Application Note

Perceptual-Based Objective Picture Quality Measurements

Introduction

In video systems, a wide range of video processing devices can affect overall picture quality. Encoders and decoders compress and decompress video content. Format converters change resolutions and aspect ratios. Devices and transmission paths add noise and delay. Any of these devices can introduce impairments in video content that reach end consumers and degrade their viewing experience.

Video equipment manufacturers want to minimize the impairments their products may introduce into the video content. They face substantial challenges in these efforts. Profitability pressures can lead to difficult tradeoffs as designers attempt to optimize performance and meet target product costs. Time-to-market pressures limit the time available for quality assurance testing.

Video broadcasters and operators of communication networks that carry video content must carefully qualify new video equipment they deploy in their networks. Once they install these products in their networks, they need to determine the system configuration that delivers the best overall picture quality. In operating networks, the engineering staff needs to conduct regular maintenance to detect system degradations before they become picture quality problems that generate viewer complaints.

Video content producers must deliver video content in an ever-increasing number of formats into a media environment that is growing more diverse. Format conversion processes, along with coloring, editing, special effects, and branding processes can introduce impairments in video content.

The consumer's ability to perceive these impairments depends not only on the type of impairment, but on display technology and viewing conditions. Viewers watching video content on an interlaced scan CRT display will see different impairments than viewers watching the same content on a progressive scan LCD display. Similarly, a viewer watching video on a large screen in the low lighting of a home theater environment will perceive different artifacts than a viewer watching the same video on the small screen of a personal video player standing on a street corner in bright sunlight.

Consumers' quality expectations continue to rise as analog video technology transitions to digital technology and standard definition transitions to high definition. Unlike analog video systems, video equipment manufacturers, broadcasters, network operators and content providers cannot rely solely on signal measurements and picture monitors to assess picture quality. They need better tools to verify that their devices, systems, or processes have not introduced impairments in video content that will affect perceived picture quality.

Subjective Assessment and Objective Picture Quality Measurement

Many organizations use an informal method of subjective picture quality assessment. When a project team needs to evaluate picture quality, it will ask a group of people to compare test video sequences to reference video sequences. Over time, one person or a small group of people will demonstrate an ability to detect video quality impairments. These are the organization's "golden eyes."

Subjective picture quality ratings by these "golden eyes" may match the end consumer's video experience. However, these discerning viewers may see artifacts that the average viewer might miss. Projects may experience delays because a "golden eyes" evaluator is not available, or project teams may be restricted to a small number of evaluations because of resource, time or scheduling constraints. Evaluation costs can become an issue, especially if the team uses a "golden eyes" evaluator from outside the organization. Subjective evaluations can easily take an hour or more. In these situations, evaluator error due to fatigue becomes a factor.

These considerations have led organizations to consider alternative approaches to subjective picture quality evaluation. Researchers have developed several different methods of conducting formal subjective picture quality assessments. The ITU-R BT.500 recommendation describes various methods, along with requirements for selecting and configuring displays, determining reference and test video sequences, and selecting subjects for viewer audiences.

Some teams may have access to internal resources that can conduct these formal subjective assessments. More typically, independent laboratories perform this subjective testing. Specifying the desired tests, gathering the required video content, recruiting and selecting the viewer audience, conducting the tests, and analyzing the results generally requires several weeks. Overall cost for these subjective picture assessments can easily reach thousand of dollars.

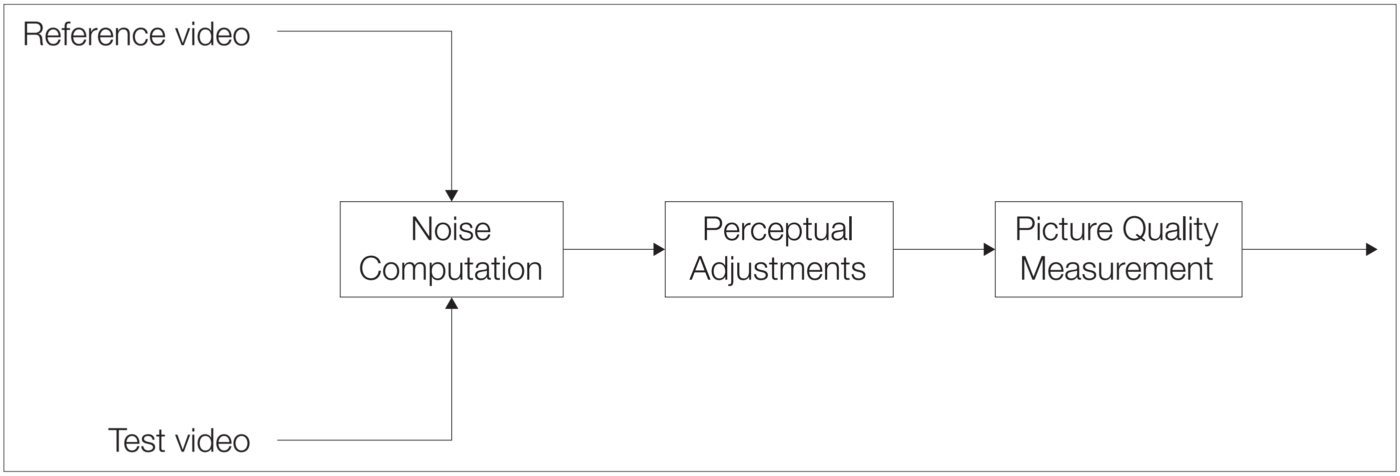


Figure 1. Noise-based Objective Picture Quality Measurements.

With this commitment of time, resources and expense, organizations will conduct a very limited amount of formal subjective picture quality assessments. If they use these methods at all, teams will generally perform this testing at very few critical milestones in a project or deployment. They cannot use these methods for frequent, repeated picture quality measurements to diagnose picture quality problems; optimize product design or system performance; and conduct extensive product, system, or content verification.

Engineering, maintenance, and quality assurance teams will turn to instruments that make objective picture quality measurements for these repeated picture quality assessments. We can initially classify objective picture quality measurements into three major groups: full-reference, reduced-reference, and no-reference.

Full-reference measurements compare a reference video sequence and a test video sequence. In the standard case, the test video is a processed version of the reference video, where the processing has introduced differences between the reference and test videos. No-reference measurements operate only on test video sequences. Reduced-reference measurements base picture quality assessments on extracted properties of the reference and test videos rather than a pixel-by-pixel comparison.

In most instances, full-reference objective picture quality measurements correspond to the subjective picture quality assessments described above. Most applications involving equipment or system design, equipment qualification, system configuration and optimization or content verification will have access to both reference and test videos. They will want to use the more capable full-reference objective picture quality measurements to assess picture quality.

