Measuring and Interpreting Picture Quality in MPEG Compressed Video Content

▶ A New Generation of Measurement Tools

Designers, equipment manufacturers, and evaluators need to apply objective picture quality methods to the equipment they are working with. The Tektronix PQA300 Picture Quality Analysis system provides fast, repeatable, quantified Picture Quality Ratings (PQR) on compressed video content.

Introduction

When dealing with technical products such as video transmission equipment, every design or evaluation project relies on accurate, repeatable measurements. Until recently, waveform and vector measurements on baseband video signals provided ample information about the signals' behavior, including its visible picture quality. Explicit parameters equate to color, brightness, contrast, and so on.

The advent of compressed digital video transmission has complicated the process of evaluating video signals, particularly with respect to their perceived picture quality. During MPEG compression a certain amount of the original content is knowingly discarded. Visible impairments such as "blockiness" are an inevitable by-product of the process. Other problems such as Gaussian noise and impulse noise (as seen in some satellite transmissions) can also affect quality. Traditional measurement techniques are no longer effective, since compression-related impairments are dependent on the video content.

The solution heretofore has been subjective picture quality tests in which human subjects viewed and judged video images. In spite of good-faith efforts to control these tests, the very nature of subjective judgement causes inconsistencies in results varying from lab-to-lab, test-to-test, and viewer-to-viewer. Furthermore, subjective tests are expensive, very time-consuming, and not very repeatable.

Fortunately a new generation of measurement tools has arrived to help designers, equipment manufacturers, and evaluators apply objective picture quality methods to the equipment they are working with. The Tektronix PQA300 Picture Quality Analysis system provides fast, repeatable, quantified Picture Quality Ratings (PQR) on compressed digital video content. Its results are expressed in numerical PQR values that show excellent correlation with human perceptual results based on subjective picture quality tests.

This application note will explain the PQR scale, and will explain how a Picture Quality Analysis system can be used to evaluate transmission system components.

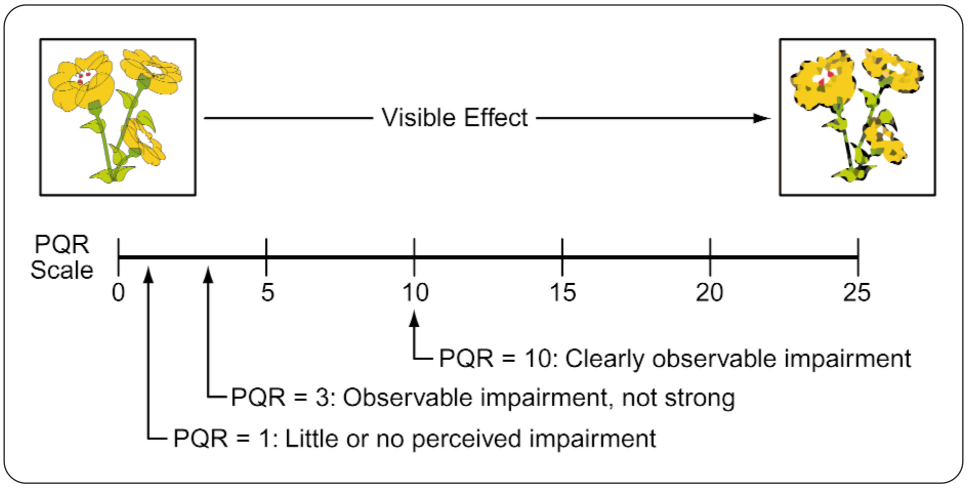


Figure 1. *The Picture Quality Rating (PQR) Scale.*

From Subjective Viewing...

