# A real time single ended algorithm for objective quality monitoring of compressed video signals

Alexander K.G. WÖRNER
Rohde & Schwarz Inc. – 7150-K Riverwood Drive
Columbia MD 21046 USA

&

Dr. Jürgen LAUTERJUNG
Rohde & Schwarz GmbH & Co KG – Mühldorfstraße 15
München 81673 Germany

## 1. ABSTRACT

Digital compression methods based on MPEG2 are now in widespread use for broadcasting of video content. Multiple television and radio programs can be multiplexed and transmitted altogether within the same bandwidth where formerly only one analog program fit. With the analog television video quality was very much a matter of the transmission chain and all the elements along the line. Once a sufficiently high video quality level was achieved, it was nearly independent from the content.

In using digital compression and transmission techniques the video quality is now strongly content and compression dependent. The changes in video quality can be very abrupt from scene to scene. A method that is capable of following all these changes in real time and displaying the picture quality, as a function of time, is required for continuous video quality monitoring.

This paper describes a single ended method that does not require a reference signal and is capable of processing the video in real time concerning picture deficiencies of a compressed video.

**Keywords:** Video Compression, DCT, MPEG2, SSCQE, Quality of Service

## 2. QOS MONITORING

To ensure a high service quality for broadcasting video content two aspects have to be looked at: Making the service available is the first aspect and making it perceivable, so that the recipient experiences the event in a positive and memorable way, is the second one. Making it available means in that context that all related data is transmitted properly and there are no errors in regards to transportation. The second thing is more difficult to be defined. It is strongly related to an individual's perception of that service. Now individuals may be very different in their personal opinions. In order to monitor the perceived video quality a process is required, that evaluates the video quality including a weighting due to subjective perception of video quality.

## 3. COMPRESSION ARTIFACTS

For a video sequence that has passed an MPEG-2 encoding / decoding chain, the picture quality depends on several parameters. Very important is the bit rate to which the encoder had been set. Nevertheless, the resulting picture quality depends even more strongly on the video material itself. Is it a sequence with much (temporal) activity or is there rather little movement in the scene? Does each single frame of the sequence have a lot of details, which is known as high spatial activity, or is the content rather flat?

What all sequences that underwent the MPEG-2 compression process have in common, is the block structure that may be visible or not but is always present. In a transmission system where the accessible interfaces between the various links of the chain are normally MPEG-2 transport stream interfaces, the most common parameter to measure for information on the quality of the transmission system is the bit error rate (BER). Although all DVB and DTV transmission systems have been defined to perform quasi error-free data transmission under normal conditions, even low BER values may result in impairments, which influence the perceived picture quality. Examples are blocks, which are overwritten with an additional DCT (Discrete Cosine Transformation) pattern, blocks that only contain a single DCT pattern, or shifted blocks.

During a research project of the Technical University of Braunschweig (Institute of Telecommunications Technologies headed by Prof. Ulrich Reimers) it became apparent that the visual impression of the MPEG-2 inherent blocking structure has the greatest impact on the picture quality in an otherwise normal video stream. Other kinds of artifacts like edge blurriness or mosquito noise, that are sometimes also observable, have been found of lower importance. Their presence is always accompanied by an inherent block structure. Therefore they must not be addressed specifically.

Based on that founding a single ended method described here in that paper is particularly addressing block structures in the analysis of video sequences. It is called the IfN/R&S algorithm.

## 4. THEORY OF THE ALGORITHM

The “Digital Video Quality Level” is computed from vectors, which contain information on the averaged differences between adjacent pixels. The MPEG-2 encoding process is based on blocks of 8 x 8 pixels and macro blocks of 16 x 16 pixels.

An analysis of the differences between all pairs of horizontally adjacent pixels shows an MPEG-2 specific characteristic. Normally the encoding process reduces the differences between adjacent pixels. In exception are the pairs of pixels across the borders of blocks or macro blocks. How this calculation process is been carried out can be understood from Figure 1.

