https://doi.org/10.2352/ISSN.2470-1173.2017.14.HVEI-131

Industry and business perspectives on the distinctions between visually lossless and lossy video quality: Mobile and large format displays

K. Brunnström(Editor)[a]
R. S. Allison[d], D. M. Chandler[b], H. Colett[f], P. Corriveau[f], S. Daly[c], J. Goel[e], J. Knopf[f], L. M. Wilcox[d], Y. Yaacob[b], S.-N. Yang[g], Y. Zhang[b]
(Authors in alphabetical order)

[a]*RISE ICT/Acreo, Stockholm, Sweden and Mid Sweden University, Sundsvall, Sweden*
[b]*Shizuoka University, Hamamatsu, Shizuoka, Japan*
[c]*Dolby Laboratories, USA*
[d]*York University, Centre for Vision Research, Toronto, Canada*
[e]*Qualcomm, Display Video Processing Group, Markham, Canada*
[f]*Intel Corp, Santa Clara, CA, USA*
[g]*Pacific University, Forest Grove, OR, USA*

Abstract

This paper will explore the mobile and business perspectives of visually lossless image quality, as well as review recent scientific advances. It is the outcome from the Special Session on Visually Lossless Video Quality for Modern Devices: Research and Industry Perspectives organized at the Human Vision and Electronic Imaging 2017 by IS&T at San Francisco Airport, Burlingame, California, USA, Jan 29 - Feb 2, 2017. It summarizes four presentations and a panel discussion.

Introduction

An often-sought goal in image and video coding is to generate images and videos which are maximally compressed, but visually lossless from the uncompressed versions. In terms of visual psychophysics, such an objective would mean that the compression distortions are below the threshold of detection. This can be loosely phrased that it cannot be seen by a majority of population most of the time.

The research area has matured and commercial applications have emerged e.g. in the mobile industry where sometimes visually lossless compression is used in order to cope with the bandwidth requirements for the visual presentation on the display. This paper will explore the mobile and business perspectives of visually lossless image quality, as well as review recent scientific advances. It is the outcome from the Special Session on Perceptually Lossless Image Quality for Mobile Devices organized at the Human Vision and Electronic Imaging 2017 by IS&T at San Francisco Airport, Burlingame, California, USA, Jan 29 - Feb 2, 2017. It will summarize four presentations (see titles and abstracts below) and a panel discussion.

Background

Common Industrial Visual Quality Assessment

In general, applications of visual quality occur in the industrial arena and have been directed toward a wide range of quality. This includes both testing methodologies as well as predictive models. For example, in the widely used International Telecommunication Union (ITU) guidelines for subjective video quality assessment, the Double Stimulus Continuous Quality Scale (DSCQS) method ITU-R Rec BT.500-13 (BT500)[1] or the Absolute Category Rating (ACR) ITU-T Rec P.910 [2] uses a 5-grade quality scale with subject input options of Excellent, Good, Fair, Poor, and Bad, as shown in Table 1 to left. Another scale listed in the BT500 [1] guidelines is the ITU impairment scale, which uses the following options: Imperceptible, Perceptual but not annoying, Slightly Annoying, Annoying, and Very Annoying, see Table 1 to the right. Note that these scales were intended for a single stimulus, but can also be paired with a known reference, as in the above mentioned DSCQS with an explicit reference or in ACR with a hidden reference. Both methods span a substantial range of visual quality, that is, they include both sub-threshold and suprathreshold visible differences. In many applications, such assessment of overall suprathreshold visual quality is exactly what is needed.

Table 1: The Quality and Impairment scales of BT500

Five-grade scale			
Quality		**Impairment**	
5	Excellent	5	Imperceptible
4	Good	4	Perceptible, but not annoying
3	Fair	3	Slightly annoying
2	Poor	2	Annoying
1	Bad	1	Very annoying

For paired comparisons, Likert scales are often used since they have a bipolar structure that enables consideration of the two stimuli, as shown in Table 2[1]. These are generally arranged in a left-to right orientation corresponding to two images being shown side-by-side. However, in some applications, the quality sought after is strictly visually lossless. That is, all visible differences (distortions) are designed to be below the human threshold and the intent of testing is to determine if this goal has been achieved. One can easily see that the five graded Quality scale in Table 1 has no ability to determine whether visually lossless quality occurs or not. The category 'Excellent' *may* imply visually lossless in some applications, and for *some* viewers but this is generally not the case. On the other hand, the impairment scale does have the ability to assess visually lossless behavior, such as the boundary between response 5 and 4. Likewise, the thresholds could possibly be determined from Likert scales using the responses -1, 0, +1, although the adjectives given are not as exact regarding threshold as does the ITU impairment scale.

Table 2: The Comparison scale of BT500

–3	Much worse
–2	Worse
–1	Slightly worse
0	The same
+1	Slightly better
+2	Better
+3	Much better

In most conceptions of visually lossless, two images (or videos) are compared, with one being a Reference and one being a Distorted version. The distortions may not mean solely deviations from realism (artifacts, such as blocking artifacts and ringing) but include any changes from the reference, even if plausible to realism (such as color shifts, tonescale shifts, blur). Terms like Original, Source, and Uncompressed are also used for the Reference, but the reference may not always be the original version, or its source, and the distortion may not involve compression so those terms do not generalize. For example, in post-production workflows, the term Mezzanine Content is used to describe content that is compressed very lightly, but is subthreshold, and is used at certain stages of the workflow. This Mezzanine content is then further compressed for distribution. So in this case, both the reference and distorted would be compressed video streams. Although there is not complete agreement on all of the details, the terms visually lossless, perceptually lossless, perceptually transparent, and visually identical are all referring to the same thing. When only visual aspects are being considered, some prefer the terms visually lossless, etc., and save analogous terms like perceptually lossless to cases when there are multiple sense dimensions involved, such as audio-visual assessment.

Thresholds and the Psychometric Function

Unfortunately, the visual threshold for most dimensions of imagery is not a step function as might be implied from the impairment table in Table 1. Rather, it is a gradual transition. Rigorous psychophysical experiments (typically, academic as opposed to industrial) tend to focus more specifically on threshold perception, and ignore the distinctions above threshold. A psychometric function is measured that finds the subject's probability of detection as a function of the strength of the parameter of interest, as shown in Figure 1 (left). For many stimuli, psychometric functions are generally of the same shape across different individuals, but exhibit varying sensitivity (causing horizontal shifts on the x-axis, Figure 1, right). For this example, a threshold may be assigned to the stimulus intensity corresponding to 50% seen (~24, pink arrow, left plot), but this is obviously just definitional, and the threshold is just a shorthand for the overall position of the psychometric function. For this plot, stimuli of strengths from 40 to 45 seem to give detectability of ~100% and are just surpassing the threshold region, which may still be considered a very slight distortion. The methods used to determine such psychometric functions do not have the ability to differentiate stimuli of strengths > 45, which is the suprathreshold region, to which the majority of the scales described above are allocated. To determine an average threshold across varying individuals, the detections thresholds from each are averaged, and a new psychometric function can be derived which describes the average subject (green curve in Figure 1, right).

One common distinction between industrial visual quality vs. academic vision science is that industrial testing tests many more viewers than academic testing, but that academic testing tests each viewer much more thoroughly. In most industrial testing, there are much fewer trials per individual (sometimes just one), as well as less stimuli allocated to the threshold region, because the stimuli are needed to span a wider range of quality differences. As a result, in most industrial testing, a psychometric function cannot be constructed per individual. But industrial testing does have much data available as a result of testing more viewers, and attempts to determine thresholds can be made by averaging all subject responses (e.g., by looking at data for responses 4 and 5 in the ITU scale) and averaging those to get a group psychometric function.

In much industrial testing, such as using the scales in Table 1, attempts are occasionally made to determine thresholds by averaging the responses across all observers. But without first determining the thresholds for each viewer, the overall psychometric function ends up being wider, and may result in a different threshold than the average threshold determined when individual psychometric functions are measured. As a result of these many factors, experiments are generally designed to either assess the threshold, or assess the full range at the expense of loss of accuracy around threshold. These design decisions involve both stimuli set as well as experimental methodologies.

From threshold to JNDs

In most terminology, Just Noticeable Differences (JNDs) are synonymous with the threshold corresponding to the 50% response (after correction for guessing) [3]. In industrial applications, JNDs tend to be used for grouped observer perception, as opposed to

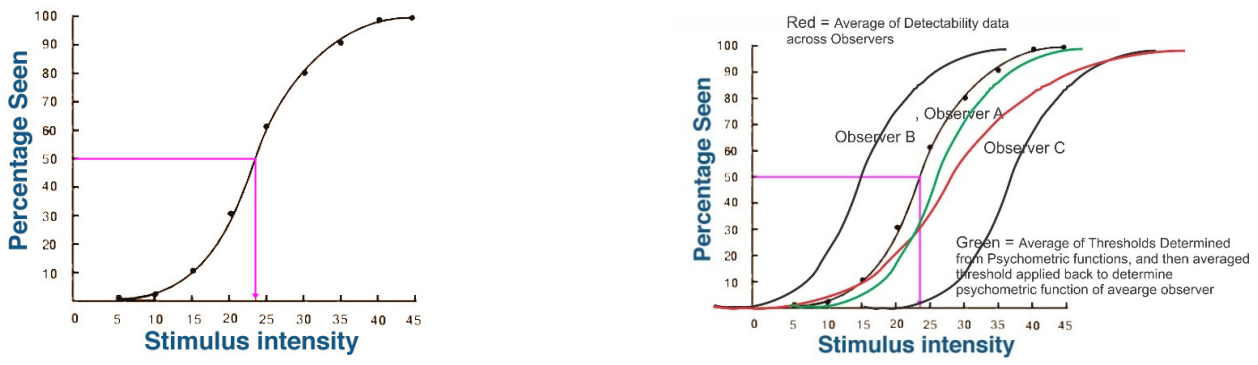


Figure 1: Left: psychometric function for an individual. Right: psychometric functions for multiple subjects and different methods to determine psychometric functions or thresholds for group behavior.

describing individuals. JNDs are often added and used as a ruler to determine quality categories. For example, it has been claimed that six JNDs correspond to a difference across subjective quality categories [3], such as from 'fair to 'good'. Another example of their usage is that one JND is not considered an advertisable difference; because it means only half the observers detect the difference. Notice that the 50% criterion is shifted from a single subject's probability of detection to the performance of a group (e.g., corresponding to the red curve in Figure 1). Unfortunately, JND summation only works for small numbers, and saturation occurs for larger visible differences. The visual system functioning as derived from JND summation is also known to deviate from that derived from appearance estimates. For example, the luminance nonlinearity derived from thresholds deviates from one derived from suprathreshold appearance steps. Various theories have been proposed and tested for such deviations [4]. Fortunately, for the goals of visually lossless quality, neither describing nor understanding large appearance differences is needed.

