# METHODS FOR EVALUATION OF DIGITAL TELEVISION PICTURE QUALITY

William Y. Zou
Public Broadcasting Service
Alexandria, VA

Philip J. Corriveau
Communications Research Centre
Ottawa, Ontario

## Abstract

With the emergence and proliferation of digital video compression in television systems, there is an increasing requirement for effective methods to evaluate picture quality. In this paper, we describe traditional and novel objective and subjective test methods. We review standard methods to subjectively evaluate picture quality, outline their limitations and suggest ways to change these subjective methods. In addition, we also discuss research into objective methods that employs perception-based models with easily measurable parameters that could give a high degree of correlation with subjective evaluation methods. Finally, we draw some conclusions about the types of research needed to further the development of methods for picture quality evaluation.

## Introduction

Improvements in technologies, particularly the use of digital compression techniques, the concatenation of diverse transmission media and a changing viewer environment, are demanding that broadcasters ensure that the quality of service is adequate and meets expectations. One problem the industry faces today is how to evaluate a new generation of artifacts.

Digital television systems introduce unique artifacts which are visibly different from those in existing analog television systems. In order to study the effect of these new artifacts on television picture quality, new approaches and methodologies may need to be developed. This paper will discuss the objective and subjective evaluation of artifacts and the problems digital television systems introduce. We will discuss a link between objective and subjective evaluation techniques. Finally, we will present some conclusions with respect to the types of research needed into methods of evaluating picture quality.

## Picture Quality Assessment

There are two ways to evaluate the quality of a video signal. The first is by using objective measurements, while the second involves subjective assessments. Recently, a novel approach of combining objective and subjective methods has emerged. In the following sections, we will examine both traditional and novel methods for picture quality assessment.

### Traditional Objective Method

Traditional objective measurement of television signal quality is performed by measuring physical parameters such as signal amplitude, timing and signal-to-noise ratio. To characterize the linear and non-linear signal distortions from signal transmission and processing, specially-designed static test signals are inserted into the video and analyzed at reception. The advantage of using static test signals is that they yield numerical values which have been correlated with subjective picture quality. The standard test signals and measurement procedures are defined in the standards such as NTC-7 and EIA RS-250.

With the introduction of digital signal processing into television, some of the traditional test signals are unable to accurately predict the signal distortions. Therefore new test waveforms have been developed specifically to characterize digital video signals. For example, to explore quantization noise a ramp signal is used instead of the staircase signal to measure the non-linear distortions. The situation is even more complex with the use of digital video compression where picture quality impairments are dependent on picture content. Therefore, traditional methods for signal quality measurement are unable to measure the perceptual picture quality; they can only measure the analog front and back ends of the system.[1]

### Perceptual Coding in Video Compression

The key principles behind video compression are (a) the removal of redundancies (spatial and temporal), and (b) the elimination of irrelevant information, i.e., information that is not visible to the human visual system (HVS). Perceptual coding, which exploits the properties of the HVS, has been employed in various video compression standards such as MPEG-2. Typically, this is done in the video pre-processing and quantization stages. In pre-processing, the source signal is filtered and sampled to reduce the information which is not visible to the HVS. This includes chroma conversion, chroma subsampling, and signal transformation. In the quantization stage, the bit allocation process allocates bits according to the perceptual importance of the various coefficients, therefore shaping the quantization noise into the perceptually less important areas. This can be accomplished by varying the relative step-size of the quantizers for the different coefficients. For example, low-spatial frequency coefficients must be quantized finely, while the less important high-frequency coefficients may be quantized more coarsely.

## Impact of Using Compression on Objective Quality Measurement

The use of digital video compression has a direct impact on objective signal quality measurement. Waveform measurements using the static test-signals fail with compressed video. First, the objective of designing a compression algorithm is not to replicate the original signal waveform as closely as possible, but to yield a perceptually equivalent approximation of the picture. Secondly, because of the constraint of bandwidth/bit rate, the resulting compressed picture quality is highly content dependent. In other words, if stressful source material (in terms of spatial detail and motion) is used, artifacts would be more visible and subjective picture quality would be degraded. In addition, the subjective picture quality is not a linear function of compression ratio or bit rate. Furthermore, unique digital transmission artifacts such as blocking, error strips and freeze frames make assessment of picture quality more difficult for digital systems than for analog systems. In addition, the time-varying nature of these impairments further complicates quality assessment of digital systems. For example, digital artifacts may be short-lived and the quality of a digital transmission may fluctuate more than that of an analog transmission.

