INTERNATIONAL ORGANIZATION FOR STANDARDAIZATION
ORGANIZATION INTERNATIONAL NORMALIZATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND ASSOCIATED AUDIO

ISO/IEC JTC1/SC29/WG11/N2825
MPEG99
July 1999

SOURCE: **Test Group**
STATUS: **Approved**
TITLE: **Report of Formal Verification Test on MPEG-4 Advanced Real Time Simple Profile (Error Robustness, Temporal Resolution Stability)**

Summary

This report illustrates the results of the verification test to evaluate the performance of MPEG-4 video error robustness tool and temporal resolution stability tool in the Advanced Real Time Simple profile compared with the Simple profile.

1. Introduction

The visual part of the MPEG-4 standard will provide a toolbox containing tools and algorithms bringing solutions to a number of functionality and covering a wide range of bit rates.
This document describes the test procedures and the results of the video verification test on the Advanced Real Time Simple (ARTS) profile. As the profile has 2 functionalities (the error robustness and the temporal resolution stability), 2 tests were executed separately. The test of the error robustness was carried out in the CCETT and FUB laboratories, and the test of the temporal resolution stability was carried out in the NTT Advanced Technology laboratory.

2. Context and Test Motivation

2.1 ARTS Profile

The ARTS profile provides some improvements over the Simple Profile. These improvements are the following.

- **Workable error condition level** – Same level QoS can be provided under about 10 times worse error conditions over the Simple Profile. This leads the expansion of the service area of the applications.
- **High speed error recovery** – The high speed recovery from the erred states can be provided with keeping high coding efficiency. Artifacts are removed soon even in case of scene changes.
- **Transmission buffering delay and temporal resolution stability** – The transmission buffering delay can be stabilized by minimizing the jitter of the amount of the coded output bits per VOP. Large frame skips are also prevented with the restriction of the small buffer and encoder can control the temporal resolution even in the highly active scene. This is an important technique to implement the low-delay system.

With these improvements, the ARTS profile is particularly suited for the following new applications and expands the related service areas.

1. Communications under the variable error conditions
   - during moving by car or train
   - during congestion on the Internet
2. Communications under bad error conditions
   - under the ground, in the buildings
   - during bad weather on the satellite networks
   - low capacity Internet with many users.
3. Communications of the wide moving scenes
   - with quick moving objects
   - with many scene changes, camera panning scenes
   - using handy camera.

2.2 Test Motivation

The ARTS profile was proposed since the Atrantic City meeting, and has been considered in N2726 ("MPEG-4 New profile under consideration"). As the evidences of the improvements were requested at the Seoul meeting (N2644: Resolution 3.1.6), the formal subjective tests were done.

3. Test of Error Robustness

3.1 System Configuration

In the test of the error robustness, the MPEG-4 video encoder/decoder was combined with the wireless system model shown in Figure 1. This model is almost same to the model in the verification test of the error robustness in Version 1 (N2604). In this system, ITU-T H.223/Annex B was used in the TransMUX layer, and the dummy data was used as the audio data. The bit rate of the audio is 8kbps for all test conditions.

As the video decoder generates the back-channel messages in the ARTS profile, these messages were multiplexed with H.223 and transmitted to the video encoder. The transmission channel errors were applied to the back-channel messages as well as the forward video and audio data.

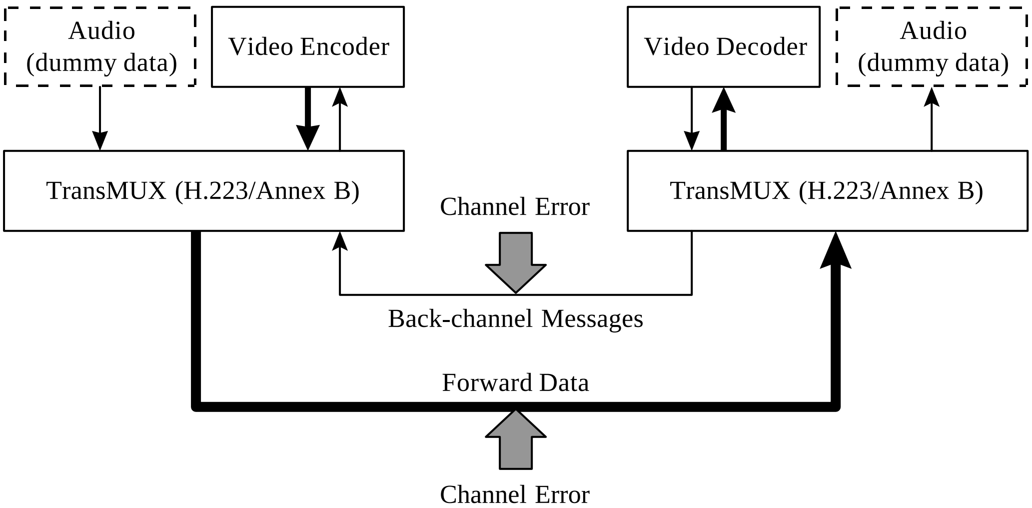


Figure 1 -- Wireless System Model

3.2 Test Conditions

3.2.1 Source Material (Test Sequences)

Two test sequences (each 1 minute long) were selected (see Table 1). They were down-sampled into the input format specified in Table 1 and both of them were handled as 60Hz format.

