PERCEPTUAL ADAPTATION OF OBJECTIVE VIDEO QUALITY METRICS

Louis Kerofsky, Rahul Vanam, and Yuriy Reznik

InterDigital Communications, Inc.
{Louis.Kerofsky, Rahul.Vanam, Yuriy.Reznik}@InterDigital.com

ABSTRACT

The limitations of objective metrics such as PSNR in evaluating video quality are well known to experts but less know to general users. A video tool which exploits perceptual phenomena may report higher subjective quality but lower objective performance than a non-perceptual tool. This presents a problem when describing the performance of a perceptually motivated algorithm to general users. We propose a method for extending existing objective metrics to account for perceptual factors such as viewing distance, ambient contrast, etc. After describing the proposed algorithm extension, we examine the objective results of the proposal and perform subjective viewing tests to confirm the behavior of the extended objective metrics.

1. INTRODUCTION

Video quality metrics are in use to evaluate quality of video compression, delivery and display. An important application is providing a summary of the performance of a video compression algorithm. Metrics range greatly in degree of sophistication from simple mean square error based comparison to a reference as with PSNR to sophisticated human visual system models, Barten [1], Visual Difference Predictor [2], Sarnoff Just Noticeable Difference [3] and others. Similarly metrics differ in their assumptions about an available reference image. The metrics mentioned above are *full-reference* requiring a full reference image for definition. The assumption of a reference image can be reduced as with *reduced - reference* metrics or entirely eliminated with *no- reference* quality metrics.

Limitations of PSNR as a visual quality metric have been well discussed in the technical community, for example in [4]. Objective metrics based on SSIM or MS-SSIM [5] have become popular recently, but the basic SSIM metric does not provide significantly more information than a basic PSNR calculation. For example, the work in [6] derives the relation between PSRN and SSIM. Despite the known limitations, PSNR is commonly used to evaluate system performance and application such as tuning encoding parameters due to the simplicity and application to a narrow range of parameter changes. In a limited application such as deciding between tools in a video codec and when used by an expert aware of its limitations, PSNR can be a valuable tool.

A problem arises when interacting with a less technical group. It is common to be asked about PSNR performance of a product when introducing it to the market place for instance. Academic references about the inadequacy of PSNR in capturing visual performance are of limited use when the customer demands a PSNR number. For example, it is well known that visual perception phenomena can be exploited to improve the performance of a compression system. Invisible but complex detail can be removed reducing the necessary bitrate to achieve similar subjective quality. As an example, the oblique effect, in which a viewer is less sensitive to diagonal frequency that to frequency in the cardinal directions, has been exploited to remove diagonal high frequency content to simplify encoding without impact on perceived quality [7]. Such perceptually invisible modifications to an image will degrade a simple objective metric such as PSNR. Thus, focus on an objective metric such as PSNR may mask the benefits of perceptual processing. We are faced with the following problem: how to use the language of traditional PSNR familiar to the customer while communicating the performance benefits of system exploiting perceptual effects. One obvious limitation of PSNR is not accounting for viewing conditions.

To address above described problem, we introduce a method for incorporating viewing conditions, particularly viewing distance, with existing objective metric calculations. This technique is applied to PSRN, SSIM, and MS-SSIM and compared to the results or subjective MOS tests performed over same content and parameters of viewing setup. It is shown that adapted versions of PSNR, SSIM, and MS-SSIM show similar behavior as MOS scores with changing viewing conditions.

The remaining sections of this paper are organized as follows. Section 2 provides a description of the modifications of traditional PSNR to account for perceptual factors of display contrast and viewing distance. Section 3 provides results of extended objective metrics computed for example sequences and viewing conditions. Section 4 presents results on subjective testing on the same data use for objective calculations. Section 5 provides our conclusions.

2. ALGORITHM DETAILS

2.1. Fundamentals

It is well known that the perception of visual quality depends upon the viewing conditions. In this paper we focus on the subjects viewing distance as a primary viewing condition factor. In psychophysics evaluations the viewing distance is often expressed as a number of picture heights rather than physical units. The contrast sensitivity function (CSF) defines for a given viewing distance the minimum contrast needed for a spatial modulation to be visible. The spatial modulation is described by a frequency in cycles per visual angle. Mathematical models of the CSF are commonly included in advanced visual models and quality metrics. A detailed mathematical model of the CSF is given in [8]. An example plot of a contrast sensitivity function is shown in Figure 1 with indications of the visible and invisible regions of contrast for various spatial frequencies measured in cycles per degree of visual angle.

