INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE NORMALISATION
ISO/IEC JTC 1/SC 29/WG 11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC 1/SC 29/WG 11 N2824

July 1999

Source: **Test and Video Group**
Status: **Approved**
Title: **Report Of The Formal Verification Tests on Advanced Coding Efficiency ACE (former Main Plus) Profile In Version 2**

Summary

This report illustrates the results of the formal subjective verification test carried out to evaluate the performance of MPEG-4 Version 2 Advanced Coding Efficiency (former Main Plus) Profile compared with MPEG-4 Version 1 Main Profile.

1. Introduction

The visual part of the MPEG-4 Version 2 standard will provide several tools to enhance the coding efficiency compared with MPEG-4 Version 1, namely Global Motion Compensation, Quarter Pel Motion Compensation and Shape-adaptive DCT. These tools are proposed for the Advanced Coding Efficiency (former Main Plus) Profile in MPEG-4 Version 2 It was recognised that an evaluation of the coding efficiency compared with MPEG-4 Version 1 is necessary to justify the inclusion of these tools.
This document describes the test procedures and the results of the Advanced Coding Efficiency profile evaluation test. The test has been conducted by FUB and carried out at the laboratories of NHK (Japan) and IRT (Germany).

2. Context and test motivation

2.1. Advanced Coding Efficiency (former Main Plus) profile

The Advanced Coding Efficiency (former Main Plus) Profile was proposed during the 46th MPEG meeting in Rome. The corresponding proposals were m4267 "Request for a new Visual Object and a new Visual Profile in Version 2: Main-Plus Object and Main-Plus Profile" and m4265 "Proposal of Acquisition Visual Object Type and Profile for MPEG-4 Visual Version 2". These contributions were made, because the contributing companies have come to the conclusion, after thoroughly having investigated the properties and specifications of the currently existing profiles of ISO 14496-2 version 1, that especially for applications with broadcast reception from mobile devices and for the process of acquiring and collecting image sequences there are certain requirements which are not fulfilled by the object types and tools of the current profiles or combinations thereof.
Therefore, the Advanced Coding Efficiency Profile is targeting at two main application areas:

a) Mobile broadcast reception
   Mobile multimedia applications like mobile TV reception, mobile video phones, access to multimedia data servers from mobile devices to name only a few, will be an important application area for MPEG-4.
   However, they face technical challenges that are significantly different from the problems typically encountered with either portable or desktop multimedia applications.
   This is because mobile receiver technologies are subject to inherent limitations such as narrow bandwidth, limited computation capacity, and unsatisfactory reliability of the transmitted media due to bad error conditions.
   To meet the resulting application requirements it is of paramount importance to provide the best coding efficiency available and to enable as much as possible combinations of coding tools and the error resilience tools of MPEG-4.
b) Acquisition of image sequences.
   The process of acquiring and collecting image sequences is an important component in many MPEG-4 applications. The MPEG-4 objects obtained by encoding image sequences acquired by a personal video camera can be used in many applications: personal homepages, transmission of acquired image sequences (e.g. from a local site to the office), video conferencing (e.g. using a notebook with a mobile communication card), personal video album, electronic presentation or surveillance systems with moving cameras.

For satisfying the needs of those applications it is necessary to code video objects having large intentional or unintentional camera motion with high coding efficiency.
However, the currently available standards do not support this functionality, i.e. it is difficult to code sequences containing global motion with high coding efficiency by using block matching.

2.2. Test motivation

These subjective evaluation tests shall give evidence of the improved coding efficiency of the Advanced Coding Efficiency Profile compared with the MPEG-4 Version 1 Main Profile as advised in resolution 2.1.4 of the 46th MPEG meeting.

