



## Case Study: Can Objective Metrics Replace the Human Eye?

*Bill Reckwerdt, CTO  
Video Clarity, Inc.*



Can Objective Metrics replace the “Gifted” Human eye when evaluating video quality? In short – no – but they have an important purpose. Formal Subjective Video Analysis is costly, time consuming, and difficult to be used regularly. Objective Metrics act as important evidence; while Subjective Analysis is the confirmation.

## Background

Over recent decades, the role of video images has grown steadily. Advances in technologies underlying the capture, transfer, storage, and display of images have created situations where communicating by means of video images has become economically feasible. More importantly, video images are in many situations an extremely efficient way of communicating as witnessed by the proverb “a picture is worth a 1000 words.”

Notwithstanding these technological advances, the current state of the art requires many compromises. Examples of these compromises are temporal resolution versus noise, spatial resolution versus image size, and luminance/color range versus gamut. These choices affect the video quality of the reproduced images. To make optimal choices, it is necessary to have knowledge about how particular choices affect the impression of the viewer. This is the central question of all video quality analysis research.

Video quality analysis research can be divided into two approaches:

1. [Subjective Video Quality Analysis](#)
2. [Objective Video Quality Analysis](#)

Video Clarity introduced ClearView to aid in Subjective Video Quality Analysis; while providing Objective Tools, which mimic the human visual system.

## Subjective Video Quality Analysis

A group of human subjects is invited to judge the quality of video sequences under defined conditions. Several recommendations are found in ITU-R BT.500.10 “Methodology for Subjective Assessment of the quality of Television Pictures” and ITU-T P.9210 “Subjective Video Quality Assessment methods for Multimedia Applications. The methods are summarized [here](#).

The main subjective quality methods are Degradation Category Rating (DCR), Pair Comparison (PC) and Absolute Category Rating (ACR). The human subjects are shown two sequences (original and processed) and are asked to assess the overall quality of the processed sequence with respect to the original (reference) sequence. The test is divided into multiple sessions and each session should not last more than 30 minutes. For every session, several dummy sequences are added, which are used to train the human subjects and are not included in the final score. The subjects score the processed video sequence on a scale (from 0 to 5 or 9) corresponding to their mental measure of the quality – this is termed Mean Observer Score (MOS).

Four serious drawbacks of this approach are:

1. The setup is difficult – rooms need to be secure, displays calibrated, etc.
2. Human subjects must be selected, screened and paid
3. It is difficult to assess why one video sequences was selected over another due to the Subjective nature of the test
4. Even though a wide variety of possible methods and test parameters can be considered, only a small fraction of the possible design decisions can be investigated due to the time consuming procedure.

All of this being said, Subjective Video Analysis is the only accurate assessment of Video Quality as seen by a group of Observers (aka: the customers).



## Objective Video Quality Analysis

Subjective Video Analysis is only applicable for development and evaluation purposes; it does not lend itself to operational monitoring, production line testing, trouble shooting, or equipment specific repeatable measurements. The need for Objective Video Quality Testing arose from the need for quantitative, repeatable video analysis.

Objective Metrics build models that describe the influences of several physical image characteristics on video quality, usually through a set of video attributes thought to determine video quality. When the influence of a set of design choices on physical video characteristics is known, then models can predict video quality. The models express video quality in terms of visible distortions, or artifacts introduced during the design process. Examples of typical distortions include flickering, blockiness, noisiness, or color shifts.

Three methods exist for Objective Video Quality Testing:

1. Full Reference: testing when the “original” and “processed” videos are both present
2. No Reference: testing when only the “processed” video is present
3. Reduced Reference: testing when information about the “original” and “processed” videos are present, but not the actual video sequences.

In general, Objective Metrics try to mimic the human visual system and return a score that indicates the MOS scores. Many metrics were measured against subjective analysis, and VQEG – [Video Quality Experts Group](#) - in conjunction with the ITU published the results as ITU COM 9-80-E. Some of the Objective Metrics are listed below:

1. PSNR – Peak Signal to Noise Ratio
2. JND – Just Noticeable Differences
3. SSIM – Structural SIMilarity
4. VQM – Video Quality Metric

Two serious drawbacks of this approach are:

1. These algorithms are measuring visible differences not video quality so if the “processed” video sequence is shifted up/down, the metric will show a difference, but the video quality is the same.
2. The best algorithms correlate to human subjects about 75% of the time.