Figure 1 diagrams one of the two categories of full-reference objective picture quality measurements. Noise-based measurements compute the noise, or error, in the test video relative to the reference video. The Peak Signal-to-Noise Ratio (PSNR) measurement is a commonly-used method in this measurement category.

The PSNR measurement is especially helpful in diagnosing defects in video processing hardware and software. Changes in PSNR values also give a general indication of changes in picture quality. However, it is well-known that PSNR measurements do not consistently match viewers' subjective picture quality assessments.

Figure 2.1. MSE=27.10 **Figure 2.2.** MSE=21.26

Figure 2. Image with Lower Mean Squared Error has Poorer Picture Quality.

Figure 2 illustrates this situation. The first step in a PSNR measurement computes the Mean Squared Error (MSE) between the test and reference video. The video frame shown in Figure 2.1 has greater MSE with respect to the original reference video than the video frame in Figure 2.2. However, the error in Figure 2.1 has high spatial frequency, while the error in Figure 2.2 consists of blocks at much lower spatial frequency. The human vision system has a stronger response to the lower spatial frequencies in Figure 2.2 and less response at the higher spatial frequencies in Figure 2.1. Subjectively, Figure 2.2 is worse than Figure 2.1, even though the noise-based PSNR measurement would assess Figure 2.1 as the poorer image.

Clearly, human visual perception is not equivalent to a simple noise detector. To produce results that match viewers' subjective assessments, a picture quality measurement that begins by computing the noise between reference and test video sequences must make adjustments that account for human visual perception (see Figure 1). Various algorithms use different approaches to calculate the noise and make the perceptual adjustments.

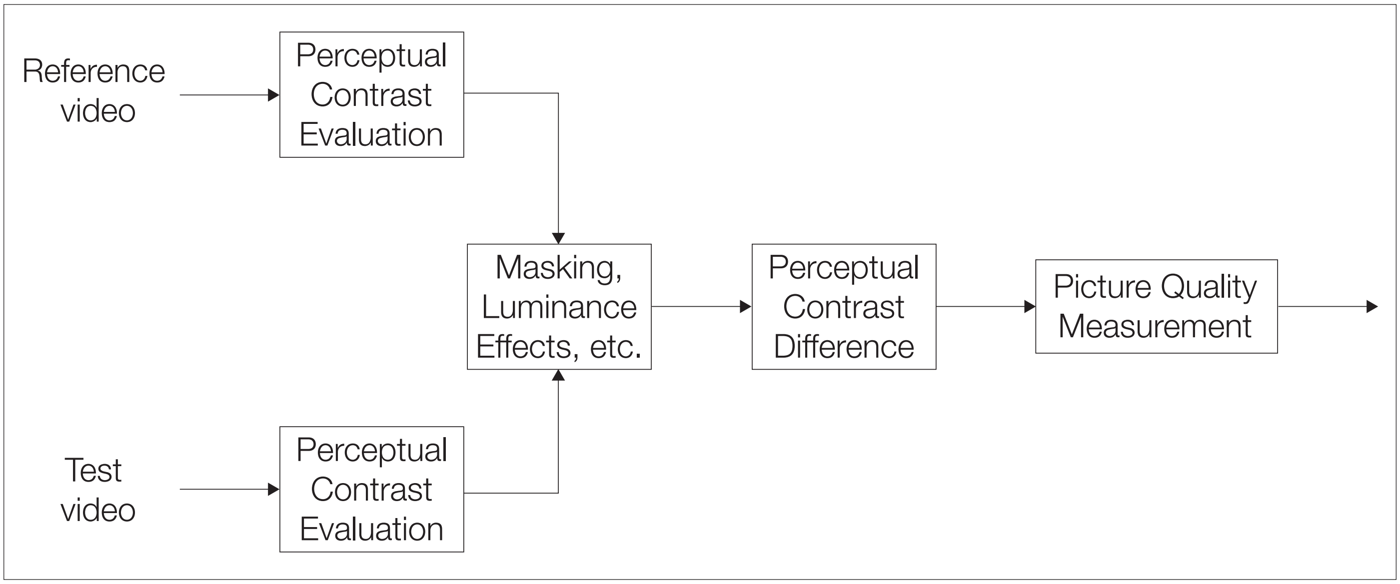


Figure 3. Perceptual-based Objective Picture Quality Measurements.

Figure 3 diagrams the second category of full-reference objective picture quality measurements. Perceptual-based measurements use human vision system models to determine the perceptual contrast of the reference and test videos. Further processing accounts for several other perceptual characteristics. These include relationships between perceptual contrast and luminance and various masking behaviors in human vision. The measurement then computes the perceptual contrast difference between the reference and test videos rather than the noise difference. The perceptual contrast difference determined by this processing is used directly in making perceptual-based picture quality measurements. With an accurate human vision model, picture quality measurements based on these perceptual contrast differences will match viewers' subjective evaluations.

The PQA500 offers both noise-based and perceptual-based picture quality measurements. The Picture Quality Rating (PQR) and the Difference Mean Opinion Score (DMOS) measurement are perceptual-based full-reference measurements. The familiar Peak Signal-to-Noise Ratio (PSNR) measurement falls into the noise-based category.

The PQA500 offers additional measurements that complement these primary measurements. These include measurements that detect video artifacts, e.g., lost edges (blurring), added edges, (ringing, mosquito noise) or blockiness. Other measurements weight the results of DMOS, PQR or PSNR with the results from these artifact detectors or from the PQA500's Attention Model[1].

The following sections describe key concepts underlying the perceptual-based PQR and DMOS measurements on the PQA500. Additional materials available on the Tektronix website cover other aspects of the PQA500's capabilities. The application note titled "Understanding PQR, DMOS, and PSNR Measurements" contains specific information on configuring, interpreting, and using these measurements. The technical brief, "An Adaptable Human Vision Model for Subjective Video Quality Rating Prediction Among CIF, SD, HD and E-Cinema" presents detailed technical descriptions of key elements in the PQA500's human vision system model. Finally, the application note titled "Picture Quality Analysis for Video Applications" has an overview of PQA500 capabilities and how these capabilities address requirements for picture quality evaluation in various video applications.

[1] The PQA500's Attention Model determines and tracks viewers' focus of attention as they watch the video content. See the PQA500's User Manual and PQA500's Technical Reference for more information on the Attention Model and attention-weighted measurements.