Due to the large extent of the cases taken into account in this verification test it has been agreed on the 47th MPEG meeting to assess the visual quality and functionality provided by this profile by performing three separate tests:

- The object-based case at low bit rates
- The frame-based case at low bit rates
- The frame-based case at high bit rates

As can be seen, the verification tests for the Advanced Coding Efficiency profile will be performed testing the frame-based coding functionality at low and high data rates, as it was also done in the coding efficiency tests for MPEG-4 Visual Version 1. In addition, to give evidence that there is no draw-back in the object-based coding mode, a test of the object-based coding functionality limited to low bit rates will be carried out.

The report of the Ad hoc Group that conducted the tests (M4698) was available on 1999 June 28th, to give national bodies the opportunity to respond to the resolution 3.1.6 of the 47th MPEG meeting. This resolution asks national bodies, when they cast their ballot on the Visual CD of MPEG-4 Version 2 (14496-2 PDAM 1), that they should take into account

- The Profiles proposed for Version 2 in N2726 (MPEG-4 Profiles under consideration)
- The results of subjective tests for these proposed profiles and that they should
- Put forward requests for new profiles whose definition they deem necessary
- Express their opinion about the inclusion of new tools in MPEG-4 Version 2

3. Time Schedule

The formal test on the Advanced Coding Efficiency Profile for MPEG-4 Version 2, described in N2712, was prepared and conducted between the 47th and the 48th MPEG meeting. The actual schedule of this test is listed in the table below.

Tasks	***From***	***To***	***Participants***
Preparation of the test materials	March 22nd	May 31st	HHI, Bosch, Univ. of Hannover, NTT, Hitachi
Test tape editing	May 27th	June 4th	FUB
Formal subjective test	June 10th	June 14th	NHK, IRT
Test data processing	-	June 14th	FUB
Final report	June 15th	June 28th	AHG

Table 1: Schedule for the formal test of the ACE Profile

4. Test Conditions

4.1. Coding Parameters

The formal test has been conducted according to the conditions reported in sections 4.1.1, 4.1.2 and 4.1.3. All the test conditions were accepted by the AhG on MPEG-4 Video Verification Tests.

4.1.1. Object Based – Low Bit rate

The test conditions for the Object Based, Low Bit Rate test are listed in Table 2.

Items	MPEG-4 V1 Main Profile vs. MPEG-4 V2 Advanced Coding Efficiency Profile	
Condition	Object Based – Low bit rate	
Sequences	Dancer, Bream, Coastguard	
Resolution	CIF (352x288)	
Bit rate	256 kbit/s	384 kbit/s
Input frame rate	12,5 Hz	25 Hz
Period of I	1st VOP only	1st VOP only
Period of P	M=1	M=1
Rate control	MPEG-4	MPEG-4

Table 2: Test conditions for the Object Based, Low Bit Rate test.

4.1.2. Frame Based – Low Bit rate

The test conditions for the Frame Based, Low Bit Rate test are listed in Table 3.

Items	MPEG-4 V1 Main Profile vs. MPEG-4 V2 Advanced Coding Efficiency Profile		
Condition	Frame Based – Low bit rate		
Sequences	MIT, Dancer, Bream Background*		
Resolution	CIF (352x288)		
Bit rate	128 kbit/s	256 kbit/s	384 kbit/s
Input frame rate	Dancer: 8,3 Hz MIT: 12,5Hz Bream: 25Hz	Dancer: 12,5 Hz MIT, Bream: 25Hz	Dancer: 12,5 Hz MIT, Bream: 25Hz
Period of I	1st VOP only	1st VOP only	1st VOP only
Period of P	M=1	M=1	M=1
Rate control	TM5	TM5	TM5

Table 3: Test conditions for the Frame Based, Low Bit Rate test.

Note: *All sequences are 10 sec.s long, but MIT that originally was 6 sec.s but was extended to 10 by a shuttle technique.

4.1.3. Frame Based – High Bit rate

The test conditions for the Frame Based, High Bit Rate test are listed in Table 4.

