INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 **N3483**
July 2000, Beijing, China

Title: **Call for Evidence Justifying the Testing of Audio Coding Technology**

Source: **Audio Subgroup**

Status: **Approved**

1 Introduction

In mid-1999 International Standard ISO/IEC 14496-3, MPEG-4 Audio Version 1 issued and in early 2000 ISO/IEC 14496-3 / AMD1, MPEG-4 Audio Version2 issued. Numerous tests have been conducted by MPEG (see references) to verify that the MPEG-4 standard contains state of the art technology. However, WG11 is always interested in new developments which may provide improvements over the existing MPEG-4 standard and which may lead to extensions of MPEG-4 or to new work items. For this reason MPEG seeks input on such new technology.

Therefore, WG11 issues with this document a call for evidence justifying the subjective testing of new audio coding technology in comparison with the MPEG-4 audio coding technology.

Interested parties are asked to provide at the 55th MPEG meeting, to be held January 15-19, 2001 in Eilat Israel, clear evidence that their technology outperforms MPEG-4 technology (see detailed timetable, below). In the spirit of MPEG-4, it is of greatest interest if the new technology demonstrates both compression and other functionality. WG11 shall judge the submitted material to assess if the proposed technology represents a significant enough improvement to warrant further quality assessment via a formal subjective test.

If there is such a need, these tests will be defined by WG11 and conducted under controlled conditions. Should a formal subjective test be conducted, proponents of the technology will be requested to underwrite the cost. Responding to this call does not imply any commitment on the part of the proposer; a decision to take part in the formal testing process can be made when the results of the Eilat meeting are available. Results of a formal subjective test will be made public, but WG11 cannot, prior to having the results of the test, commit to any course of action regarding the proposed technology.

In order to prepare for evaluations of proposed technologies at the January 2001 meeting, proposers are kindly requested to do the following:

Registration:

Register by 18th October 2000 an intention to compete with MPEG-4. Register by sending an email to Schuyler Quackenbush (Chairman of the MPEG Audio Subgroup, srq@research.att.com). Email should indicate contact names, company and the technology in MPEG-4 Audio that will be considered (e.g. audio coding, scalable coding or wideband speech coding).

Documentation:

Submit by 1st December 2000, the following:

the bitstreams, decoders and decoded sound files (*.wav) associated with the proposed algorithms. At the same time the corresponding items for the MPEG-4 technology will be submitted and made available. Decoders shall be delivered as executables on any commonly available computing platform.

Submit by 20th December 2000, the following:

the documents that describe the performance of proposed algorithms in comparison to MPEG-4 technology.

The proposer's documents should be written in Microsoft Word97 and submitted via email to Schuyler Quackenbush. These documents will be uploaded to the MPEG document site as an input to the January MPEG meeting. The bitstreams, decoder executables and *.wav files should be uploaded to an FTP site indicated to a proposer in response to their registration. These files will be made accessible to members of WG11 as they become available.

Proposers should base their evidence on "reference quality" MPEG-4 encoders, similar to those that have been used in the MPEG verification tests. In most cases the publicly available MPEG-4 encoder software is not able to deliver "reference quality." Proposers should contact Schuyler Quackenbush for information on how to obtain both test material and "reference quality" encoded test material. There may be a reasonable fee associated with access to the reference quality encoder or encoded material. Every effort will be made to share these expenses amongst all the proposers that benefit.

It is required that the MPEG-4 coders used in the comparison be compliant MPEG-4 coders and that they adhere to test conditions and bitrates described below.

WG11 must emphasize that evidence presented as part of this call should not be interpreted as definitive subjective quality assessments. Such interpretations require an appropriately designed and conducted formal subjective test.

Participation:

Attend the MPEG meeting, 15-19 January 2001 in Eilat Israel. It is strongly urged that experts familiar with the proposed technology attend in order to allow discussions on details of the proposals. Proposers should bring to the meeting the decoded audio material associated the comparison between proposed technology and MPEG-4 technology on some appropriate media (e.g. DAT tape, audio CD or *.wav files) along with hardware to play that media for evaluation purposes.