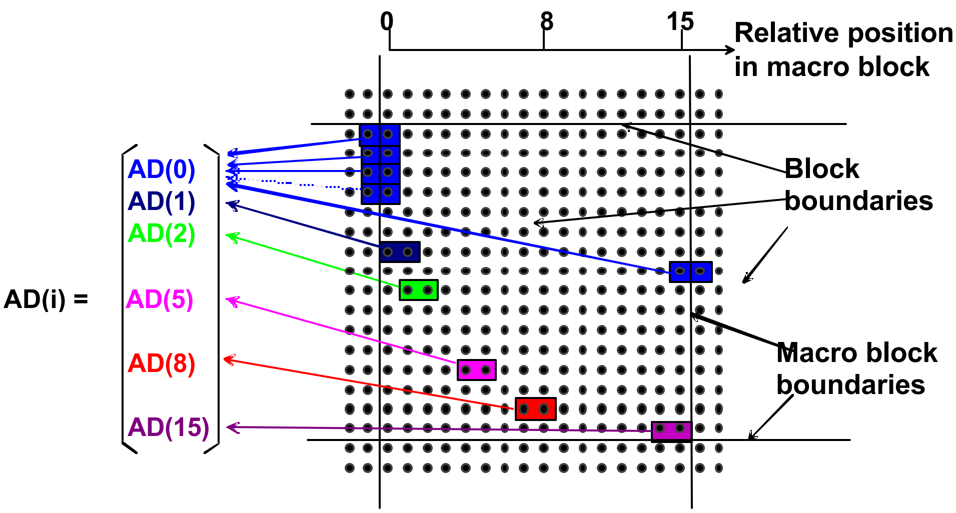


**Figure 1: Calculation process for the amplitude differences of adjacent pixels (only some horizontal sample pairs shown)**

16 different vector elements AD(i) with an index “i” from 0 to 15 are been built. They represent the average amplitude difference of all pixel pairs with the same relative distance inside the macro block grid of a frame. Figure 1 shows only individual sample pairs that contribute to the relevant vector element.

Figure 2 displays the results of the calculation of the average differences carried out on the original 'Flowergarden' sequence. The calculated values for all pairs of pixels (i.e. vector elements AD(i) ) are very close. Their overall average represent the spatial activity of each frame.