Items	MPEG-4 V1 Main Profile vs. MPEG-4 V2 Advanced Coding Efficiency Profile		
Condition	Frame Based – High bit rate		
Sequences	Mobile & Calendar, Flower & Garden, Disk		
Resolution	CIF (352x288)		
Bit rate	512 kbit/s	768 kbit/s	1024 kbit/s
Input frame rate	25 Hz	25 Hz	25 Hz
Period of I	N=24	N=12	N=12
Period of P	M=3	M=3	M=3
Rate control	TM5	TM5	TM5

Table 4: Test conditions for the Frame Based, High Bit Rate test.

4.2. Display format

The decoded video sequences in CIF format (352x288) are upsampled to the ITU-R BT.601 frame format (720x576 pel) using an extended version of the mpeg4_filter program by Andreas Hutter, TU Munich (W1552) that is available by anonymous ftp at ftp://ftp.tnt.uni-hannover.de/pub/MPEG/mpeg4-seqs/mpeg4_filter.c. The upsampling is done in FUB according to N0322.

5. Formal verification test

5.1. Test method

The formal subjective tests on Advanced Coding Efficiency Profile, which is described in the document entitled "New MPEG-4 Profiles Under Consideration" (N2726), have been organised into three separate test sessions to assess the three cases under consideration:

- Object Based (@ low bit rate)
- Frame Based (@ low bit rate)
- Frame Based (@ high bit rate)

The test sessions and the test tapes have been prepared at FUB.
The formal subjective tests have been conducted in NHK (Japan Broadcasting Corporation) and IRT (Institut of Rundfunktechnik GmbH in Munich).
The Low Bit Rate tests (Frame Based and Object Based) have been performed using the Double Stimulus Impairment Scale (DSIS) test method.
A DSIS test session is conducted in a way that the assessor always see an original sample of the sequence to be tested (not impaired), then the sequence under test (coded according to the coding method that has to be evaluated).
The assessors are requested to assess the quality of the sequence under test, having the reference sequence in mind.
In Appendix A the instructions to the assessors for the DSIS test method are reported.
The Frame Based, High Bit Rate test has been performed using the Double Stimulus Continuous Quality Scale (DSCQS) test method.
A DSCQS test session is conducted the assessor in a way that the assessor always see twice a pair of sequences. One of them is the original; the other is the sequence under test. The order of presentation is not known by the subjects and is usually varied in a pseudo random way to have a balanced set of "stimula" (i.e. the reference should be seen as first as many time the sequence under test has been).
The assessors are requested to assess the quality of both of these sequences and not the difference in quality between them.
In Appendix B the instructions to the assessors for the DSCQS test method are reported.
Both DSIS and DSCQS test methods were designed by ITU to assess the quality of television pictures.
The tests conducted to assess the Advanced Coding Efficiency Profile have been designed with a minor modification. This modification has been introduced to improve the performance of both test methods in the evaluation of multimedia video material, in which the medium level of quality is not so high (as instead usually is the case in the evaluation of TV video signals).
The modification is to use as original material the sequences up-sampled from the CIF format.
In this way a modest degradation is introduced also in the "reference" video material; the aim of this processing is to decrease the gap in quality between the coded and the reference material. This allows to avoid the compression of votes towards the lower end of the evaluation scale.
The results obtained demonstrate that, with this small modification of the test method, the assessors have been able to properly evaluate even small differences.
In order to compensate for the session order effect, the tests at the IRT started with the high bit rates, while NHK started with the low bit rate tests.

5.2. Laboratory set-up at NHK

5.2.1. General viewing conditions

Compliant with ITU-R BT.500-7
Date of tests: 11 Jun, 1999

5.2.2. Other viewing conditions

Monitor: Shibasoku CM291 (29" CRT professional studio monitor)
Peak luminance of the display: 70 cd/m^2
Viewing distance: 4 H (four times the height of the monitor)

5.2.3. Observers

22 non-expert observers (male: 9, female: 13, age between 24 and 40, normal visual acuity)

5.3. Laboratory set-up at IRT

5.3.1. General viewing conditions

Compliant with ITU-R BT.500-7
Date of tests: 14 Jun, 1999

5.3.2. Other viewing conditions

Monitor: 19" CRT professional studio monitor

Peak luminance of the display: 80 cd/m^2
Viewing distance: 4 H (four times the height of the monitor)

5.3.3. Observers

21 non-expert observers

6. Analysis of test results

6.1. Statistical analysis

The data provided by the testing sites has been statistically processed to obtain the Mean, the Standard Deviation and the Confidence Interval.