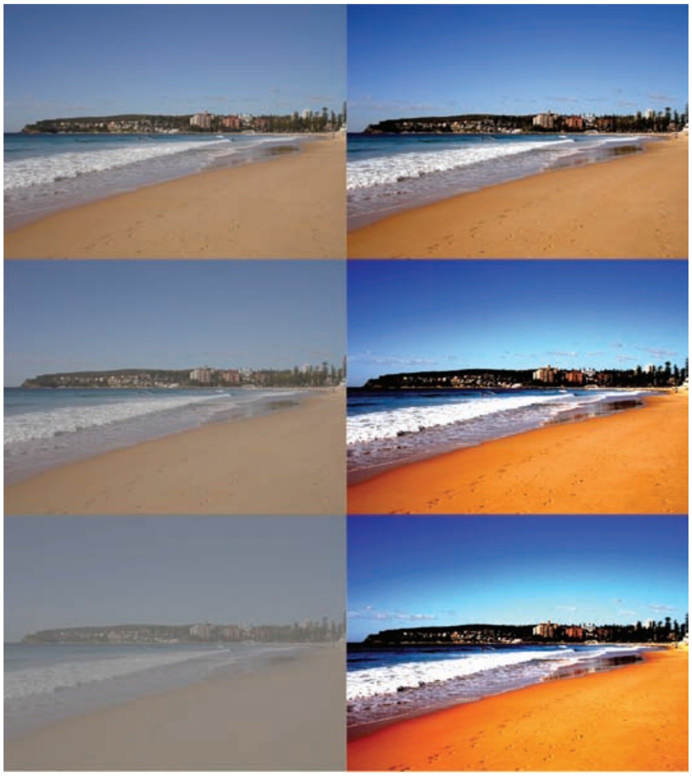


Figure 4. Contrast Makes Objects Distinguishable.

Contrast and Perceptual Contrast

In assessing picture quality, viewers need to distinguish objects in the images and detect impairments. Contrast is the difference in visual properties that makes objects distinguishable from each other and from the background (Figure 4).

Luminance dominates human perception. Hence, perceptual contrast is most strongly related to luminance contrast. Various definitions of luminance contrast are used for different situations, but they are all ratios of a luminance difference and an average luminance. Contrast is expressed as a percentage value.

Michelson contrast is commonly used when the scene has a roughly equal number of bright and dark regions.

$$\text{Michelson Contrast} = \frac{L_{max} - L_{min}}{L_{max} + L_{min}}$$

where:

L_{max} = the maximum luminance in the scene

L_{min} = the minimum luminance in the scene

Luminance is a photometric measurement of light intensity. Ratios like Michelson contrast values are similarly photometric measures determined with a light meter.

However, the human vision system does not respond as a light meter. Most critically, people's ability to perceive a luminance difference, their contrast sensitivity, depends on the luminance distribution in space and time. A human vision system model used in perceptual-based picture quality measurement must accurately capture the spatial and temporal frequency response of perceptual contrast sensitivity.

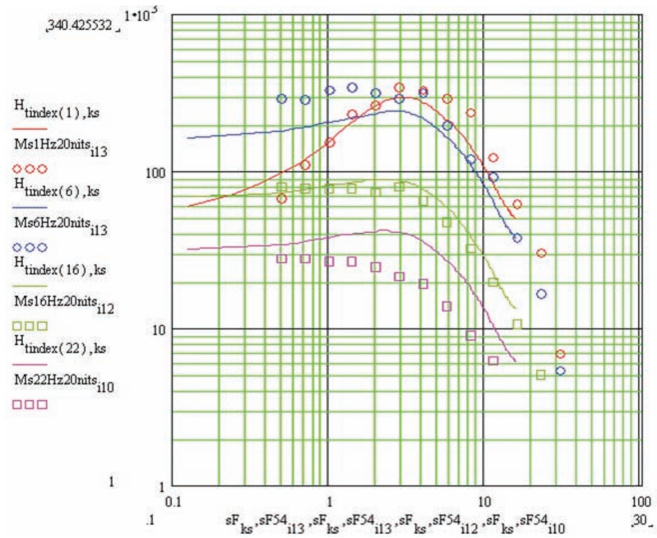


Figure 5.1. Spatial Frequency Response.

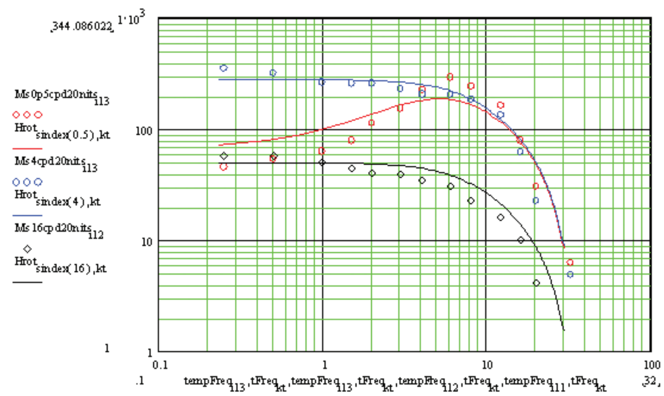


Figure 5.2. Temporal Frequency Response.

Figure 5. Interdependence of Contrast Sensitivity Spatial and Temporal Frequency Responses.

Perceptual Contrast Sensitivity

Numerous studies have shown that the human vision system has a band-pass response to spatial and temporal variation in contrast. These studies use sine-wave gratings, essentially alternating patterns of light and dark bars, to determine these contrast sensitivity functions.

In spatial frequencies, at moderately low luminance levels and with little or no temporal variation, the human vision system has its maximum contrast sensitivity at approximately 4 cycles per degree (cpd) with a cutoff frequency of approximately 60 cpd (see red plot in Figure 5.1). In temporal frequencies, at moderately low luminance levels and with little or no spatial variation, the human vision system has a maximum response at approximately 8 Hz with a cutoff of approximately 50 Hz (see red plot in Figure 5.2).

The plots in Figure 5 show another important property. The human vision system's response to spatial variation in contrast depends on temporal variations (Figure 5.1). Similarly, the human vision system's response to temporal variation in contrast depends on spatial variation (Figure 5.2). In particular, the spatial frequency response changes from band-pass to low-pass as the temporal variation in luminance difference increases in frequency. Temporal frequency response also changes from band-pass to low-pass as the spatial variation in the luminance difference increases in frequency.

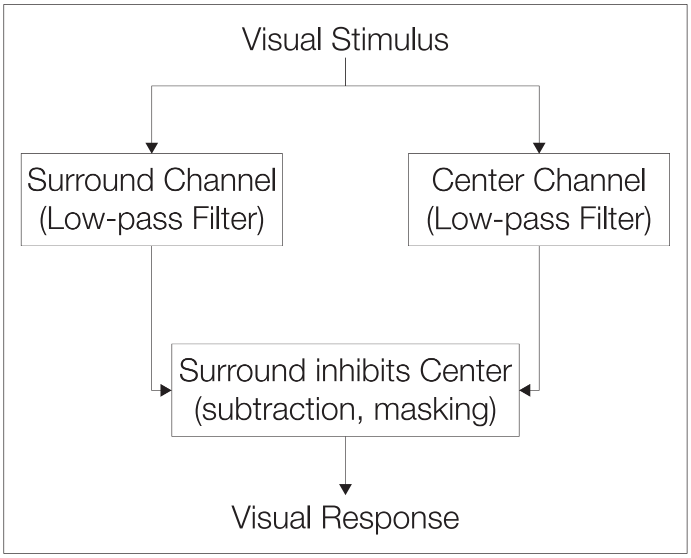


Figure 6. Center and Surround Channels.