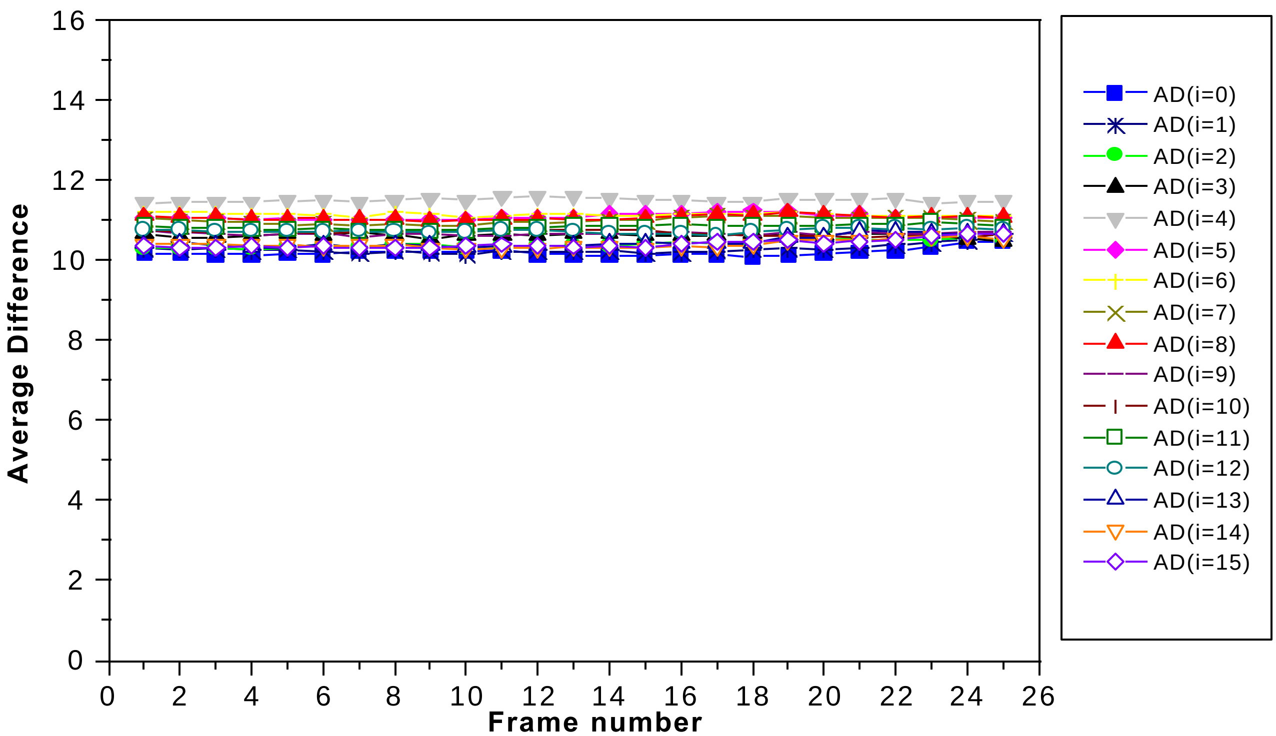


**Figure 2: Averaged pixel amplitude differences, sample sequence “Flowergarden” (uncompressed)**

After encoding the same sequence with 2 Mbit/s and subsequent decoding, the average differences of amplitudes of adjacent pixels show a particular pattern. (Figure 3)

In this diagram the values for the elements AD(i=0) and AD(i=8) are enhanced. They characterize the influence of the blocking structure in an MPEG-2 decoded picture. The “Digital Video Quality Level Un-weighted” (DVQL-U) is calculated from these values.