Name of the sequence	Copyrights owner	Description of the content	Source format (BT.601)	Input format	Target Bit-Rate
Australia	CCETT	News Show	625/50	CIF	64/128kbps
Overtime	NTT DoCoMo	Video Phone	525/60	QCIF/CIF	32/64kbps

Table 1 -- List of Source Materials

3.2.2 Coding Conditions

The following coding tools were used in this test.

Tools	ARTS Profile	Simple Profile
I/P-VOP	✓	✓
AC/DC Prediction	✓	✓
Slice Re-synchronization	✓	✓
Data Partitioning	✓	✓
Reversible VLC	✓	✓
Header Extension Code (HEC)	✓	✓
NEWPRED	✓	
Adaptive Intra Refresh (AIR) and Cyclic Intra Refresh (CIR)		✓

Table 2 – Coding tools

The coding parameters of this test are described in Table 3.

The QP values were not changed in one sequence to evaluate only the degradation by the errors although the rate control was used in this test.
The coding frames were skipped to prevent the overflow of the buffer.
The ARTS profile has higher coding efficiency than the Simple profile, because the NEWPRED can recover from the error states without the forced Intra MBs.
To evaluate this feature, the same QP values were used both in the ARTS profile and the Simple profile. Therefore the video coding rates of the ARTS profile were smaller than that of the Simple profile, and the stuffing bits were inserted to keep the same channel bit rate in the TransMUX layer.
As the errors on the stuffing bits did not cause any degradation, the total number of the errors on the video sequence became small. As a result, the high coding efficiency could be evaluated as the high error robustness under the same QP.
The assumed round trip delay, which is the time from sending the video packet to receiving the back-channel message, was about 300ms for the ARTS profile. This value is appropriate for the typical mobile systems.

	ARTS profile			Simple profile		
	32kbps	64kbps	128kbps	32kbps	64kbps	128kbps
Video Coding Rate	19kbps	48kbps	67kbps	20kbps	51kbps	112kbps
Resolution	QCIF	CIF	CIF	QCIF	CIF	CIF
Target Frame Rate	6fps	6fps	10fps	6fps	6fps	10fps
QP value	10	12	12	10	12	12
Rate Control	Frame skip (fixed QP)			frame skip (fixed QP)		
Video Packet Size	11MBs	44MBs	44MBs	480bits	600bits	600bits
HEC insertion	all video packets			2nd video packet only		
Period of I-VOP	1st VOP only			1st VOP only		
Forced Intra MBs (AIR/CIR)	0/0	0/0	0/0	3/1	10/2	40/4
Round Trip Delay	300ms			-		

Table 3 -- Coding Parameters for Error Robustness Test

3.2.3 Error Conditions

The error conditions of this test are described in Table 4. W-CDMA is one of the promising candidates for the next generation of mobile communication systems, and the length of the burst errors is said to be shorter than that of the conventional systems. Assuming the same bit error rate, the number of the burst errors in the short burst error system is larger than that of the long burst error system, and the packet-error-rate of the short burst error system is higher than that of the long burst error system. Therefore the short burst error conditions were evaluated as "very critical error condition" in this test.

Name	Avaraged Bit Error Rate	Length of Burst Error
Critical Error Condition	10^{-3}	10ms
Very Critical Error Condition	10^{-3}	1ms

Table 4 -- Error Conditions

Error sequences were generated using software supplied by NTT DoCoMo. As the image degradation by the errors depends on the error pattern, 25 kinds of error patterns were simulated for each condition in this test. The typical error pattern was automatically selected so that the produced PSNR was the nearest one to the averaged PSNR over all error patterns without any manual adjustment, and the typical error pattern was finally used for the subjective evaluations.

3.2.4 Display Format

The decoded video sequences with transmission channel errors and without errors (reference) were displayed in CIF format side by side in the center of a mid gray ITU-R BT.601 frame (60Hz). The reference was located on the left hand side, and the sequences decoded in presence of errors were located on the right hand side, as shown in Figure 2 The display format was always CIF. The QCIF sequences were up-sampled into CIF format.