For years, an International Telecommunications Union document known as ITU-R BT500 has defined standards for subjective picture quality measurement. Although its origins preceded the widespread use of video compression, the standard has evolved to include two methodologies for evaluating video impairments:

- **The double-stimulus impairment scale (EBU) method**. Test subjects view a reference image, then three seconds of blank mid-gray, then the test image. They rate the difference on a scale of 1 (Very Annoying) to 5 (Imperceptible)
- **The double-stimulus continuous quality scale method (DSCQS).** Test subjects view test and reference images, then rate both on a continuous quality scale ranging from 0 (Bad) to 100 (Excellent). Viewers do not know which image is the test and which is the reference.

There are a number of standardized test sequences with varying amounts of motion and color. While ITU-R BT500 viewer tests offer a reassuring connection with real audiences, results are subject to variations such as the viewers' distance from the screen, adherence to instructions, and individual visual acuity.

...to Objective Measurements

The Human Vision Model and the Picture Quality Rating (PQR) scale were developed to provide fast, repeatable objective picture quality measurements. The PQA300 system uses the Human Vision Model as the basis for its evaluation of the differences between every pixel of every frame of two signals – source and processed content – in a video sequence. It then computes an overall PQR value for the sequence. The PQA300 produces a video picture quality evaluation that can be used to guide development and selection of equipment for content production and transmission.

The result is a quantified measurement that tends to remain constant over many repetitions for any specific sequence, bit rate, and codec (assuming the codec itself behaves similarly over multiple runs). The standard "Football" sequence processed through a video encoder at 7.0 Mbit/sec receives the same PQR value today as it did yesterday. It is important to note here that the codec may produce differing PQR values for different test sequences, though.

What affects PQR? Codecs comply with certain standards, of course, but individual implementations vary among manufacturers. One vendor's codec may handle motion very well, but not color. Another may handle a landscape well, but not a busy sporting event. This is why it's necessary to evaluate codecs and equipment with a series of test sequences. Used in combination, the most common set of standard test sequences applies a mix of stresses to the codec.

Objective picture quality methods are growing in importance. Not only are the measurements less costly and easier to carry out than subjective trials, they are the quantified result of a dispassionate "viewing." Quality of Service guarantees are becoming a cornerstone of Service Level Agreements (SLA) among producers, carriers, and distributors of video content. The PQA300 delivers a concrete number that relates closely to overall QoS. Its controlled, quantifiable measurement regime is the best way to track compliance with SLA terms.

Interpreting PQR

PQR values range from 0 (a perfect transmission) to a grossly distorted 25. The continuum of PQR values is shown in Figure 1. In general, ratings of 10 or more indicate annoying compression artifacts or other impairments. Perhaps most importantly, PQR results correlate well with subjective tests on the same material. That is, a sequence that is highly rated in human viewing tests will also achieve a favorable (low) PQR value. Interestingly, the PQR rating is actually more "accurate" than any human eye because it can detect picture impairments that are below the visible threshold. This effect is taken into account when the overall PQR result is calculated.

Figure 2 depicts a sample set of results from a test comparing DSCQS subjective scores with PQR values for the same material. The graph shows the results from two different coders. The "Barcelona" test sequence is a colorful parade with low motion and fine detail. The "Flower" sequence is a flower garden viewed from a moving vehicle. Like Barcelona, it offers a high-color, low-motion challenge to the coders. Other tests (not shown on this graph) depict cartoon figures, soccer (football) games, and a radio announcer standing in front of a stone wall. Each sequence stresses the codec in a different way. In the graph, five different bit rates are represented: 2.0, 3.0, 4.5, 7.0, and 10.0 Mb/s.

