OBJECTIVE PICTURE QUALITY EVALUATION FOR THE DIGITAL BROADCASTING BY INTRODUCTION OF CRITICALITY

NHK Science & Technical Research Laboratories

Picture quality in digital coding depends heavily on such picture characteristics as spatial detail and motion. Statistical analysis is thus required to estimate picture quality in digital broadcasting systems. A new definition of " criticality" is proposed as a quantitative measure of difficulty for MPEG-2 video coding, to analyze the picture quality of television programs statistically. Equipment is developed to measure criticality and the relationship between criticality values ands assessment test scores is examined for test sequences. The results show that the definition of criticality provides sufficient accuracy to estimate picture quality for a large number of pictures statistically. Criticality is useful not only for estimating picture quality distribution characteristics but also for selecting a set of test sequences.

The criticality is defined as a number of output bits per pixel from a MC + DCT encoder with a fixed quantizer. Using the criticality, a method to obtain statistical distribution of picture quality for television programs in digital broadcasting includes definition of criticality (Step 1), criticality measurement of a large number of samples taken from typical television programs (Step 2), subjective assessment of the quality achieved by a system under test (Step 3), and derivation of picture quality distribution characteristics by combining the results of Step 2 and Step 3 (Step 4). The definition of criticality and the procedure to derive statistical picture quality distribution have been included in a draft revision of ITU-R Rec. BT. 1210-1 (Test materials to be used in subjective assessment) and a draft revision of ITU-R Rec. BT. 1129-1 (Subjective assessment of standard definition digital television (SDTV) systems), respectively.

The ANNEX to this document describes the above in detail.

Contact person
Name: Yukihiro NISHIDA
Tel: +81 3 5494 2227
Fax: +81 3 5494 2309
E-mail: ynishida@strl.nhk.or.jp

ANNEX

Derivation of picture quality distribution characteristics

1 Introduction

Picture quality of digitally coded images highly depends on picture contents such as spatial details and motion. Therefore, it is not easy to confirm how well a system under test performs for broadcast television programs only by subjective assessment tests, where a limited number of test sequences are used. By introducing " criticality" as a quantitative measure of difficulty of digital coding, and by establishing a method to derive picture quality distribution characteristics using criticality, it will become possible to evaluate picture quality statistically.

The purpose of this Annex is to give a definition of criticality and to specify a method to derive picture quality distribution characteristics. Further study may be desirable to describe better the characteristics of broadcast images.

2. Definition of criticality

Criticality is defined as " the number of output bits per pixel from a hybrid DCT encoder with a fixed quantizer" . Criticality is averaged over each whole frame. Figure 1 shows the configuration of the criticality measurement equipment based on the definition. This definition corresponds to the bit-rate required to obtain an almost constant picture quality for various sequences, because a critical sequence requires a higher bit-rate to maintain picture quality to the level of the non-critical sequences.

Although the current definition may not reflect local degradation within each frame, the definition provides sufficient correlation with picture degradation to allow estimation of picture quality distribution characteristics of digital coding systems.

An example of criticality measurement based on the definition described here is shown in Appendix 1 to this Annex.

3 Method to derive picture quality distribution characteristics

A procedure to derive picture quality distribution characteristics is as follows:

Step 1 Conduct a subjective assessment of picture quality produced by a coding system under test.

Step 2 Measure criticality of the test sequences used in the subjective assessment.

Step 3 Derive a relationship between criticality and subjective picture quality from the results of Step1 and Step2.

Step 4 Measure criticality distribution of broadcast programs for a long time period.

Step 5 Derive picture quality distribution characteristics (quality vs. frequency of occurrence) by combining the results of Step 3 (criticality vs. quality) and Step 4 (criticality vs. frequency of occurrence).

This method enables to evaluate the performance of coding systems in terms of frequency of occurrence of certain picture quality for overall broadcast programs. The procedure described above is depicted in Figure 2.

An example of derivation of picture quality distribution characteristics based on this procedure is described in Appendix 2 to this Annex.