## Designing New Objective Measurement

In past, new methods were proposed to remedy the inadequacies of traditional objective signal quality measurement for assessing digital television signals. They can be grouped into three categories: 1) using synthetic dynamic test signals and analyzing them to measure the dynamic performance of compressed video, 2) performing distortion measurement to assess how well the compressed video replicates the original, 3) using natural video test scenes and analyzing a set of features that correlate well with subjective picture quality. The first approach is an extension of the traditional objective test methods that use signal waveforms. The second one is based on fidelity comparison. The third approach, called perception-based objective measurement, to be discussed later in the paper, aims to bridge objective and subjective quality measurement and leads to a new direction for compressed picture quality assessment.

### Dynamic Test Signal Measurement

To assess the dynamic performance of video compression systems, time-variant test signals have been designed. Examples are test signals to measure edge busyness, slope overload, dynamic nonlinearity, temporal response and dynamic response measurements.[2] To examine the loss of picture quality due to compression, techniques such as measuring step-response rise time for image blur, step-response jitter for edge busyness, peak-to-peak signal to minimum quantizing error for rise time of a moving edge for temporary image blur, and maximum noise amplitude for a "dirty window" effect were proposed.[3] Using newly developed multi-dimensional test signals such as zone plates, measurements of resolution,[4] zone-plate loading,[5] noise loading[6] and dynamic noise[1] were also proposed. The industry has not yet widely accepted or used these proposed methods. The missing ingredient in these methods is that the HVS is not incorporated; there is no link between the measured objective values and the related subjective picture quality.

## Distortion Measurement

In image and video coding, the root-mean-squared error (RMSE) and peak signal-to-noise ratio (PSNR) are widely used as a measure of distortion. These metrics are popular largely because of their analytical tractability. However, a global measures such as these cannot predict the perceptibility of distortions. Spatially localized and time-varying distortion is far too complex to be summarized by a global distortion measure.[7]

Bit error rate (BER) has long been used for data communications. For digital transmission, BER measurement is effective to assess the integrity of the transmission system. In digital television systems, transmission errors manifest themselves as freeze frame or block distortions. The relationship between error rate and subjective picture quality is further complicated by the error concealment scheme at the receiver. Like RMSE, error rate does not accurately predict the perceptual impact of transmission errors.

## Subjective Quality Assessment

Subjective assessment of picture quality attempts to quantify the response of the HVS to selected images. This is done by having viewers rate the subjective quality of images using pen and paper tests or computerized tests. Recommendation 500[8] of the International Telecommunications Union, Radiocommunications Section (ITU-R), defines guidelines and procedures for such tests.

The recommended testing procedure for subjective quality assessment is the double-stimulus continuous-quality scale (DSCQS) method. In this method, a 10 cm graphical scale is divided into five equal intervals. In the middle of each interval the following quality terms are associated from top to bottom (Excellent 100-80, Good 79-60, Fair 59-40, Poor 39-20 and Bad 19-0). The 10 cm scale provides a "continuous" rating system to help reduce quantization errors in viewer responses. Selected test sequences are presented twice in pairs sequentially in time to allow for first viewing and then evaluating. Sequence duration is usually 10 seconds although variation from 5 to 15 seconds is not uncommon. Viewers are instructed to assess the quality of both the reference and test picture, though they are unaware of the order of presentation for each trial. A maximum of 30 minutes of constant viewing is suggested in order to avoid having the effects of fatigue and boredom introduced to the results. This does not mean that a total test has to be limited to 30 minutes. If a design calls for more than 30 minutes of viewing, there should be a break between sittings.

Recommendation 500 provides detailed specifications for viewing distance, display parameters and viewing environment. But it leaves certain design factors unspecified, and it is the experimenter's responsibility to choose the levels of these factors. The two most important factors that are left up to the experimenter’s discretion are selection of video sequences and viewer characteristics (expert, non-expert, age, etc.). Clearly, these factors cannot be set in advance and are unique to the requirements of each evaluation.