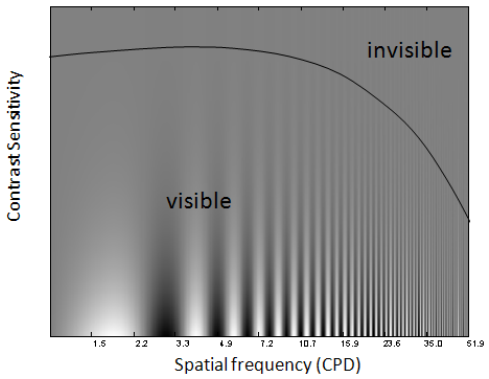

Figure 1 Contrast Sensitivity Plot

2.2. System Diagram

A diagram of the algorithm defining the extension of an objective metric to include viewing condition dependence is shown in Figure 2. We assume an objective metric $M(.,.)$ for computing a full reference image quality metric is given for instance M could be PSNR or SSIM. A description of viewing conditions is given in the form of a display contrast ratio C accounting for ambient and viewing distance D. The contrast is used to determine a cut-off frequency above which image modulations are invisible since the display is unable to achieve contrast high enough for the visual system to distinguish. Using the viewing distance, the cut-off frequency is converted from cycles per visual degree to cycles per pixel. A low pass filter, F, is designed with this cut-off frequency. The reference image I_{ref} and the image under evaluation $\boldsymbol{I_{test}}$ are each filtered to produce modified images.

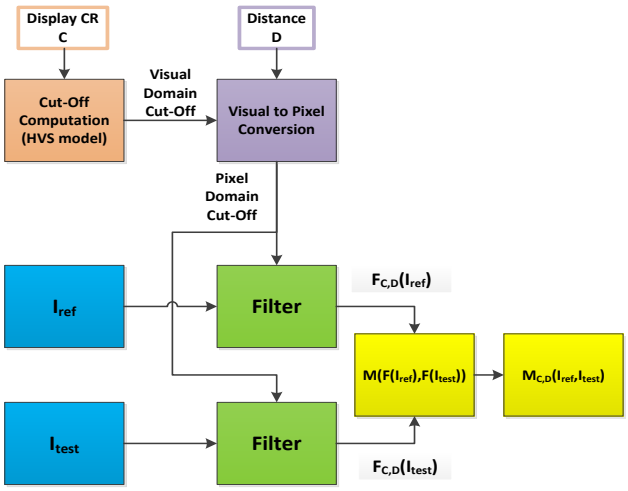

Figure 2 System diagram extending metric M to include viewing conditions in metric M_P

The proposed extended objective metric $M_{C,D}(.,.)$ is defined as the given objective metric $M(.,.)$ evaluated on the filtered reference and filtered test images. This process is summarized in Eq. (1).

$$M_{C,D}(I_{ref}, I_{test}) \equiv M\left(F_{C,D}(I_{ref}), F_{C,D}(I_{test})\right) \quad (1)$$

In the above formula, $\boldsymbol{F_{C,D}(I)}$ is the result of filtering the image $\boldsymbol{I}$ with a low pass filter $\boldsymbol{F_{C,D}}$ designed based on the viewing conditions described by contrast and viewing distance.

2.3. Cut-off frequency determination

One of the important characteristics of video reproduction setup is the contrast ratio achievable by the display under certain ambient lighting conditions. Given a finite contrast ratio, the highest frequency which is visible under a given CSF and viewing distance can be determined. The contrast ratio defines a maximum contrast achievable for any image shown on the display. The limit on display contrast determines a lower bound on the contrast sensitivity achievable by the display. Spatial frequencies above the cut-off frequency require contrast levels greater than that achievable on the display to be visible. Thus given a display contrast ratio we can determine a cut-off frequency beyond which detail will not be visible. This concept is illustrated in Figure 3. Three regions of pairs (spatial frequency, contrast sensitivity) are indicated. The region above the CSF curve is invisible to the viewer. The region below the CSF curve is visible. The visible region with sensitivity below the minimum display sensitivity cannot be achieved on a display and is indicated as infeasible. The minimum sensitivity and the highest visible spatial frequency are linked by the CSF. Mathematically, the upper bound on the display contrast C determines a lower bound on the achievable contrast

sensitivity. The CSF then relates the lower bound on sensitivity to an upper bound on visible frequency.