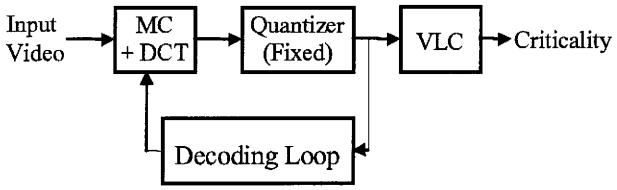

FIGURE 1 Configuration of criticality measurement equipment

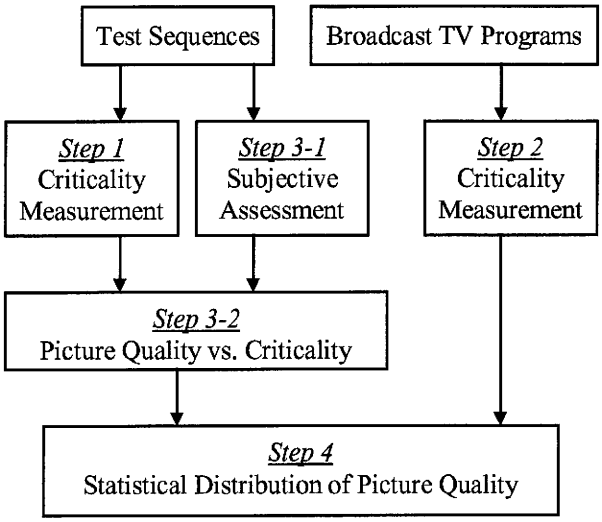

FIGURE 2 Procedure to derive picture quality distribution characteristics

APPENDIX 1 TO ANNEX

Results of criticality measurement

1 Introduction

To provide necessary information for Step 2 and Step 4 described in the Annex criticality of test sequences used for a subjective assessment and criticality distribution of broadcast television programs have been measured. A measurement equipment has been developed to conduct the criticality measurement. Detailed description of the equipment and the measurement results are given in this Appendix.

2 Criticality measurement equipment

The equipment employs MPEG-2 coding algorithm with some exceptions. Only field-based prediction is used, and no bi-directional prediction is employed. Intra macroblock refreshment, having cycles of 0.5 seconds, is used. This smoothes away fluctuations of the output bits introduced by intra/inter coding modes. The quantizer characteristics are linear and comply with the MPEG-2 Standard. The parameter value of " quantizer_scale_code" , which gives a quantizer step closely related to picture quality degradation, can be selected arbitrarily from 1 through 31. The number of output bits is measured in each whole frame and is converted into criticality dividing by the number of pixels in a frame.

3 Criticality of test sequences

Criticality of test sequences used for the subjective assessment described in Appendix 2 has been measured. A list of test sequences and their contents are provided in Table 1 of Appendix 2. Figure 3 describes the fluctuation of criticality for three different sequences over five-second interval. " Nintama" (cartoon) is a less critical sequence, with high criticality where there is a cut change. Criticality of the " Mobile (Fade)" sequence changes during the fade interval, fading in from the black level in the beginning and fading out in the end of the sequence. Criticality of the " Pop-jam" sequence fluctuates significantly due to the rapid lighting changes.

Figure 4 shows the mean and standard deviation of each sequence for five-second interval. Most sequences have criticality measures from 0.8 to 1.4 bits/pixel. " Susie" and " Nintama" show low criticality, and " Green Leaves" and " Mobile (Super)" present high criticality. The results coincide well with degree of spatial detail or motion. " Mobile (Fade)" , " Sprinkling" , and " Pop-jam" have a large standard deviation because the picture content varies significantly over the five-second interval.

4 Criticality of broadcast television programs

Criticality distribution of broadcast television programs has been measured for one week, a total of 130 hours, from 15 through 22 February 1995. In the measurement, composite NTSC signals were converted into component Y/C signals and quantizer_scale_code was set to 6. The frequency of occurrence of criticality for television programs was calculated every 5×10^{-3} bits/pixel. The measurement was conducted for the NHK General Service, which includes a wide range of program genres such as drama, sports, news, and so on. The statistical distribution of criticality for overall television programs is shown in Figure 5. This figure also shows criticality for the test sequences used for the subjective assessment.

The mean of the distribution is approximately 0.44 bits/pixel and the mode is about 0.3 bits/pixel. By converting the figure into a cumulative frequency it is found that the majority of the sequences used for the subjective assessment test are located in the range of upper 10% of the criticality distribution. At the same time rather non-critical sequences, such as " Susie" and " Nintama" , are also included into the test sequences used. Criticality is useful not only for the derivation of picture quality distribution characteristics but also for the selection of a set of test sequences that conforms to the description of " critical, but not unduly so" in Recommendation 500-7.

The distribution characteristics for different program genres have also been measured. Criticality distribution for classified television programs is shown in Figure 6. Sports programs, which often include rapid motion and fine details, are more critical than other programs, and drama programs, which usually include slow movement and close-ups, are less critical. This means that a digital coding system that performs well for drama does not necessarily perform well for sports programs. Criticality helps selecting appropriate parameters of a digital coding system for various kinds of program genres.