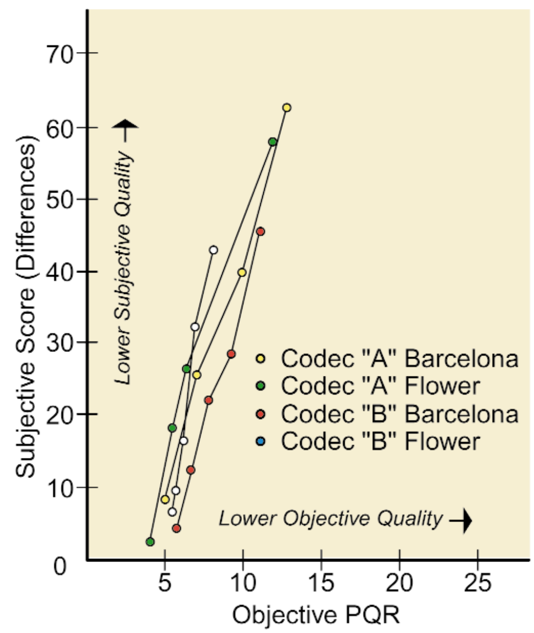


Figure 2. *A comparison of subjective and objective picture quality measurements. In this test, the two methods showed a high level of correlation.*

Remember, in the DSCQS test two images are separately viewed and rated on a scale of 0-100. A greater numerical difference indicates poorer picture quality. In Figure 2, the increasing difference values in the subjective tests are consistent with rising PQR values.

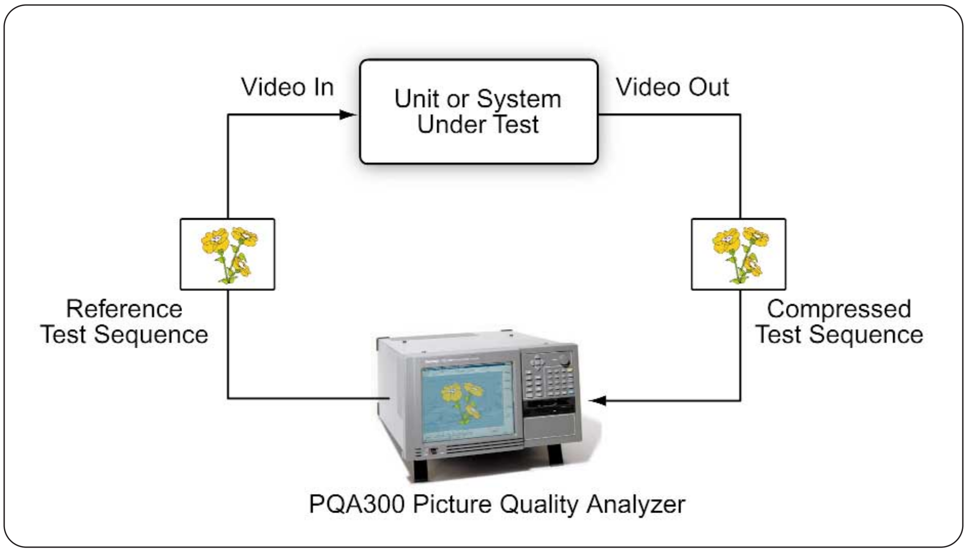
Measuring PQR: Test Setups and System Configurations

The PQA300 is an uncompromised reference-based measurement system. Its internally-stored test sequences can act as the stimulus source for the system under test (SUT). That same content is used as the comparison reference for the compressed video content acquired from the SUT. Figure 3 shows a basic system connection.

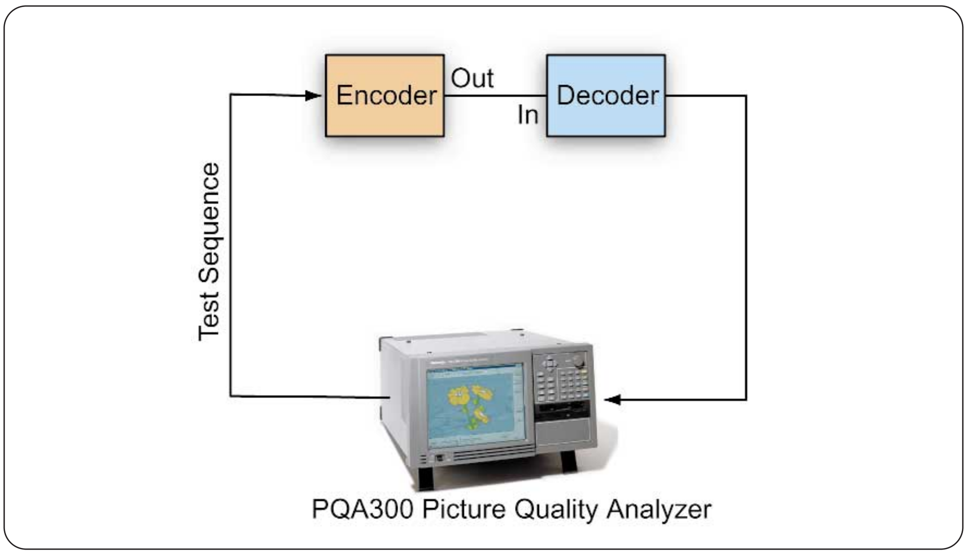
This is a simple, effective configuration for PQR measurements on new component and equipment designs. The obtained PQR is a reliable predictor of the design's performance in a system or network context.