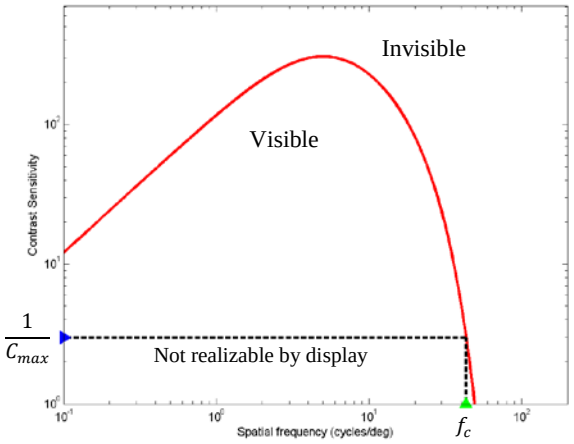

Figure 3 Relating minimum sensitivity and maximum visible frequency via the CSF

An appendix provides details on the inversion of the CSF model. We use an approximation of the inverse CSF model for determining the cut-off frequency f_c from the maximum contrast C_{max} achievable on the display in a given ambient environment.

2.4. Visual frequency conversion

The spatial frequency $f_c'(s)$ given by the inverse CSF is in units of cycles per visual degree. For use in image processing these need to be converted to pixel specific units. The conversion between visual angle and pixels relies upon the viewing distance and display pixel density i.e. ppi. The relevant geometry is shown in Figure 4.

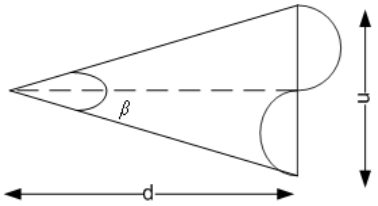

Figure 4 Geometry relating visual angle, viewing distance and cycle length

The spatial frequency f of a sinusoidal grating with cycle length n pixels can be computed as:

$$f = \frac{1}{\beta}\ [cpd], \beta = 2\tan^{-1}\left(\frac{n}{2d\rho}\right), \qquad (3)$$

where the distance from viewer to display is d inches. The visual angle corresponding to this cycle is β degrees. The display density is ρ ppi. When the viewing distance is expressed in pixels this conversion from inches to pixels using the pixel density is not needed.

2.5. A synthetic example

A clear limitation of PSNR can be seen in comparing the images shown in Figure 5. Both are full HD images, 1920x1080, consisting of columns of full white or full black pixels. Image A, consists of alternating full black and full white columns beginning with white. Image B consists of alternating columns of full black and full white columns beginning with black. Thus image B differs from image A by a 1-pixel horizontal translation. The difference is invisible unless carefully controlled viewing with the images flipping between A and B is done. For practical purposes A and B are subjectively identical. Example images of A and B are included in this document but cannot be distinguished.

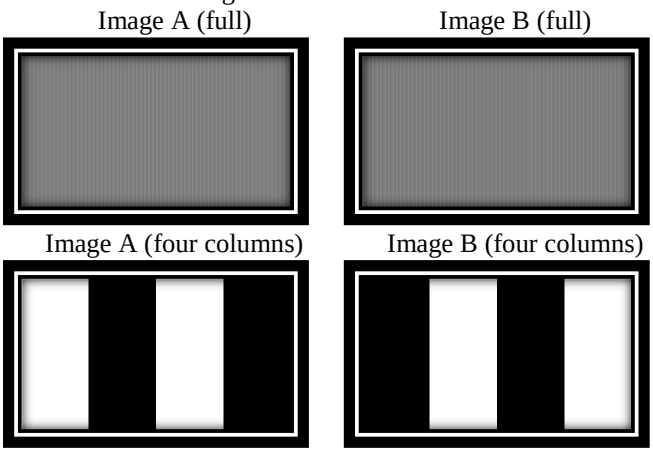

Figure 5 Images A and B full size and zoomed

Computing the PSNR between A and B gives a different conclusion. The MSE difference between A and B is maximum i.e. no other image pair can have higher MSE. Thus the PSNR between A and B is the absolute smallest possible for an image pair. If using 8-bit images the MSE becomes maximal 255^2 and the PSNR becomes zero.