The advantages of this approach are

1. It is repeatable
2. The Video Quality Testing tool price is the major cost
3. Small differences are detected anywhere within the video sequence
4. A quantitative score can be generated per frame or for a series of frames

## What is ClearView

Video Clarity's ClearView system offers a set of video quality analysis tools for software developers, hardware designers, QA/QC engineers, video researchers, and production/distribution facilities. ClearView plays, records, displays, and analyzes video sequences. The device can capture video content from virtually any source-file or digital or analog source such as SDI, HD-SDI, DVI, VGA, HDMI, component, composite, or S-video. Regardless of the input, ClearView can ingest and convert it to fully uncompressed 4:2:2 Y'CbCr, 4:4:4 RGB, ARGB, or RGBA. This allows CODECs to be compared.

The simplest use of the Video Clarity solutions is to compare two video sequences on the same display under identical conditions, and decide subjectively which one looks better. The videos can be displayed in split-screen, seamless-split, split-mirror (butterfly), or A-B (source minus result) modes where the split is either horizontal or vertical. Playback supports zoom, jog, shuttle, and pause for in-



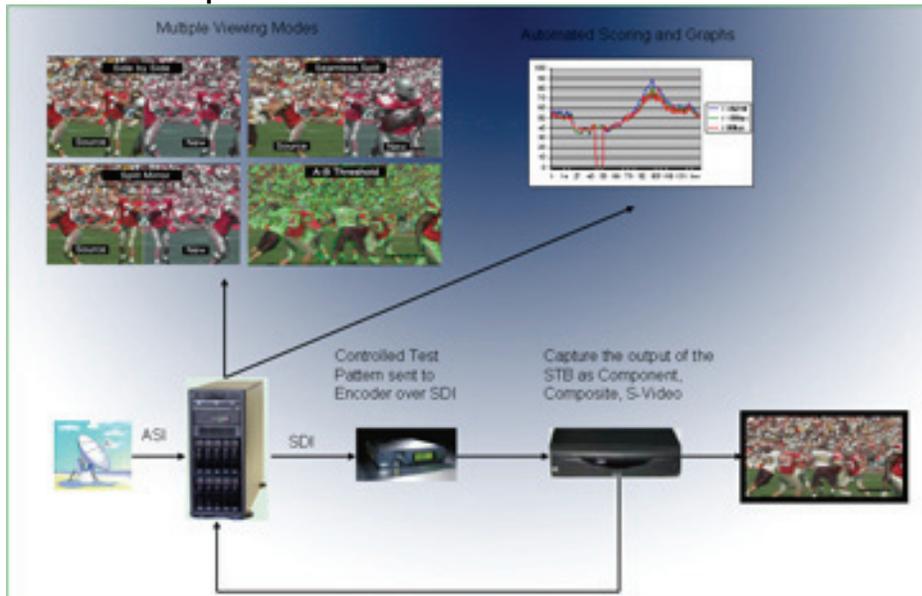
depth analysis. The device has a multi-clip playlist capability so combinations of video sequences can be played. Some of the display modes are shown below.

### Subjective Analysis Display Modes (Vertical Split)



To simplify the work flow, any video sequence can be played while capturing another video sequence, thus, combining the video server, capture device, viewer and video analyzer into one unit. By doing this ClearView controls the test environment, which allows for automated, repeatable, quantitative video quality measurements.

### Work Flow Setup



To maintain quality as compression standards change, everyone along the chain needs to constantly measure the performance of the compression engines on which day-to-day operations depend. ClearView includes Full Reference, No Reference, and Reduced Reference video metrics.