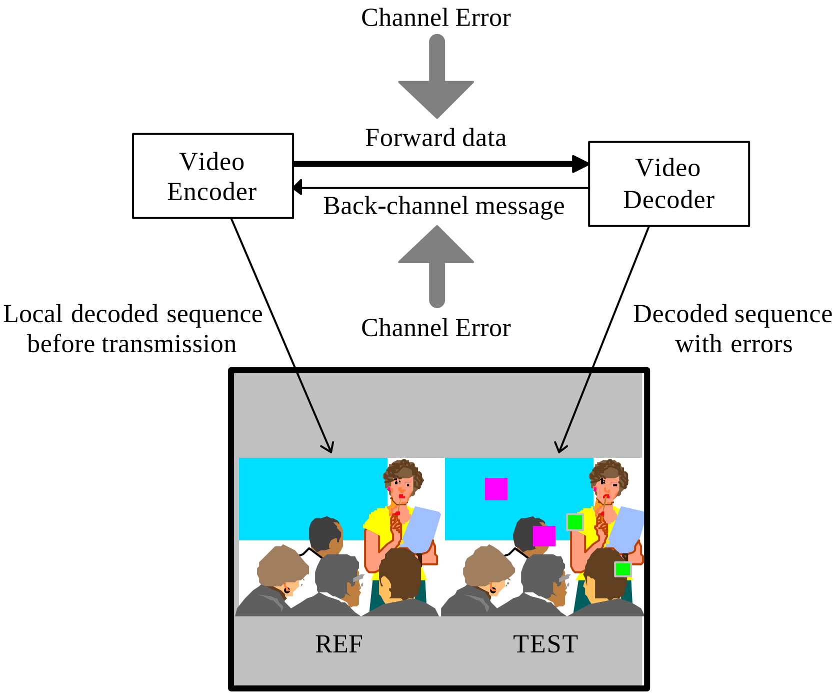


Figure 2 -- Example of Display Format for Error Robustness Test

3.3 Test Method

The "Error Robustness" functionality of the MPEG-4 ARTS profile has been subjectively assessed using the DSCQE (Double Stimulus Continuous Quality Evaluation) test method.

This test method was proposed to MPEG in previous documents to evaluate the error robustness at very low bit rate, and it has been already successfully used to assess the "Error Robustness" functionality of the MPEG-4 Simple profile in Version 1.

3.4 Continuous Evaluation

The continuous evaluation was brought to the ITU-R because the previous methods presented some inadequacies to the video quality measurement mainly in the case of digital compression schemes. The previous test methods are based on the test sequence limited to 10 seconds (not enough to represent a real service). Digital artefacts are strongly depending on spatial and temporal content of the source image. This is true for the compression schemes but also concerning the error resilience behaviour of digital transmission systems. Until now it was very difficult to choose representative video sequences, or at least to conclude about their representativeness. The probability of occurrence of the impairments was not a controlled feature.

When low bit rate coding is applied it is usually important to evaluate the *fidelity* of the transmitted information more than the basic quality of the images, as in the cases of recovery of strong errors on the network. To obtain the desired fidelity evaluation, the panel of viewers watches at two displays (or at two windows into the same display), one of them shows the encoded-decoded video sequence without any transmission errors, the other shows the same encoded-decoded video material under conditions of transmission errors. The viewers are asked to check the differences and to judge the difference in quality between the left (reference) and the right frame.

This evaluation is done by each viewers by moving a slider, ranging from *'imperceptible'* (i.e. top of the scale, coded as 100) to *'very annoying'* (i.e. bottom of the scale, coded 0).

A typical DSCQE test session is made of one or more test sequences, each of them standing for at least 1 minute. Before the test session begins, the viewers receive a detailed explanation of the test protocol and of how they have to assess the differential fidelity of the two sequences displayed.

The votes are sampled every 500 ms the status each slider. The collection of the data was done automatically using a PC, properly connected to the slider.

3.5 Laboratory set-up

The set-up of the test laboratories has to be done according to the recommendations BT.500-7 (ITU-R). One or more 19" professional monitor was used (e.g. Sony BVM1910) to display the sequences.

The viewers were located in front of the monitor at distance equal to 4 times the height of the screen.

Each test was done using a panel of 18 non-expert viewers. Three of them were located in front of each monitor.

3.5.1 Laboratory set-up at CCETT

The subjective assessment environment consists of two different rooms. The first is the "viewing room" where tests are performed, while the second one is the "technical room" containing VTRs and other equipment necessary to perform tests.

The equipment

Test sequences were recorded on a D1 tape and displayed to the observers on a Sony PVM 2054 QM using a D1 Sony VTR controlled by the IQ++ platform including hardware and software and described bellow. The monitors were calibrated using a Minolta Luminance meter, a Photo Research spectrophotometer and the standard Pluge signal.
Votes were collected using the IQ++ sliders.
Six groups of 3 non-expert observers were participated to the test, organised into 2 sessions. Observers were different for 50 and 60 Hz tests. Prior to the test, subjects were screened for visual acuity by using a Monoyer Optometric Table. Besides, test for normal colour vision was performed using Ishihara's tables. These tables are designed to detect colour blindness or strong colour vision deficiencies; that is to say, to check ability of the observers in discriminating colours. The subject who is not able to pass the test is discarded.
All the viewing conditions were compliant with the ITU-R B.T. Rec. 500.

The software

The software platform, called IQ++, has been developed and commercialised by CCETT in the framework of the RACE/MOSAIC project. It allows processing of the results of subjective evaluations based on the most used methodologies described in ITU-R Rec. B.T. 500. The high level of modularity in the conception of IQ++ allows easy update and build-up of new test methods, which may be as an adaptation of existing methods, or the result of brand new research in methodology.
IQ++ is composed of three modules: the 'Test Preparation' module, the 'Test Driver' module and the 'Result Processing' module, running on Windows'95.

"Test Preparation" Module

This module deals with the building-up of a test, from:

- The selected method (if known, otherwise a new one can be created)
- The list of test conditions(algorithms, codecs),
- The list of images or video sequences.

From this data the software produces the test, creating in particular, the random list of different cells. (This list must be compliant with the rules associated to the chosen test method).
This module also produces a file which is able to drive a VTR for the automatic editing of a video tape, according to the test characteristics.
Finally, this programme provides the possibility to manipulate the test frameworks. These frameworks describe the test method and are abstractions of the tests themselves.
Standardised methods are provided in the form of predefined test frameworks, but the user can edit these structures to create his own modifications. This module is capable of managing these new methods in the same way it does the standardised methods. A personal library of methods can easily be built.

