green and blue. The white/warm composite always splits first, yielding white and red/yellow. The black/cool composite might either split into (black and green/blue) or (green and blue/black). The second process derives secondary basics from fuzzy intersections of the primary basics (e.g., orange is a mix of red and yellow). The five secondary basics are gray, brown, purple, pink, and orange. Gray enters language at a greater variety of points, making it somewhat of a wild card.

There is some evidence that new color terms enter language based on maximizing the added value of the new term—that is, the new color term's focus is maximally distant from existing color terms. Highly saturated regions of the color space are more perceptually important and thus enter language earlier (e.g., red, yellow), while less distinctive areas enter later (e.g., blue, green).

The four primary hues (red, yellow, green and blue) are elemental, in that they cannot be perceptually subdivided into more basic hues. This leads naturally to an understanding of hues as two opposing pair processes (red/green and blue/yellow) with an assumption of a four-basic-hues model. This model is widely assumed in much of color vision literature (i.e., a three dimensional space with axes black/white, red/green and blue/yellow).

From a psychological standpoint, the perceptual distance between red and blue is too large. A five-basic-hues model is more balanced (red, green, yellow, blue, purple). The Munsell color space uses this five-basic-hues model, which puts red in opposition to blue-green. A study of opponent colors supports the five-basic-hues model over the four-basic-hues model.

A different way of understanding the basic color terms is through a linked map. Naming tests of color chips can be used to identify areas where color chips are equally likely to be assigned to one of two different color terms (e.g., green and yellow). As an example, a linked map of red, green, brown and gray shows bridges between red/brown, brown/green, green/gray and brown/gray. Note that no bridge exists between red/gray. The linguistic path from red to green goes through brown—not gray, as is implied by the four-basic-hues model.

A consequence for color perception in video is that we would expect a perceptually uniform color space to reflect the importance of basic color terms. The significance of magenta and insignificance of grey within Fig. 3 is troubling.

5. EXPLORING COLOR IN ART

The oldest and most mature field in which to study color is art. From an engineer's perspective, this is also the most foreign. Art is taught visually and kinesthetically, and thus the knowledge can be less accessible.

5.1. Color Palettes and Color Schemes

Exploring Color [7] describes how artists use color. This section provides a brief overview.

Artists use a six-basic-hues model called the color wheel (see Fig. 5). All six colors are given equal importance and placed equidistant in a circular configuration. The entire color wheel is modified to create different palettes by shifting all of the colors. The palette creates a cohesive group of compatible colors that are well matched in intensity, tinting strength, and opacity. The artist's color wheel and palettes more closely match the linguistic model of color perception than does YCbCr (compare Figs. 3, 4 and 5).

Notice that the color wheel does not contain neutral colors (white, black, and grey). In art, neutrals are considered achromatic or colorless. Neutrals are used to understand the *value* of a color (i.e., the luma component) and to modify colors. For example, the pastel palette in Fig. 3 is shifted toward white, while the Old Masters' palette is shifted toward a warm brown by substituting Raw Sienna, Burnt Sienna and Payne's Gray for yellow, red and blue respectively.

A color scheme establishes a subset of the colors to be used. A color scheme unifies the artwork through harmony and contrast:

- **Monochromatic**: one color has no discord
- **Analogous**: two adjacent colors produces harmony (e.g., red and orange)
- **Complementary**: opposing colors are the most dynamic (e.g., yellow and purple)
- **Triad**: a triangle of colors is bold and energetic (e.g., purple, orange and green)
- **Pure hue against a neutral color scheme**: produces the most striking contrast (e.g., yellow against black)

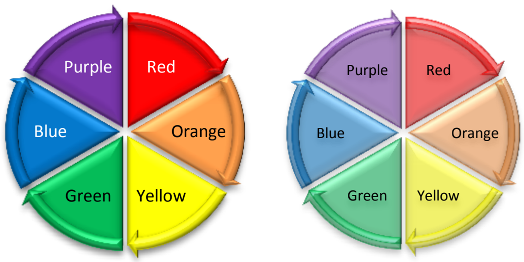

Figure 5. The entire color wheel (left) is shifted to create a pastel palette (right).