It is well known that the quality of the output image from a CODEC is significantly influenced by the criticality of the video sequence. Thus, the selection of appropriate image sequences is an important process. One of the first considerations is that material be chosen from several sources.

For example, material should be drawn from film, video and computer graphics. This is important because film material tends to be less stressful on a CODEC than a video. Both moving and still sequences should be used, with some motion sequences containing scene cuts and fades in order to assess CODEC performance. The practice of taking a subset of ITU-R accepted sequences and a set of sequences representing regular programming has proven to be useful.

Generally, non-expert viewers are used to evaluate image quality. Non-expert viewers are people who have no prior professional or extensive personal experience in dealing with video display systems or devices. These viewers are recruited through a reliable source and screened for normal or corrected to normal visual acuity, normal contrast sensitivity and normal color vision.

Below, we review a variety of improvements that have been proposed to the methods described in Recommendation 500. Suggestions for changing the design of the rating scale, duration of sequences and overall viewing environment have been made.

## Quality Rating Scale

The Double Stimulus Continuous Quality Scale is a 10 cm line that corresponds to numerical ratings between 0 and 100. Viewers evaluate the subjective quality of images using scales with the assistance of associated word indicators. These indicators (Excellent, Good, Fair, Poor and Bad) currently break the continuous scale into five equal 20 mm sections. Some researchers state that interpretation of the descriptors, may lead viewers to divide the scale into psychologically unequal sections. For example some may rate Good and Excellent closer in perceived quality than Good and Fair. Viewers have in the past been asked to write the adjectives beside a numerical scale but it was impossible to conclude whether the adjectives were truly equidistant. To complicate the issue further the concern of language translation becomes a factor. Extreme effort was taken by the ITU-R to select adjectives that are easily translated from English but other researchers such as Kees Teunissen suggest alternate scale design. Kees Teunissen suggests a 10 term scale rather then the current 5 term scale.[10] This concept is illustrated in Figure 1, where both the current and proposed scales are sketched.

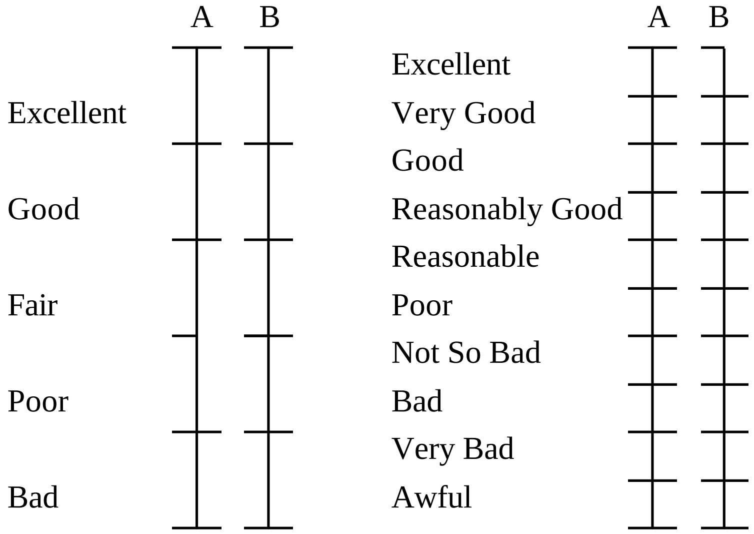

Figure 1 Proposed Changes to ITU-R DSCQS

However, the addition of more descriptors to the current scale creates the possibility of increased confusion. Doubling the number of adjectives associated with the rating scale increases the complexity of the judgment task. For example, non-expert viewers may tend to become distracted from the actual task of evaluating picture quality if there are too many categories to distinguish between. Research should be performed to fully evaluate the implication of changing the current scale and to ensure that any scale used achieves the intended task.