Sub-threshold explorations

In this century, research in quality assessment has been directed to understanding sub-threshold vision. Motivations range from frustrations with the visual quality task interfering with the overall quality of experience to observations that many viewers may not be aware of visual distortions that are still considered important to the product. An example of the former is that in determining quality of experience of differing display capabilities in conveying the emotions of a narrative movie, natural viewing of the movie with audio from beginning to end is required. However, such requirements pose extreme difficulties to traditional psychophysical testing methods. The common methods of viewing, comparing, and rating video clips of 10-20 sec duration put the viewer in a completely different state of mind than when actually watching and following the story. Examples of the latter are numerous in cases where those involved in the professional workflow of content notice far more details relating to their craft than the consumer viewer. Rather than assuming what the viewer does not notice is not important, the presumption is that the net total of experience with the craft affects the viewer in a number of ways, e.g. honing their attention to specific attributes. These highly trained observers may be unaware of this impact. For example, those in the craft readily use vertical camera angle placement to show dynamics of character subordination/ dominance [5], but how many consumer viewers notice such changes? Another example occurs for studies of discomfort, such as for stereoscopic displays or virtual reality, where the viewer may not notice signs of impending discomfort until it is too late.

Rather than use traditional psychophysical testing (whether industrial or academic), physiological measurements can be used. They can allow for the studying naturalistic viewing, as well as the subthreshold region. Turn-key research equipment now enables eye-tracking, EEG measurements, galvanic skin responses, facial thermal emission imaging, and visible facial expression and reaction imaging. Such techniques are now currently being used to assess levels of emotional engagement as a result of technical display differences [6] or in causing stress on the oculomotor visual system, see Section Colett et al, below.

Detection of Compression Artifacts on Laboratory, Consumer, and Mobile Displays

Y. Zhang, Y. Yaacob and D. M. Chandler

As outlined above, a range of parameters have been evaluated in threshold-based approaches to quality assessment (using forced-choice procedures and calibrated displays). However, practitioners often find that such thresholds are much lower than commonly visible in many applications, particularly when display characterization is not performed, see Wilcox et al below. In addition to the impact of the task demands, three candidates for such discrepancy are the display, the signal, and the viewing distance/angle. In the case of the display, factors such as contrast loss due to tonescale variations, ambient light and display reflectivity, motion blur due to temporal response, loss of high frequencies due to spatial MTF, and dynamic range variations are considered the most likely. Regarding the signal, the content's noise level and texture are the primary suspects in elevating thresholds due to masking. Lastly, psychophysical thresholds have strong frequency dependence, so viewing distance miscalibration can shift expected frequencies to higher values where the thresholds are generally higher. Off-angle viewing can significantly lower the contrast displayed with LCD technologies, thus lowering the contrast of the distortion from that expected using threshold data.

Consequently, it remains unclear whether such thresholds are valid when measured for true broadband compression distortions in actual images/videos presented on mobile and consumer-grade displays. In this section, we discuss our explorations of the display portion of the issue. Specifically, we asked:

1) Can thresholds measured on mobile devices yield the same results as those measured on laboratory and desktop displays when viewing conditions and display EOTFs (Electrical to Optical Transfer Functions) are kept constant?
2) How are the thresholds affected when EOTFs change on mobile displays, and do such changes agree with model predictions?
3) How do the variabilities in thresholds due to (1) and (2) compare to the variability across subjects, content, and gaze location?

Here, we present some preliminary findings of a pilot experiment designed to shed light on these issues. We measured contrast detection thresholds for HEVC compression distortions in small images using a mobile device (Apple iPad), and a forced-choice procedure. We discuss how these thresholds compare to similar thresholds measured on other displays, on the same display but with different display settings.

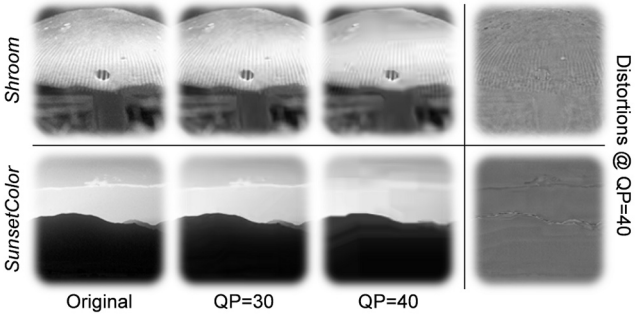


Figure 2 Stimuli used in the study—original and HEVC-compressed image segments from the CSIQ image quality and masking databases[7].

Effect of Display Type: Mobile vs. Desktop vs. Laboratory

Contrast detection thresholds for High Efficiency Video Coding (HEVC) [8] compression distortions were measured for crops from two images from the CSIQ masking database [7]; images *Shroom* and *SunsetColor* (see Figure 2). The compressed images were generated by using the reference HEVC encoder and by adjusting the QP value from 1-51.

The thresholds were measured on three displays:

- a Display++ LCD monitor from Cambridge Research Systems,
- a consumer-grade LCD monitor from I-O Data, and
- an Apple iPad Air 2.

All three displays were adjusted to have similar EOTFs and identically sized stimuli (3x3 degrees). The EOTFs were measured by using a DataColor Spyder5 in a darkened room. Figure 3 shows the measured EOTFs. The solid lines denote fits of the function:

$$L = a + (b + kV)^{\gamma}$$

to the measured data. Here, L denotes luminance, and denotes 8-bit pixel value; the measured parameters are shown in the legend of Figure 3 for each display.

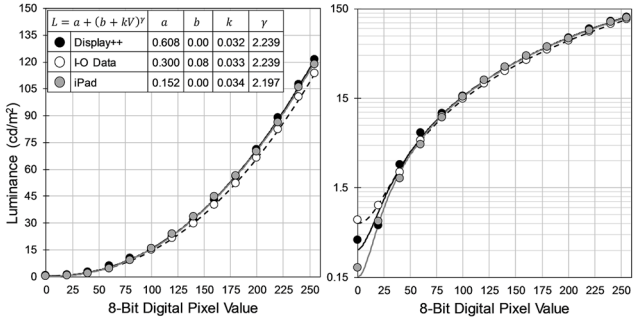


Figure 3: EOTFs of the three displays on linear and logarithmic luminance scales.

The contrast thresholds were measured by using a three-alternative forced-choice procedure guided by a Quest staircase with a fixed 48 trials, a 10 ms time-limit per stimuli presentation, and audio feedback. Three trained male adults with normal or corrected-to-normal vision (YZ, YY, and DC, the three authors of this section) served as subjects in the experiment.

Figure 4 shows the resulting thresholds. For subjects YZ and YY, an ANOVA revealed that the display type had a non-significant effect on the thresholds (YZ: p = 0.1822; YY: p = 0.08596). For subject DC, the effect of the display type was significant at the 0.02 level, but non-significant at the 0.01 level (p = 0.0155). For all three subjects, the effect of the image was significant ($p < 0.01$). The suggestion is that, for a given subject on a given image, there is just as much variation in the thresholds between different trials as there is between the different monitors. Moreover, these variations are all much less than the variation between subjects. For image *Shroom*, the standard deviation across displays was approximately 1.5 dB (averaged across subjects); whereas the standard deviation across subjects was approximately 3 dB (averaged across displays). For image *SunsetColor*, the standard deviation across displays was approximately 2 dB (averaged across subjects); whereas the standard deviation across subjects was approximately 1 dB (averaged across displays).

These results suggest that thresholds measured in the laboratory setting (either by using a specialized display or a modern desktop monitor) are still fruitful when the content is viewed on an iPad. Similarly, for the stimuli used in this study, thresholds can be measured directly on a mobile display. These suggestions, of course, assume that viewing conditions and EOTFs remain similar.

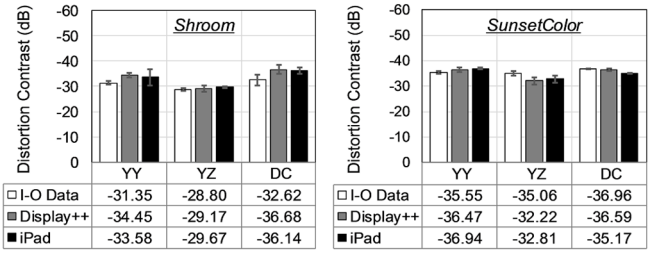


Figure 4: Contrast detection thresholds on different displays. Error bars denote ±1 standard deviation of the mean. Note that the vertical axis is reversed, and thus taller bars represent lower thresholds.

Effect of Display Setting

The ability to measure thresholds directly on a widely used mobile device such as the iPad, enables the possibility of measuring thresholds via crowdsourcing. However, subjects might erroneously adjust the iOS "Brightness" setting, thereby affecting the EOTF and ultimately the affecting the thresholds. Similarly, subjects might mistakenly perform the experiment in a non-darkened room, thereby affecting the thresholds.

Thus, in a follow-up pilot experiment, we measured thresholds on the iPad under three iOS "Brightness" settings: 0%, 50%, and 100%; and at 50% in a room lit by daylight (as opposed to a darkened room). The stimuli and procedures were identical to the previous experiment. Only the third author of this section (DC) participated in this pilot experiment.

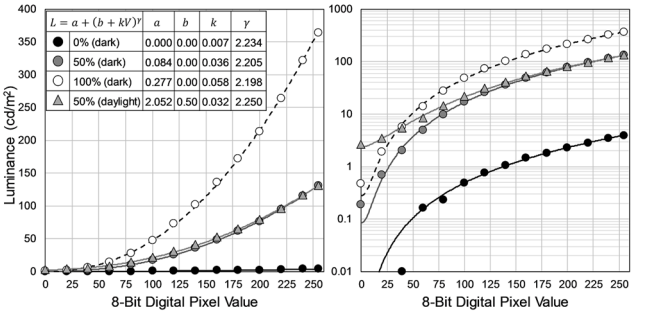
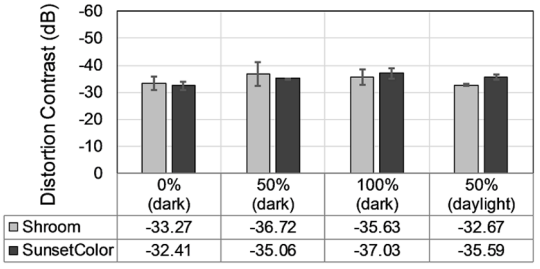


Figure 5: EOTFs of the iPad with different iOS "Brightness" settings and in darkened vs. daylight room settings on linear and logarithmic luminance scales.

Figure 5 shows the EOTFs of the iPad under these different settings. Observe that the iOS "Brightness" setting primarily affects the slope on a linear luminance scale (vertical offset on a logarithmic scale); this is captured in the fits by the parameter k. However, the "Brightness" setting also has a small effect on the minimum brightness (parameter a). Similarly, changing the room illumination from a darkened room to a room lit by daylight primarily raises the low end of the curve with negligible effects for

larger luminances; this is captured by changes to parameters and *a* and *b*.



	0% (dark)	50% (dark)	100% (dark)	50% (daylight)
Shroom	-33.27	-36.72	-35.63	-32.67
SunsetColor	-32.41	-35.06	-37.03	-35.59

Figure 6: Contrast detection thresholds on the iPad under different settings/room illuminations. Error bars denote ±1 standard deviation of the mean. Note that the vertical axis is reversed, and thus taller bars represent lower thresholds.

The resulting thresholds are shown in Figure 6. Lowering the iOS "Brightness" setting to 0% raised the thresholds for both images ($p = 0.031$). We suspect that this effect is due to noise masking (increased variance of the internal decision variable): The reduced range and contrast of the display made it difficult to see the both the distortions and image [9]. Although the reduced visibility of the image may very well have reduced the amount of contrast masking, for correlated distortions, oftentimes the mask (image) and target (distortions) are visually captured as a single percept, and therefore subjects often look for mangled content rather than for separate distortions [10, 11]. Thus, a 0% "Brightness" setting generally made everything harder to see; thus giving rise to greater noise (internal) masking.