6.2. Results

The tables below show the results of the three test cases obtained from the results of both NHK and IRT laboratories, using a total of 43 assessors.

In the tables V1 stands for Main Profile, V2 for Advanced Coding Efficiency Profile.

		256		384	
		V1	V2	V1	V2
Bream	Mean	3.44	3.91	3.53	4.00
	SD	1.05	0.81	0.85	0.82
	95%CI	0.31	0.24	0.26	0.24
Coast Guard	Mean	1.88	2.26	1.53	2.23
	SD	0.96	0.79	0.80	1.00
	95%CI	0.29	0.24	0.24	0.30
Dancer	Mean	1.35	1.58	2.05	2.07
	SD	0.78	0.70	0.87	0.91
	95%CI	0.23	0.21	0.26	0.27

Object Based - Low Bit rate

		128		256		384	
		V1	V2	V1	V2	V1	V2
Bream (back-ground)	Average	2.74	4.53	3.86	4.56	4.28	4.58
	SD	0.88	0.67	0.77	0.85	0.83	0.88
	95%CI	0.26	0.20	0.23	0.26	0.25	0.26
MIT	Average	1.93	2.49	2.67	3.23	3.19	3.77
	SD	0.88	0.91	0.81	0.92	1.01	1.07
	95%CI	0.26	0.27	0.24	0.28	0.30	0.32
Dancer	Average	1.26	1.33	1.84	2.67	3.16	3.79
	SD	0.69	0.52	0.81	0.81	0.69	0.86
	95%CI	0.21	0.16	0.24	0.24	0.21	0.26

Frame Based - Low Bit rate

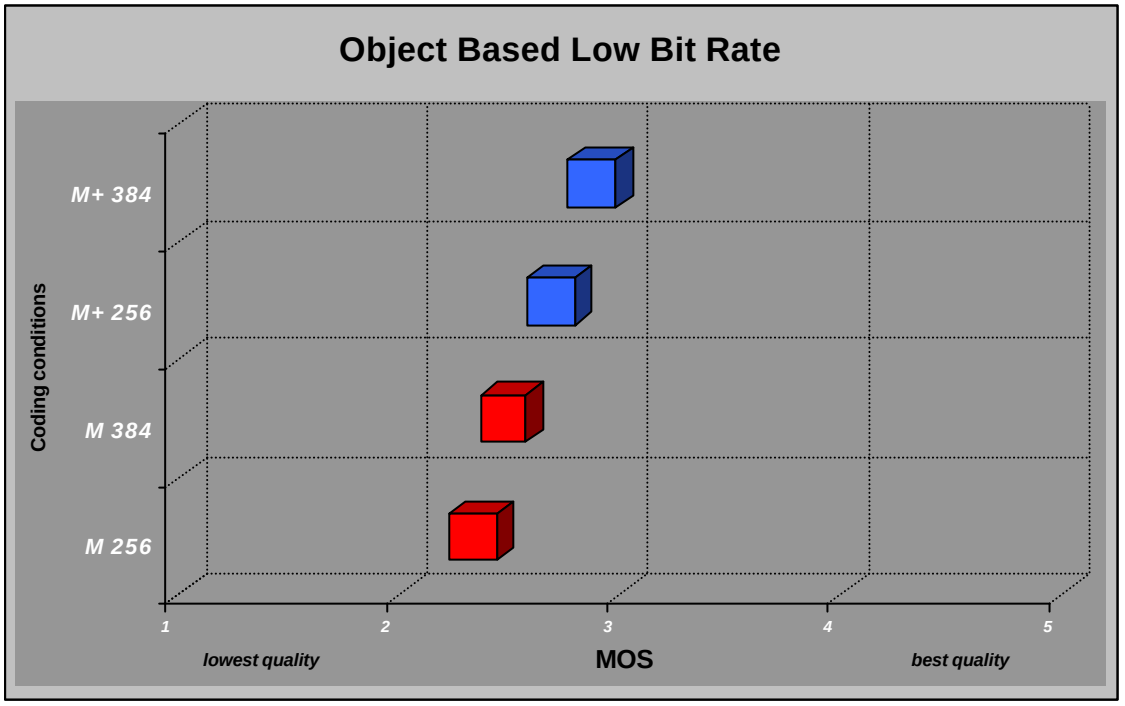
		512		768		1024	
		V1	V2	V1	V2	V1	V2
Mobile Calendar	Mean	52.36	43.74	42.19	36.05	38.57	33.02
	SD	19.72	17.42	20.66	19.62	21.73	18.25
	95%CI	6.04	5.33	6.32	6.00	6.65	5.59
Flower & Garden	Mean	31.50	27.17	21.19	12.00	18.86	12.69
	SD	21.28	19.34	18.69	17.68	14.80	14.59
	95%CI	6.52	5.92	5.72	5.41	4.53	4.47
Disk	Mean	17.55	17.60	14.86	17.64	14.62	4.71
	SD	20.50	13.87	15.95	13.94	13.75	14.00
	95%CI	6.28	4.24	4.88	4.14	4.21	4.29