Other techniques can be used to increase or decrease contrast:

- **Intensity contrast**: varying the color saturation
- **Temperature contrast**: warm colors appear to advance, while cool colors recede
- **Quantity contrast:** the size and use frequency of a color impacts its perceived contrast (e.g., large areas have greater color impact; a color broken into many small areas creates energy and movement)
- **Gradation:** gradual color changes reduce contrast and divert attention, while abrupt color changes increase contrast and draw attention

Artists choose color to draw your attention toward some areas and away from others.

5.2. Painting Lessons

Painting What You Want To See [8] demonstrates a traditional progression of art lessons for watercolor and oil painting:

1. Line drawings: contour drawings, outline shapes
2. Value drawings: grayscale or color, showing local values (object color) without shadows
3. Shadows
4. Mixing different colors (e.g., skin tones, rich darks, shades of green)

Fig. 6 shows example artworks that depict the first three elements separately. Children's artwork captures line and sometimes value, but shadow is omitted. To understand how natural this omission is, look back at Figs. 1-3 and 5. The highlights and shadows improve the figure artistry without detracting from the color representation.

Conversely, the luma plane characterizes line, shadow and some parts of the value drawing—while omitting other parts of the value drawing. From an artistic perspective, the luma plane representation of an image overemphasizes shadows. Compare the luma plane in Fig. 7 with the shadow and grayscale drawings in Fig. 6.

Figure 6. From left to right, shows an example contour drawing, value artwork, shadow painting, and grayscale drawing (by the author).

Figure 7. Look at the man's skin tones and the sign's backdrop. In color (left), the light yellow of the sign contrasts with the man's skin tones. In the luma plane (right), the sunlit and shadowed areas match.

Figs. 1-3 and 5 use darker colors to denote shadows. This darker colored shadow choice appears in *The Calling of the Apostles* by Domenico Ghirlandaio on the northern wall of the Sistine chapel.

Reid [8] often uses color opposites to create shadows (e.g., create watercolor skin tones by mixing cadmium red, cerulean blue and cadmium yellow light). This technique can be seen in the far more famous paintings by Michelangelo Buonarotti on the ceiling of the Sistine chapel, such as *The Creation of Adam*. The aesthetic impact is more compelling, and thus probably a more accurate rendering of shadow perception.

6. SHADOWS AND OPTICAL ILLUSIONS

The lighting conditions we experience change in intensity, color temperature, and angle. Vary the light conditions, and an object's shadows, hue, saturation, and lightness change. The human visual system mitigates the impacts of these changes to reveal the object's color. From the perspective of linguistics and art, this color more accurately describes the object than does the range of luma values spanned by its shadows.

Edward H. Adelson [9] describes how the human visual system compensates for illumination and other viewing conditions. Stated loosely, the visual system partially removes shadows to better characterize the object behind the shadows. While the Y plane of YCbCr reflects a single measurement of luma for the entire image using one white balance, the visual system makes multiple measurements. By analogy, the visual system uses several white balances at once and defines Y conditionally, depending upon the surroundings. This allows us to differentiate a light surface in dim light from a dark surface in bright light.

Fig. 8 demonstrates the *simultaneous contrast effect*, by which surrounding colors influence lightness perception. The lightness or darkness of each square is judged within the local context of the neighboring square. Additionally, the human visual system tends to ignore gradual changes in light level, such as the soft edge of the shadow. This allows the human visual system to break the image down into meaningful components, here a checkerboard and cylinder [9].[2]

These cognitive processes may partially explain the behavior of basic color terms seen in Section 5. Our visual system modifies colors in response to context and illumination. The broad range of colors spanned by each basic color term may reflect an understanding that the set of just noticeably different colors represent the same color impacted by different illumination.