Going from 50% to 100% "Brightness" did not have a significant effect on the thresholds. Again, we attribute this finding to the fact that subjects often look for mangled content rather than for separate distortions. At 100% "Brightness," the images are capable of greater contrast masking, but also greater visibility of mangled features (i.e., greater ability to see the effects of the target on the mask).

A bit surprisingly, when comparing the thresholds measured at 50% "Brightness" in a darkened vs. day lit room, the effect was not significant ($p = 0.12$). Additional trials and subjects are needed before any conclusions can be drawn.

Summary

A longstanding unknown in regards to quantifying visual losslessness in compressed images and videos is the applicability and reliability of such measurements, especially in regards to mobile displays. In this preliminary work, we have shown that contrast detection thresholds for HEVC distortions in 8-bit images are similar when measured (via a forced-choice procedure) on an iPad Air 2 as compared to when measured on desktop and laboratory displays. This finding assumes that the EOTFs are similar; however, for the limited stimuli used in this study, the thresholds are surprisingly robust to reasonable variations in the iOS "Brightness" setting and/or room illumination.

Business Perspectives on Visually Lossless and Lossy Quality

S. Daly

One of the key factors in favoring an accurate visually lossless descriptor as opposed to a wider ranging quality descriptor is the maturity of the technology used in the business. Businesses with mature technologies have products that are often extremely high quality, with no distortions noticeable in their product. However, they still do not want to waste effort or incur higher costs delivering a physical quality higher than visually noticeable. On the other hand, businesses with developing technologies have products where distortions are visible, but the customer accepts that due to other factors such as convenience, expectation level, cost, etc. In general, the developing businesses are continuously improving their technology and ramping up their quality, and need to keep track of quality improvements that are nevertheless still in the visually lossy realm. As mentioned in the background, the need for visually lossless assessment or wide ranging quality assessment will affect the distribution of stimuli, as well as the methodology, such as two alternative forced choice, paired comparisons, or comparative rating via scales. In addition to those methodology choices, the way the imagery is presented to the viewer for comparison is critical. For convenience of discussion, this section will use the term *video* to include digital video, digital cinema, as well as still imagery.

Different methods of video comparison

Three key video comparison methods are sequential comparison, simultaneous comparison, and oscillation. 'Simultaneous' is more generally referred to as side-by-side (SBS), and oscillation is more generally referred to as toggling (also as flicker). For completeness in encompassing all the methods of quality assessment, a fourth could be included, which is no comparison. That is, a single stimulus presentation (with no reference). Visually lossless in the truest sense cannot be done with single stimuli. Some distortions can indeed be assessed in a single stimulus presentation if their appearance looks entirely synthetic (e.g., blocking artifacts) or violate laws of physics (e.g., contains scene lighting incongruences due to image compositing [12]). These cases can be generalized to where the distortions' spatiotemporal statistics are inconsistent with the reference imagery statistics. However, many other distortions that are consistent with the reference imagery statistics cannot be assessed without a comparison. Examples of these include blur, contrast, color and texture. If someone's hat changes from cyan to green as a result of a tonescale compression algorithm, the viewer would not be able to detect that difference without a comparison image, since both colors are plausible to a third-party viewer. A better term than visually lossless for the indistinguishable distortions as assessed by single stimulus testing is *plausibly lossless*

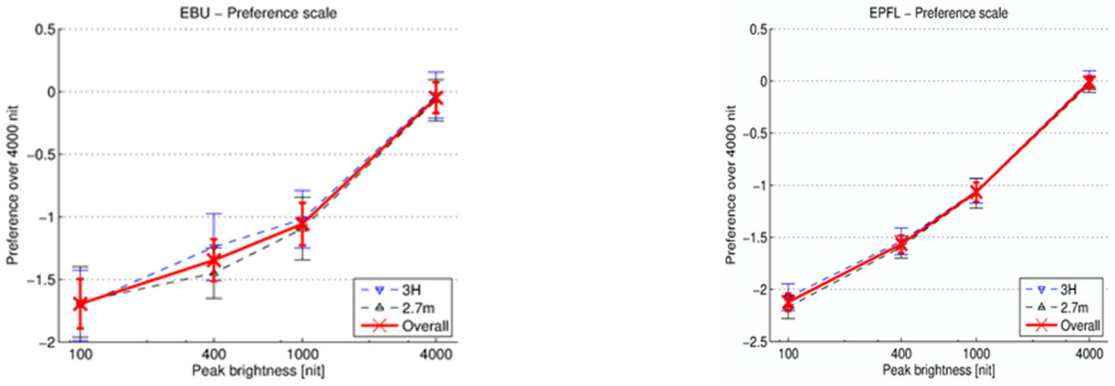


Figure 8: Comparison between sequential (L) vs. side by side (R) comparisons for the same stimuli, displays, and subjective task (preference). The sequential testing results were conducted by the EBU while the side-by-side were conducted by EPFL.

For the traditional test video clips of 10-15 sec duration, it is known that it is much easier to see differences when the video clips are shown side-by-side than when they are shown sequentially. A recent study verified this by directly comparing the two methods[13]. The experiment was identical for both cases, including display, stimuli, and task. The experiment tested one parameter of display capability: maximum luminance for HDR (high dynamic range). In the sequential testing, one Dolby professional reference display (Pulsar) was used. For the side-by-side testing, two Pulsar displays were used. The resolution of each was full HD (1920x1080), the diagonal was 42", the bit-depth was 12 bits RGB, the color gamut of the signal was 709, the black level remained constant at 0.005, and the ambient was 20 lux. A hidden upper anchor was used for each comparison. The viewer's task was to rate the quality (according to their own personal preference) of each of the two stimuli shown using a Likert scale. The maximum luminances tested were 100, 400, 1000, and 4000 cd/m^2 (nits). Six different HDR video clips were used, where two different max luminances were compared in each trial. The main point of the results (shown in Figure 8) is that sequential comparisons are more difficult that the side-by-side. This shows up both in terms of the confidence intervals and the shape of the curves. The confidence intervals are clearly seen to be on average 2x larger for the sequential comparison task, and the range of quality is reduced. For example, there is not a significant distinction between the 400 and 1000 nits versions in the sequential testing, while there is a clear distinction across all four tested stimuli parameters in the side-by-side methodology.

To better understand why the side-by-side comparisons gives more pronounced quality distinctions, it is worth noting that any image comparison requiring a viewer's response is a task, involving various stages of visual memory and mental mapping. Figure 7 shows some of the key processes for the rating comparisons as used in the mentioned experiment. Both of the

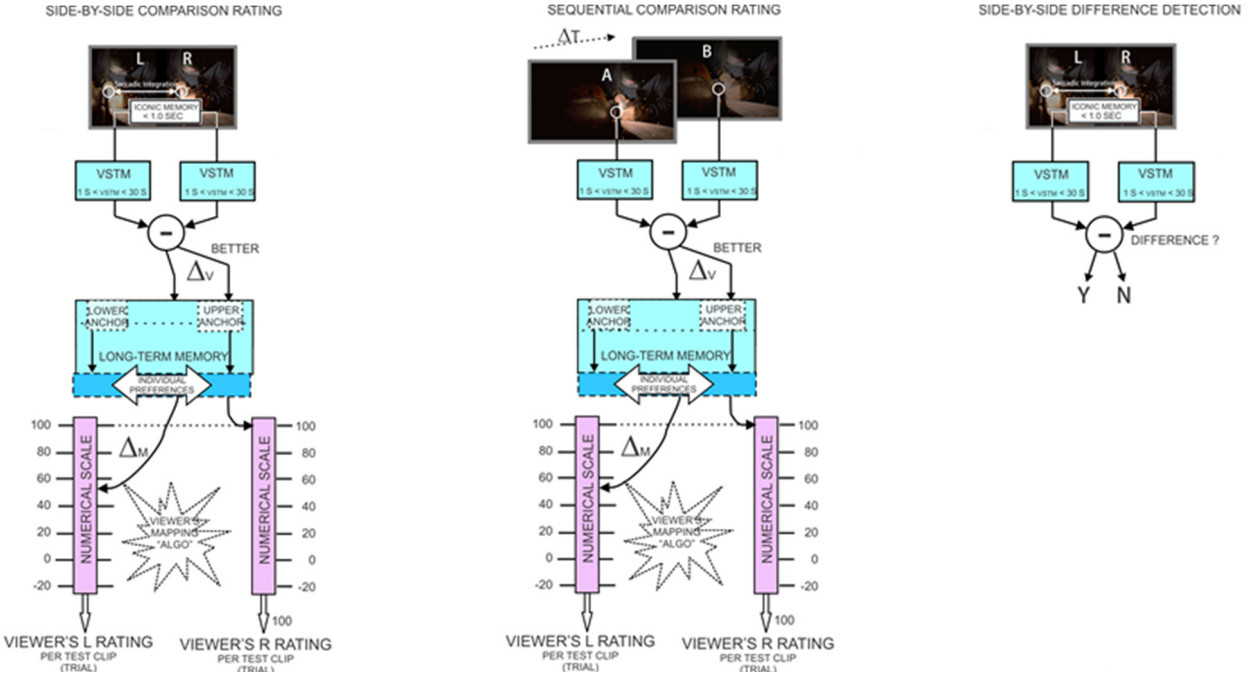


Figure 7: Key memory and mapping stages for an SBS rating task (left), a sequential rating task (middle) and an SBS visually lossy detection task (right). Note: sequential is referring to viewing one entire test video clip, followed by another one (the other half of a pair with differing parameters).