Note: Two observers were rejected.

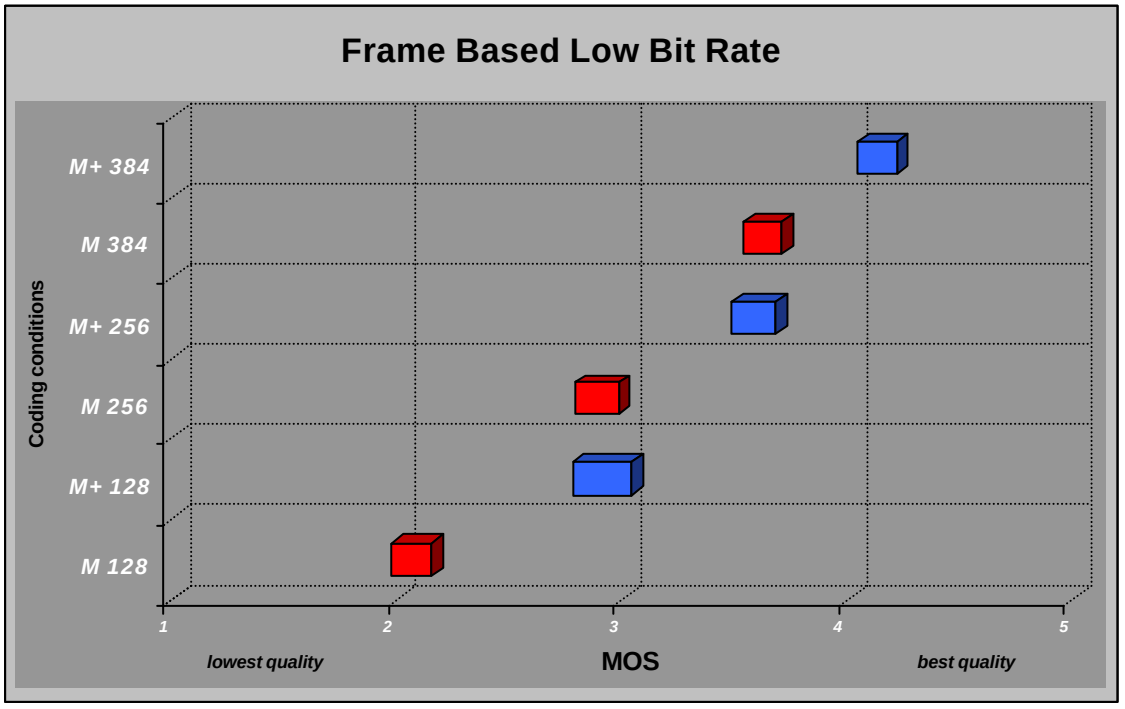
Frame Based - High Bit rate

Graph 1, Graph 2 and Graph 3 provide a visual indication of the performances of the Advanced Coding Efficiency Profile compared with the Main Profile. Here, non-overlapping bars indicate statistically significant differences in subjective image quality.
In both the Object Based Low Bit Rate test and the Frame Based Low Bit Rate test the DSIS test method was used.
Graph 1 and Graph 2 should be read as follows:

- The visual quality (MOS) increases from the left to the right
- Bars of the same colour correspond to the same profile; the Advanced Coding Efficiency (former Main Plus) Profile