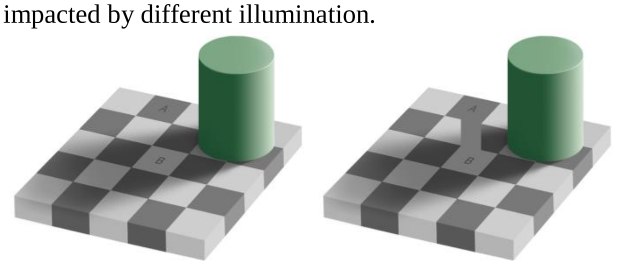

Figure 8. Same color illusion (left) and proof (right) by Edward H. Adelson.[2] The light squares in shadow (B) have the same luma value as the dark squares in light (A). This visual illusion demonstrates how the visual system compensates for shadows.

7. COLOR SPACES & SUBJECTIVE DISTANCES

The subjective video quality test “NTIAcolor” was conducted according to ITU-T Rec. P.913. The stimuli were two colors displayed side-by-side as large rectangles on a computer screen. The background surround was mid-level gray.[3] Subjects were asked to rate the color difference on a discrete 5-level scale with labels imperceptible, subtle, mediocre, obvious and striking. The LCD monitor was color calibrated and the room lighting was dim (≈20 lux). The test was conducted in a sound isolation booth at 3H viewing distance, as shown in Fig. 1 top within [10].

The test included two pools of subjects. Pool 1 focused on comparisons among the eleven basic colors (see Section 5 in [10]) and color placed between these basic colors. Pool 2 focused on random color comparisons and comparisons between colors that ranged across one dimension (e.g., from white to purple). Pool 1 contains 273 comparisons, and pool 2 contains 214 comparisons.

Subjects were screened using the post screening method in Annex A.1 of ITU-T Rec. P.913, with a Pearson correlation threshold of 0.3. This eliminated two subjects from pool 1 and two subjects from pool 2. Pool 1 contained two colorblind people, each of whom missed 10+ color plates. These subjects were retained, as the subject screening threshold did not indicate any unusual behavior. After screening, pool 1 contained 29 subjects and pool 2 contained 11 subjects. These pools were combined without any scaling, which adds noise to this analysis. The four comparisons that appeared in both sets were treated as separate data points (e.g., white to black).

Table 1 shows the Pearson correlation between color distance measured in the following two ways:

- Subjectively, reported as mean opinion score (MOS)
- Linear distance, measured using Euclidean distance

Linear distance is measured for each plane separately and for all three planes. Column “All Planes” contains the square root of the sum of the squared distances from each plane, measured separately. One limitation of these statistics is that the dataset ignores frequency of different color combinations. Table 1 should be considered a preliminary analysis.

Table 1. Pearson Correlation between Subjective Edge Distance MOS and Color Space Distance

Color space	Each Plane			All Planes
RGB	R =0.41	G =0.48	B =0.35	0.44
YCbCr	Y =0.51	Cb=0.37	Cr=0.44	0.69
XYZ	X =0.50	Y =0.51	Z =0.36	0.64
CIELAB	L*=0.42	A*=0.33	B*=0.15	0.24

Table 2 shows distance MOSs for most combinations of basic colors using the 29 subjects in pool 1. Table 3 shows the same distances measured in the luma plane of YCbCr. Notice that some obvious differences in Table 2 are negligible in Table 3. The nonlinearities of YCbCr can be seen by trying to add distances within Table 2. For example, white/gray (3.7) plus gray/black (3.6) is greater than white/black (6.3 versus 5.0). Tables 2 and 3 demonstrate how poorly the luma plane represents the human visual system.

[2] http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html

[3] Judd [5] recommends a surround that averages the luma of both samples, for decreased bias.

Table 2. Subjective Edge Distance[4]

	Black	Wh.	Red	Gr.	Yel.	Blue	Gray
Black		5.0	4.6	4.3	4.6	4.2	3.7
White	5.0		4.6	4.5	4.1	4.3	3.6
Red	4.6	4.6		4.3	4.1	4.3	4.2
Green	4.3	4.5	4.3		3.9	3.9	3.6
Yellow	4.6	4.1	4.1	3.9		4.1	4.1
Blue	4.2	4.3	4.3	3.9	4.1		4.0
Gray	3.7	3.6	4.2	3.6	4.1	4.0	
Brown	3.8	4.2	3.9	3.8	4.1	4.1	3.4
Orange	4.5	4.2	3.7	4.0	3.4	4.4	4.0
Purple	4.0	4.2	4.3	4.0	4.1	3.7	3.5