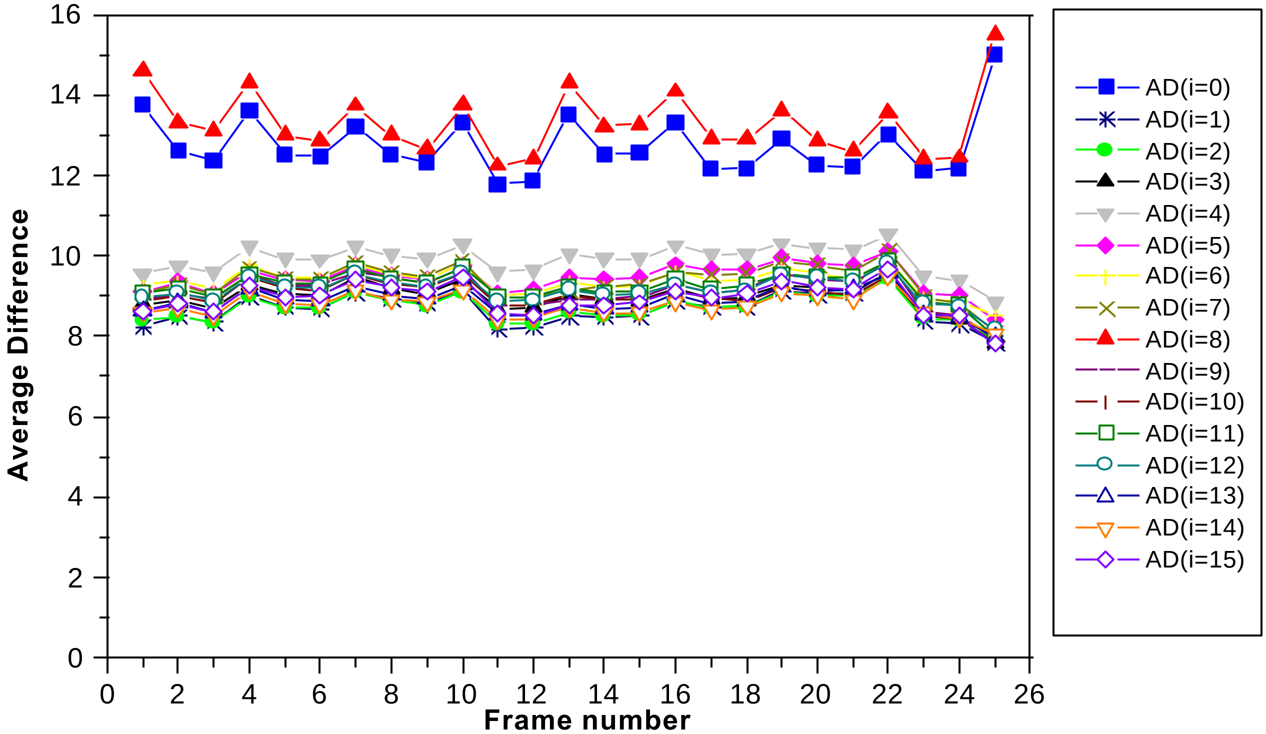


**Figure 3: Averaged pixel amplitude differences sample sequence "Flowergarden" (MPEG-2 coded/decoded with 2 Mbit/s)**

As long as masking effects that may result from spatial activity and / or temporal activity, are not considered the whole parameter is very sensitive to any blocking structure even far below the threshold of visibility. In this sense it can be compared with the measurement of the signal-to-noise-ratio (S/N) as it is applied on analog video signals.

The spatial activity is computed by averaging all pixel amplitude differences regardless of their position within the frame, whereas the temporal activity is calculated by averaging the amplitude differences of the same pixels in subsequent frames. A later version of the algorithm uses amplitude averages of pixel groups prior to comparing with the same group of pixels in a subsequent frame rather than single pixels values. This enabled a more accurate assessing of the temporal activity especially in case of slow moving background or slow camera panning.

If the appropriate masking due to subjective perception is incorporated, the resulting "Digital Video Quality Level Weighted" (DVQL-W) delivers the equivalent of the Mean Opinion Score (MOS) of subjective tests but as a predicted value. The results are displayed on an SSCQE (single stimulus continuous quality evaluation) scale acc. ITU-R Rec.500 [1].

With the masking included, the algorithm shows an excellent correlation of about 82 % with subjective assessments (Figure 4). This correlation result compares very well to nearly all double-ended algorithms, which had been proofed in recent VQEG comparison tests to be within the same range.

In using different picture resolutions the block width can be somehow different from 8 and the macro blocks are not 16 pixels wide. Therefore a pre-analysis is been carried out before the actual real time quality evaluation is started to find the correct block width and adjust the algorithm to that value. For this adjustment the same computation is employed using increasing block widths in a trial and error method until the correct width of the inherent block structure is been detected.

Since the calculation of the DVQL-W parameter is mainly based on DCT related effects, the same algorithm can be applied to other DCT-based compression systems.

## 5. IMPLEMENTATION

Rohde & Schwarz has implemented the IfN/R&S algorithm for the calculation of the Digital Video Quality Level on a platform, that allows for the real-time computing of this parameter. The device is called DVQ.

An important feature of the applied algorithm is that it makes use of the decoded video stream only. Because the unit incorporates an MPEG-2 decoder, the algorithm can therefore be carried out at any point in a network where a decoder interface is available. These are in principle all Transport Stream interfaces. In bypassing the internal decoder also an already decoded and uncompressed SDI signal (Serial Digital Interface standard, ITU-R B.T. 601 or SMPTE 259M) can be evaluated in the same way.

By using the SDI and any of the MPEG-2 inputs simultaneously, analysis of the actual and the original signal can be performed at the same time. That way allows for a referenced measurement where degradations within the original signal are been taken into account calculating the quality values of the actual signal. The results then represent the difference in degradations.