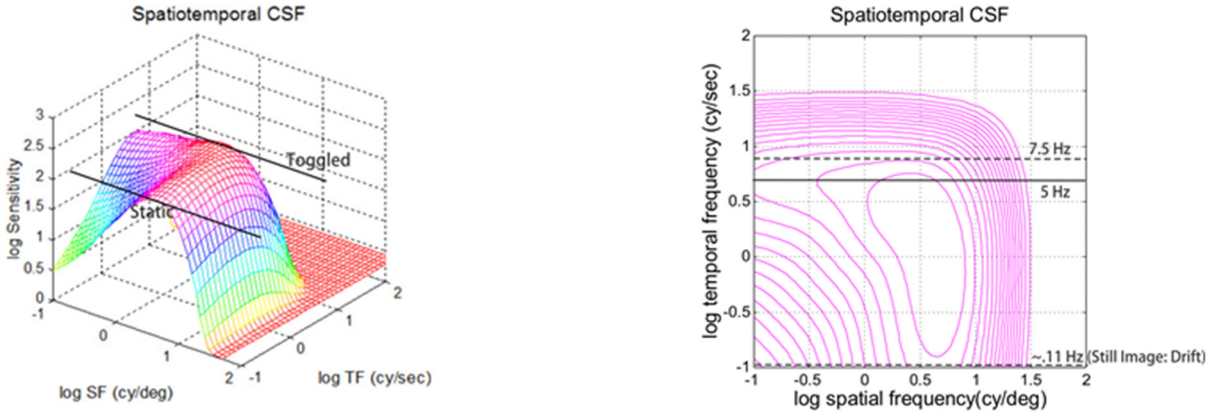


Figure 9: Spatiotemporal CSF (at ~light adaption level of 10 cd/m^2) showing the general effect in a surface plot (left) and more specific changes in sensitivity in the contour plot for the oscillation techniques (right). Contours deltas are 0.25 log10 in sensitivity. task (preference).

compared stimuli cannot be foveated at the same time[1], so in the side-by-side method (leftmost plot), saccadic eye movements are required to compare the Left and the Right stimuli. Iconic memory is the term for the portion of visual memory that integrates imagery across saccades and enables us to build up a mental picture of the world having a wider field of view (FOV) than the fovea's mere 4-6 degrees[14, 15]. In the side-by-side methodology, the iconic memory is used for an additional purpose than building up a mental image; it is also used for comparing similar image regions. Regardless of its end purpose, it is still limited to be less than 1 second. These visible differences are registered in the Visual Short-Term Memory (VSTM), and its duration limits come into play [16, 17]. These can be considered to hold the visible representations in the range from about 1 sec to 30 seconds. This upper limit suggests why video clips of duration less than 15 secs are preferred in the testing community. The visible differences, ΔV, are noted from those in the VSTM. To go from these visible differences to a subject's numerical rating, these visible differences must be mapped into that rating range. This requires memory of previous stimuli being shown, which would have occurred further back in time than the limits of the VSTM. In addition, if upper or lower anchors are not used (the experiment in Figure 8 had only an upper anchor), long term memory of video quality over perhaps years or decades may be involved. Further, individual preferences on which image features are more important (contrast vs. color vs. sharpness vs. texture, etc.) act as biases on the long-term memory. Lastly, from this internal range of magnitude of visible differences, visual quality must be mapped into a numerical scale. This involves higher level cognition than the previous steps, and is susceptible to even greater subject variability. To no surprise, the higher accuracy memory functions have the shorter durations. So in terms of accuracy, the iconic is best, followed by VSTM, and then long-term[2]. The case for sequential comparison is shown in the middle. The temporal delta would be greater than 10-15 sec for typical video quality testing. That methodology deprives the visual system of the iconic memory being able to input localized visible comparisons to the VSTM, because many foveations to different portions of the image would have occurred before the other paired stimuli is seen. That is the most likely source of the larger confidence intervals and range compression in Figure 8 for the sequential method.

Let us now consider the simpler task of assessing visual fidelity (i.e., whether something is visually lossless or lossy), as shown in Figure 7 for a side-by-side comparison (right). Since there is no rating required, a simple yes or no response can be given. Thus the task removes the inaccuracy and biases of long term memory, as well as individual variations in mapping their visual memory to a rating scale. Fortunately, for the businesses where visually lossless is the most relevant criteria, their use of experiments designed around a visually lossless criterion are able to obtain much more consistent and accurate data.

The third approach mentioned, toggling, reduces the internal processing and memory load of the viewer even further. Toggling has been used since digital imaging systems with frame buffers were available in the late-seventies. The term comes from a toggle switch, and the technique is still commonly used by image processing algorithm developers to look for differences in their resulting images. It is generally used for still images. It has also been used for video clips as well, but with less success. The two images to be compared are displayed in register (i.e., to the exact pixel position) on the screen, and the viewer toggles3 as desired between the two images. The change occurs in-place and with no interstimulus interval, or blanking field which might cause masking. Spatial and amplitude differences thus pick up an additional temporal modulation. Differences that would previously be below threshold using just side-by-side comparisons often become visible. This occurs for several reasons. One is that detecting visible differences in an image requires a search over the two compared images for differences. It can take a substantial amount of time to scan and foveate an entire image, particularly for detailed imagery that may be displayed with a FOV as large as 67 degrees (4k display viewed at the specified distance of 1.5 picture

[1] And thus a reason to use the term *side-by-side* over the term *simultaneous*

[2] Excluding rare eidetic individuals

[3] In current systems, left or right keyboard arrows are often used to swap (or toggle) images being displayed, as well as the spacebar. The toggle switch traditionally had two positions and allowed instantaneous swapping of its inputs, and these features are preserved with the newer methods, such as using a keyboard. Occasionally, toggling is referred to as *sequential* when still images are toggled, but the majority of work in this field doesn't use sequential to refer to the *rapid* alternation used in oscillation or toggling.

heights). The imposed temporal modulations caused by toggling enables better detection in the periphery (which while having poorer spatial resolution, has better temporal bandwidth and sensitivity than the fovea), aiding the viewer to find and then foveate regions formerly in the near or far periphery. So the toggling substantially aids the search task. In addition, the lack of needed eye movements for SBS comparisons (once a region having difference is found) aids in the detection of small spatial phase shifts that would be lost across a saccade. A third reason is that even in the fovea, the addition of temporal modulation at the right frequency can improve detection. Shown in Figure 9 is the spatiotemporal CSF (contrast sensitivity function). The temporal frequencies caused by toggling can shift the spatial frequencies of the distortion to a more sensitive part of the CSF as compared to what occurs with a static image comparison (shown in general on the left). While the highest spatial frequencies do not change that much between the two cases, there is a noticeable change at the frequencies near the peak, and a substantial change for spatial frequencies that are lower.

While toggling was originally an ad hoc technique, it has recently been made more rigorous [18] by removing the viewer's control and having the images automatically oscillate in place at a specific frequency. For the CSF at the light adaptation shown in Figure 9, it can be seen how an oscillation of 5 Hz maximizes the sensitivity to all visible spatial frequencies, as compared to a static, or still image comparison. Since the eye does not hold steady when foveating a region (there are always drift eye movements), the temporal frequencies for a static image comparison are not at zero Hz. An estimate of the temporal frequencies involved for static image viewing is shown as around 0.11 Hz in the diagram, although it is better to describe these drift eye movements in terms of velocity. The difference between the 5Hz and the 7.5 Hz as suggested in [18] are relatively minor and a change in CSF light adaptation level going upward in cd/m^2 would likely put the 7.5 Hz value on the CSF peak and ridge. A related approach for imposing motion on distortions to make them more salient has been used for studying amplitude quantization by phase shifting the quantization interval as a function of time [19]. These techniques result in the best ability of the visual system to see differences, and can also speed up the search time, but may not be relevant to the business application as will be described later.

Calibration to the display

Calibration is needed because while it is possible to determine the contrast required for detection of a given frequency component of a distortion, the contrast per code value depends on the luminance calibration (generally referred to as the display's EOTF, electro-optical transfer function). Increasing a display's contrast and using the same signal quantization results in an increase in the contrast per code value. If that increase is large enough, a previously subthreshold frequency will become visible. A recent example of this is that the increased dynamic range of HDR displays required an increase from the previously acceptable 8 bits/color to 10 bits for consumer usage and 12 bits for professionals. Similar phenomena also occur for the other image and perceptual dimensions listed above. Of the various visual behavior relating to thresholds, masking is the most impervious to lack of calibration, since once it rises above absolute threshold (i.e., no masking) it almost follows a linear SNR behavior. For systems where color, dynamic range, resolution, frame rate, etc. are approximately fixed, then prediction of masking can provide a strong visual foundation for quality prediction, such as shown by uncalibrated models [20-22]. However, most display ecosystems are moving away from that situation and are trending toward more variability along these key display capability dimensions. At present, current visual models that can be calibrated to calibrated displays [23-25] have been shown to perform better in cases where display capabilities are not fixed, such as HDR[13].

While there were many businesses unable to design for visually lossless quality, there were niche applications where it was indeed possible to quantify most of these parameters, or at least limit them to specific ranges. This particularly occurred in closed systems where the product included the display, the proprietary image format, and the encoding. Examples of these include some defense imaging systems (e.g., aerial image analysis), some medical systems, high-end graphic arts WYSIWYG systems, and cinematic post-production. In other applications, while there were some unknown calibration dimensions, visually lossless criteria could be used in the design by assuming standardized specs and ideal or worst-case parameters (such as 3 picture height viewing and a specified EOTF for HDTV[26]). For handling the unknowns of display reflectivity and ambient light, which have a strong interaction on the black level, techniques like the Pluge signal were developed.

Fortunately, the current trends are that the display is becoming more knowable and quantifiable, and thus enabling closer adherence to visually lossless goals. For one, the displays are much more stable than they have been in the past, especially TVs which had much thermal drift causing color and convergence errors in the CRT era. More importantly, there are standardized pathways for the display to communicate its capability to the delivery system. As an example, EDID metadata that is exchanged from a display to a graphics card (and advanced OTT delivery services) contains information about the display's primaries, its tonescale EOTF (electro-optical transfer function [27], of which gamma is a legacy example [28]), and its pixel resolution. More advanced metadata is now being used in a number of applications where these values are augmented by the minimum and maximum luminances, bit-depth, and other parameters of the content[29]. Further, dynamic metadata is being used to pass essential signal information to the display in order to aid tone-mapping and gamut mapping algorithms, motivated because the color volume of displays can now vary so substantially [30]. Ambient light sensors are becoming more advanced, having Vλ sensitivity to match the eye, and can be used for display's internal algorithms to tailor the signal to the resulting black level changes. Even the key weakness in spatial calibration, i.e., the viewing distance, has a pathway to be solved with presence detectors (motivated by energy conservation) and depth sensors (motivated by interactivity) which are making continual headway into display products. Finally, the burgeoning HMD displays for VR have the fortunate advantages that the viewing distances are exactly known (as designed for in the optics) and the ambient can be easily controlled (generally kept dark). Thus, the video content delivery system can tailor the signal sent to the display so that the advanced visual model approaches aiming for visually lossless quality, which require such calibration, can finally be used to their theoretical intention.

Business considerations

A famous sign in many service businesses is "Cheap, Fast, Good – Pick Any Two". It is likely obvious to any reader that increasing quality comes with a cost. In display hardware, there is

a general struggle against physics to increase quality, offered initially at a higher cost, and then gradually the manufacturing efficiencies and scale of production can bring the costs down. Similar constraints are involved in the compression and video chip business. Rarely does one see a quality improvement and a cost reduction being introduced at the same time. For those wanting both, they must wait and essentially be late adopters. In this section, we will start with an anecdotal example so that concrete details can be discussed, and then we will describe some general issues.

The plot in Figure 8 was from an experiment [13] to provide data on whether the TV industry should develop a new ecosystem for High Dynamic range (HDR). There are a number of key attributes involved in HDR, including bit-depth, black level, local contrast, mid-tone contrast, compression technique, average luminance level, and maximum luminance. While high dynamic range includes increasing the range at the dark end as well as the bright end, one of the unique attributes of high dynamic range is more accurate rendering of highlights than traditional video. Such highlights include both specular reflections as well as emissive objects (visible light sources) and can require very high maximum luminance [31]. A study was designed to specifically probe this aspect in comparison to existing TV standards, known as standard dynamic range (SDR), and standardized in ITU-R Rec BT.709[32], with an EOTF subsequently defined in ITU-R Rec BT.1886 [26]. Most viewers watching SDR see only 8bits/color video that is compressed. One aspect of HDR is that it requires a higher bit-depth than SDR, and details of whether 10 or 12 bits/color are needed depend on viewing conditions. Currently in television systems, HDR is generally bundled with an increase in spatial resolution and color gamut as well, for example to going from the BT.709 (sRGB) color gamut to the DCI P3 gamut or even wider with the ITU-R Rec. BT.2020 gamut [33]. But in order to focus solely on the parameter of maximum luminance, the study used uncompressed videos at 12 bits, all with a BT.709 color gamut and an HDTV pixel dimensions (1920x1080). Four maximum luminance values were studied. They were placed approximately on a logarithmic luminance scale based on general visual system properties. The four luminances were 100, 400, 1000, and 4000 cd/m^2. Deviations from strict logarithmic spacing were motivated by practical existing television systems and displays.