"Test Driver" Module

This module allows to drive a test using the information from the previous module.
The test structure, based on key-concepts such as session, presentation and results, gives the user broad flexibility. It is therefore possible carry out the test several times for different observers, e.g. multi results test. It is possible also to split a test into several sessions and to carry it out, session by session, for different groups of observers, e.g. multi session test, or to repeat the same test for the same observers, e.g. multi-presentation test. The combination of these different functionality can provide a solution for all possible configurations.
During the test the VTR is set on the position "remote control" and it is fully controlled by the IQ++ software, but the test can also run asynchronously, if the playing system has not a remote control with the appropriate protocol.

"Results Processing" Module

This module deals with the processing of the results
The format of the input data file is compliant with the last developments of the recommendations 500. This obviously means that the output data file format of the previous module "test driver" is also compliant.
The "Results Processing" module allows the processing of raw data and the presentation of the results in the simplest possible manner, notably owing to the use of graphs. This module is very easy to be used because, for example, it is based mainly on contextual menus. It allows different form of processing to be associated to the same test. Each form of processing is itself composed of a series of operations. Parameters are included in each operation, such as scale range modifications, thresholding, etc.
From raw data the user is able to choose the "by default" form of processing, based on the type of test, if standardised, or to generate the processing list himself by selecting different operations.

The results are displayed by a piece of external graphic software. The current software used Microsoft Excel, but it is possible to create a link with other software.

3.5.2 Laboratory set-up at FUB

The subjective assessment environment consists of four different rooms. The first and the second rooms are the "listening/viewing room" where tests are performed, the third one is the "audio technical room" containing a digital audio recorder and the other equipment necessary to perform audio tests and the fourth one is the "video technical room" containing the DVTR two video disks and the other equipment necessary to perform tests video and audio visual tests.

The equipment

Test sequences were recorded on four D1 tapes and displayed to the observers on two Sony BVM 20E1/E grade A professional studio monitor using a D1 BTS/Philips DVTR controlled by the IQ++ platform including hardware and software and described in the previous section. The monitor were calibrated using a Sony custom probe and the standard Pluge signal. Votes were collected using the IQ++ sliders.

Six groups of 3 non-expert observers were participated to the test, organised into 2 sessions. Observers were different for 50 and 60 Hz tests. Prior to the test, subjects were screened for visual acuity by using a Monoyer Optometric Table. Besides, test for normal colour vision were performed using Ishihara's tables. These tables are designed to detect colour blindness or strong colour vision deficiencies; that is to say, to check ability of the observers in discriminating colours. The subject who is not able to pass the test is discarded. All the viewing conditions were compliant with the ITU-R B.T. Rec. 500.

The software

The software platform is the IQ++, developed and commercialised by CCETT and described in previous section. The results have been further processed using Microsoft Excel to obtain some comparison of the influences among session and presentations.

3.6 The training phase

The training phase was a crucial part of this test, since subjects could misunderstand their task. Written instructions (cf. Annex A) were provided to be sure that all the subjects receive exactly the same information. They included explanation about what the subjects were going to see, what they had to evaluate (i.e. difference in quality) and how they had to express their opinion. Any question from the subjects were answered in order to avoid as much as possible any opinion bias from the test administrator.

After the instructions, a training session was run. In this way subjects were made acquainted both with the voting procedures and kind of errors. During the training session a number of representative conditions were shown. They included the following test conditions:

- no transmission error (i.e. encoded/encoded at 32 kbit/s) - 0.5 minutes
- typical errors at 128 kbit/s bitrate - 0.5 minutes
- critical errors at 384 kbit/s bitrate - 0.5 minutes
- critical errors at 32 kbit/s bitrate - 0.5 minutes

The training sequences and their presentation order were the same in all the test sites. The sequences must be different from those used in the test and they should be played one after the other without any interruption.

When the training session is finished, the experimenter should check that in the case of no errors the evaluations are close to one hundred, if they are not he should repeat the explanation and repeat the training session.

3.7 Experimental design

The experimental design of this test was done taking into account the experience gained in the previous test performed to assess the "Error Robustness" functionality in the MPEG-4 Simple profile.

To improve the results of the test and to make the test less tiring to the subjects, it has been decided to reduce the single video segment length to one minute; at the same time it has been decided to have at least one coding condition common to the two test material used (see Table 5).

The sequences were recorded on a D1 tape one after the other without any interruption between two subsequent sequences. The table below shows the presentation order used in the test.

Presentation order	*Sequence*	*Bit rate*	*Error condition*
1	Australia	64	ARTS Very Critical
2	Overtime	32	ARTS Critical
3	Australia	64	Simple Very Critical
4	Australia	128	ARTS Critical
5	Overtime	32	ARTS Very Critical
6	Overtime	64	Simple Critical
7	Australia	128	Simple Very Critical
8	Overtime	32	Simple Critical
9	Overtime	64	ARTS Critical
10	Overtime	32	Simple Very Critical
11	Australia	128	ARTS Very Critical
12	Australia	64	Simple Critical
13	Overtime	64	Simple Very Critical
14	Australia	64	ARTS Critical
15	Overtime	64	ARTS Very Critical
16	Australia	128	Simple Critical

Table 5 – Presentation Order Used in ER Test

A total of 36 subjects participated in the test in two different test site; the test were run at 60 Hz .