The simplicity of the quality scale lies in the fact that viewers are asked to evaluate quality for both a reference and test sequence. This provides the experimenter with raw mean scores for both reference and test sequences. For example, the rated quality of the reference could be 85 out of 100 and the quality of the test could be 65 out of 100. From this it can be determined that the difference is -20 (Test Score - Reference Score). When interpreting quality data, this difference measure is one of the more interesting facets. The distance between stimuli is a very useful tool when determining the quality of a system and some other methods do not allow for this comparison. Rather than asking the viewer to determine the difference between the reference and test, the easier task of evaluating individual quality is requested. The difference data is then extracted using the above described method. This allows the experimenter to first examine the stability of scores given to reference sequences, and then extract difference data. However, one study questioned the ability of the continuous scale method employed in quality tests to yield difference information.[11]

## Test Sequence Duration

Current testing methods use 10 second sequence segments. Viewers are asked to evaluate the overall quality of each 10 second segment. Presumably each viewer attempts to average their impression of the quality over the duration of the sequence. However, digital artifacts that are short-lived and temporally spaced may not be captured by a 10 second segment. It might be more appropriate to increase the duration of test sequences to properly evaluate these time-varying artifacts. An experiment was performed to evaluate an increased duration of the image sequences to 30 seconds. It was found that subjects were strongly influenced by what they saw in the last section of the sequence (a recency effect). For example, if a sequence had an artifact in the first 15 seconds and then recovered to be a better quality, a subject would assign a high score. On the other hand, if a visible artifact occurred in the latter part of the sequence, the viewer would lower the rated value. It was observed that subjects formed a weighted temporal average of the instantaneous quality variations during the sequence, giving maximum weighting to the part of the sequence seen most recently.[12] This effect can lead to a difference in rated quality of about 10 points on a scale of 100, or half a grade on a scale of 5. It is clear from this result that viewers form an averaged value of perceived quality.

The problem of averaged picture quality ratings could be overcome by increasing the sequence duration and sampling viewers more frequently. This principle was applied to a study where the sequence duration was increased and a sliding scale was used for continuous assessment of picture quality. The mechanical scale sampled the position of the slider every second. This experiment demonstrated that it was possible to measure the quality in a continuous and reliable manner.[13] The notion of continuous rating over a longer duration has lead to a proposal by T. Alpert and J.P. Evain called the Single Stimulus Continuous Quality Evaluation Method (SSCQE).[14] The SSCQE has been designed to allow for the evaluation of picture and service quality over a wide range of areas. This method allows for subjective evaluations in conditions near to the ‘home’ environment. Assessors are asked to evaluate the quality of long sequences (20-30 minutes) with content representative of the service being provided. The votes are collected continuously (twice per second) from sliding devices that each assessor can manipulate during the entire test session. This allows for the interpretation of the entire program stream and for a better evaluation of time-varying impairments.

With these new methods of evaluation, the number of sampled points gathered increases by a factor of more than 20. Of course, the reporting and analysis of this additional data causes its own difficulties. However, the ability to analyze sectional data will be gained. In theory, these methods will allow for a better understanding of time-varying impairments and a true sense of overall subjective quality for longer program material.

## The Viewing Environment

In order to ensure the reliability and repeatability of test results, viewing conditions are carefully described in ITU-R Recommendation 500. However, some researchers[15-17] feel that these conditions do not reflect those of a typical ‘home’. For example, factors such as viewing distance and display size, ambient lighting and display parameters may differ in the ‘home’ than those specified in Recommendation 500. Deviations between the 'home' environment and Recommendation 500 are not a serious problem, provided that the condition of the viewing environment are at least as critical as the ‘home’ environment. In this situation tests conducted under Recommendation 500 will always guarantee a satisfactory performance in the ‘home’.

### Viewing Distances and Display Size

Currently, the recommended viewing distance is 3H (3 times the picture height) for HDTV and 5H for all SDTV systems (NTSC, PAL and SECAM). It is known that the seat position that viewers chose can vary from the distance specified in the recommendation. One study found that by using larger HDTV displays, there is increased convergence of the desired (currently used) and the preferred (chosen by viewers) viewing distances. Essentially, the finding is that with increased monitor size, the viewers are not likely to move farther from the display, but to remain in the same seating position. This could be attributed to the fact that with increased resolution, the line structure is less visible. Thus a viewer can sit closer to the set even at 3H without perceiving the line structure. Based on this work, a display size and distance of 3H could be standardized for laboratory use that would keep both the preferred and desired distances relatively equal, and ensure critical viewing conditions.