The existing SDR TV system was designed for ~ 100 cd/m^2 as the maximum[4] and in calibrated studios, the reference monitors are set very close to this value. This is true for both episodic and live broadcast video content, and is the maximum luminance that is seen by individuals involved in the approval process (cinematographer, colorist, director, producer for episodic content, and the video shader and producer for live content) before distribution occurs. The ambient lighting followed the industry production specs of producing a surround of 5 cd/m^2 to 10 cd/m^2. The next value, 400 cd/m^2, was selected as a typical higher-end consumer TV max luminance at the time of the study. As a reminder, the content seen at 100 cd/m^2 by the approvers is generally stretched upwards in most TVs. The value of 1000 cd/m^2 was selected to represent the capability of the first generation of consumer HDR TVs. Lastly, the 4000 cd/m^2 value was selected because that was the maximum luminance capability of the professional HDR displays used in the experiment.

Initial attempts at using the BT500 5-point rating scale (excellent, good, fair, poor, bad) in pilot studies were inconclusive because a majority of viewers rated the lowest capability value (100 cd/m^2) as excellent, and there was no headroom on the scale to indicate higher quality than that. This was partially due to their inexperience seeing uncompressed 12-bit video, as well as a reference display (such as having lower noise, better uniformity, etc.). As a result of lack of useful guidance from the BT500 document, it was decided the experiment needed to explore testing options as well as the maximum luminance parameter. Two key comparison methodologies were agreed upon, a sequential and a side-by-side comparison. Video clips of 10-15 seconds were used based on common video testing, so the sequential method meant that one version of a video clip was shown, followed by a version with a different max luminance, all being shown on a single HDR reference display, and then followed by the viewer's rating. For the side-by-side testing, two identical displays were arranged so that viewers could compare both at the same time and arranged so each was seen with an approx. orthogonal viewing angle to the display screen. This approach has traditionally been avoided for rigorous studies in the past due to difficulties in getting two displays to have the same color, tonescale, and black level. However, modern digitally driven reference displays with internal light sensors, thermal regulation, and compensatory image processing can enable such displays to appear identical. Randomization of the various content with known parameters was used in case there might be a small physical bias, despite being physically immeasurable. After presentation of the video test pair, the viewer was asked for a preference rating comparison. For the side-by-side testing, the Likert scale shown below was used. For the sequential testing, it was modified to replace L and R with A and B, where A was explained to be the first instance of the sequentially shown pair, see Figure 10.

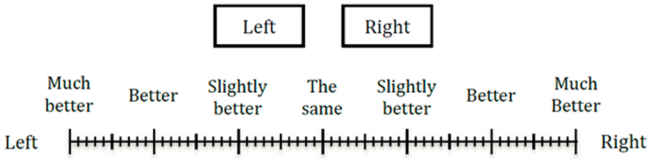


Figure 10: Spatiotemporal CSF (at ~light adaption level of 10 cd/m2) showing the general effect in a surface plot (left) and more specific changes in sensitivity in the contour plot for the oscillation techniques (right). Contours deltas are 0.25 log10 in sensitivity.

The results have been discussed earlier in this section with the side-by-side having better confidence intervals than the sequential, as well as having a larger range of preference. What visual memory and cognitive processes are involved in each methodology have also been discussed. Let us now discuss some key business aspects. For a new television ecosystem, both the televisions and the video signals need to be updated. These involve two key different industries: the television set manufacturers and the broadcasters. For television makers' customers, the majority of TV sales involve side-by-side viewing of competing TV products arranged in a store at the time of the purchasing decision. Some customers may be influenced by written ratings, descriptions, and recommendations in either mainstream or more technical press, but most of the time, a side-by-side viewing is involved. The

[4] In many systems, Reference White, which is generally the diffuse white maximum luminance is set to 100 cd/m^2 and the peak luminance (the maximum luminances) is set to 120 cd/m^2.

broadcasters have a different situation since it is generally not possible for their consumers to view their service compared to a competitor's (e.g., a different network) in a side-by-side manner. Rather, comparison is made by the consumer in a sequential manner by changing the channel.

The plot shows the viewers using sequential comparisons were not able to show preference differences for the 400 and 1000 cd/m^2 parameters confidently. This is very important for business considerations in 2015-2017, as HDR TVs are being introduced. SDR TVs are typically 300-500 nits, and the 1st generation of HDR TVs is typically around 1000 nits. The sequential testing doesn't give any confidence to the preference of the new 1000 nits HDR TVs over the current SDR TVs, while the side-by-side testing does give substantial confidence. The sequential results directly relevant to the broadcasting business would not be able to indicate with confidence that a change to a 1000 cd/m^2 system would be worthwhile, whereas the side-by-side results that are directly relevant to the TV set makers does conclude with confidence that change would be preferred to the viewer. However, both businesses involved in the ecosystem need the other business to agree to a similar upgrade. Assuming the trend of increasing maximum luminance can continue, and ranges closer to 4000 nits will eventually be reached, a future-oriented decision might be for both business segments to agree to move forward with HDR. Another way to look at the results, however, is that the side-by-side gets closer to the true perceptual experience of the viewer, whether or not they can see the comparisons directly. Of course, a critical customer of many broadcasters is the advertising industry, and their professional viewers would likely be able to see side-by-side comparisons in a production suite. As a result of these many factors, the broadcasting industry in several key regions decided to go ahead with HDR transmission. It is not clear if it was the future capability considerations or the benevolence to the viewer that was the dominating factor.

General business considerations regarding visually lossless or lossy quality are specific to the business. For example, visually lossless criteria are relevant for mature businesses already delivering a high quality, with examples being those that have a six-sigma defect strategy. Visually lossless is also relevant for businesses with high-end products and high cost ranges. Examples in printing and video include most of the production workflow. An example of visually lossless compression includes what is known as mezzanine compression, having low compression ratios below 2-3:1 and yet still use advanced techniques like wavelet or DCT. Businesses where visually lossy quality ratings are more relevant include newly developing businesses, developing products offering new features and conveniences, and businesses specializing in lower-cost products. For example, new businesses arising to compete with mature businesses usually begin with a lower quality and increase it as they expand their market. Streaming is a good example of a service business that initially had very low quality (circa 2006) whose quality weaknesses included not only the customers' bandwidth but also color and tone miscalibration. Now however, there are streaming services of the highest quality, with 4k resolution at 10 bits/color and visually lossless performance for three pictures heights.

For the businesses where visually lossless is the most relevant, each of the three comparison methods are suited toward different applications. Toggling (in particular the automatic alternation techniques known as flicker), is most suited to imaging applications that are information-task based, where small features and minute phase shifts may be important, and the localization shortcut aspect of toggling can be a surrogate for a strenuous search process. Particularly when it is unknown which elements of the imagery are most critical to the task. Examples include products and services for forensic, histology, aerial imaging, scientific visualization, medical, etc. A special case is for products within the video path where the customer is a technical person that uses such a toggling technique for assessment, even if the end customer of the entire video path is a non-expert consumer. Applications where results from side-by-side testing are most relevant include products that are generally purchased in stores, and competitor products are available. Televisions fall in this category, as well as mobile displays to a lesser degree. Lastly, applications where sequential testing methodology is most suited include most consumer services, such as broadcast, cable, and internet delivery (OTT) of video. However, particular companies aiming for the highest levels of quality may decide on one of the other methods if their philosophy is to deliver the best quality to their customer (even if the customer doesn't notice it; see physiological testing discussion in the background for such motivations).

The issues of viewer variability (both from viewer-to-viewer as well as individual consistency) are discussed in more detail in other sections of this paper. One business aspect related to viewer variability is the customer market segment. As mentioned, vision science analyzes the JND and psychometric function per subject, but industry must extend the JND concept to the overall population. In industry, one JND corresponds to 50% group detectability and is thus considered not advertisable, and it has been suggested that 2 JNDs is the minimum for advertising quality differences [3]. However, this ignores market segmentation and specialty products where the market being sought after is already known to be a subset of the overall population. For example, tuning a visually lossless product to the average viewer's sensitivity essentially means that half of the populations will see the distortion, and half will not. That is a significant loss of potential market, so some engineering criteria include analysis of cumulative distribution functions and aim for a more demanding viewer/customer than the average observer. For example, a design might be tuned so that all but the upper 10% most critical sensitive viewers would experience visually lossless quality.

Subjective assessment and the criteria for visually lossless compression

L. M. Wilcox, R. S. Allison and J. Goel

Objective metrics of image quality have the advantage of repeatability and are suitable for automatic assessment and monitoring of image quality. Such metrics are in high demand, given the increasing requirements for real-time image compression needed to deal with the bandwidth requirements of high-resolution image transmission [21, 34]. However, it is clear that while subjective testing is labor intensive and costly, it remains the only reliable means of evaluating the impact of image compression and the visibility of artefacts. As described in preceding sections (Background, Zhang et al and Daly), a wide range of qualitative methodologies are available: both threshold and suprathreshold methods have been widely employed to assess image quality (and the success of compression algorithms). Forced-choice threshold methods are often used to establish if a compression algorithm is

visually lossless as they are sensitive measures of the visibility, that are less impacted by bias and amenable to statistical analysis.

In 2015, ISO/IEC published evaluation protocols based on forced-choice procedures that could be used to evaluate images across display platforms. Their protocol describes two variants: one normative (Annex A) and the other (Annex B) based on a flicker paradigm proposed by Hoffman and Stolitzka [18]. In the normative approach the original image is presented as a reference and, in another part of the display, the observer is presented both the original image and the processed image side-by-side (randomly ordered) and required to choose which of these pair matches the reference image. This is a classical forced-choice procedure intended to measure sensitivity to artefacts in the processed image.

However, there are several issues that suggest that other procedures might be more appropriate. First, while there may be salient artefacts in the image observers may not know where to look. Until attention is brought upon these areas, the literature suggests that even large changes in the image are often not seen [35, 36]. This issue is partly addressed in the ISO/IEC protocol because instead of using large full screen images, the stimuli are crops of a pre-defined size that can be selected to include potentially problematic sub-regions. Techniques to highlight the changes could make for more sensitive and efficient detection (see Section by Colett et al below). Also important use cases involve dynamic detection of visibility. For instance, in video or interactive content frame-to-frame differences in quality might be noticeable. The second variant described in ISO/IEC 29170-2:2015 (Annex B) uses direct temporal comparison in a flicker/toggle paradigm [37] (see also Section by Daly above). In this procedure, two image sequences are shown side-by-side as illustrated in Figure 11 On one side of the display the original image is shown alternating with the processed image (at a rate of 5Hz), while on the other side the original image alternates with itself (i.e. does not change).