3.8 Test Results

The result of the test of the "Error Robustness" functionality are her below represented in four separate figure (from figure 3 to figure 6). Figure 3 and Figure 4 show the results of the tests done in Lab. A and B for the "Very Critical" error condition. Figure 5 and Figure 6 show the results of the tests done in Lab. A and B for the "Critical" error condition. In all the four graphs the red line represents the behavior of the Simple profile, while the blue line represents the behaviour of the ARTS profile.

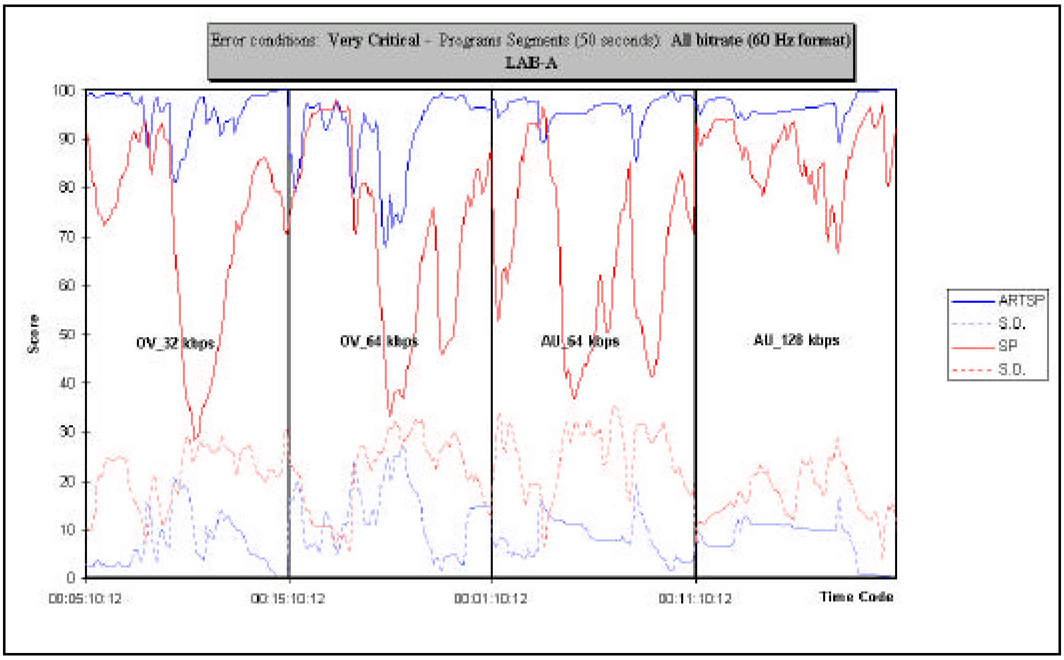


Figure 3 -- Results of ARTS Profile vs. Simple Profile Test (ER-1)

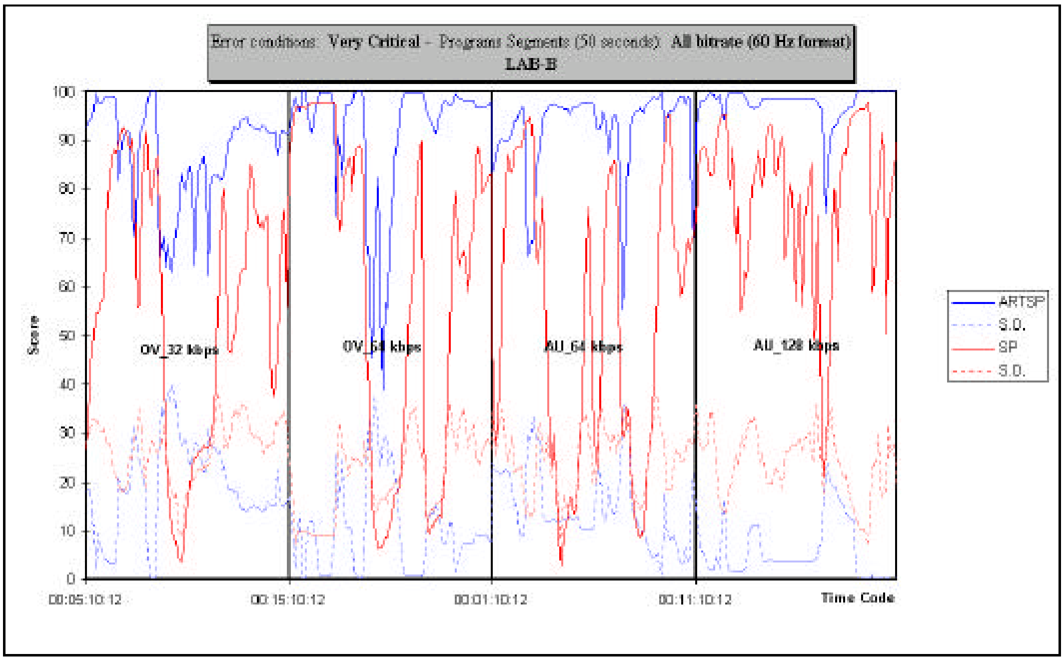


Figure 4 -- Results of ARTS Profile vs. Simple Profile Test (ER-2)