Table 1 Objective metric values

Frequency	PSNR	SSIM	MS-SSIM
1. 0000	0.0000	-0.9964	-0.9964
0.8408	31.8282	0.6443	0.9805
0.7072	37.7853	0.9815	0.9991
0.5946	39.2811	0.9821	0.9991
0.5000	40.4581	0.9824	0.9991

For purposes of illustration we apply low pass filters of various cut-off frequencies prior to calculation of the PSNR, SSIM and MS-SSIM using the implementation provided at [11]. The value calculated for metric at each cut-off frequency is given in Table 1. If we compute the objective metrics for various low-pass cut-off frequencies, we quickly converge to a more representative difference for images which are subjectively identical.

For this extreme synthetic example, the viewing condition adaptivity restores some meaning to existing objective metric calculations by increasing to high quality as the viewing distance increases where the images are subjectively identical. We note that although MS-SSIM includes several low-pass filter stages to produce multiple resolutions, MS-SSIM reports a low value on this image pair unless preprocessing is used.

3. OBJECTIVE EVALUATION

We demonstrate example calculation of P-PSNR on sample content and varying viewing conditions. An effective display contrast of 100:1 was used with different viewing distances.

3.1. Methodology

Source material was generated by using two well-known full HD, 1920x1080, sequences from the video coding community "ParkScene" and "Kimono". Each was encoded using the open source x264 encoder [9] with three fixed quantization levels QP = 26, 32, 38 giving a range of compression quality and artifacts. Each sequence was decoded to provide six sample video sequences, two content at three encoding levels each.

A range of viewing distances in picture heights was selected to sweep the distance parameter. For each viewing distance, an appropriate cut-off frequency was selected based on a CSF model and assumed display contrast ratio of 100:1. The list of distances and corresponding cut-off frequencies is shown in Table 2.

Table 2 Distance Cut-Off Frequencies

Distance (PH)	Cut-Off Frequency
1	1.0000
3	1.0000
5	0.7646
7	0.5461
9	0.4248
11	0.3475
13	0.2941

Three objective metrics PSNR, SSIM, and MS-SSIM were computed at each distance between filtered original and filtered decompressed images giving the perceptual extensions P-PSNR, P-SSIM and P-MS-SSIM respectively.

3.2. Results

The extensions of PSNR, SSIM and MS-SSIM are plotted as a function of viewing distance for three different compressed versions of the "ParkScene" sequence and "Kimono" sequences in Figure 6, Figure 7 and Figure 8 respectively. We observe that the metrics all increase with viewing distance for both of these sequences under all three coding conditions QP = 26, 32, 38.

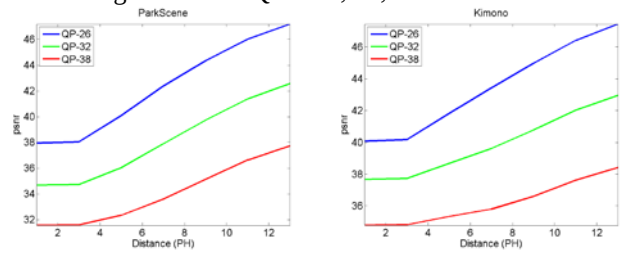

Figure 6 P-PSNR versus distance

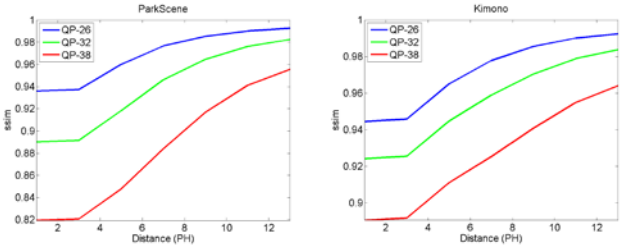

Figure 7 P-SSIM versus distance

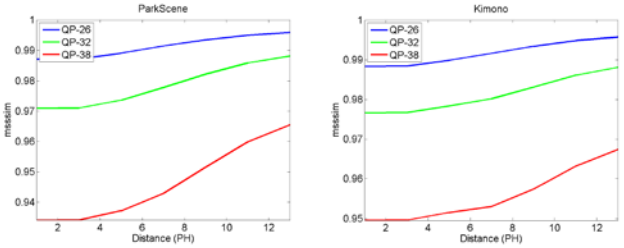

Figure 8 P-MS-SSIM versus distance

We note that MS-SSIM contains filtering and subsampling which appear similar to use of the viewing condition driven filtering prior to the objective metric calculation. Note however that MS-SSIM does not have an adaptation to viewing conditions rather the same multi-resolution derivation of calculations used regardless of viewing conditions.