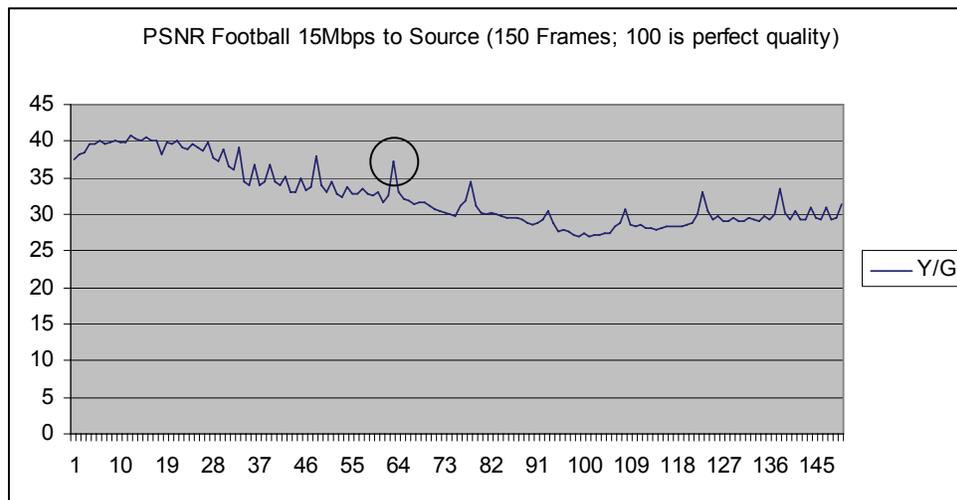
The Full Reference Objective Metrics compare the “original” to the “processed” video and include:

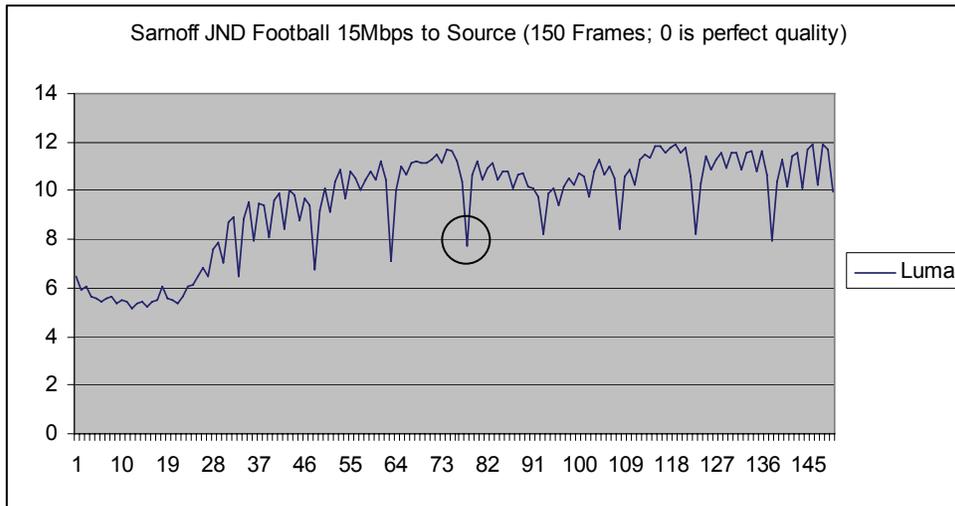
1. [PSNR](#)
2. [Sarnoff JND](#)

The Video Clarity solution compares two video sequences, temporal & spatially aligns them, normalizes for color & brightness shifts, and measures the difference against a threshold generating a pass/fail condition. For example, this tool would be used in QA to run tests based on an encoder/decoder, which you had established as your standard based on its subjective and objective scores "Gold", and compare other encoders/decoders "Device Under Test".

As an example, an uncompressed, 1080i football sequence is compared to a 15Mbps, MPEG-2 processed sequence. The video sequence includes high motion and crowd pan. The PSNR and Sarnoff JND scores are listed below and plotted per frame. The average JND score is 9.5 (scale 0 to 100; where 0 is perfect) and the average PSNR score is 32.6 (scale 0 to 100; where 100 is perfect). Generally, these scores are both good.

It can be seen that the video scored well when the scenes were easy to compress, and progressively got worse as expected based on the complexity of the content. The spikes in the score are I-frames. ClearView computes a score for each component and frame separately, and then produces an average per component over a defined time period. The PSNR and Sarnoff JND scores are shown below. ClearView generates a pass/fail based on a threshold and allows the viewer to display any video frame for their own subjective analysis.