The basic PQR measurement also has its place in everyday network operations and in installation and maintenance (I&M) work. It can be used to verify new installations, to evaluate bandwidth allocation schemes, and to set the most cost-effective compression levels for all types of content.

The "Unit or System Under Test" block in Figure 3 symbolizes a wide range of hardware, software, networks, and transmission elements, of course. Often it is desirable to evaluate the combined effect of two particular elements, such as an encoder and a decoder, that are intended to transport the same compressed signals. By pairing the two elements and isolating them from other parts of the transmission chain (Figure 4), it is possible to determine a PQR value that reflects the composite impact of encoding and decoding, without the additional effect of network transmission.



▶ ***Figure 3**. Basic picture quality measurement configuration.*



▶ ***Figure** 4. Evaluating the combined effects of the encoder and decoder only.*

The configuration in Figure 4 is also the basis for design evaluation of codecs. By keeping one element, say, the decoder, constant you can substitute a series of coder implementations, feed the test sequence through the pair, and compare the PQR values that result. This will aid the refinement of encoding algorithms, ensuring that the end-product delivers high-quality video through the system. Similarly, you can use a known encoder and substitute a selection of decoders to evaluate decoder performance.

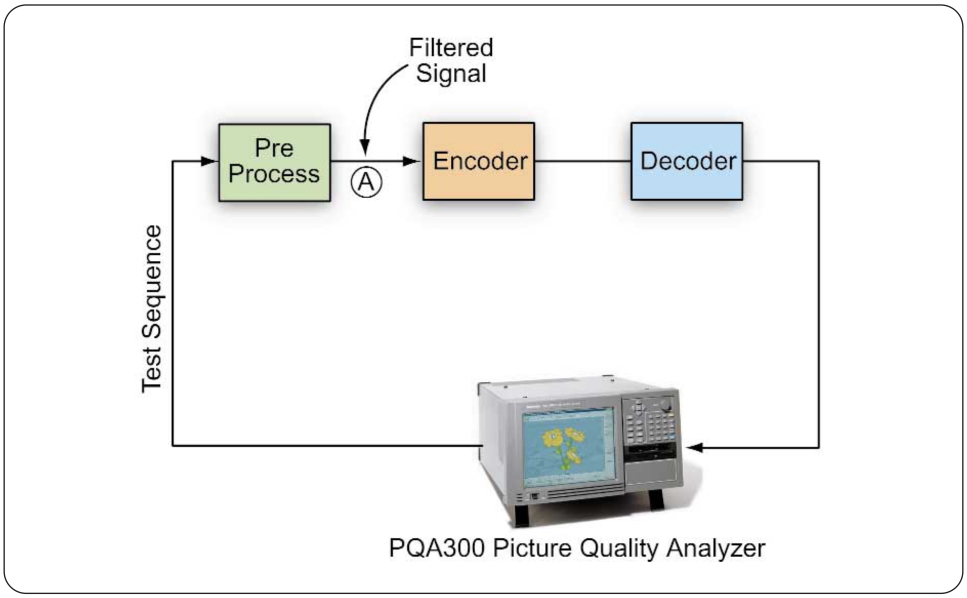


Figure 5. *Pre-processing improves visible picture quality but may degrade PQR.*

Pre-Processing Effects

As explained earlier, video compression knowingly discards some of the content – hopefully, only that which will not create perceptible differences in the processed video. Coders do their best to handle motion, color, and detail but any coder's operation has some unwanted side-effects. Essentially, the more "work" the coder has to do, the greater the side effects are likely to be. Therefore, one way to reduce compression artifacts is to decrease the coder's workload.

That means reducing the amount of information the coder has to evaluate and respond to. One proven way to accomplish this is to pre-process the video content before it goes into the encoder, as shown in Figure 5. The pre-processor may be a separate external component, or it may be integrated into the encoder section. The action of the pre-processor is similar to filtering.

Importantly, this added step has its own effect on PQR measurements. The filtering tends to alter the video content enough to drive PQR values higher (worse), even though the viewed image quality may appear to be better (more defect-free). In this case the image may look "smoothed" and less sharp. The pre-processing trades off some sharpness and clarity with the goal of reducing more objectionable MPEG artifacts.