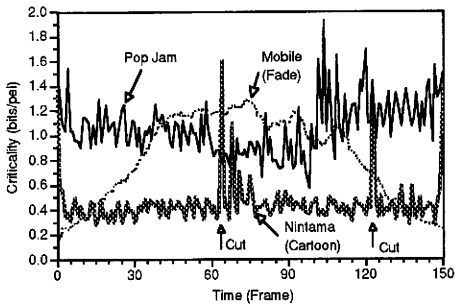

FIGURE 3 Fluctuation of criticality for five-second interval

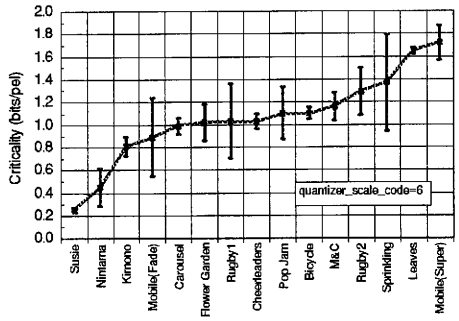

FIGURE 4 Means and standard deviations of criticality of test sequences

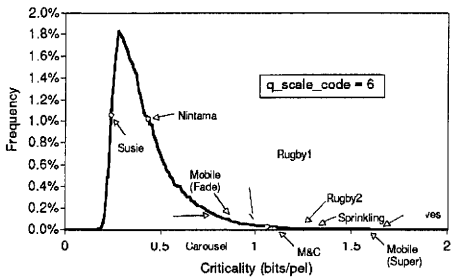

FIGURE 5 Distribution of criticality for television programs and criticality of test sequences

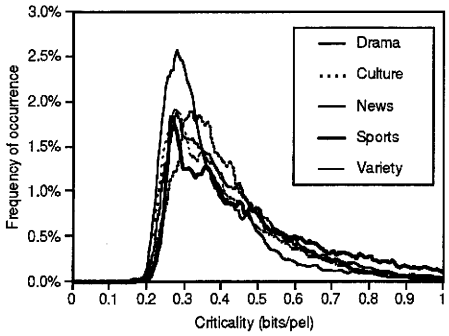

FIGURE 6 Criticality distribution for classified television programs

APPENDIX 2 TO ANNEX

An example of derivation of picture quality distribution characteristics

1 Introduction

The procedure to derive picture quality distribution characteristics that is described in Annex is shown in this Appendix. As an example digital coding of the standard television using MPEG2 MP@ML (Main Profile at Main Level) video coding algorithm has been analyzed.

2 Subjective assessment

A subjective assessment for MPEG2 MP@ML has been conducted as Step 1 described in Annex. The "Test Model", which is a software codec developed in MPEG, was used to encode test sequences. Bit-rates tested were 4, 6, 9, and 15 Mbps. 15 test sequences, which are component SDTV signals listed in Table 1 were used. DSCQS (the double-stimulus continuos quality-scale) method specified in Recommendation 500-7 was used for the assessment.
The result of the subjective assessment is presented in Figure 7. Quality difference (DSCQS %) in the figure represents the degradation from the reference, original 4:2:2 component picture.

3 Criticality measurement

Criticality of the test sequences and the criticality distribution for a large amount of broadcast programs for a period of one week have been measured. These correspond to Step 2 and Step 4 in Annex. The results of the measurements are presented in Figure 3 in Appendix 1 to Annex of this document.

4 Correlation between criticality and picture quality

The correlation between criticality and the scores of the assessment test has been analyzed as Step 3 in Annex. The information in Figure 4 in the Appendix 1 to Annex and Figure 7 are used for the analysis. Figure 8 shows the relationship between criticality and quality difference. Assuming linearity in the relationship between criticality and picture quality, regression lines were derived using the least squares method. The regression line at each bit-rate is illustrated in the figure.

5 Derivation of picture quality distribution characteristics

Based on the information regarding the criticality distribution of television programs as well as the relationship between criticality and picture quality, picture quality distribution characteristics of digitally coded television programs have been derived (Step 5). The picture degradation in broadcast television programs is converted into cumulative frequency of occurrence as shown in Figure 9.
Probability of occurrence over DSCQS 12% and 18% for overall television programs are shown in Table 2. It has become clear that the degradation exceeding DSCQS 12% is about 1% at 9Mbps and about 9% at 6Mbps for overall TV programs when MP@ML algorithm is employed and the programs are viewed at the distance of 6H.