### Ambient Lighting

For most viewing conditions the ambient lighting in the ‘home’ tends to be quite low. For example, to fully enjoy the effects of a movie, most viewers will reduce the amount of light in the room. The light level specified in Recommendation 500 is quite low and probably very close to that of the average 'home'. Ambient lighting conditions provide an environment in which a viewer's attention is not drawn away from the display. This allows for viewers to evaluate picture quality without being distracted by other stimuli.

### Display Parameters

Luminance, contrast and resolution interact in varying proportions to result in a perceived level of image quality. With changes in monitor technology, such as increased peak luminance, the current Recommendation 500 specifications may need to be revised to higher values of these parameters.

## Correlation Between Objective and Subjective Methods

Subjective quality assessment plays such an important role in both analog and digital television environments. However, subjective evaluation is both costly and time-consuming which has motivated researchers to explore objective measurements that are correlated with the perceptual characteristics of the HVS. This goal leads to the development of perceptual models.

With the advances in the knowledge of the HVS, new methods for making correlation between objective and subjective measures are emerging.  These new methods rely heavily on perception-based modeling.

### Perception-based Objective Measurement

In theory, perception-based objective picture quality measurement can quantify the subjective quality of a digitally compressed picture without having human subjects rate the picture. Perception-based methods compare a pair of images/sequences, usually natural video scenes, to predict the visibility of the difference similar to subjective method. A reference measurement is made on the input (source) signal and an identical measure is made on the output of a system under test. The parameters (feature statistics and comparison) are defined as combinations of an input measurement with a corresponding output measurement according to explicit formulas. The selection of parameters is based on the HVS. Such formulas produce results that effectively quantify the comparability of the input and output measures. In general, analysis can include measurements that quantify information in the time and/or the spatial-frequency domain.

It should be noted that an additional complication arises from the use of natural scenes since the severity of compression artifacts is a function of both the system and the picture content.

## Selection of Objective Features

The success of assessing compressed video relies on the selection of an appropriate set of features which can be measured and used to characterize video information. To model the HVS, features can also be extracted in the spatial, temporal and chromatic domains. Researchers have found that the features which incorporate HVS processing lead to higher correlation with subjective scores performed by human observers.

### Spatial-frequency Domain

Luminance transitions, which define edges are an important feature in HVS modeling. As discussed earlier, perceptual coding in video compression occurs at the pre-processing and quantization stages. High-frequency spatial components are often eliminated for the purpose of data reduction. Loss of spatial-frequency might result in blurring or smearing. Coarse quantization may introduce blockiness into the picture. The perceptual block distortion is a result of false spatial-frequency components or edges introduced by compression. Therefore, spatial features are good indicators of the amount of spatial information or edges in the compressed picture.

Various edge detection schemes such as the Sobel, Roberts and line edge detectors (filters) can be used to extract spatial information. Another approach is to compute the local entropy of the neighborhood of each pixel.[18] The histogram of each neighborhood is generated and the entropy is then computed. Pixels of interest are those for which the value exceeds a specified threshold. Unlike convolution techniques that depend on region contrast to detect edges, the local entropy operator does not. Therefore, very faint edges found in low contrast images will be detected just as edges with higher contrast.

### Temporal Domain

Like in the spatial domain, redundancy in the temporal domain can be reduced if not completely eliminated using motion estimation and compensation. However, the limitations of search range, accuracy, bit rate and the selection of Group of Picture (GOP) affect coding efficiency and subjective performance. Artifacts such as temporal noise, blockiness and jerkiness are the result of compression in the temporal domain.

The temporal features are indicators of the amount of temporal changes, or degree of motion, in video scenes from one frame to the next. Temporal features can be measured by the inter-frame difference energy.