The pre-processor changes the image, which is bound to change the PQR. The material going into the encoder is a modified copy of the reference test sequence, and the PQA300 sees these modifications as "errors." Particularly if it is integrated with the encoder, it is difficult to determine the pre-processor's effect on the content.

If, however, there is a means of isolating test point "A," then it is possible to estimate the impact of pre-processing. By analyzing the preprocessor separately (connecting test point "A" directly to the PQA300's input), you can get a general idea of the component's effect.

Unfortunately, the PQR result at point "A" cannot be factored directly into the overall system PQR. Because all of these processes – filtering, encoding, and decoding – are non-linear, mathematically combining the individual components will not produce a meaningful result. But the test can produce a measurement of the relative effect of differing preprocessor implementations.

Summary

Objective picture quality measurements are becoming an increasingly important part of equipment design and evaluation, bandwidth allocation, and QoS compliance programs. The PQR scale has achieved wide acceptance as a means of interpreting picture quality. The Tektronix PQA300 Picture Quality Analyzer delivers repeatable, quantified PQR values, and can be connected to make measurements across complete transmission systems or individual components. With tools such as the PQA300, a clear understanding of test setups, and familiarity with the PQR scale and its implications, it is possible to achieve consistent improvements in compressed video transmission.

Contact Tektronix:

ASEAN Countries (65) 356-3900

Australia & New Zealand 61 (2) 9888-0100

Austria, Central Eastern Europe,
Greece, Turkey, Malta & Cyprus +43 2236 8092 0

Belgium +32 (2) 715 89 70

Brazil and South America 55 (11) 3741-8360

Canada 1 (800) 661-5625

Denmark +45 (44) 850 700

Finland +358 (9) 4783 400

France & North Africa +33 1 69 86 81 81

Germany + 49 (221) 94 77 400

Hong Kong (852) 2585-6688

India (91) 80-2275577

Italy +39 (2) 25086 501

Japan (Sony/Tektronix Corporation) 81 (3) 3448-3111

Mexico, Central America, & Caribbean 52 (5) 666-6333

The Netherlands +31 23 56 95555

Norway +47 22 07 07 00

People's Republic of China 86 (10) 6235 1230

Republic of Korea 82 (2) 528-5299

South Africa (27 11) 254 8360

Spain & Portugal +34 91 372 6000

Sweden +46 8 477 65 00

Switzerland +41 (41) 729 36 40

Taiwan 886 (2) 2722-9622

United Kingdom & Eire +44 (0)1344 392000

USA 1 (800) 426-2200

For other areas, contact: Tektronix, Inc. at 1 (503) 627-1924

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 "Resources For You" at **www.tektronix.com**

ISO 9001

0401 TD/XBS 25W-14675-0

Tektronix