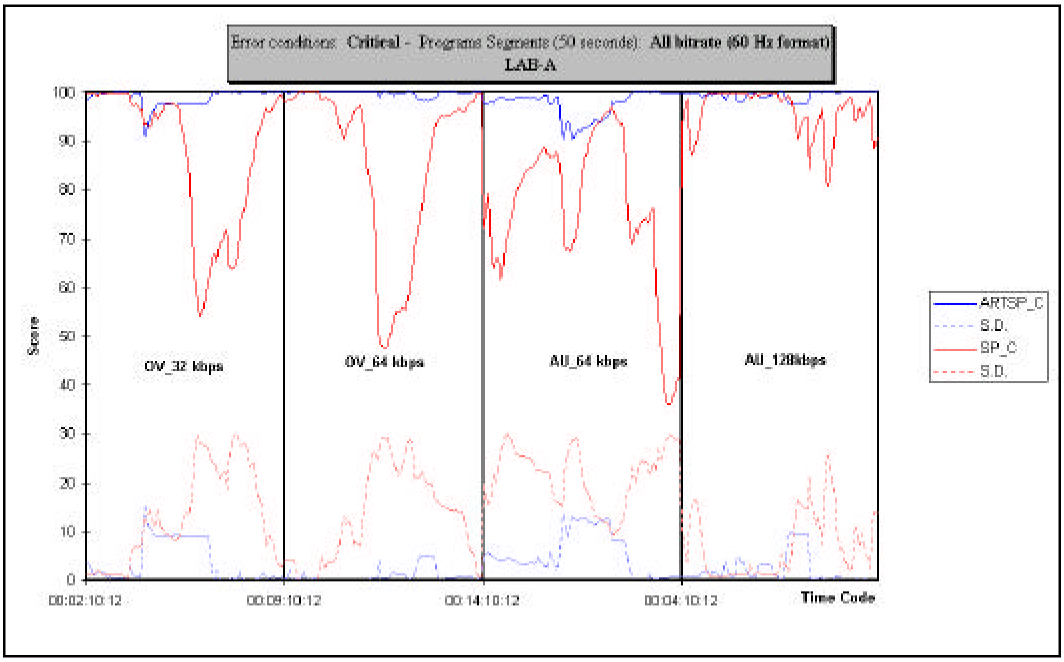


Figure 5 -- Results of ARTS Profile vs. Simple Profile Test (ER-3)

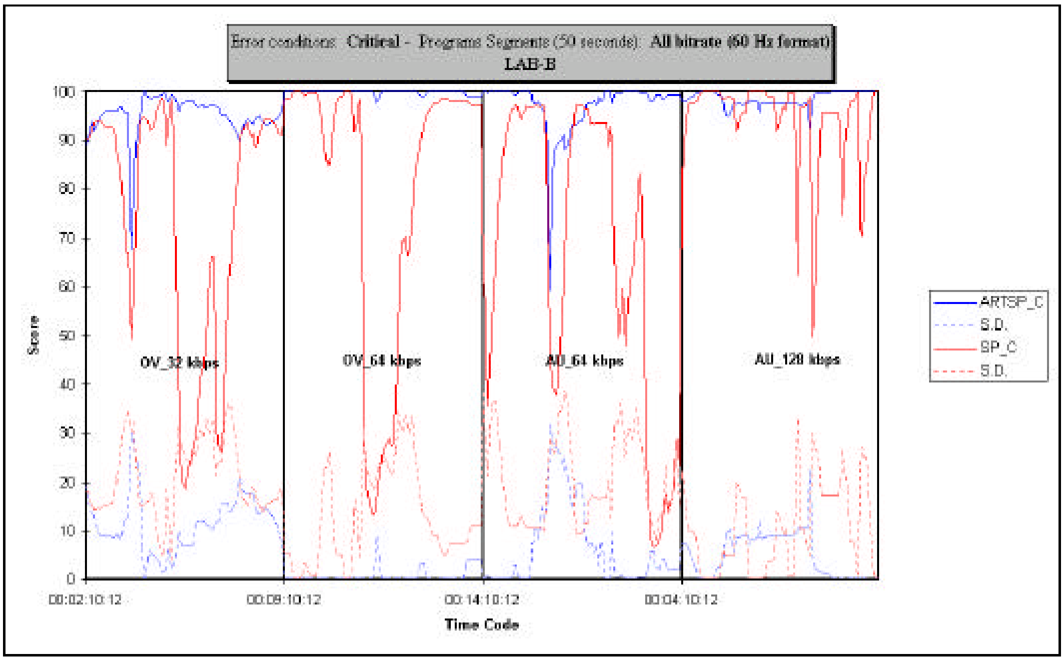


Figure 6 -- Results of ARTS Profile vs. Simple Profile Test (ER-4)

In all the figures the dotted lines represents the behaviour of the associated Standard Deviation.
The values close to 100 indicate that no (or few) transmission errors have been detected (good behaviuor of the Error Robustness algorithm); lower values indicate higher degradation of the images due to transmission errors.