### Feature Statistics and Comparison

The use of natural video scenes for objective picture quality measurement raises the issue of how to deal with the variation of measured values due to the constant change of picture content and the dynamic performance of compression systems. Using statistics provides the best way to characterize features from different frames at different times in a sequence. The primary statistics of these features can be the mean and the standard deviation of the pixel values across a frame/field. To better present the distortions caused by compression and other signal processes, the ratios of input and output features can be calculated based on various comparison functions. Some proposed functions include the maximum and minimum (over time) of the log ratios, the maximum and minimum (over time) of the error ratios, root mean square (over time) of the error ratio, the peak signal-to-noise ratio, the minimum (over time) of the peak signal-to-noise ratio and root mean square (over time) of the peak signal-to-noise ratio.[19]

## Perceptual Models for Picture Quality Assessment

The basic approach of developing perceptual models is 1) the selection of a set of features which can be objectively measured, 2) the establishment of a model which can simulate the HVS for the processing and analyzing of the selected features. By weighting individual features, perceptual models can predict not only specific attributes but also a global picture quality.

Development of perceptual models has two major obstacles. First, the HVS[20] is extremely complex and not fully understood. Secondly, complex models are expensive to implement, and also time consuming for real-time applications.

In general, a simple HVS model can be viewed as having several levels, one of which is the peripheral level.[21] Most of the peripheral level can be described in signal processing terms such as filtering, sampling and coding. More specifically, the current understanding of early vision can be characterized by modules diverting a common flow of information into specific areas such as color, motion and shape detection. For video quality assessment there is another high level process that sits on top of the HVS. This is the process that a human uses to make a quality judgment. For example, the HVS might be able to detect an impairment but the human judgment might not consider it significant enough to influence the quality decision.

Perceptual model were first developed for machine vision, target detection and photograph analysis. Recently, perceptual models for color video and compressed video including HDTV picture quality assessment, have been studied. Various systems have been proposed and implemented for different applications.[22-27] These systems in general select spatial and/or temporal features, which are extracted by filtering, and manipulate these features through processes such as squaring, averaging, weighting, and normalization. To some extent, these processes simulate some functions of the HVS. The last stage is to detect and construct numerical results or graphic maps.

Using numerical indexes is convenient and compatible with the traditional subjective methods. The most popular quality indexes are the ITU-R quality and impairment scales. Other scales such as Just Noticeable Difference (JND) are also used for picture quality measurement. One JND corresponds to a 75% probability that an observer viewing two images multiple times would perceive a difference. JND’s can also be plotted as a map showing the probability, as a function of position on the images, that an observer would be able to detect differences between the images. By averaging JND values, a single mean JND value can be generated for image quality scale.[26]

### Example of Perception-Based Objective Measurement System

Wolf et al proposed a perception-based objective measurement system developed for video and television signal.[27,28] They developed an automatic video quality measurement system which is composed of two sub-systems: one for source and one for destination. By extracting a set of features that can be used to predict perceptual video quality changes, an objective quality estimate of the system performance can be obtained by comparing the source and corresponding destination features. The system is based on both spatial and temporal features using the Sobel filtered image for spatial domain and the motion difference image for time domain analysis. In order to produce an single score/index of picture quality performance, a formula composed of a set of spatial and temporal features is determined using two subjective test experiments.[29] Experiments were carried out on the CODECs operated at 56 kbps to 1.5 Mbps for videoconferencing and 45 Mbps for program contribution. They also include channel impairments such as burst errors. The system had a 0.92 correlation with the subjective scores.[30] However, the database did not include a critical range of MPEG-based compression systems operating at bit rates between 1.5 and 20 Mbps. This bit rate range is believed to be the operation model for the future digital television service including HDTV. Based on the algorithm, a computer system was developed to perform the quality rating as well as distortion measurements. The system allows the user to view video side-by-side with plots of measurement results on the computer display.[31]