These characteristics occur because the human vision system creates its overall response by combining input from a center channel and a surround channel. Both channels act as low-pass filters on the video stimulus. The low-pass filter in the surround channel has a lower cutoff frequency. In creating the overall response, the surround channel inhibits the center channel response through a combination of subtraction and masking (see Figure 6). This reduces low frequency components in the visual stimulus, improving people's ability to detect edges, individual objects, and scene details. This also creates the typical band-pass response seen in contrast sensitivity plots.

The surround channel integrates the visual stimulus over both space and time. This combined surround channel response has only low spatial and temporal frequencies. Thus, the surround channel inhibits lower frequencies of the combined spatial and temporal center channel response. As higher frequency temporal variation is added to the visual stimulus in the contrast sensitivity experiments, the temporal component of the surround channel has diminished effect and the spatial frequency response changes from band-pass to low-pass (see blue, green and magenta plots in Figure 5.1).

Likewise, the temporal frequency response remains band-pass as long as spatial variations in luminance difference remain at low frequencies. As spatial variations increase in frequency, the spatial component of the surround channel has a diminished effect and the temporal frequency response transitions to a low-pass character (see blue and black plots in Figure 5.2).

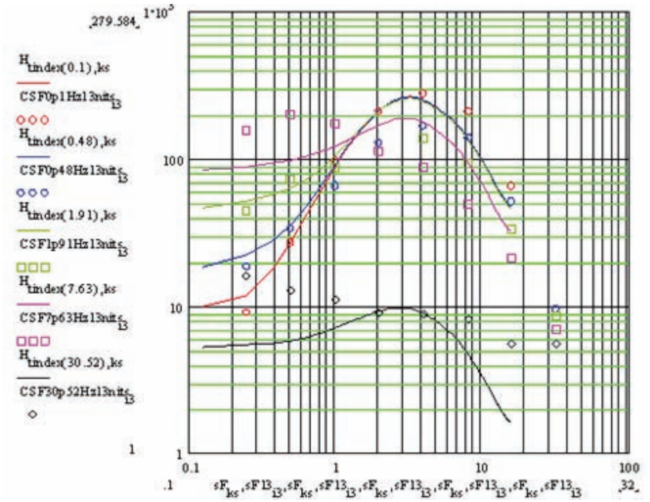


Figure 7.1. Spatial frequency response.

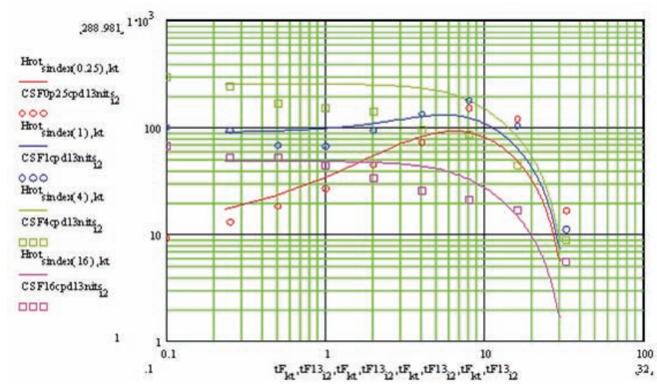


Figure 7.2. Temporal frequency response.

Figure 7. Interdependence of Contrast Sensitivity Spatial and Temporal Frequency Response at 13 candela/m^2.

Perceptual contrast sensitivity also depends on the overall luminance of the visual stimulus as illustrated in Figure 7. These plots show the spatial and temporal frequency responses at 13 candela/m^2 of overall luminance. A candela/m^2 is commonly called a "nit."

Figure 5 plots the contrast sensitivity functions at 20 nits in overall luminance.

The two plot sets do not use identical parameters. However, they illustrate how overall luminance affects contrast sensitivity. For example, compare the spatial frequency response at 7.63 Hz temporal variation and 13 nits (magenta plot in Figure 7.1) to the response at 6.0 Hz and 20 nits (blue plot in Figure 5.1). The peak of the band-pass occurs at a lower frequency for the lower mean luminance case. The impact on these plot characteristics is less for the temporal frequency response, although the change in overall luminance does shift data point values for comparable plot parameters.

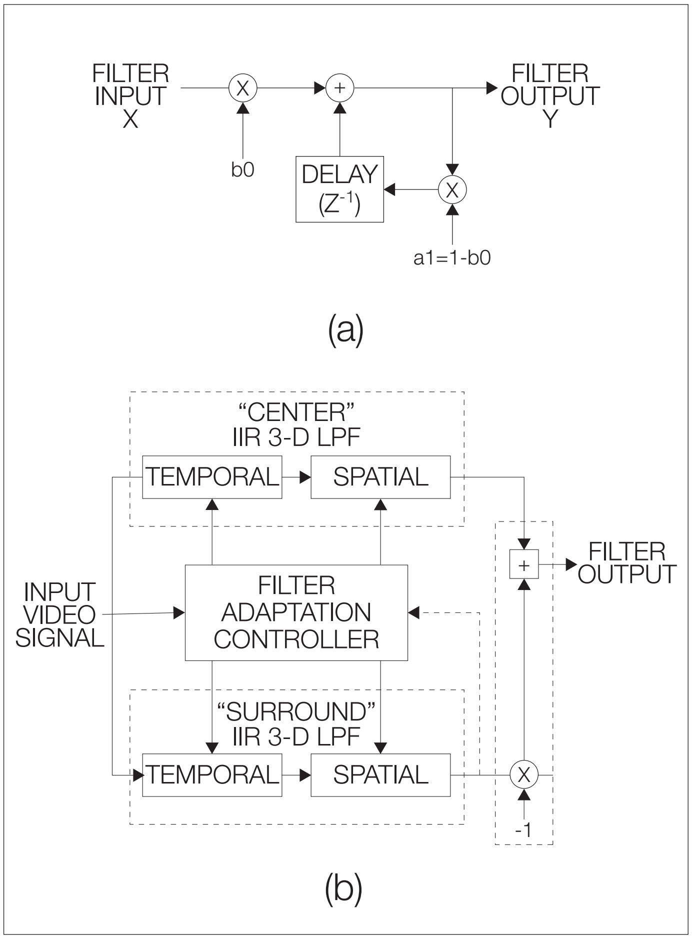


Figure 8. Adaptive Filtering in PQA500's Human Vision System Model.