is labeled as M+ (blue) and the Main Profile is labeled as M (red). The results are ordered in descending quality.



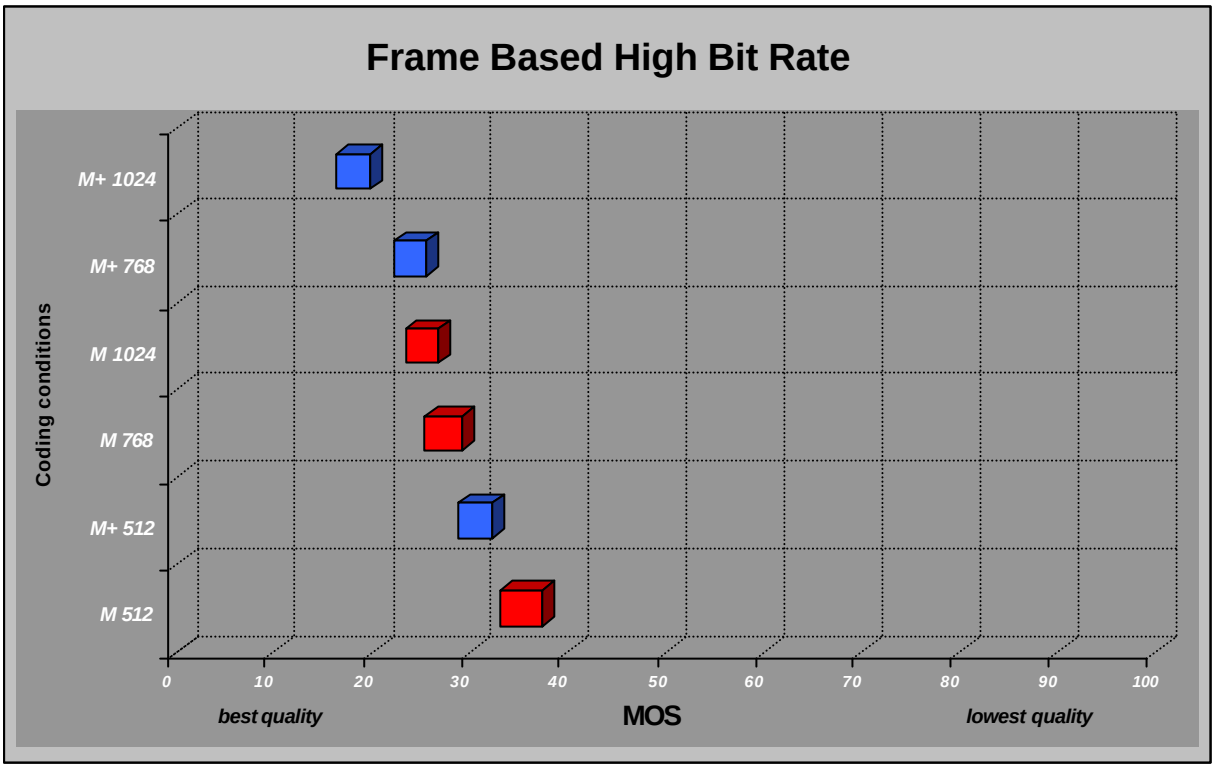
Graph 1 - Graphic representation of the results of the Object Based Low Bit Rate test



Graph 2 - Graphic representation of the results of the Frame Based Low Bit Rate test

In the Frame Based High Bit Rate test the DSCQS test method was used.
Graph 3 should be read as follows:

- The visual quality (MOS) decreases from the left to the right
- Bars of the same colour correspond to the same profile; the Advanced Coding Efficiency (former Main Plus) Profile is labelled as M+ (blue) and the Main Profile is labelled as M (red).