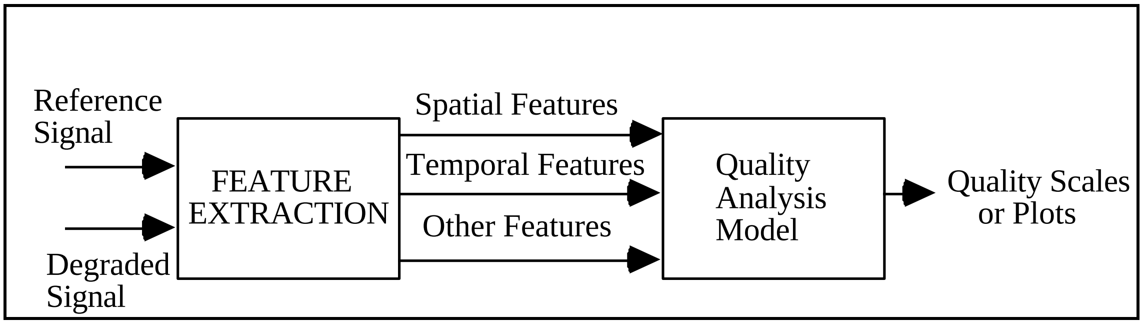

Figure 2 Perception-Based Objective Picture Quality Measurement System

## Conclusions

For quality monitoring and performance evaluation of digital video, perception-based as well as objective measurements are essential. Practical and efficient methods and equipment must still be developed. It should be noted that, so far, no perceptual model is able to replace subjective methods for picture quality assessment. Even the most sophisticated models are unable to discriminate between two or more similar systems.

One key component for perception-based objective measurement as well as subjective evaluation is the establishment of standard test material to be used in all the laboratories. Some sequences have already been accepted by the international community (as found in the 'Test Materials to be Used in Subjective Assessment' - ITU-R Document 11/6-E)[32] but these sequences lack the ability to address some of the current issues. The largest hurtle to clear is the ability to choose sequences that evaluate digital artifacts and represent typical viewing material.

The search for a universal objective measure of picture quality will fascinate researchers for a number of years to come, and the link between the subjective and objective measures will take longer. However, the constant march of technology will drive us to our goal regardless.

## Acknowledgment

The authors wish to thank S. Wolf of NTIA/ITS for reviewing the paper and providing valuable comments. The authors also wish to thank, Lew Stelmach, Andre Vincent and Metin Akgun of the Communications Research Center for their efforts in reviewing and commenting on this paper.

## References

1. W. Y. Zou, "Performance Evaluation: From NTSC to Digitally Compressed Video," SMPTE Journal, Dec. 1994.

2. Technical Report 16: "DS-3 Transport for Contribution Application of System M-NTSC Television Signal - Analog Interface and Performance Objectives," Committee T1-Telecommunications, May, 1991.

3. Report 1206: "Methods for Picture Quality Assessment in Relation to Impairments from Digital Coding of Television Signals," Reports of the CCIR, 1990, Annex to vol. XI - part 1.

4. E. S. Kohn and G. W. Beakley, "Resolution Testing of An HDTV CODEC with Moving Zone Plates," 134th SMPTE Technical Conference, Toronto, Nov. 1992.

5. G. W. Beakley, C. R. Caillouet, E. S. Kohn and J. L. Van Pelt, "Testing of a Digital HDTV CODEC Through the NASA Communications Network," 1992 HDTV World Conference Proceedings, NAB, April 1992.

6. G. W. Beakley and E. S. Kohn, "Transmission of Compressed Digital High-Definition Video Through the NASA Communications Network," Proceedings of SPIE Conference, San Diego, July 1992.

7. P. Teo and D. Heeger, "Perceptual Image Distortion," 1st. IEEE Intl. Conf. on Image Proc., vol. 2, pp. 982-986, Nov. 1994.

8. Recommendation ITU-R BT. 500-6, "Methodology for the Subjective Assessment of the Quality of Television Pictures," ITU-R 1974-1994.

9. W. Y. Zou, K. Ellsworth, J. A. Kutzner and P. J. Hearty, "Subjective Testing of Broadcast-Quality Compressed Video," SMPTE Journal, Dec. 1994.

10. K. Teunissen, "The Validity of CCIR Quality Indicators Along a Graphical Scale," SMPTE Journal, vol. 105, No. 3, March 1996.

11. B. L. Jones and P. R. McManus, “Graphic Scaling of Qualitative Terms,” SMPTE Journal, November 1986.

12. R. Aldridge, J. Davidoff, M. Ghanbari, D. Hands, and D. Pearson, "Recency Effect in the Subjective Assessment of Digitally-coded Television Pictures," Rapport No. 1071, Proceedings of the MOSAIC Workshop Advanced Methods for the Evaluation of Television Picture Quality, Institute for Perception Research, Sept. 11, 1995.