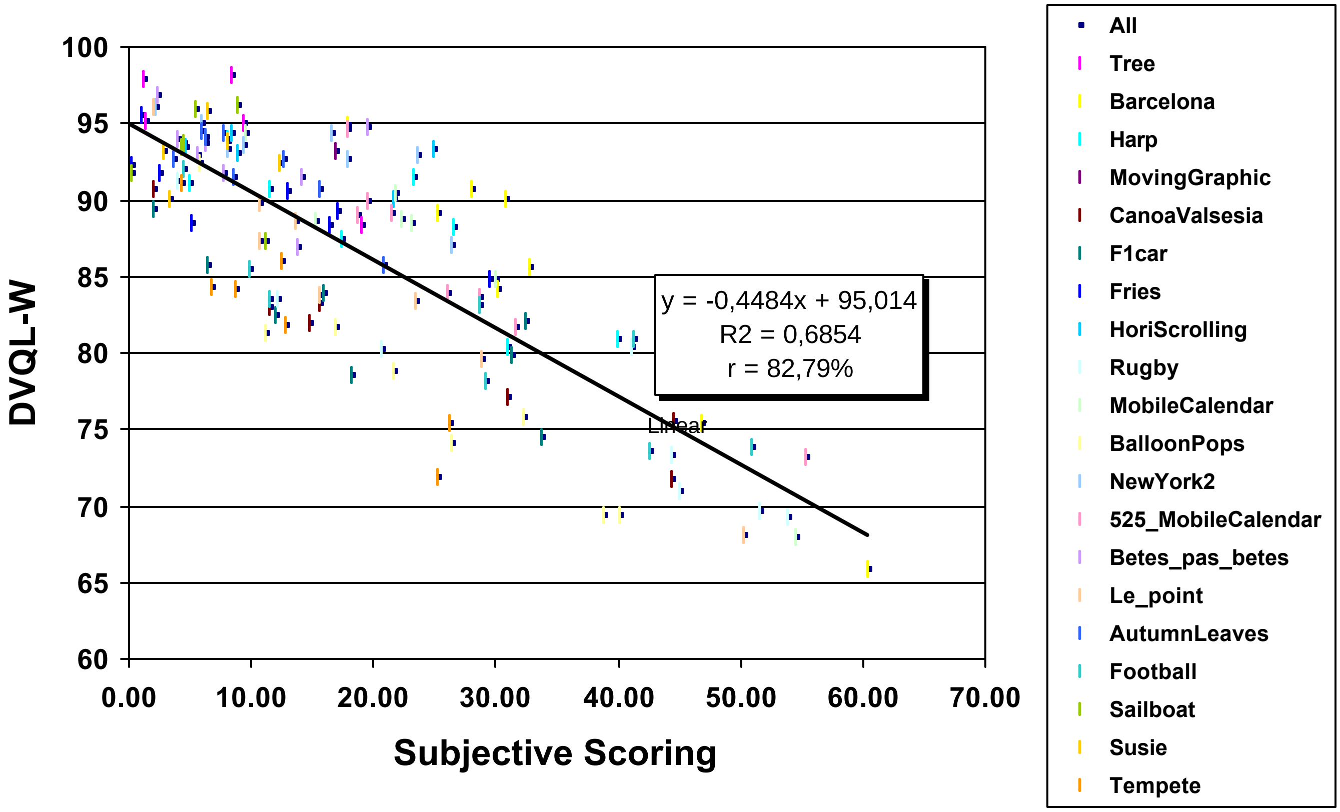


**Figure 4: Correlation between DVQL-W and subjective scoring**

The same instrument is capable of detecting errors like freeze frames and video or audio outages. Those errors cannot be assigned a quality value. They are kept track of by storing their occurrence together with corresponding information including event time and duration. Further events where the actual quality value is below an adjustable quality threshold are stored in the same non-volatile report. This extension of the described video quality analysis is also very important for a complete quality of service monitoring.

In general, the instrument can determine the influence of the encoder and the transmission system on the picture quality. If the transmission system does not produce any additional impairment, the measured picture quality at the output of the network should correspond to the picture quality at the input of the network.

The measurement device DVQ is on the market with great success for about two years. Rohde & Schwarz has been awarded an EMMY in 2000 for "Pioneering development of equipment to provide objective measurement of perceptible picture quality in digital television systems ".

## 6. APPLICATION AREAS

The most important application is monitoring of broadcast transport streams. Possible quality deficiencies concerning video quality such as blocking, freeze frames or complete outages can be detected in real-time and remedial action taken instantaneously. For reporting and billing purposes quality values can be recorded and archived. Together with a comprehensive protocol monitoring of the transport stream layer a complete quality of service monitoring can be established. Various kinds of equipment are available to complement the video quality monitoring with the IfN/R&S algorithm and sufficiently fulfill that task in a complete monitoring systems solution.

Another application of the algorithm might be of interest for service providers and/ or network operators is the control of encoder parameters such as bit rate and GOP length by the DVQL-W parameter. The same seems to work for the management of bit rates in a statistical multiplex.