Human vision system models with fixed spatial and temporal filtering cannot capture these complex relationships. Adaptive filters can model these relationships successfully. The human vision system model Tektronix developed for the PQA500 uses adaptive filters to reproduce this behavior, including the inter-adaptation between the surround and center channels. Figure 8a shows the adaptive integrator used as the primary building block in the PQA500's spatiotemporal filters. The block diagram in Figure 8b shows the connections in the model that replicate the human visual system response. The solid lines in the plots in Figures 5 and 7 show the output from this human vision system model. These results correlate well with the actual results from tests with human subjects shown as square data points on the plots.

The complex interactions in the human vision system lead to non-linear effects that produce optical illusions. Well-known illusions include:

- Afterimages: An afterimage or ghost image that continues to appear after the exposure to the original image has ceased. One of the most common afterimages is the bright glow that seems to float before the eyes after staring at a light bulb or a headlight for a few seconds.
- Frequency doubling illusion: An apparent doubling of spatial frequency when a sinusoidal grating is modulated rapidly in temporal counter-phase.
- Phantom pulse illusion: An observer sees successive flashes of light. The time between flashes is decreased until the observer sees only a single flash. As the time between flashes decreases, some observers see a phantom third flash between the two flashes.

Optical illusions can affect the differences viewers perceive in subjective evaluations. For example, an after-image can have a masking effect. Viewers experiencing an afterimage from an earlier visual stimulus may be less able to discern details in newly presented video. To ensure a good match with subjective evaluations, objective picture quality measurements should account for optical illusion effects.

Optical illusions are also useful tests for human vision system models. If the model input corresponds to the visual stimulus for an optical illusion, the model output should correspond to the optical illusion. In addition to accurately capturing the temporal and spatial characteristics of perceptual contrast sensitivity, the PQA500's human vision system model can correctly simulate optical illusions, including the illusions described above.

$$\frac{\Delta L}{L} = k$$

(a)

(b)

Figure 9. Weber's Law.

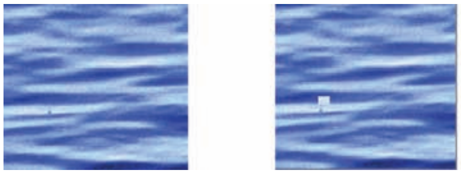
Detecting Perceptual Contrast Differences

Full-reference objective picture quality measurements correspond to subjective evaluations which ask viewers to detect differences between reference and test video sequences. As noted earlier, perceptual-based full-reference picture quality measurements use a human vision model to directly determine the perceptual contrast differences between the reference and test video sequences. Accurately modeling perceptual contrast sensitivity forms the foundation for detecting these perceptual contrast differences. However, the human vision system model must account for several other factors that affect perceptual contrast.

These factors include Weber's Law, which holds at moderate contrast levels, and variations from Weber's Law that occur at low and high contrast levels. The perceptual mechanisms underlying Weber's Law also create the related Steven's Effect and a relationship between perceptual threshold and area. Finally, a human vision system model should account for various masking mechanisms that affect viewers' ability to detect differences between the reference and test videos.

Weber's Law concerns the relationship between luminance difference and luminance of the surrounding area. Weber's Law states that as luminance surrounding an object increases, the luminance difference needed to distinguish the object from the surrounding area must also increase. Figure 9a shows the mathematical form of Weber's Law. In the formula, L is the background luminance, ΔL is the incremental luminance of an object above the background luminance (L) that is needed to make the object just noticeable. The value k is a constant. Thus, if the background luminance is increased, the object's incremental luminance (ΔL) must increase by the same amount to remain just noticeable.

Figure 9b illustrates the increased difficulty in distinguishing objects as the surrounding luminance increases. Notice how the objects on the tablecloth become more difficult to see as the luminance level is increased. In effect, Weber's Law indicates that constant contrast, rather than constant luminance, has more importance in determining the information viewers' perceive in video.



Reference Test

Figure 10.1. Low Spatial Variance Video.

Reference Test

Figure 10.2. High Spatial Variance Video.

Figure 10. Variance Masking.

a. Reference b. Test 1 c. Test 2

Figure 11. Similarity Masking.

The **Steven's Effect** also concerns luminance changes. In this effect, people perceive an apparent increase in luminance when the luminance changes. The human vision system model used in the PQA500 accounts for Weber's Law and the Steven's Effect. In particular, modeling the Steven's Effect requires the luminance sensitive, adaptive, spatiotemporal band-pass filtering described earlier.

The PQA500 also accounts for the **Area Threshold** behavior in perception. The human vision system will not perceive a small contrast difference in a small area. It can perceive that same small contrast difference when it appears over a larger area.

Perceptual contrast must rise above a threshold before the human vision system perceives the contrast. Thus, very low levels of perceptual contrast in the reference and test video sequences are masked by this threshold behavior. The PQA500's human vision system model includes a **Noise Masking** stage to account for this factor.

At moderate to high contrast levels well above the perceptual contrast threshold, called the supra-threshold region, spatiotemporal sensitivity tends to be flatter across the frequency range than in the plots shown in Figures 5 and 7. This threshold to supra-threshold transition accounts for some more important exceptions to Weber's Law. The PQA500's human vision system model includes mechanisms that adapt to different contrast levels in addition to the luminance level adaptations.

The image content surrounding a perceptual contrast difference can affect viewers' ability to perceive the difference. A perceptual contrast difference is much easier to see when it occurs in a region of low spatial variance, e.g., regions of sky or slow moving water. The same contrast difference is much harder to perceive when it occurs in a region of high spatial variance, e.g., grass or tree leaves. Regions of high spatial variance mask perceptual contrast differences between reference and test videos.

Figure 10 illustrates this **Variance Masking** behavior. In the low spatial variance scene of flowing water shown in Figure 10.1, viewers can quite easily perceive the constant luminance block in the test video. Viewers have much more difficulty perceiving the constant luminance block within the high spatial variance scene of tree leaves in Figure 10.2 (circled in red in the second test video image).

Viewers will also have difficulty detecting differences if these differences strongly correlate with elements in the video image, a perceptual factor called **Similarity Masking**. Figure 11 presents a simplified illustration of this concept.

The Test 1 image in Figure 11b contains a difference relative to the Reference image in Figure 11a. However, the Test 1 image is highly correlated with the Reference image. In other words, the Test 1 image and the Reference image are similar.

The Test 2 image in Figure 11c contains the same magnitude of difference, e.g., the same MSE, but the Test 2 image has lower correlation with the Reference image. The Test 2 image in Figure 11c is less similar to the Reference image than the Test 1 image in Figure 11b.

Viewers can more easily perceive the difference in the Test 2 image. The similarity between the Test 1 image and the Reference image has a significant masking effect on the difference.