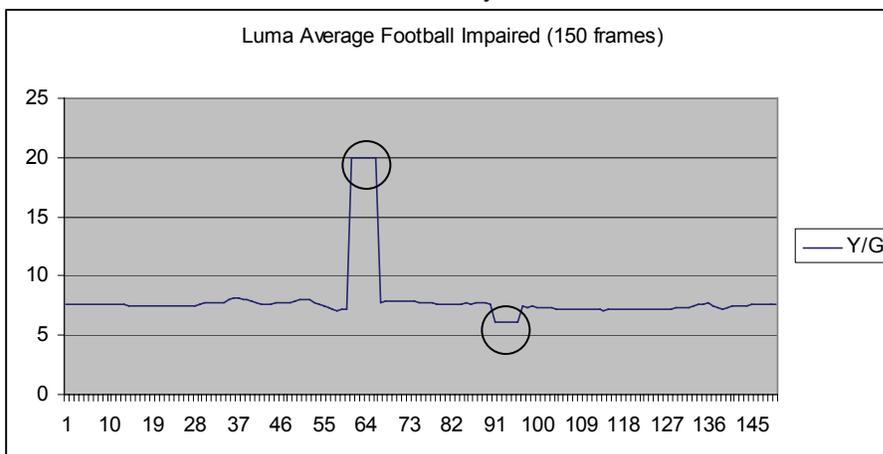
Many times the “original” material is not present. In this case, the model measures the amount of overall disturbance based on a combination of signals: blockiness, temporal distortion, luma and color levels. This type of model is termed No Reference Objective Metrics. It estimates the visible distortions directly from the “processed” video; instead of comparing it to the “original”. ClearView includes:

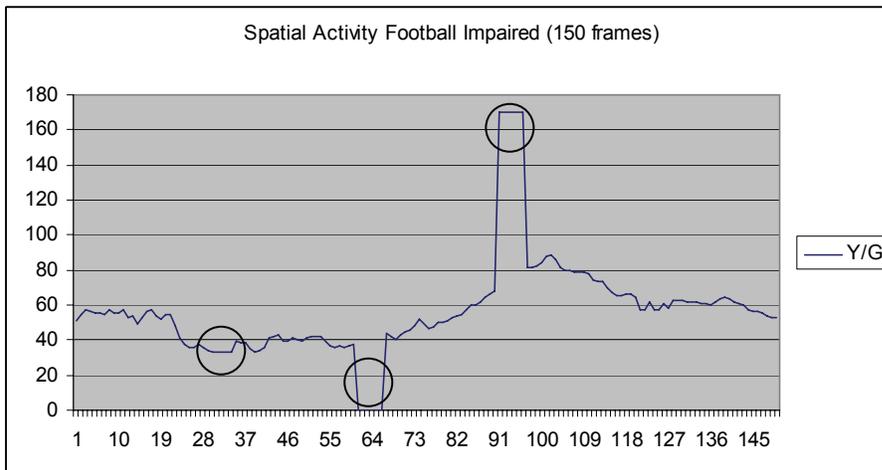
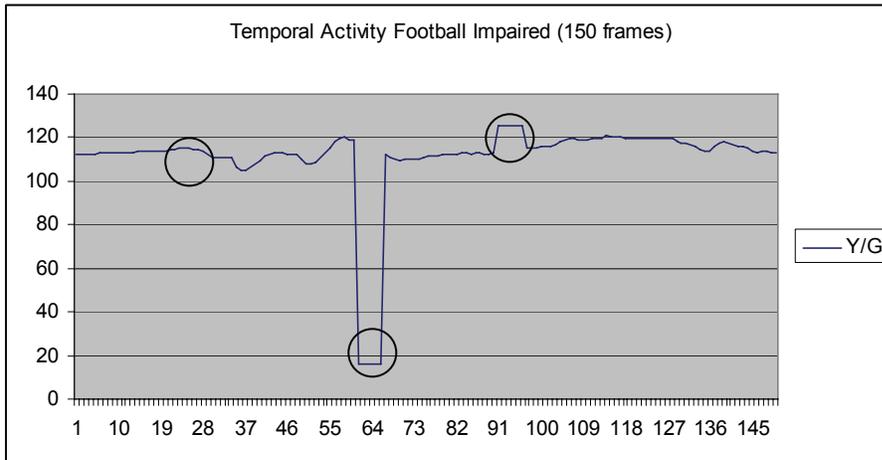
1. [Luminance Average \(PSNR to Black\)](#)
2. [Chrominance Average \(PSNR to Black\)](#)
3. [Spatial Activity](#)
4. [Temporal](#)