Set top box manufacturers may be interested in testing their products with impaired sequences. In such a case not only the decoders could be integrated in the tests but also the efficiency of any implemented concealment strategy could be verified.

## 7. EXPERIENCES AND RESULTS

Since the algorithm was released two years ago a lot of experience could be gathered since then. This chapter describes sample measurements that have been performed by T-Nova Berkom, a wholly owned subsidiary of Deutsche Telekom [3]. The scope of the tests was to compare the performance of a fixed versus a statistical multiplex. The results are not surprising but rather somehow intuitive and encourage the use of the IfN/R&S algorithm for video quality measurements and monitoring.

Two identical environments comprising of a series of video and audio encoders plus a transport stream multiplexer have been used. One multiplex was using fixed data rates, the other one was a statistical multiplexer. The encoders were fed with the same set of broadcast programs in parallel, in order to have comparable conditions.

At the output of both multiplexers quality values have been measured simultaneously with the IfN/R&S algorithm by means of two separate devices DVQ and recorded onto a PC. The following different program multiplexes were used:

- 3 programs @ 3.5 Mbit/s (video @ 3.2 Mbit/s)
- 4 programs @ 3.5 Mbit/s (video @ 3.2 Mbit/s)
- 3 programs @ 4.7 Mbit/s (video @ 4.4 Mbit/s)

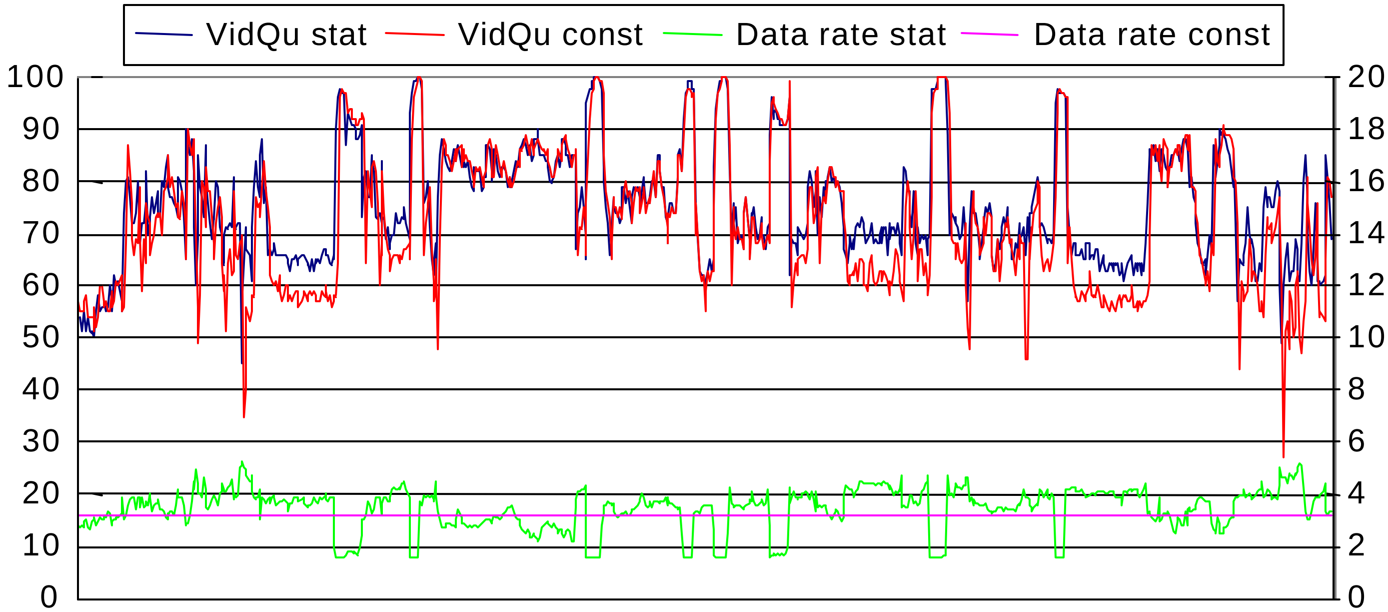


**Figure 5: Digital Video Quality Level DVQL-W (left scale) and data rate in Mbit/sec (right scale) over time for constant and statistical multiplex**

The total time period was about one hour. The measured quality values of one particular program both with fixed and statistical multiplex are displayed in figure 5 for an extracted time period of six minutes. While the quality ratings in general follow each other very closely, it is also noticeable, that the quality values of the program transmitted with statistical multiplex don't drop so sharply for short periods of time like the ones with constant data rate do. No negative quality peaks could be encountered with statistical multiplex.