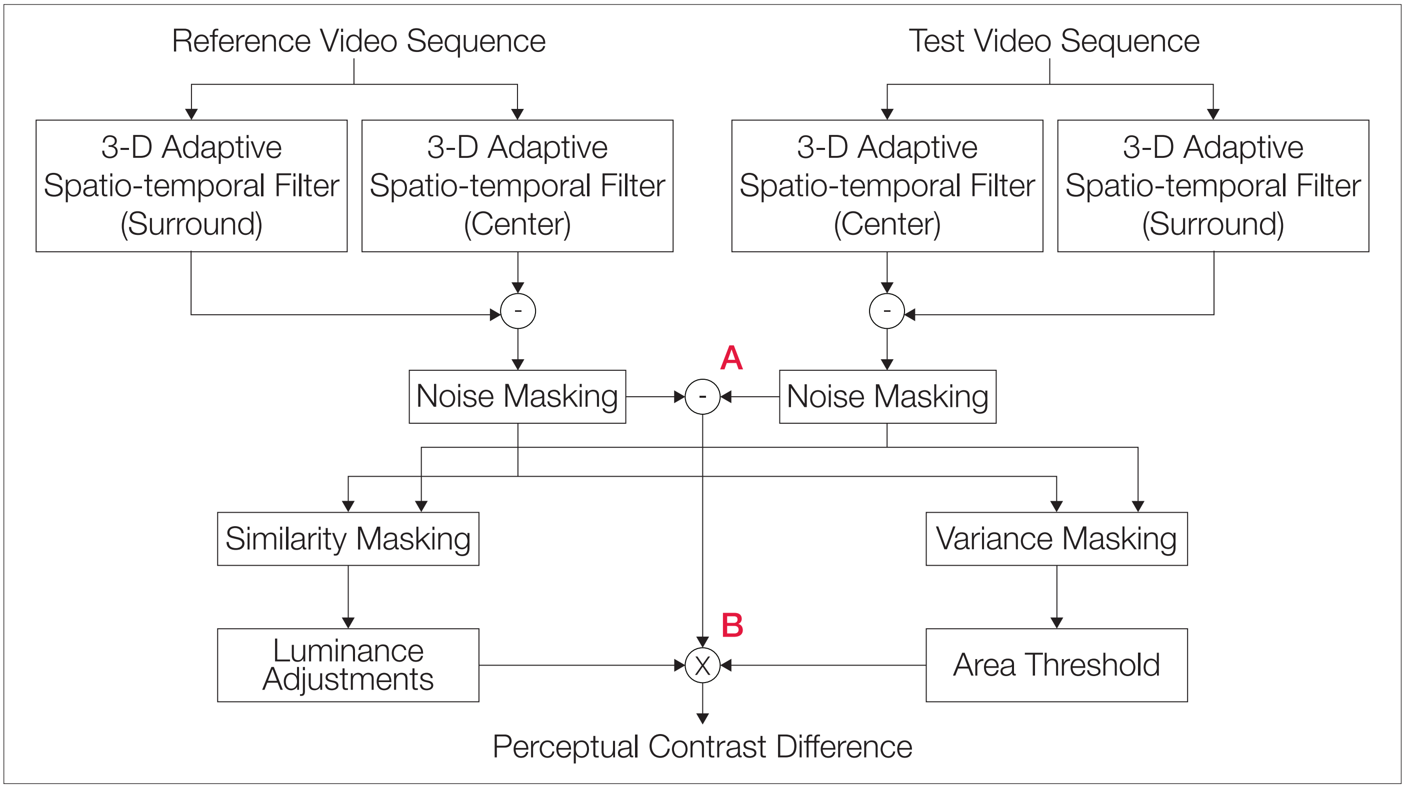


Figure 12. Human Vision System Model Block Diagram.

The PQA500's human vision system model accounts for the perceptual factors described above and accurately models the spatial and temporal characteristics of perceptual contrast sensitivity functions. Figure 12 shows a very simplified block diagram of the model. The technical brief available on the Tektronix website presents a more detailed description of the model's components and operations.

The PQA500 first converts data contained in the reference and test video files into the light values required as the model's input. It then determines the perceptual contrast of frames in the reference and test video sequences, using adaptive filtering to appropriately model the spatial and temporal contrast sensitivities.

After noise masking, the PQA500 subtracts the perceptual contrast of reference video pixels from the corresponding pixels in the test video sequence. This creates the initial version of the perceptual contrast difference map, recording the perceptual contrast differences at every pixel in each frame of the test video sequence (point A in the diagram).

The PQA500 continues processing the reference and test video sequences, evaluating the other factors that impact perception. This analysis produces individual weightings for each pixel in every processed frame that either diminishes (masks) the perceptual contrast difference at the pixel or enhances this difference. These weightings are applied to the initial perceptual contrast difference map (point B in the diagram) to create the final perceptual contrast difference map used in the perceptual-based picture quality measurements.

(a) Reference (b) Test

(c) Unsigned PCD Map (d) Signed PCD Map

Figure 13. Perceptual Contrast Differences (PCD) Map.

Figure 13 shows a perceptual contrast difference map generated by the PQA500 from reference video content shipped with the instrument (Figures 13a and 13b). In the default configuration (Figure 13 c), the PQA500 computes an unsigned perceptual contrast difference map. In this version, the perceptual contrast differences of the brighter pixels on the display have larger absolute values than the perceptual contrast differences of darker pixels on the display.

However, perceptual contrast differences are actually signed values. In some cases, the perceptual contrast of the reference video may exceed the perceptual contrast of the test video. In other cases, the reverse may occur. The signed perceptual contrast difference map in Figure 13d shows this information. In this version, pixels with no perceptual contrast difference appear as a neutral gray. Pixels with negative perceptual contrast difference appear dark, while pixels with positive perceptual contrast difference appear bright.

With the perceptual contrast difference map, the PQA500 has information on the perceptual contrast differences viewers can detect in test video sequences relative to reference video sequences. The PQA500 can convert this information into the perceptual-based PQR and DMOS measurements that match viewers' subjective assessments of the test videos. The next section discusses basic concepts in performing these measurements using the information in the perceptual contrast difference map, and describes aspects of interpreting and using these measurements. See the application note titled "Understanding PQR, DMOS and PSNR Measurements" for additional information.

$$\text{Minkowski}\,(f_n) = \left(\frac{1}{N_v N_h} \sum_{j=0}^{N_v} \sum_{i=0}^{N_h} \left| \text{PCD}(i,j,f_n) \right|^{p} \right)^{\frac{1}{p}}$$

(a)

$$\text{Minkowski}_{seq} = \left(\frac{1}{M} \sum_{n=0}^{M} \text{Minkowski}\,(f_n)^{q} \right)^{\frac{1}{q}}$$

(b)

Figure 14. Minkowski Metrics Used in the PQA500.