Graph 3 - Graphic representation of the results of the Frame Based High Bit Rate test

7. Conclusions

Graph 1, Graph 2 and Graph 3 show a clear superiority of the Advanced Coding Efficiency Profile compared with the Main Profile. Furthermore:

- Graph 1 shows that, for the Object Based at Low Bit Rate case, the quality provided by the Advanced Coding Efficiency Profile at 256 Kbps equals the quality provided by Main Profile at 384 Kbps.
- Graph 2 shows that, for the Frame Based at Low Bit Rate case, the quality provided by the Advanced Coding Efficiency Profile at 128 Kbps and 256 Kbps equals the quality provided by the Main Profile at 256 Kbps and 384 Kbps respectively.
- Graph 3 shows that, for the Frame Based at High Bit Rate case, the quality provided by the Advanced Coding Efficiency Profile at 768 Kbps equals the quality provided by Main Profile at 1024 Kbps.

Thus, the evidence of a clear superiority of the Advanced Coding Efficiency Profile compared with the Main Profile is given.

8. Participants

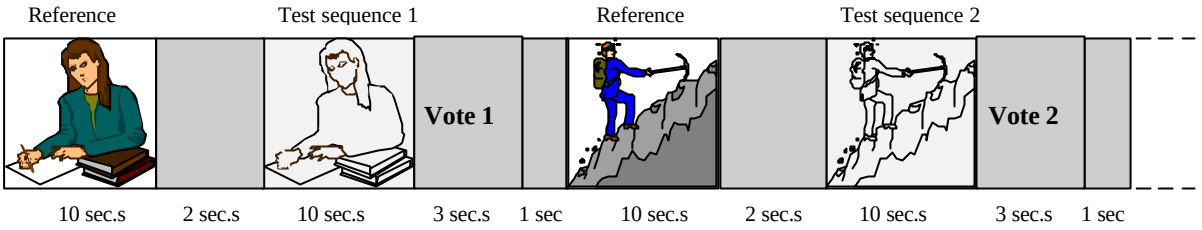
Test material preparation		
Frame Based Low bit rate	S. Takamura; Y. Nakaya, Y. Suzuki	NTT, Hitachi
Frame Based High bit rate	S. Bauer ; U. Benzler	Bosch, Univ. of Hannover
Object Based Low bit rate	G. Heising ; A. Kaup	HHI , Siemens
Conduct of tests	H. Imaizumi, S. Sakaida; A. Schertz	NHK, IRT
Test design, tape editing, result processing	V. Baroncini	FUB
Report	U. Benzler, V. Baroncini	Univ. of Hannover, FUB

Appendix A: Double Stimulus Impairment Scale (DSIS) Method Instructions

Dear Subjects,

Thank you for participating in this test.

In the DSIS tests, a series of pairs of sequences will be displayed on the monitor. This figure describes what you will see and hear during a DSIS test session:



For each pair, the first sequence displayed is the *reference* sequence, and the second is the *test* sequence. Your task is to evaluate the **degradation of the test sequence with respect to the reference sequence**. You do that by marking one and only one box in the following rating scale:

Imperceptible	☐
Perceptible but not Annoying	☐
Sligthly Annoying	☐
Annoying	☐
Very Annoying	☐

Your evaluation must reflect **your opinion of the global degradation** of the **whole** test sequence. Therefore, vote only after the end of the test sequence and base your evaluation on the entire sequence.