The PSNR rises without bound as the distance increases. The SSIM and MS-SSIM metrics increase toward saturation as the distance parameter increases. This saturation of quality is more representative of the expected property. The lower quality versions increase more rapidly than the higher quality versions of the sequence.

4. SUBJECTIVE EVALUATION

We investigate the subjective quality of the same compressed video sequences in a viewing test. Sixteen (16) viewers were used to rate the subjective quality of the video under different viewing distances. Nine (9) were imaging experts while the remaining seven (7) were viewers without specific experience in image quality evaluation.

4.1. Methodology

For testing, each of the compressed sequences was used without filtering. Subjects were asked to rate the quality of video using the five point scale Table 3. Prior to scoring, subjects where shown a video sequence encoded with low quantization QP=20 and told this was an example of excellent quality. The same video sequence encoded with high quantization QP=40 was shown as an example of Bad quality.

Table 3 Quality Scale

5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

During evaluation subjects were placed at three different viewing distances 3H, 5H and 9H from the display. At each distance, the subjects were shown the six compressed sequences in a random order and asked to provide a score. Sony LMD-941W reference monitor was used in our tests.

4.2. Results

Results of the subjective voting are shown below. Results for individual sequences are shown with subjective quality versus viewing distance plotted for three quantization levels QP 26, 32, 38.

We observe some trends in this subjective data. Looking at the ParkScene results in Figure 9, the subjective quality scores converge with increasing viewing distance. For low quality sequences at QP=38 the subjective score rises with increasing viewing distance presumably because compression artifacts are less visible and distract less from the quality. The higher quality sequences both decline slightly in subjective quality but tend to converge at the extreme viewing distance. Looking at the Kimono results in Figure 10, the lowest quality encoding, QP=38, increases in subjective score as the viewing distance increases. The higher quality encodings are nearly constant with the note that the qualities converge at the farthest viewing distance. Interestingly, the subjective quality of high quality encodings falls as the viewing distance increases. The extremely high P-PSNR values > 40 don't translate to improvement in subjective quality for images already having high subjective quality. In this case, significant compression artifacts are not visible even at the closest distance. When viewed at a higher distance, the content has less visible detail and hence is rated lower. An interesting question is how subjects account for different distance when providing subjective ratings. We did not make an attempt to investigate this in detail as we are focusing on use cases where compression artifacts are noticeable at close viewing distance. For both sequences, the subjective quality scores of fine and coarsely quantized data converge at longer viewing distance suggesting inability to differentiate between different levels of compression at these distances.

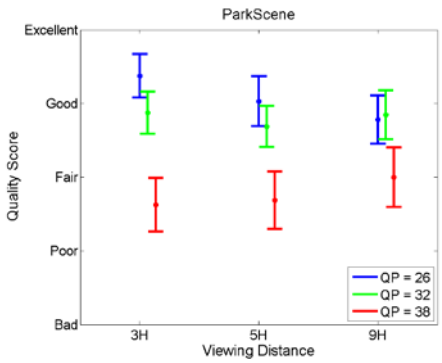

Figure 9 Subjective quality versus viewing distance "ParkScene"

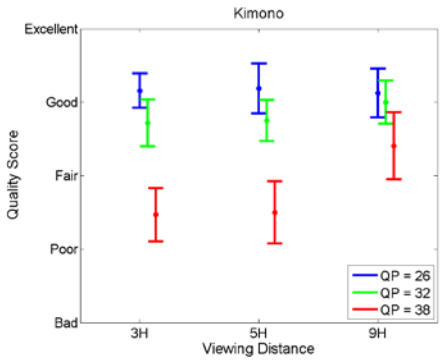

Figure 10 Subjective quality versus viewing distance "Kimono"

5. CONCLUSIONS

A method for incorporating viewing conditions with a full reference objective video quality metric was proposed. The method relies on using a low pass-filter to account for viewing distance. This can be effective in cases where the metric does not contain the ability to vary viewing conditions or when the viewing condition parameters cannot be modified in a particular implementation.