As mentioned earlier, the PQA500 has several capabilities beyond the principal picture quality measurements it offers. The PQA500's artifact detection measurements and Attention Model also use the human vision system model and perceptual contrast information. Using the results of these measurements to weight the basic picture quality measurements, evaluators can account for viewers' focus-of-attention or tolerance for different types of artifacts in assessing picture quality. This application note does not discuss these topics further. However, more information is available in the PQ500 User Manual, the PQA500 Technical Reference, and the application note titled, "Picture Quality Measurement for Video Applications."

Perceptual-Based Objective Picture Quality Measurements

To compute objective picture quality measurements for each frame in the test video, and for the overall test video sequence, the PQA500 must "pool" the individual perceptual contrast differences at every pixel. Potential pooling methods include averaging the perceptual contrast differences of the pixels in the video frame, or computing the root mean square (RMS) of these values.

Experimental work in visual sciences has shown that pooling these values using a more generalized mean, often called the Minkowski metric, yields the best results. Figure 14a shows the Minkowski metric used by the PQA500 to pool the perceptual contrast differences in a video frame. Negative and positive differences have equal impact. Thus, the metric pools the absolute value of the perceptual contrast differences (PCD) for every pixel in the video frame. In this formula, N_h is the number of pixels in a video line and N_v is the number of lines in the video frame. Calibration with subjective data determined that a value somewhat larger than 2.0 for the exponent, p, produces optimal results for this per-frame pooling.

Figure 14b shows the Minkowski metric used to determine the overall perceptual contrast difference of the entire test video sequence. Here, the Minkowski values for each video frame are pooled to compute a Minkowski value for the sequence. In this formula, M is the number of video frames in the video sequence. Calibration with subjective data determined the exponent, q, should have a value almost double the value used for exponent p in the per-frame Minkowski.

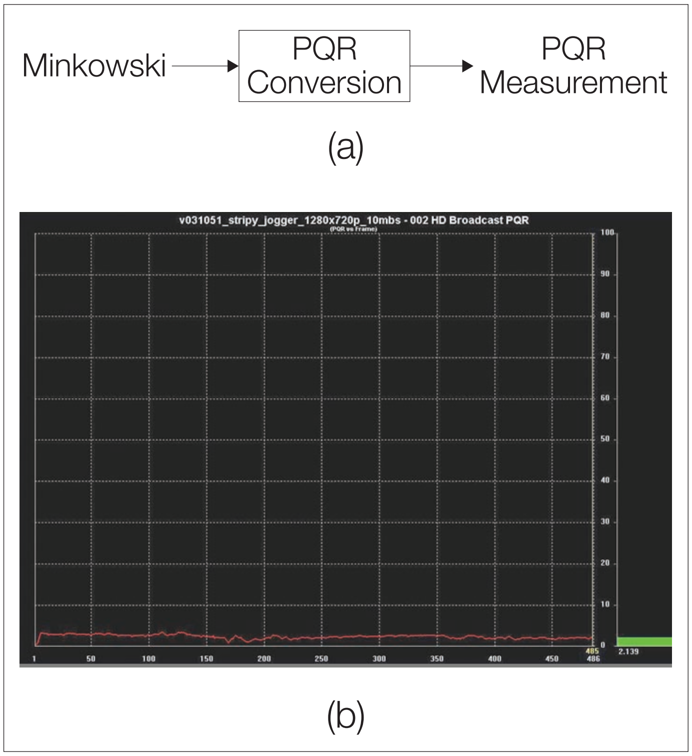


Figure 15. Picture Quality Rating Measurement.

After computing the Minkowski metrics, the PQA500 knows the perceptual contrast difference between the reference and test video, on both a per-frame basis and for the overall test video sequence. It uses this information to make two different types of perceptual-based picture quality measurements. Both types of picture quality measurements use the Minkowksi metrics computed from the perceptual contrast difference map. They differ in the conversion process they apply to these metrics. The conversion processes correspond to different subjective evaluation methods.

The Picture Quality Rating (PQR) measurements correspond to perceptual sensitivity measurements that determine Just Noticeable Differences (JNDs). In these measurements, the PQA500 directly converts Minkowski values to PQR values (Figure 15a). The conversion sets the value of 1 PQR to approximately 0.1% aggregated perceptual contrast difference between the reference and test videos. This is the perceptual contrast difference that perceptual sensitivity studies associate with 1 JND between the reference and test videos. Thus, 1 PQR equals 1 JND.

With this amount of perceptual contrast difference, most viewers can barely distinguish the test video from the reference video in the forced choice pair-wise comparisons used in these perceptual sensitivity experiments.

PQR values less than 1 indicate that viewers cannot detect differences between the reference and test videos. Viewers will perceive the test video as having equal quality to the reference video. If the reference and test videos are identical, the perceptual contrast difference map will be completely black and the PQR measurement results will equal 0.

As the perceptual contrast difference between the reference and test videos increases, the PQR measurement returns larger values. Viewers will have an easier time distinguishing differences.

We can treat each integer PQR value as an integer JND level. For most viewers, a test video that generates a PQR greater than 2 is noticeably different than the reference video. Visible MPEG encoder artifacts typically correspond to PQR measurements in the 2-4 range. Overall, higher PQR values represent degradation in test video quality relative to the reference video. Figure 15b shows a typical PQR measurement on the PQA500.

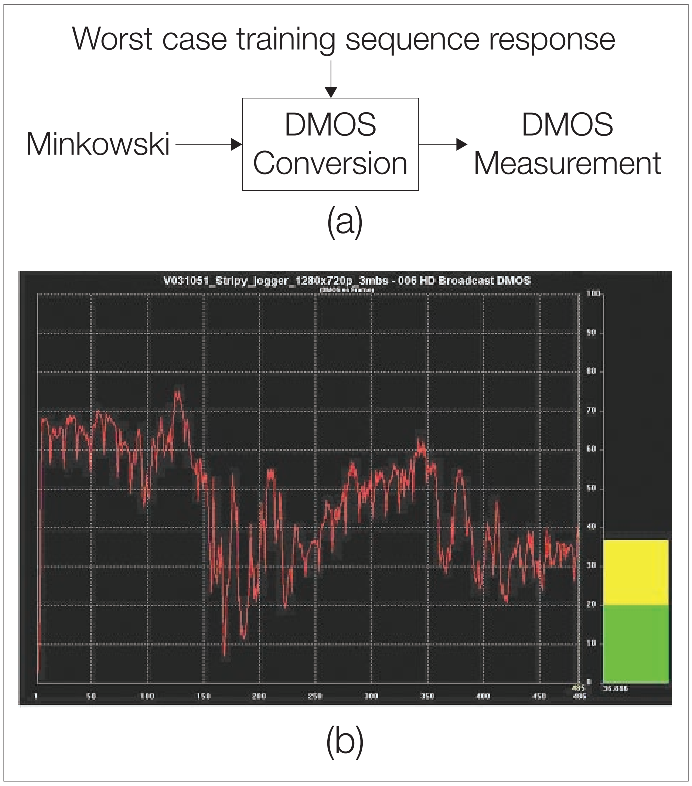


Figure 16. Difference Mean Opinion Score Measurement.