Do not hesitate to rate a sequence either at the top or bottom of the scale, if that is how you believe it should be rated.

A voting form will be distributed before this session. On this form will be a series of rating scales like the one above, one scale for each sequence in the test session. All the scales are numbered. Use scale 1 for the first test sequence, scale 2 for the second one and so on.

After you have seen the test sequence N, you will see the message “VOTE N”. Look at the scoring sheet and check for the correct number. Then mark the box corresponding to the **quality level** you have chosen

During these tests do not comment on the sequences you have seen or talk with other assessors.

Before recording your vote, always check to be sure you are using the correct scale on the score sheet.

Finally, it is important that you keep your concentration throughout the test session.

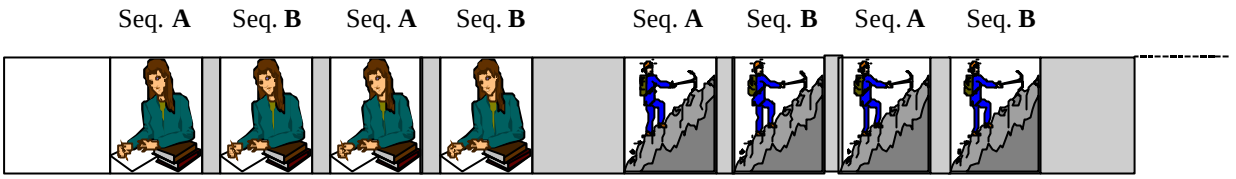
Now try this evaluation procedure in a practice session. You will see a series of sequences using the exact same timing as will be used during an actual test session. This will allow you to become familiar with the timing of the test and to practice using the rating scales.

If you have any questions, please ask them now.

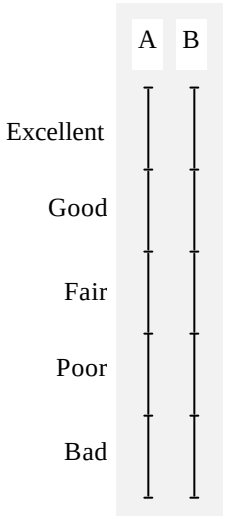
Appendix B: Double Stimulus Continuous Quality Scale (DSCQS) Method Instructions

Dear subjects,

Thank you for participating in this test. In the DSCQS tests, a series of two pairs of sequences will be displayed on the monitor. This figure describes what you will see and hear during a DSCQS test session:



Your task is to evaluate the QUALITY of BOTH test sequences in each pair. You do that by marking one point on each of the following rating scales:



The rating scale on the left is for grading the first sequence (A), the rating scale on the right is for grading the second sequence (B). Please use an horizontal mark (no crosses or other signs that could make your vote ambiguous). Each evaluation must reflect **your opinion of the global quality** of the **whole** sequence. Therefore, only vote after the end of the second sequence and base your evaluation on the entire duration of each sequence. Do not hesitate to rate a sequence either at the top or bottom of the scale, if that is how you believe it should be rated. A voting form will be distributed before each test session. On this form will be a series of rating scales, like the ones above, one scale pair for each sequence. The scale pairs are numbered sequentially. Use scale pair 1 for the first sequence pair, scale pair 2 for the second sequence pair and so on.

The first sequence you will be see will be announced by the message “A” (standing for 2 seconds) then the second sequence you will be see will be announced by the message “B” (standing for 2 seconds) Then the two sequences will be shown again and the messages will change into “A*” and “B*”. Finally you will see the message “VOTE N”. Look at the scoring sheet and check for the correct number. Then mark the box corresponding to the **quality level** you have chosen

During these tests do not talk with other assessors or comment on the sequences you have seen. Before recording your vote, always check to be sure you are using the correct scale on the score sheet. Finally, it is important that you keep your concentration throughout the test session. Now try this evaluation procedure in a practice session. You will see a series of pairs of sequences using the exact same timing as will be used during an actual test session. This will allow you to become familiar with the timing of the test and to practice using the rating scales.

If you have any questions, please ask them now.