4. Test of Temporal Resolution Stability

4.1 Test Conditions

4.1.1 Source Material (Test Sequences)

Four test sequences of Australia, Overtime, Foreman and Crowd (background of Akiyo with Crowd sequence) were selected for this temporal resolution stability test. In this case the Australia and Overtime sequences have been obtained extracting a 10 seconds spot of highly active scene form the original long sequences already used in the previous error robustness tests (for Australia sequence it was selected the scene in which a man stands up and goes to the map, and for the Overtime sequence it was selected the scene in which a man comes back and sits down). All of them were handled in 60Hz format.

4.1.2 Coding Conditions

The used coding tools are listed in Table 6 .

Tools	ARTS Profile	Simple Profile
I/P-VOP	✓	✓
AC/DC Prediction	✓	✓
Dynamic Resolution Conversion(DRC)	✓	

Table 6 -- Coding tools for Temporal Resolution Stability Test

The temporal resolution stability test for ARTS profile vs. Simple profile was performed according to the coding conditions specified in the table here below.

Sequences	Australia, Overtime, Foreman, Crowd		
Resolution	CIF (352x288)		
Bit rate	64 kpbs	96 kbps	128 kpbs
Target frame rate	10 Hz for Australia and Overtime, 5 Hz for Foreman and Crowd	15 Hz for Australia and Overtime, 5 Hz for Foreman and Crowd	15 Hz for Australia and Overtime, 7.5 Hz for Foreman and Crowd
Period of I	1st VOP only		
Rate control	MPEG-4 MB based rate control with frame skip		

Table 7 -- Coding Conditions for Temporal Resolution Stability Test

4.1.3 Display format

The decoded video sequences were displayed in CIF format and inserted in the center of a mid gray ITU-R BT.601 frame (see Figure 7).

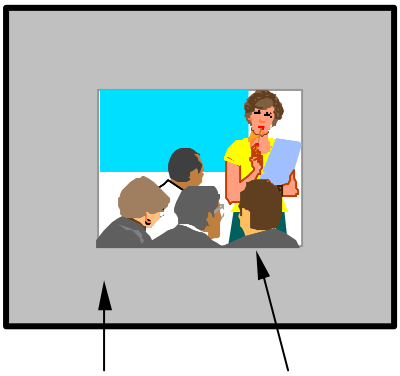


Figure 7 -- Example of Display Format for Temporal Resolution Stability Test

4.2 Test Method

To perform the test of Temporal Resolution Stability the Single Stimulus test method has been selected.
The test has been performed at the NTT-AT test laboratories.
The laboratory set-up has been designed according to the Recommendation ITU-BR 500-7.
A total of 18 subjects have participated to the test.

The equipment

Test sequences were recorded on a D1 tape and displayed to the observers on two Sony BVM 2011 studio monitor using a D1 Digital VTR. The monitor were calibrated using the standard Pluge signal.
Votes were collected manually using a voting sheet.
Six groups of 3 non-expert observers participated to the test, organised into 3 sessions. Prior to the test, subjects were screened for visual acuity by using a Monoyer Optometric Table. The subject who is not able to pass the test is discarded.
All the viewing conditions were compliant with the ITU-R B.T. Rec. 500-7.

Instructions to the subjects

Before starting the tests the subjects are properly instructed on the task they are supposed to do. To avoid any possible bias, the instruction are read from a printed paper (see example in Appendix B). In this way all the subjects receive the same instructions.

The training phase

A subjective assessment test must include a Training phase, during which the subjects try a short test session that reproduces the same condition of the real test.
When the training section is finished, the experimenter should check eventual errors and answer to the subjects question, if any.

Test organisation

The test session was made up of all the possible combinations of sequences and coding conditions; further more 5 dummy combinations have added at the beginning of each session, to allow the stabilisation of the opinions of the subject. The votes collected during the stabilisation phase has been discharged.

4.3 Test Results

Figure 8 shows the results of the test of the Temporal Resolution Stability functionality of the Advanced Real Time Simple Profile.
For each bit rate a different colours has been used.