Difference Mean Opinion Score (DMOS) measurements correspond to subjective picture quality evaluations conducted using procedures defined in ITU-R BT.500. Because the perceptual contrast difference map contains the information on differences viewers will perceive between reference and test videos, the PQA500 can predict the DMOS values the test video would receive if evaluated as described in the ITU recommendation. Unlike testing with people, the PQA500 can produce a DMOS result for each frame in the sequence as well as the overall video sequence.

Like the PQR measurement, the PQA500 converts the Minkowski values computed from the perceptual contrast difference map into a DMOS measurement (Figure 16a). However, the DMOS measurement uses an entirely different conversion function. Built on data gathered from subjective testing, the DMOS conversion function contains a mapping between a given amount of perceptual contrast difference and the DMOS values viewer audiences typically give reference and test videos with that amount of difference.

If the reference and test videos are identical, the perceptual contrast difference map will be completely black. The predicted DMOS results (more simply the DMOS result) will be 0 in this case. Small amounts of perceptual contrast difference between the reference and test videos will produce DMOS values in the lower range of the 0-100 scale. These values correspond to the “Excellent” and “Good” positions on the ITU quality scale. Greater perceptual contrast difference between the reference and test videos will produce DMOS values in the upper range of the scale. These values correspond to “Poor” and “Bad” positions on the ITU quality scale. Figure 16b shows a typical DMOS measurement on the PQA500.

Figure 16a illustrates an additional difference in the conversion function used in DMOS measurements. DMOS measurements include a configuration parameter called the worst case training sequence response. This parameter corresponds to the training step associated with the subjective evaluation procedures defined in the ITU-R BT.500 recommendation. Before viewers evaluate any video, they are shown training video sequences that demonstrate the range and types of impairments they will assess in the test. ITU-R BT.500 recommends that these video sequences should be different than the video sequences used in the test, but of comparable sensitivity. In other words, the training video sequences cover the range from the “best case” to the “worst case” video the viewers will see.

Without the training session, viewers’ assessments would vary widely and change during the test as they saw different quality video. The training session ensures coherent opinion scores. However, this means the DMOS results for test sequences depend on the video content shown in the training session. Similarly, the results of a predicted DMOS measurement depend on the worst case training sequence response associated with the measurement.

As described in earlier sections, the PQA500 accurately models the characteristics of the human vision system. Similarly, data from subjective tests were used to calibrate the PQR and DMOS measurement conversion functions. The combination of an accurate human vision system model and calibrated conversion functions ensures measurement results match viewers' subjective assessments.

Engineering, verification, and quality assurance teams can apply these objective picture quality measurements to a wide range of applications. Examples include designing and optimizing video codecs, qualifying video equipment for deployment, configuring video systems to achieve optimal quality with minimal bandwidth, and verifying quality of re-purposed video content.

Over a wide range of impairments and viewing conditions, the DMOS measurement helps these teams determine how differences between the reference and test videos affect subjective quality ratings. The PQR measurement helps these teams determine how much viewers will notice differences between the reference and test videos, especially in the critical case of high-quality video when differences are near the visibility threshold.

This application note describes key elements underlying perceptual-based objective picture quality measurements to provide a conceptual foundation for interpreting and using PQR and DMOS measurements. Additional materials available on the Tektronix website provide further information on these measurements and other PQA500 capabilities.

Conclusion

In most cases, engineering and quality assurance teams that need to assess picture quality cannot afford the time and expense associated with recruiting viewers, configuring tests, and conducting subjective viewer assessments. They need accurate and repeatable objective picture quality measurements that can make these assessments more quickly than subjective evaluations, and at a lower cost. However, these objective measurements should match subjective evaluations as closely as possible.

Viewers are not noise detectors. They will perceive some differences between the reference and test videos, but not others. Full-reference picture quality measurements based on detecting the noise differences between the reference and test videos need to make adjustments to account for the characteristics of human perception.

Perceptual-based full-reference picture quality measurements take a different approach. Using a human vision system model, they directly compute the perceptual contrast differences between the reference and test videos. They use these perceptual contrast differences to produce results that match subjective viewers' ratings of video quality.

The PQA500 offers two perceptual-based picture quality measurements, the PQR and DMOS measurements. They use a human vision system model with the adaptive filtering needed to accurately capture the spatial and temporal characteristics of contrast sensitivity. This model also accounts for other key factors affecting viewers' ability to perceive differences in the test video. Careful calibration of the human vision system model and the conversion functions used to produce the PQR and DMOS measurements ensure results well-matched to subjective picture quality evaluation.

Contact Tektronix:

ASEAN / Australasia (65) 6356 3900
Austria +41 52 675 3777
Balkans, Israel, South Africa and other ISE Countries +41 52 675 3777
Belgium 07 81 60166
Brazil +55 (11) 3759 7600
Canada 1 (800) 661-5625
Central East Europe, Ukraine and the Baltics +41 52 675 3777
Central Europe & Greece +41 52 675 3777
Denmark +45 80 88 1401
Finland +41 52 675 3777
France +33 (0) 1 69 86 81 81
Germany +49 (221) 94 77 400
Hong Kong (852) 2585-6688
India (91) 80-22275577
Italy +39 (02) 25086 1
Japan 81 (3) 6714-3010
Luxembourg +44 (0) 1344 392400
Mexico, Central, South America and Caribbean 52 (55) 54247900
Middle East, Asia and North Africa +41 52 675 3777
The Netherlands 090 02 021797
Norway 800 16098
People's Republic of China 86 (10) 6235 1230
Poland +41 52 675 3777
Portugal 80 08 12370
Republic of Korea 82 (2) 6917-5000
Russia & CIS +7 (495) 7484900
South Africa +27 11 206 8360
Spain (+34) 901 988 054
Sweden 020 08 80371
Switzerland +41 52 675 3777
Taiwan 886 (2) 2722-9622
United Kingdom & Eire +44 (0) 1344 392400
USA 1 (800) 426-2200
For other areas contact Tektronix, Inc. at: 1 (503) 627-7111

For Further Information

Tektronix maintains a comprehensive, constantly expanding collection of application notes, technical briefs and other resources to help engineers working on the cutting edge of technology. Please visit **www.tektronix.com**

03/09 EA/ 28W_21916_1