We have shown the effectiveness of the proposed method by producing perceptually-adapted versions of PSNR, SSIM, and MS-SSIM and by comparing their outputs to MOS scores produced by human observers. It is shown that with our proposed modifications those metrics correlate better with the MOS scores. The variations in scores with viewing condition are otherwise absent.

6. REFERENCES

[1] P. Barten, Contrast sensitivity of the human eye and its effects on image quality, Vol. 72. SPIE press, 1999.

[2] S. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity." SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology. International Society for Optics and Photonics, 1992.

[3] J. Lubin and D. Fibush. "Sarnoff JND vision model." (1997): 97.

[4] Z. Wang and A. Bovik. "Mean squared error: love it or leave it? A new look at signal fidelity measures." Signal Processing Magazine, IEEE 26.1 (2009): 98-117.

[5] Z. Wang, A. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process. 13, pp. 600–612, Apr. 2004.

[6] A. Hore and D. Ziou. "Image quality metrics: PSNR vs. SSIM." Pattern Recognition (ICPR), 2010 20th International Conference on. IEEE, 2010.

[7] Y. Reznik and R. Vanam. "Improving the coding and delivery of video by exploiting the oblique effect." Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE. IEEE, 2013.

[8] P. Barten, "Formula for the contrast sensitivity of the human eye." Electronic Imaging 2004. International Society for Optics and Photonics, 2003.

[9] VideoLAN x264 encoder project described at http://www.videolan.org/developers/x264.html.

[10] R. Vanam and Y. A. Reznik, "Perceptual pre-processing filter for user-adaptive coding and delivery of visual information," in Proc. Picture Coding Symposium (PCS), pp. 426 – 429, 2013.

[11] Multimedia Signal Processing Group MMSPG at EPFL http://mmspg.epfl.ch/vqmt

APPENDIX: CALCULATION OF CUT-OFF FREQUENCY

A practical display has a limit on the achievable contrast under particular ambient viewing conditions. A CSF model relates this minimum contrast sensitivity to a maximum visible spatial frequency defined as the highest frequency where the CSF exceeds this minimum sensitivity.

The cut-off frequency is computed as follows:

1) Determine the maximum contrast achievable on the display on the display under existing conditions, C_{max}
2) The minimum sensitivity $S_{min} = 1/ C_{max}$
3) Find the solution of $S'(u) = S_{min}$. This defines the cut-off frequency.

To determine the frequency at which S_{min} is achieved, a mathematical model of the human contrast sensitivity function developed by Barten [8] is summarized below. The sensitivity threshold of a spatial frequency of u cycles per degree is given by:

$$S(u) = \frac{Ae^{-Du^2}}{\sqrt{(B+u^2)\left(C + \frac{1}{1-e^{-0.002u^2}}\right)}}$$

Where constants A, B, C, D are defined below in terms of the object luminance L, the object size X_0.

$$A = \frac{5200E}{\sqrt{0.64}} \qquad B = \frac{1}{0.64}\left(1 + \frac{144}{X_0^2}\right)$$

$$C = \frac{63}{L^{0.83}} \qquad D = 0.0016\left(1 + {}^{100}/_{L}\right)^{0.08}$$

$$E = exp\left(-\frac{ln^2\left(\frac{L_s}{L}\left(1+\frac{144}{X_0^2}\right)^{0.25}\right) - ln^2\left(\left(1+\frac{144}{X_0^2}\right)^{0.25}\right)}{2\,ln^2(32)}\right)$$

For large u, this can be approximated by

$$S_1(u) = \frac{Ae^{-Du^2}}{\sqrt{(B+u^2)(C+1)}}$$

This model is a function of the viewing conditions. For given viewing conditions, the display brightness and size determine the constants L and X_0. Thus the function $S_1(u)$ is determined.

Function $S_1(u)$ can be analytically inverted to give:

$$u = S_1^{-1}(s) = \sqrt{\frac{LambertW\left(\frac{2DA^2e^2DB}{(C+1)s^2}\right)}{2.D}}$$

where $LambertW(z)$ is a solution of equation:

$$LambertW(z).e^{LambertW(z)} = z$$

This relates the minimum sensitivity achievable on the display to the highest visible frequency through the CSF model. A plot summarizing this inverse relation is shown below. Three regions are identified: points above the CSF cannot be seen by the viewer, points below the CSF are visible, points to the left of the minimum sensitivity are infeasible, limits on the display contrast prevent these sensitivity levels from being displayed.