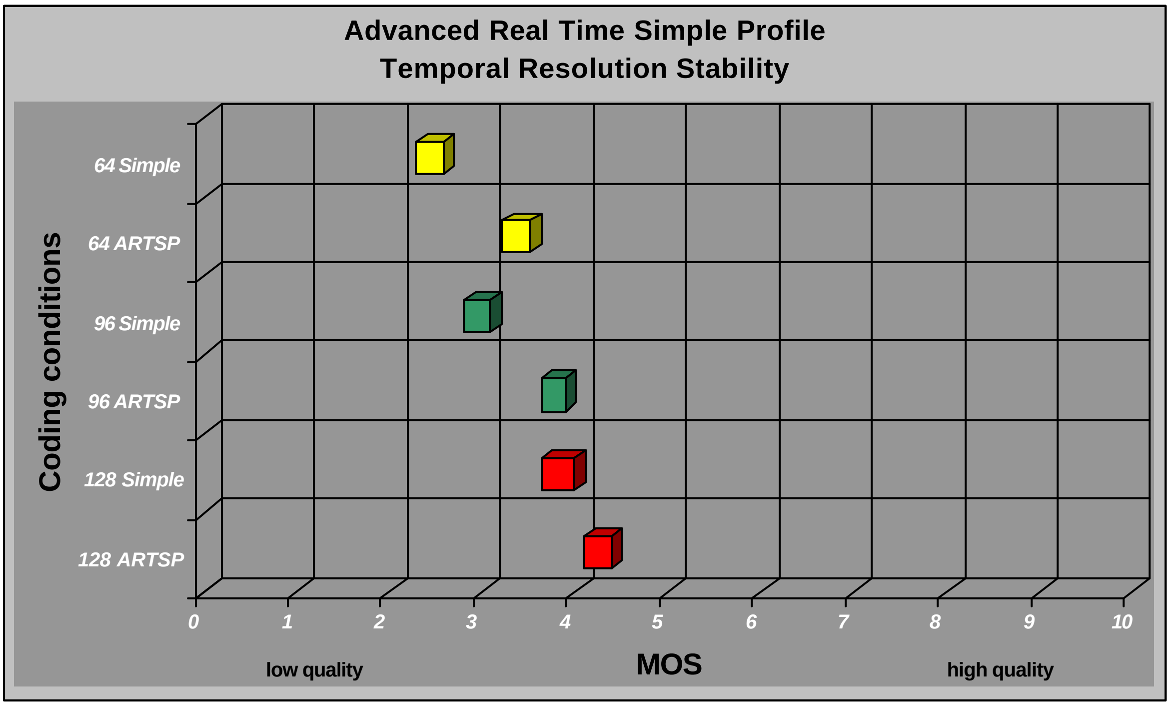


Figure 8 -- Results of ARTS Profile vs. Simple Profile Test (TRS)

5. Conclusions

5.1 Error robustness results

The four graphs reported above show a clear superiority of the Advanced Real Time Simple Profile compared with the Simple Profile. This superiority was clear in both laboratories and for both the error cases ("critical" and "very critical"). More in detail the ARTS Profile outperforms SP in the recovery time from transmission errors. Furthermore ARTS Profile in the "Critical" error condition provides results that for most of the test time are close to a complete transparency, while SP is still severely affected by errors.

5.2 Temporal resolution stability results

There is evidence of difference in quality between ARTS Profile and Simple Profile; furthermore the results show that ARTS Profile at 64 Kbps outperforms Simple Profile at 96 Kbps and that ARTS Profile at 96 Kbps performs as well as Simple Profile at 128 Kbps.
The results reported above provide a clear indication of the superiority of Advanced Real Time Simple Profile when compared with the Simple Profile.

6. Participants

Test material preparation	S. Fukunaga, H. Kimata	Oki, NTT
Test tape editing	V. Baroncini	FUB
Formal subjective tests	S. Pefferkorn, V. Baroncini	CCETT, FUB
Results processing	S. Pefferkorn, V. Baroncini	CCETT, FUB
Report	V. Baroncini, S. Fukunaga, H. Kimata	FUB, Oki, NTT

Table 8 -- Participants for Error Robustness Test

Test material preparation	E. Morimatsu	Fujitsu Laboratories
Test tape editing	E. Morimatsu	Fujitsu Laboratories

Formal subjective tests	I. Yoroizawa	NTT-AT
Results processing	V. Baroncini	FUB
Report	V. Baroncini, E. Morimatsu	FUB, Fujitsu Laboratories

Table 9 -- Participants for Temporal Resolution Stability Test

Annex A - DSCQE Method Instructions[1]

Dear Subjects,

This experiment requires the evaluation of the differences that you will be able to see among two video scenes put on the left and right windows of the screen respectively.

The left windows is labelled as "REF" and the right windows is labelled as "TEST".

Both REF and TEST sequences are affected by coding impairments, more or less visible. Furthermore the TEST sequence is sometimes affected by extra impairments (namely those due to the transmission of the video over different media).

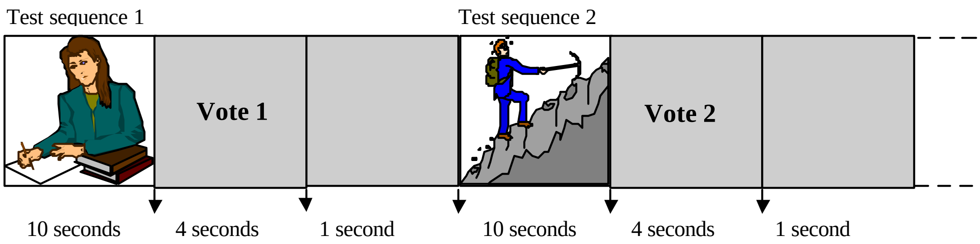
The actual test will engage you for about 30 minutes, during this time you will have to move a slider from the top to the bottom, and viceversa, to express the PERCENTAGE OF DIFFERENCES that you see between REF and TEST sequences.
If you do not see any difference move the slider to the top. As far as you see a difference, move the slider down proportionally to the amount of the impairments that you see.
Now we will show some sequences similar to those that you have to evaluate and we will illustrate what kind of errors you can expect to see.
Then a short training experiment will help you to familiarise with the test, with the use of the sliders and with the evaluation process.

1 This instructions were translated to the languages of the countries where the test labs sites are located

Annex B - Single Stimulus (SS) Method Instructions

Dear Subjects,

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



Your task is to evaluate the **<u>quality</u> of each sequence,** by marking one and only one box in the following rating scale:

	10	☐
Excellent	**9**	☐
	8	☐
Good	**7**	☐
	6	☐
Fair	**5**	☐
	4	☐
Poor	**3**	☐
	2	☐
Bad	**1**	☐
	0	☐

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