In this example, the uncompressed, 1080i football sequence was altered to create anomalies as described below:

1. After 1 second, 5 frozen frames were inserted
2. After 2 seconds, 6 frames of black were inserted
3. After 3 seconds, 6 frames of unrelated video were inserted
4. After 4 seconds, 5 frames of video were degraded (compressed)

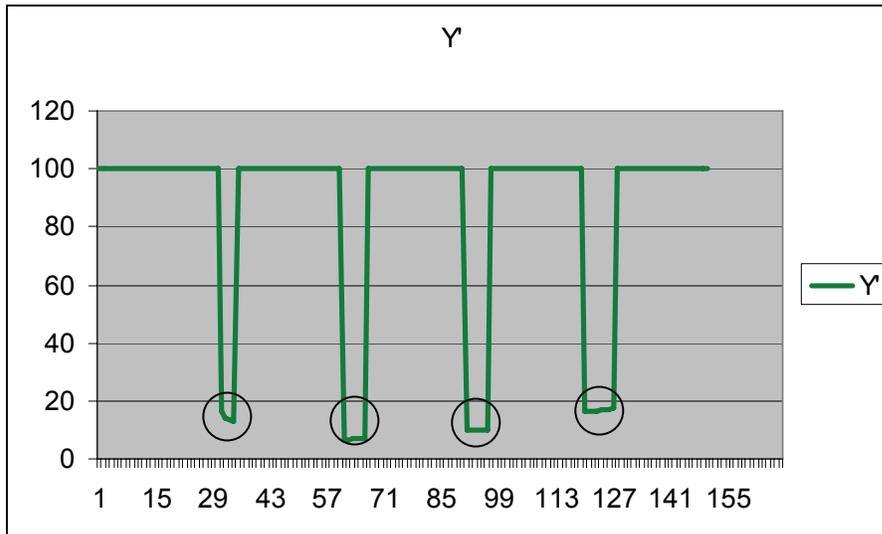
With this video sequence, the average luminance and chrominance numbers show a disturbance for anomalies #2 and #3. The Spatial and Temporal show a flat line for anomalies #1, #2, and #3. None of the metrics were able to detect anomaly #4.





From the earlier definition, if a No Reference Objective Metric on the “original” video and a No Reference Objective Metric of the “processed” video are compared, this is termed Reduced Reference Objective Video Quality Analysis. This type of testing helps when the “original” is present somewhere in the network, but cannot be streamed for Full Reference analysis.

Using the same impaired video in the second example and the uncompressed video sequence from the first example, a correlation of their No Reference Objective Metric scores generated the following information. All four anomalies are detected.



## Conclusion

Any operation, large or small, from hardware/software designers to end user-producers who work with any kind of compression in standard or high definition, should subjectively and objectively analyze compression performance to evaluate how their technical and aesthetic choices will look to end-users.

Video Clarity provides a system-level solution – ClearView – both for Subjective Viewing and for Objective Scoring. The simplest use of the Video Clarity solutions is to compare two video sequences side-by-side on the same display under identical conditions, and let the viewer decide which one looks better. However, this is all too dependent on the viewer; that is, it is highly subjective and inherently unrepeatable. ClearView includes many Objective Metrics to mimic the human visual system. These metrics return an automated, repeatable, and inherently objective pass/fail score, which can be correlated to any one frame of video or series of frames.

The Objective Metrics can be applied to compare two video sequences, or to evaluate a single sequence when a reference sequence is not available. In either case, the objective score is a strong predictor of the subjective score that you would get when averaged of a large number of viewers.

## The Author

Bill Reckwerdt has been involved in digital video since the early 90's from digital compression video on demand, to streaming servers. He received his MS specializing in Behavioral Modeling and Design Automation from the University of Illinois Urbana-Champaign. He is the VP of Marketing and the CTO for Video Clarity, which makes quantitative, repeatable video quality testing tools - <http://www.videoclarity.com>.