13. R. Hamberg and H. de Ridder, "Continuous Assessment of Perceptual Image Quality," Rapport No. 1071, Proceedings of the MOSAIC Workshop Advanced Methods for the Evaluation of Television Picture Quality, Institute for Perception Research, Sept. 11, 1995.

14. T. Alpert and J. Evain, "SSCQE (Single Stimulus Continuous Quality Evaluation) A New Subjective Assessment Method Introduced in ITU-R Recommendation 500-7 Presentation and Results," Proceedings of HDTV'96, Los Angeles, CA, Oct. 1996.

15. M. Ardito, M. Gunetti and M. Visca, "Preferred Viewing Distance and Display Parameters," Rapport No. 1071, Proceedings of the MOSAIC Workshop Advanced Methods for the Evaluation of Television Picture Quality, Institute for Perception Research, Sept. 11, 1995, Rosemont, Il, June 1995.

16. M. Ardito and M. Gunetti, "The Impact of Display Parameters on the Quality Perceived by the Viewers," IEEE 1995 International Conf. on Consumer Electronics, Digest of Technical Papers.

17. S. Rihs, "The Influence of Audio on Perceived Picture Quality and Subjective Audio-video Delay Tolerance," Rapport No. 1071, Proceedings of the MOSAIC Workshop Advanced Methods for the Evaluation of Television Picture Quality, Institute for Perception Research, Sept. 11, 1995.

18. J. Barba, P. Fenster, H. Jeanty and J. Gil, "Edge Detection in Cytology Using Local Statistical Properties," SPIE vol. 974 Applications of Digital Image Processing XI (1988).

19. ANSI standard T1.801.03-1996, "Digital Transport of One-Way Video Signals - Parameters for Objective Performance Assessment.

20. R. A. Young, "Oh say, can you see? The physiology of vision," SPIE vol. 1453 Human Vision, Visual Processing, and Digital Display II (1991).

21. J. Lim, "*Two-dimensional Signal and Image Processing*," Prentice Hall, 1990.

22. C. R. Carlson and R. W. Cohen, " A Simple Psycho-physical Model for Predicting the Visibility of Displayed Information," Proc. Soc. Infor. Display, vol. 21. pp. 229 (1980).

23. P. C. Teo and D. J. Heeger, "Perceptual Image Distortion," Proceedings of the First International Conference on Image Processing, vol. 2 pp. 982-986, Austin, Nov. 1994.

24. H. Hamada and S. Namba, "A Study on Objective Picture Quality Scales for Pictures Digitally Encoded for Broadcast," IEICE Trans. com vol. 177 B No.12, Dec. 1994.

25. J. A. Saghri, P. S. Cheatham and A. Habibi, "An HVS-Based Image Quality Measure," SPIE vol. 974, Applications of Digital Image Processing XI (1988).

26. J. Lubin, "A Visual Discrimination Model for Imaging System Design and Evaluation," David Sarnoff Research Center, Feb. 1995.

27. S. Wolf, "Features for Automatic Quality Assessment of Digitally transmitted Video," NTIA Report 90-264, US Department of Commerce, June, 1990.

28. A. Webster et al., "An Objective Video Quality Assessment System Based on Human Perception," SPIE Human Vision, Visual Processing, and Digital Display IV, vol. 1913, San Jose, CA, Feb. 1993.

29. S. Wolf and A. Webster, "Objective and Subjective Video Performance Testing of DS3 Rate Transmission Channels," ANSI T1A1 contr. No. T1A1.5/93-060, April, 1993.

30. N. B. Seitz, S. Wolf, S. Voran and R. Bloomfield, "User Oriented Measures of Telecommunication Quality," IEEE Comm. Magazine, vol. 32, No. 1 Jan. 1994.

31. C. P. Cressy and G. W. Beakley, "Computer-based Testing of Digital Video Quality," IS&T/SPIE's Symposium on Electronic Imaging Science and Technology, San Jose, CA, Feb. 1994.

32. Draft temporary document “Test Materials to be Used in Subjective Assessment” - ITU-R Document 11/6-E, 1996 sessions.