Figure 6 shows the quality ratings averaged over the whole test period, fixed and statistical multiplex of each program next to one another. It proofs that statistical multiplex in average increases the picture quality compared to fixed multiplex if the quality level is comparable low. This effect becomes even stronger, if there are more programs to source bandwidth from like with the configuration 4 x 3.5 (middle columns in figure 6) or if the actual program is running at lower quality at all (program "Eurosport"). However if the average quality is already quite high, the statistical multiplex doesn't really worsen the picture quality through sourcing bandwidth from those programs. The slight decreases in the average values noticeable with the MTV program are within the statistical resolution.

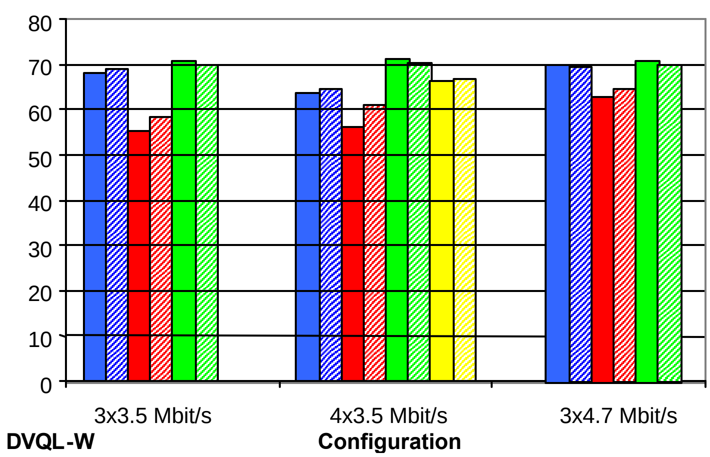


**Figure 6: Average video quality levels [DVQL-W] for constant (filled bars) and statistical multiplex (shaded bars)**

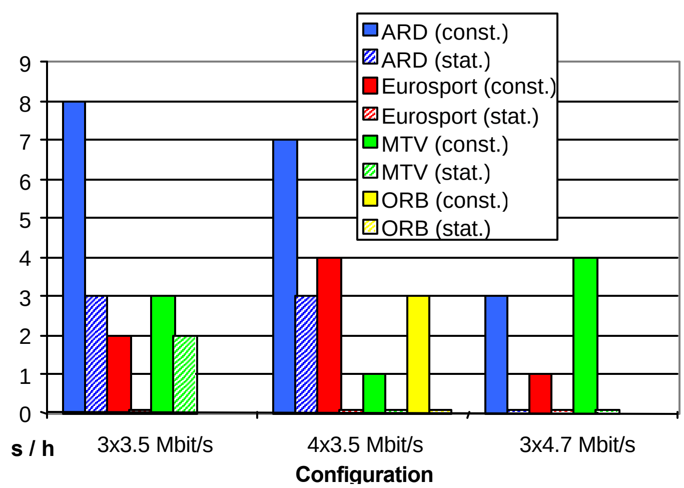


**Figure 7: Occurrences of video quality level drops [DVQL-W < 20] for constant (filled bars) and statistical multiplex (shaded bars)**