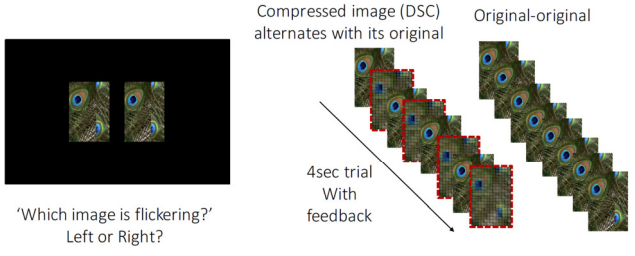


Figure 11: Illustration of the ISO/IEC 29170-2:2015 flicker protocol (Annex B). To the left is the observer's view of the stimuli, illustration of the image alternation (5Hz). The reference location is randomized on each trial, viewers have 4 sec. to view the image sequences and are given feedback.

In this procedure the reference is provided sequentially in the same location as the processed image and thus the image differences should be extremely salient due to sensitive motion and change detectors in the visual system. The flicker paradigm is also relevant to cases where transient image artefacts may occur such as video and interactive content.

In a large scale trial (N=120) conducted at York University we implemented the ISO/IEC 29170-2:2015 flicker protocol [37] to assess the qualitative effectiveness of the Video Electronics Standards Association (VESA)[5] display stream compression standard (DSC1.2) using a wide range of image content, including known challenges to the algorithm (see Figure 12 below).

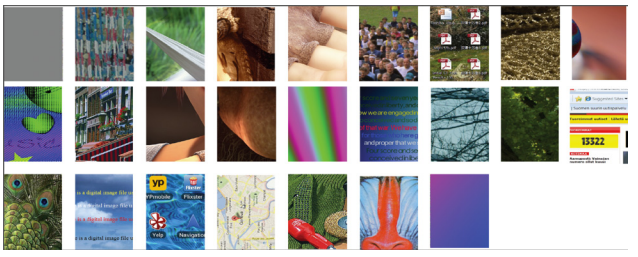


Figure 12: Thumbnails of images used in the assessment of VESA DSC1.2. A wide range of compression parameters were applied to each image (chroma subsampling, lines per slice, bits per channel).

As specified by the ISO/IEC protocol, in addition to the test conditions of interest, a number of obviously degraded control conditions were evaluated. These 'control' conditions provided encouragement to participants and who otherwise were performing at threshold most of the time. In addition, performance on these trials was used to exclude observers who were not paying attention; observer data was only included if an individual scored >= 95% on control trials. Each condition was tested multiple times to arrive at a detection probability for each stimulus condition. For each observer, and each condition, the proportion correct detection was calculated. The ISO/IEC standard recommends reporting of summary graphs in the format shown in Figure 13. Here the mean proportion correct is plotted across observers with ± 1 standard deviation and symbols indicating the best and worst performing observers (downwards and upwards oriented triangles respectively).

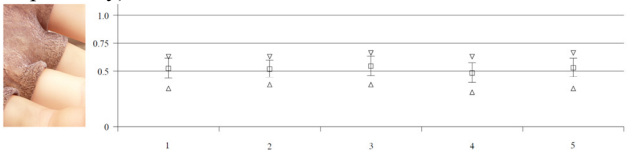


Figure 13: The graph shows the proportion correct for a given image (thumbnail to left) under different compression conditions (coded as numbers 1-5). Open squares represent the group average, error bars indicate ± 1 standard deviation and downward and upwards triangle symbols indicate the best and worst performance. Each data set represents a different level of compression.

As discussed above, the criterion used to define visually lossless is critical and is under debate. Given that the ISO/IEC 29170-2:2015 standard is based upon detection; a detection threshold approach is used. Following psychophysical convention, a 75% correct criterion (midway between guessing and perfect responses in the two-alternative task) is recommended. Specifically, the standard proposes that lossless performance occurs when no observer detects the compressed reference on greater than 75% of the trials (although the standard allows for modification of the criterion). In our study, the large majority of test conditions met these strict criteria for visually lossless (see Figure 13 above). However, in some instances this criterion may result mis-categorization of reference conditions as lossy. The results shown in Figure 14 illustrate this phenomenon. In this graph, under all compression conditions average observer performance is clearly at chance (50% for this two alternative

[5] the Video Electronics Standards Association, is an active industry trade group in the video display industry. (www.vesa.org)

task); however in condition 3 one observer detected the flicker on more than 75% of the trials.

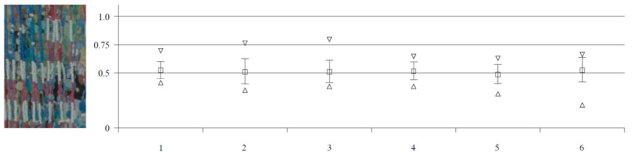


Figure 14: As in Figure 3, results are shown for here for 6 compression conditions for the image shown in the inset. Open squares represent the group average, error bars indicate ± 1 standard deviation and downward and upwards triangle symbols indicate the best and worst performance.

It is clear that this criterion places considerable emphasis on potential outliers in the data set. In their original implementation of the flicker protocol Hoffman and Stolitzka identified and selectively tested a set of 19 (out of 35) highly sensitive observers in their data set [18]. They suggest that given the potential impact of such observers that the criterion for lossless could be increased to 93%, but just for these sensitive individuals. However, this approach introduces a potential bias to the test protocol: it is left to the experimenter to define the sensitive observers who will be held to a different standard. Another approach would be to consider the results of all observers, but to adopt a visually lossless criterion based on their average performance and the associated standard deviation (for example using the estimated 95th percentile rather than the sample maximum). Statistical techniques based on the variance could be used to identify highly sensitive observers or outliers, and, if appropriate for the use-case, remove them from the data set [38].

Another factor that contributes to the sensitivity of the ISO/IEC protocol is the extent to which practice on a limited image set can, over time, contribute to the creation of highly trained observers. Such observers could learn to attend to specific image regions, and as a result be able to better detect artefact-related flicker. As noted in previous sections, side-by-side presentation makes it more likely that observers will make direct comparisons between reference and original image regions, improving detection rates. Furthermore, in the paradigm implemented here, cropped rather than full screen, image regions are viewed. At the recommended viewing distance each image is within highly sensitive foveal vision, which further enhances the probability of detecting the flicker created by alternating the reference and original images. These factors, combined with the sensitivity of the human visual system to spatio-temporal variation within this range, will draw attention to compression-related distortions. Over time and trials viewers will be attuned to specific regions and artefacts that would otherwise be undetectable. These training effects could be reduced by using a large pool of observers on a limited set of conditions, but reduction of test content will in turn reduce the generalizability of the evaluation.

The results of our evaluation of VESA DSC1.2 using the ISO/IEC 29170-2015 flicker protocol show that this forced-choice paradigm is a highly effective means of evaluating sensitivity to image differences [37]. The design of this test protocol and the visually lossless criterion applied is extremely sensitive, and in particular emphasizes the most sensitive viewers. It is arguable that this protocol is too sensitive, and that the results of the side-by-side flicker task highlight artefacts that would not ever be visible under 'normal' viewing conditions. As outlined elsewhere in this paper, there are other candidate approaches to the assessment of image quality. We argue that the appropriate methodology depends on the objectives and the use case. For example, if the goal is to conservatively evaluate the possibility that a compression artefact might be visible under any situation, then the flicker paradigm is a viable approach as it highlights differences between images regardless of whether they are noticeable in the absence of a reference.

Usage perspectives on visually lossless and lossy quality and Assessment

H. Colett, J. Knopf, P. Corriveau and S.-N. Yang,

The display industry today struggles to distinguish between visually lossy and lossless encoding, as visual experiences are determined by various factors such as usages, form factor and content. The debate becomes even further entangled when one tries to define and quantify visually lossless based on empirical measurements, when the chosen measurement protocol and stimuli can critically impact the outcomes.

As outlined in the previous section Wilcox et al, VESA has presently adopted a testing protocol and procedure for evaluating visually lossless encoding. From a practical perspective, this is only one particular way to investigate whether a process or algorithm can potentially produce a result that is truly visually lossless. Such a specific approach was reasonable when research first started on this topic, as the problem space in which we needed to investigate was clear, and the usages and context for the definition could be clearly carved out. However, the increasingly complex ecosystem in electronic displays warrants a re-investigation and re-definition.

Innovations in testing methods, and their underlying theories, open the possibility of using different types of techniques, such as gaze tracking, to potentially augment and improve existing methodologies by emphasizing the visual functions in specific usages.

The proposed research emphasizes the need of refining the definitions of visually lossless and delineating the testing procedure based on the usages and viewing context.

Classical visually lossless (VL) compression of display content is defined as the loss of image quality induced by the compression algorithm that can't be perceived by a user [39]. This definition by its nature is very vague, as the quality of perceived image can be affected by individual visual characteristics, viewing conditions and display apparatus. Thus, there is no unified position on what the definition of visually lossless truly is. Conversely, in the display industry it has been deemed that each company, organization or institution can set thresholds based on their products and desired experiential delivery. What this means at its core is that the statistical requirements for user studies can be set at different levels and different means of detection. To develop a commonly accepted definition of visually lossless, it becomes necessary to provide a unified principle of assessing these outcomes that takes into considerations of the above factors.

In 2014, VESA published a new standard that uses visually lossless image compression to increase the rate of data transmission carried by a display interface data rate, thus saving power while maintaining viewer's experiences[37]. As part of the new standard, visually lossless with regard to visual psychophysics requires that any content distortion caused by compression be below the threshold of conscious detection. Therefore, visually

lossless can be defined as any loss of data due to file compression that is not detectable to a 'typical' user on a 'typical' display under 'typical' viewing conditions. This has been implemented as a side-by-side comparison of control (original to original) and target (original to compressed) images alternating (flickering) at 2.5 Hz. A user. is asked to discern the target image from the control image within 30 seconds of viewing. Outcomes of such method can be affected by factors such as viewing distance, screen luminance, pixel density and viewer's visual acuity, etc.

Not surprisingly, the use of the ambiguous term 'typical' has been proven problematic when it comes to verification testing. The definition and assessing methods of visually lossless compression has been guided by the ecology of display industry. When the first discussions of visually lossless compression began, the ecosystem of media was relatively simple as the usage did not vary and the form factor uniform. With the pervasiveness of media content in the world today, the fundamental definition of visually lossless or "lossy" needs to be examined at a contextual level. When one moves from consumption of content on a phone to a large format TV or to new AR/VR environments, the nature of the changed viewing environments dictates that the definition of visually lossless will need to be expanded from passive perception of visual environment to active visuomotor interaction with it.

With the adoption and deployment of virtual reality devices, the issue becomes even more complex due to the artificial binocular delivery of the stimuli to the visual system. With it vision is not just the result of stimulus-derived representations, but also that of interaction between visuomotor processes. Many feel that this will cause the visual experiences to be different in different usages and the definition of lossless vision to be bifurcated between monocular and binocular devices. These would suggest a need of different methodologies for testing visually lossless compression according to involved visual imagery as well as underlying visuomotor functions.

In the previous section it is mentioned that the definition of visually lossless has been left to the manufacturer of the devices, which opens an interesting competitive angle in determining who has the best performance and who will or will not claim performance based on any given standard implementation. The authors feel that as the definition of visual experiences has evolved, there is a large grey area that needs to be explored around performance in order to have contextually-correct definitions. There is still room for debate around what could be standardized versus not and whether there are more generalized testing methods based on not specific devices but utilized human visuomotor processes.