TABLE 1 Outlines of test sequences

Name	Origin	Contents
Susie	ITU-R	Woman telephoning, slow motion
Flower Garden	ITU-R	Panning on detailed colored flowers and houses, slow pan
Mobile & Calendar	ITU-R	Miniature train circling before a colored setting, slow motion
Cheerleaders	ITU-R	Cheerleaders dancing before spectators, random motion
Bicycle	ITU-R	Two women riding bicycles in a woods, pan/zoom
Carousel	MPEG	Moving carousel, complex, fast movement, mainly horizontal
Sprinkling	ITE	Woman under synthesized sprinkler, random motion
Green Leaves	ITE	A path in a woods, zoom/random motion
Kimono	NHK	Clothes with detailed color and texture, pan
Rugby 1	NHK	Stadium overview, pan/cut
Rugby 2	NHK	Rugby players in a field in front of spectators, cuts, random fast motion
Pop-jam	NHK	Two singers on stage, rapid lighting changes
Nintama	NHK	Cartoon, cut
Mobile (Super)	Syn	Mobile & Calendar with captions scrolling vertically
Mobile (Fade)	Syn	Mobile & Calendar with fade in/out

ITU-R: ITU-R conventional television test sequences (Rec. ITU-R BT. 1128)

MPEG: MPEG test sequences (ISO/IEC JTC1/SC29/WG11 N0763)

ITE: down-converted sequences from ITE HDTV test sequences (Rec. ITU-R BT. 710-2)

NHK: down-converted sequences from NHK HDTV programs

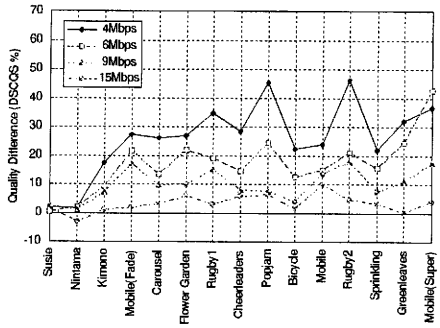

FIGURE 7 Result of subjective assessment (MP@ML at 6H)

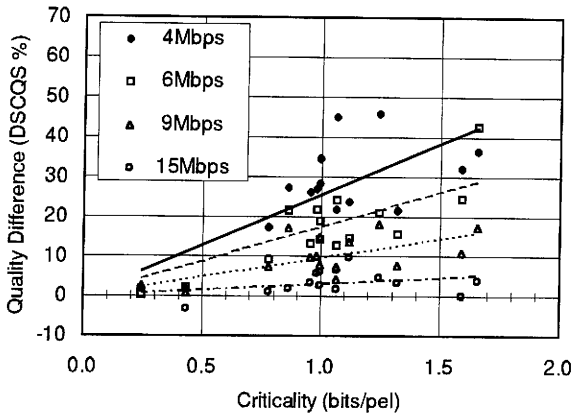

FIGURE 8 Relationship between criticality and result of assessment (MP@ML at 6H)

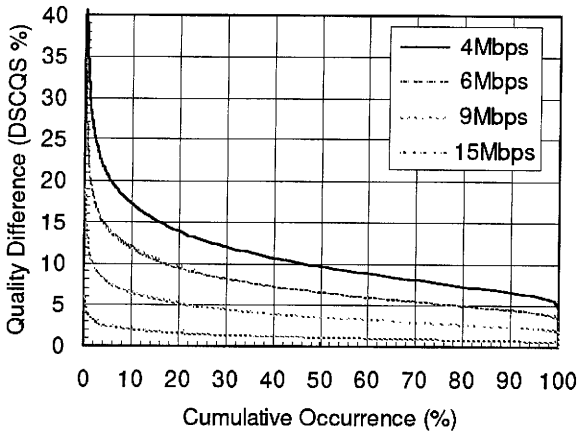

FIGURE 9 Cumulative frequency of occurrence of picture degradation (MP@ML at 6H)

Table 2 Probability of occurrence of degradation (MP@ML at 6H)

Bit-rate	Over 12 (DSCQS %)	Over 18 (DSCQS %)
4 Mbps	30% (17 - 44%)	8.5% (3.9 - 14.7%)
6 Mbps	9.2% (4.4 - 16.1%)	1.9% (0.9 - 3.5)
9 Mbps	0.9% (0.3 - 2.0%)	0.1% (0.02 - 0.4%)

( ) : 95 % confidence interval

REFERENCES

1. Recommendation ITU-R BT.500-7, " Methodology for the Subjective Assessment of the Quality of Television Pictures" , 1994.

2. Nakasu, E., Aoki, K., Yajima, R., Kanatsugu, Y., and Kubota, K., Statistical Analysis of Picture Quality for Digital TV Employing MPEG-2 Video Standard. SMPTE Journal, **105**, 11, pp.702-711 (1996).