Table 3. Objective Distance in Luma Plane of YCbCr

	Black	Wh.	Red	Gr.	Yel.	Blue	Gray
Black		219	65	118	194	25	112
White	219		154	101	25	194	107
Red	65	154		53	129	40	47
Green	118	101	53		76	93	6
Yellow	194	25	129	76		169	82
Blue	25	194	40	93	169		87
Gray	112	107	47	6	82	87	
Brown	63	156	2	55	131	38	49
Orange	146	73	81	28	48	121	34
Purple	71	148	6	47	123	46	41

8. THE COLOR DISTANCE CHALLENGE

Objective video quality models currently rely upon the YCbCr color space. Consequently, measurements of color difference differ from human visual system perceptions of color difference. Objective measurements ignore higher cognitive phenomena, such as the eleven basic color terms and the simultaneous contrast effect. Objective video quality model performance is unlikely to improve without at least partially addressing these problems.

Objective video quality metrics do not necessarily need an optimized color space. The first need is for a distance measure for extremely large color differences. Errors on the order of a just noticeable difference will be subsumed by the error present in objective metrics and subjective scores. This solution has the advantage of simplicity: a look-up table would suffice, given sufficient subjective data. The accuracy of the distance measure could be demonstrated using a modified version of PSNR (e.g., calculated on color distance between pixels of the original and degraded frames).

The second need is for an improved color edge detector. Ideally, such an edge detector would incorporate knowledge of the visual system from the multiple perspectives presented in this paper. My experience gained over two decades of analyzing subjectively rated video sequence indicates that people care more about large edges than small. At 3 × 3 pixels, the Sobel and Laplace filters overemphasize noise. Thus, the color edge detector should measure large edges (e.g., a 13 × 13 pixel spatial information filter [11]).

The core challenge is that the solution will be more complex than a simple matrix manipulation of an existing color space. Any technique that is substantially more perceptually uniform than those seen in Tables 1 and 3 will require subjective testing and a nonlinear mapping.

The NTIAcolor subjective dataset provides a starting point for the development of a color distance metric or color edge detector. The MOS scores is available at http://www.its.bldrdoc.gov/resources/video-quality-research/data.aspx. Due to the fast and simple task, crowdsourcing may be an excellent mechanism to improve upon this database. The NTIAcolor database could be used for subject screening.

[4] Abbreviated column labels are white (Wh.), green (Gr.) and yellow (Yel.)

9. REFERENCES

[1] C. Poynton, ColorFAQ.pdf. Available from www.poynton.com

[2] R. G. Kuehni and A. Schwartz, *Color Ordered: A Survey of Color Order Systems from Antiquity to the Present*, Oxford University Press, Inc., 2008.

[3] R. G. Kuehni, *Color Space and Its Divisions: Color Order from Antiquity to the Present*, John Wiley & Sons, Inc., 2003.

[4] A. H. Munsell, *Atlas of the Munsell Color System*, 1913. Available: http://munsell.com/color-blog/atlas-color-system-charts/

[5] D. Judd, "Idea color space," *Contributions to Color Science*, National Bureau of Standards (NBS) Special Publication 545, 1979, p. 637-653. Reprinted from *Color Engineering*, 8(2), 1970.

[6] C. L. Hardin and Luisa Maffi, *Color Categories in Thought and Language*, Cambridge University Press, 1997.

[7] N. Leland, *Exploring color*, Northern Light Books, 1998.

[8] C. Reid, *Painting what you want to see*, Watson-Guptill Publications, 1983.

[9] E. H. Adelson, "Lightness Perception and Lightness Illusions," *The New Cognitive Neurosciences, 2nd Ed., Cambridge*, MA: MIT Press, pp. 339-351 (2000), 2000.

[10] M. Pinson et al., "The Influence of Subjects and Environment on Audiovisual Subjective Tests: An International Study," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 6, No. 6, October 2012, pp. 640–651

[11] "Spatial Information (SI) Filter," Available at http://www.its.bldrdoc.gov/resources/video-quality-research/guides-and-tutorials/spatial-information-si-filter.aspx