Copious research has demonstrated that human vision is not a replica of the visual world, but an outcome of interactive visuomotor processes [40]. Scientists have commonly identified two types of human vision: featural and spatial [41, 42]. Features such as shape, color and complex object categories are encoded in the ventral aspect of human cerebral brain [43], whereas different spatial representations such as retina- and body-centric ones are encoded in the dorsal aspect. Featural perception dominates the conscious vision and is generated as much by bottom-up visual stimuli as by top-down insertion and creation of visual imagery. The spatial information is utilized by the brain for forming perception and guiding complex actions, but is not directly accessible to the perception [44, 45]. When viewers process displayed visual content, much information is utilized by the brain but not consciously identified. Visual attention serves to combine the two types of vision and makes some aspect of it available to visual consciousness [44, 46]. Hence, functionally lossless vision should be defined as unimpeded visuomotor processes in maintaining such interactive representations of visual world. The purpose of the visual tasks, the predominant context (static or in motion) and the level of focused attention determine the threshold of lossy vision.

We suggest it is useful but insufficient to simply compare compressed images to uncompressed ones for determining lossless vision. Such outcomes need to be obtained in a realistic task consistent with the form factor and usage, in which the task goals determine what should be attended and at which conscious level. Visually lossy should then be defined as either the detection of visual degradation or impeded execution of visual tasks.

To discern lossy vision, we propose to utilize a well-recognized method of gaze-contingent image degradation (GCID). GCID is achieved by switching between an original image and degraded image during eye fixations or saccadic eye movements. The perception of stable visual world is the result of visuomotor integration across eye fixations, where relevant visual percepts are maintained and unrelated features discarded. Instead of comparing side-by-side flickering images, in GCID the original image was switched to the compressed ones during eye fixation or saccades. The visuomotor performance and eye behaviors during selected eye fixations with control (original to original) and target (original to compressed) images are compared. This allows the effect of difference in visual imagery to be separated from artificial stimulation such as by persistent image flickering. Since the switched image is always processed with foveal vision, the effect of visual attention is assured and maximum visual acuity is utilized regularly by the viewer. Furthermore, as the eye movements are guided by internal processes to search for and utilize necessary visual information, impeded vision is readily identified with a change in eye movement pattern [18, 37]. Lossy vision is present when the visuomotor process is slowed or altered, as determined by increased eye fixation duration, reduced saccade length and higher frequency refixation when the compressed image was present during eye fixation.

Our preliminary data has demonstrated the effectiveness of such a paradigm. The data was obtained with a single switch between original and degraded (blurred) image at 100ms after the onset of eye fixation, Figure 15 shows that the location of degraded image (subtle blur) was detected at a level beyond chance (93%). In addition, the fixation duration (or latency of saccadic eye movements) near the blurred areas were significantly increased (250 to 269 ms) and saccade amplitude decreased (3.05 to 2.71 degrees). Figure 16 shows that saccade initiation probability was reduced when a blur was present, suggesting impeded visuomotor processing. Such eye movements and eye behaviors are tightly linked to the task on hand and the stimulus being process. Therefore, this method can be useful to assess lossless images by measuring the proper baseline responses and change in them caused by lossy images. Such a paradigm can be utilized to evaluate lossy vision involving different ocular demands (e.g., performing a visual task at a close distance) or methods of image rendering (e.g., VR/AR displays).

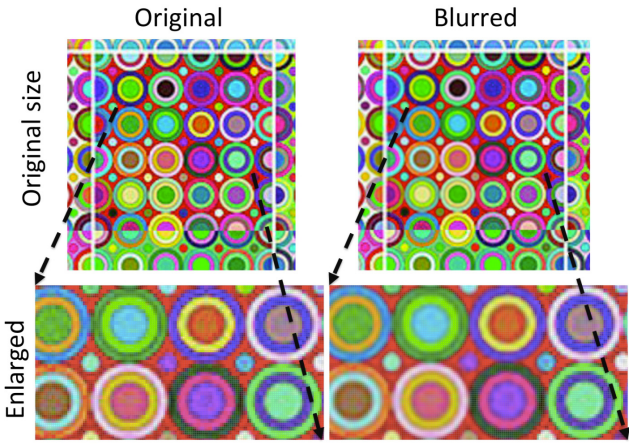


Figure 15: Example original (left) and degraded images (right, subtle blur). Human subjects were asked to survey the image and reported the area of blur image. The degraded image was present for alternate fixation by switching the images during fixation at 100ms after fixation onset. Human subjects were also to detect the blur despite of the subtle change (93%).

To conclude, the authors propose that there is a need for several methodologies that complement each other in evaluating visually lossless compressions. The expansion of usages and the fact that the same compression algorithms will be deployed across products and platforms lead to the involvement of different visuomotor processes. A method for utilizing eye tracking and contextual response feedback from users is proposed to assess both conscious and unconscious vision. Further results will be published once the implementation is complete and testing is done in comparison to the standard VESA testing.

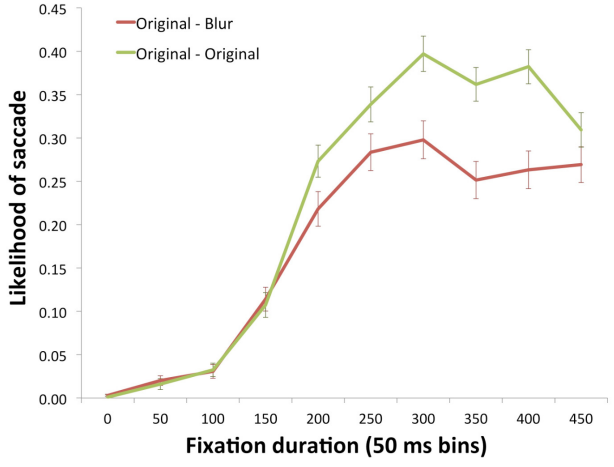


Figure 16: Saccade probability calculated from the change in fixation duration using hazard function analysis. For fixations with degraded (blurred) images, a large proportion of saccades were delayed. The detection of blurred image (subtle blur, square symbols) has a latency of 200 msec. (i.e., 100 msec. after image switch), when the saccade probability began to be reduced compared to without image degradation (same image, triangle symbols).

Panel discussion

The special session also included a panel discussion moderated by Prof Kjell Brunnström, Acreo Swedish ICT AB, Kista, Sweden. The panel consisted of

- Associate Prof. Damon Chandler, Shizuoka University, Hamamatsu, Shizuoka, Japan
- Scott Daly, Dolby Laboratories, , USA
- Adjunct Prof. Phil Corriveau, Intel Corp, Santa Clara, CA, USA & Pacific University
- Prof. Edward Delp, Purdue University, West Lafayette, Indiana, USA
- James Goel, Qualcomm, Display Video Processing Group, Markham, Canada
- Prof Laurie Wilcox, York University, Toronto, Canada

The discussion was mainly focused around the standardized flicker protocol by ISO/IEC and VESA[37]. A strong opposing view was given by Dr Andrew Watson, saying that the flicker paradigm is very sensitive, measuring something that is may be irrelevant when measuring visibility of artifacts that never ever occur in still images, because still images do not flicker.

It was explained, mainly by James Goel, that one industrial use of visually lossless compression is in the final stage in the mobile devices where just before sending the pixel information to the display internally compression is performed and at the display the pixel data is decompressed. Since this compression stage needs to be very conservative not to introduce additional artifacts, the industry has adopted this overly strict method. He also asked whether this would be more relevant when looking at applications such as image scrolling.

Scott Daly pointed out that there is engineering usage for the paradigm for instance for debugging purposes when there is a need to find areas where algorithms give different results, see also Daly's Section above.

The audience and panel members also discussed the definition of the term visually lossless. It was suggested that structural lossless may be a better definition, since a small difference in a complex texture may not be possible to notice. A countering view was that such small differences can be important for certain applications (such as relating to information-task, like medical imaging), and that one must be careful in generalizing across different businesses. Quantifying individual variations and their effect on visually lossless criteria is important and may also be dependent on the specific business. Another suggestion was that we really need to talk about visually experience lossless.

Bernice Rogowitz was surprised that the problem was solved with a subjective paradigm rather than using objective metrics. It was pointed by Scott Daly that perhaps HDR-VDP2.2 and HDR-VQM [24, 25], could do it, but they are still to computationally expensive to use.

Conclusions

The main conclusion from this special session is that visually lossless is not yet clearly defined. It can be very different depending on what the usage would be for the imagery. For example, it is crucial for in medicine that all relevant details are there to make a correct diagnosis, but on the other hand for video consumption on a smart phone will tolerate quite substantial differences.

Although the research area has had a long tradition, the current interest from the mobile industry has put renewed focus on the topic. It may be that rather than converging on a single definition of visually lossless, industry will instead adopt criteria and methodologies that best suit their use cases and objectives.

Acknowledgements

Acreo's work was funded by Vinnova (Sweden's Innovation Agency), which is hereby gratefully acknowledged.

References

[1]. ITU-R. (2012). *Methodology for the subjective assessment of the quality of television pictures* (ITU-R Rec. BT.500-13). International Telecommunication Union, Radiocommunication Sector.

[2]. ITU-T. (1999). *Subjective video quality assessment methods for multimedia applications* (ITU-T Rec. P.910). International Telecommunication Union, Telecommunication standardization sector.

[3]. Thiel, R., P. Clark, R.B. Wheeler, P.W. Jones, M. Riveccie, and J.-F. Dupont, *Assessment of Image Quality in Digital Cinema Using the Motion Quality Ruler Method.* SMPTE Motion Imaging Journal, 2015. **116**(2-3): p. 61-73.

[4]. Hillis, J.M. and D.H. Brainard, *Do common mechanisms of adaptation mediate color discrimination and appearance? Contrast adaptation* Journal of the OSA A, 2007. **24**(8): p. 2122-2133.

[5]. Harris, A. *The Hidden Side of Silence of the Lambs' Most Famous Scene*. Slate's Culture Blog 2014 [cited 16/02/2017; http://www.slate.com/blogs/browbeat/2014/10/15/silence_of_the_lambs_video_essay_from_tony_zhou_and_every_frame_a_painting.html.

[6]. Darcy, D.P., E. Gitterman, A. Brandmeyer, S. Daly, and P. Crum. *Physiological capture of augmented viewing states: objective measures of high-dynamic-range and wide-color-gamut viewing experiences*. in *IS&T International Symposium on Human Vision and Electronic Imaging, February 14-18, 2016*. 2016. San Francisco, CA, USA: Society for Imaging Science and Technology. p. HVEI126.

[7]. Alam, M.M., K.P. Vilankar, D.J. Field, and D.M. Chandler, *Local masking in natural images: A database and analysis.* Journal of Vision, 2014. **14**(8): p. 22-22.

[8]. *ITU-T Rec. H.265 - High efficiency video coding*. 2016.

[9]. Watson, A.B., R. Borthwick, and M. Taylor. *Image quality and entropy masking*. in *SPIE Electronic Imaging*. 1997.

[10]. Nachmias, J. and R.V. Sansbury, *Grating contrast: discrimination may be better than detection.* Vision Research, 1974. **14**(10): p. 1039-42.

[11]. Chandler, D.M., K.H. Lim, and S.S. Hemami. *Effects of spatial correlations and global precedence on the visual fidelity of distorted images*. in *Proc. SPIE 6057, Human Vision and Electronic Imaging XI, 15 - 19 January 2006* 2006. San Jose, CA, USA: SPIE.