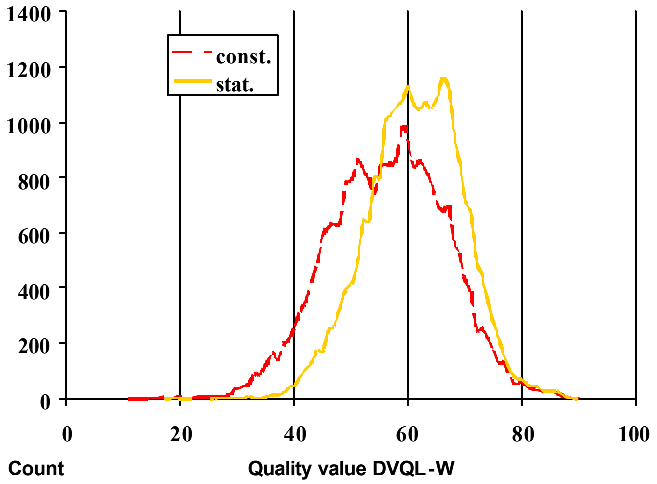


**Figure 8: Distribution of digital video quality values (configuration 4 x 3.5 Mbit/s)**

| | Mean | Std.Dev. | Count |
|---|---|---|---|
| Constant Multiplex.: | 56.4 | 10.37 | 24448 |
| Statistical Multiplex: | 61.2 | 8.39 | 24440 |

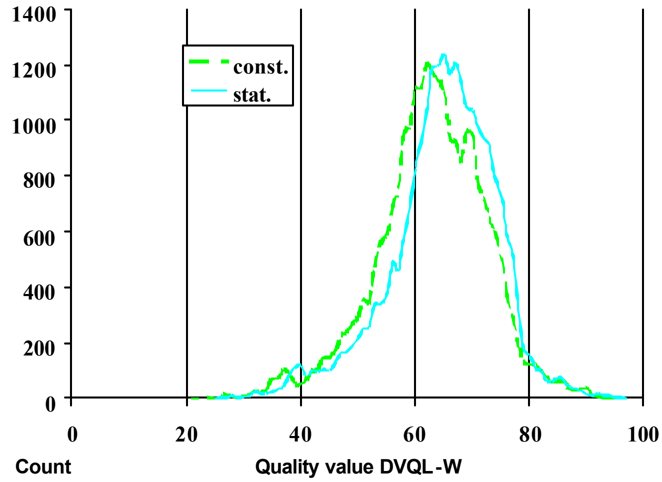


**Figure 9: Distribution of digital video quality values (configuration 3 x 4.7 Mbit/s)**

| | Mean | Std.Dev. | Count |
|---|---|---|---|
| Constant Multiplex: | 62.8 | 9.55 | 24454 |
| Statistical Multiplex: | 64.8 | 9.38 | 24430 |

In figure 7 the individual counts of how often the actual digital video quality value (DVQL-W) dropped below a threshold of 20 during the test period is displayed for statistical and fixed multiplex next to each other. A value of 20 marks the separation line between poor and bad on the quality scaling acc. SSCQE [1].

Both figures 8 and 9 show the distribution of quality values throughout the whole test for the program “Eurosport” but for different multiplex configurations. This particular program was the one that benefit most by using statistical multiplex instead of constant bit rate due to its high temporal active sports content. The two diagrams not only proof, that when using statistical multiplex the standard deviation of quality values decreases accompanied by an increase of the overall average quality. In other words: statistical multiplex is able to minimize the range of quality values so that the variation of the video quality during a running program is somehow smaller. The overall gain in quality is larger in a configuration of four programs than with only three programs, since the average quality is lower and there are more resources within the multiplex to pool data rate from.

## 8. CONCLUSIONS AND OUTLOOK

Furthermore these two diagrams give good proof, that the quality evaluation algorithm described here in that paper is an essential tool for long-term and real time assessments of video quality. It gives an ideal continuous, on-the-fly prediction of subjective video quality due to human perception regarding digital compression artifacts. With a statistical analysis of the results further indications about the quality performance of video processing equipment can be obtained.

The largest field of applications is foreseen in standard definition environments of MP@ML and 422P@ML applications. Nevertheless the inclusion of HP@HL is both a potential useful as well as challenging extension.

## 9. REFERENCES


[1] ITU-R Draft Recommendation BT.500-8: „Methodology for the Subjective Assessment of the Quality of Television Pictures“

[2] Dr. Jürgen Lauterjung: „Picture Quality Measurements”, IBC Conference Proceedings, 1998

[3] Dr. Manfred Kühn, Dr. Jochen Antkowiak: „Statistical multiplex – what does it mean for DVB-T”, FKT 4/2000