[12]. Tan, M., J.-F. Lalonde, L. Sharan, H. Rushmeier, and C. O'Sullivan, *The Perception of Lighting Inconsistencies in Composite Outdoor Scenes.* ACM Transactions on Applied Perception, 2015. **12**(4): p. article 18.

[13]. Hanhart, P., P. Korshunov, T. Ebrahimi, Y. Thomas, and H. Hoffman, *Subjective Quality Evaluation of High Dynamic Range Video and Display for Future TV.* SMPTE Motion Imag. J., 2015. **124**(4): p. 1-6.

[14]. Neisser, U., *Cognitive psychology*. 1967: Appleton-Century-Crofts.

[15]. Dick, A.O., *Iconic memory and its relation to perceptual processing and other memory mechanisms.* Perception & Psychophysics, 1974. **16**(3): p. 575-596.

[16]. Sperling, G., *The information available in brief visual presentations.* Psychological Monographs: General and Applied, 1960. **74**(11): p. 1-29.

[17]. Magnussen, S., *Low-level memory processes in vision.* Trends in Neurosciences, 2000. **23**(6): p. 247-251.

[18]. Hoffman, D.M. and D. Stolitzka, *A new standard method of subjective assessment of barely visible image artifacts and a new public database.* J. of the Society for Information Disp., 2015. **22**(12): p. 631-643.

[19]. Froehlich, J., G.M. Su, S. Daly, A. Schilling, and B. Eberhardt. *Content aware quantization: Requantization of high dynamic range baseband signals based on visual masking by noise and texture*. in *2016 IEEE International Conference on Image Processing (ICIP)*. 2016.

[20]. Sheikh, H.R. and A.C. Bovik, *Image information and visual quality.* IEEE Transactions on Image Processing, 2006. **15**(2): p. 430-444.

[21]. Wang, Z., E.P. Simoncelli, and A.C. Bovik. *Multiscale structural similarity for image quality assessment*. in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*. 2003.

[22]. Wang, Z., A.C. Bovik, H.R. Sheikh, and E.P. Simonelli, *Image quality assessment: From error visibility to structural similarity.* IEEE Transactions on Image Processing, 2004. **13**(4): p. 600-612.

[23]. Mantiuk, R., K.J. Kim, A.G. Rempel, and W. Heidrich, HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. ACM Trans. Graph., 2011. **30**(4): p. 1-14.

[24]. Narwaria, M., R. Mantiuk, M. Silva, and P. Le Callet, *HDR-VDP-2.2: A calibrated method for objective quality prediction of high dynamic range and standard images.* J. of Electronic Imaging, 2015. **24**(1).

[25]. Narwaria, M., M. Perreira Da Silva, and P. Le Callet, *HDR-VQM: An objective quality measure for high dynamic range video.* Signal Processing: Image Communication, 2015. **35**: p. 46-60.

[26]. ITU-R. (2011). *Reference electro-optical transfer function for flat panel displays used in HDTV studio production* (ITU-R Rec. BT.1886). International Telecommunication Union, Radiocommunication Sector.

[27]. Miller, S., M. Nezamabadi, and S. Daly. *Perceptual Signal Coding for More Efficient Usage of Bit Codes*. in *The 2012 Annual Technical Conference & Exhibition*. 2012.

[28]. Poynton, C.A. *Rehabilitation of gamma*. in *Proc. SPIE 3299, Human Vision and Electronic Imaging III*. 1998. San Jose, CA, USA:

[29]. SMPTE. (2014). *ST 2086:2014 - SMPTE Standard - Mastering Display Color Volume Metadata Supporting High Luminance and Wide Color Gamut Images*.

[30]. SMTPE. (2016). *ST 2094-1:2016 - SMPTE Standard - Dynamic Metadata for Color Volume Transform — Core Components*.

[31]. Daly, S., T. Kunkel, X. Sun, S. Farrell, and P. Crum. *41.1: Viewer Preferences for Shadow, Diffuse, Specular, and Emissive Luminance Limits of High Dynamic Range Displays*. in *SID Symposium Digest of Technical Papers*. 2013. Blackwell Publishing Ltd. p. 563-566.

[32]. ITU-R. (2002). *Parameter values for the HDTV standards for production and international programme exchange* (Rec. ITU-R BT.709-5). International Telecommunication Union, Radiocommunication Sector.

[33]. ITU-R. (2012). *Parameter values for ultra-high definition television systems for production and international programme exchange* (ITU-R BT.2020). International Telecommunication Union,

[34]. Bovik, A.C., *Automatic Prediction of Perceptual Image and Video Quality.* Proceedings of the IEEE, 2013. **101**(9): p. 2008-2024.

[35]. Mack, A. and I. Rock, *Inattentional blindness: Perception without attention*, in *Visual attention*, R.D. Wright, Editor. 1998, Oxford University Press: New York. p. 55-76.

[36]. Simons, D.J., *Attentional capture and inattentional blindness.* Trends Cogn Sci, 2000. **4**(4): p. 147-155.

[37]. ISO/IEC. (2015). *DIS 29170-2, Evaluation procedure for visually lossless coding.* International Organization of Standards.

[38]. Tukey, J.W., *Exploratory data analysis*. 1977: Addison-Wesley.

[39]. Karam, L., *Lossless Image Compression*, in *The Essential Guide to Image Processing*, A.C. Bovik, Ed. 2009, Elsevier Acad. Pr. p. 385-417.

[40]. Churchland, P.S., V.S. Ramachandran, and T.J. Sejnowski, *A critique of pure vision*, in *Large-scale neuronal theories of the brain*, C. Koch and J.L. David, Editors. 1993, MIT Press. p. 23.

[41]. Ansorge, U., W. Kunde, and M. Kiefer, *Unconscious vision and executive control: How unconscious processing and conscious action control interact.* Consciousness and Cognition, 2014. **27**: p. 268-287.

[42]. Goodale, M. and D. Milner, *Sight unseen: An exploration of conscious and unconscious vision*. 2013: OUP Oxford.

[43]. Treisman, A.M. and G. Gelade, *A feature-integration theory of attention.* Cognitive Psychology, 1980. **12**(1): p. 97-136.

[44]. Najemnik, J. and W.S. Geisler, *Optimal eye movement strategies in visual search.* Nature, 2005. **434**(7031): p. 387-391.

[45]. Yang, S.N., *Effects of gaze-contingent text changes on fixation duration in reading.* Vision Res, 2009. **49**(23): p. 2843-55.

[46]. Koch, C. and N. Tsuchiya, *Attention and consciousness: two distinct brain processes.* Trends Cogn Sci, 2007. **11**(1): p. 16-22.

Author Biography

Robert Allison is a Professor at York University and a member of the Centre for Vision Research. He obtained his PhD, specializing in stereoscopic vision in 1998 and did post-doctoral research at York University and the University of Oxford. His research enables effective technology for advanced virtual reality and augmented reality and for the design of stereoscopic displays. He is recipient of the Premier's Research Excellence Award in recognition of this work.

Kjell Brunnström, Ph.D., is a Senior Scientist at Acreo Swedish ICT AB and Adjunct Professor at Mid Sweden University. He is an expert in image processing, computer vision, image and video quality assessment having worked in the area for more than 25 years. Currently, he is leading standardization activities for video quality measurements as Co-chair of the Video Quality Experts Group (VQEG). His current research interests are in Quality of Experience for visual media in particular video quality assessment both for 2D and 3D, as well as display quality related to the TCO requirements.

Damon M. Chandler received the B.S. in Biomedical Engineering from The Johns Hopkins University (1998); and the M.Eng., M.S., and Ph.D. in Electrical Engineering from Cornell University (2000, 2004, 2005). From 2005-2006, he was a postdoc in the Department of Psychology at Cornell. From 2006-2015, he was on the faculty at Oklahoma State University. He is currently an Associate Professor at Shizuoka University, where his research focuses on modeling properties of human vision. He is as an Associate Editor for the IEEE TIP and the Journal of Electronic Imaging.

Hannah Colett, BS is a Human Factors Engineer and Software Developer at Intel. Her specialty is the development, design, and implementation of applications for collecting subjective metrics from users for competitive assessment. Currently Hannah works in the Sales and Marketing arm of Intel working on the development of methodologies for measuring, collecting, predicting, and benchmarking the user experience associated with various competitive technologies. She also works extensively with her alma mater Oregon State University to bridge academia and industry needs.

Philip Corriveau, BSc. is a Principal Engineer and Director of UX at Intel Corporation and Adjunct Professor at Pacific University. In January 2009 he was awarded a National Academy of Television Arts & Science, Technology & Engineering Emmy® Award for User Experience Research for the Standardization of the ATSC Digital System. Philip manages a multi-disciplined team investigating user experience metrics for Intel products and technologies. Currently Philip works in the Sales and Marketing arm of Intel

Scott Daly is currently with Dolby Laboratories, and is working on High Dynamic Range (HDR), High Frame Rate (HFR), and VR imaging systems, with a focus on perceptual issues. He has degrees in Electrical Engineering and Bioengineering. His previous accomplishments include key contributions to the DICOM medical imaging standard, a technical Emmy for a video transceiver, and the Visible Differences Predictor (VDP) while at Kodak. He holds an Otto Schade award from SID from his accumulated work at Kodak and Sharp, and while at Dolby was a co-author of the cone-based Perceptual Quantizer nonlinearity (SMPTE 2084.).

James Goal, is employed as a Director of Engineering and Technical Standards by Qualcomm, since 2011. He has a B.Sc. in Applied Science and Electrical Engineering from the University of Waterloo, Canada, 1992.

Juliana Knopf holds a BA in History and Anthropology from University of California, Irvine and a MA in Anthropology from California State University Fullerton. She is a User Experience Researcher and Human Resources professional she also holds an MS in Human Resources Management from Chapman University. She currently works at Intel Corporation in the Human Resources group, conducting research on the organization's culture and doing program management for qualitative user studies and fieldwork.

Laurie Wilcox Ph.D., is a Professor of Psychology at York University, Toronto. She is a long-standing member of the Centre for Vision Research, and of the graduate program in Biology. In addition to fundamental research on stereoscopic depth perception she collaborates with industry partners on applied projects related to 3D cinema, image quality assessment, and electronic display systems. Her work is funded by several sources, including the Natural Sciences and Engineering Council of Canada.

Shun-nan Yang, Ph.D., is an Associate Professor of Optometry and the director of Vision Performance Institute at the Pacific University College of Optometry. His research specialty is the cortical and subcortical control of eye movements in complex tasks such as reading and scene viewing. He is involved in the updating of Human Factor Ergonomic Standard in the United States. Dr. Yang studies how visual information is processed in the cortex and utilized to initiate cognitive control of eye movement in complex visuomotor sequences. He has published peer-reviewed articles in electronic imaging, vision, optometry, and neurophysiology-related journals.

Yi Zhang received the B.S. and M.S. degrees from Northwestern Polytechnical University, Xi'an, China, in 2008 and 2011, respectively, and the Ph.D. degree from Oklahoma State University, Stillwater, OK, in 2015, all in electrical engineering. He is currently doing the Postdoctoral research in Shizuoka University in Japan. His research interests include 2D/3D image processing, machine learning, pattern recognition, and computer vision