2 Test Material & Test Methodology

For the demonstration of evidence, the same quality assessment methodology shall be used as was successfully developed and employed for the MPEG-4 Core Experiment process (N1748). This is described in the following guidelines:

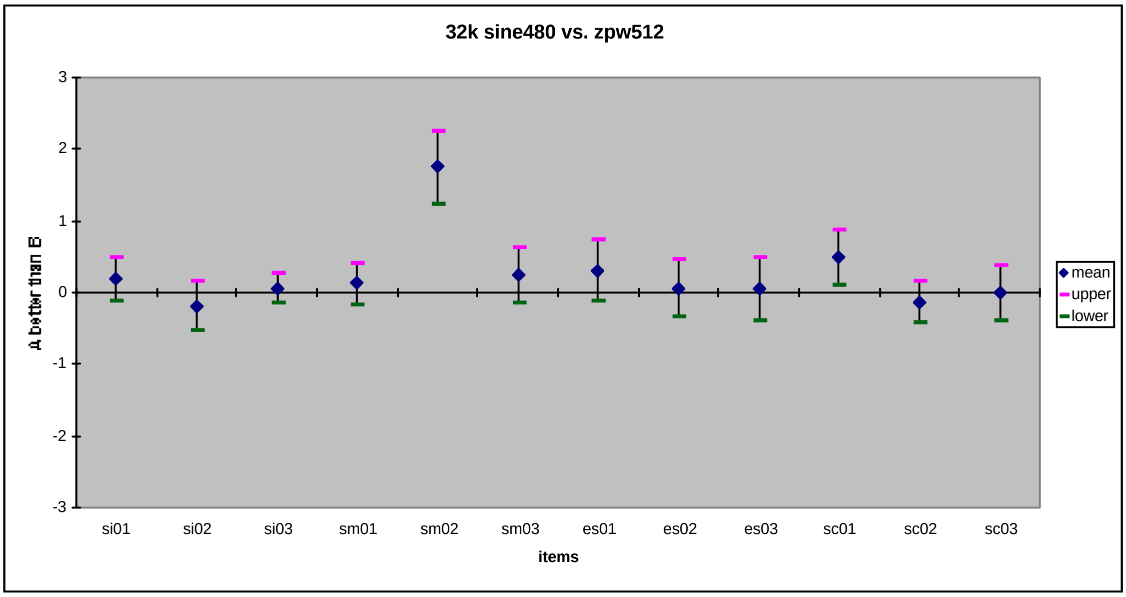
- The test methodology uses the Comparison Mean Opinion Score (CMOS) test. The sequence played to the listeners for each trial is Ref/A/B, Ref/A/B, where Ref is the original uncoded signal and A and B are both coded signals. For each test item, if A is the signal coded using the proposer's codec then B is the signal coded with the MPEG-4 reference codec, or the converse. For speech coders, each trial is A/B rather than Ref/A/B, Ref/A/B. For audio coders, the bandwidth of the reference signal (Ref) should be chosen such that it does not exceed the bandwidth provided by the coders under test by an unacceptable degree.

- Manual tuning is not permitted (i.e. there shall be no adaptation of coding parameters or algorithms for specific test items).
- The assignment of codecs to positions A and B is randomized on a per-item-basis and is unknown to the listener (“blind test”).
- To compensate for positional effects, each pair of signals is presented twice such that the signal A in the first comparison is presented as signal B in the second comparison.
- The seven-grade comparison scale is used (attributes: “A is much better than B”, better, slightly better, equal, slightly worse, worse, much worse). The listeners are asked to give integer grades (i.e. not to use decimal places). For speech coders, only the range of –2 ... 2 is used.

Comparison of the Stimuli	Score
B is much better than A	+3
B is better than A	+2
B is slightly better than A	+1
B is the same as A	0
B is slightly worse than A	–1
B is worse than A	–2
B is much worse than A	–3

Seven point comparative grading scale

- The playback should be done using Stax Lambda Pro and Stax Lambda Nova headphones in a controlled (acoustically isolated) environment.
- A minimum of 8 listeners is required to support a basic level of statistical significance.
- Training is required to make listeners familiar with the test procedure and with the range of distortions that are representative of the processed test set.
- The results of the listening tests are to be given by the average scores and the 95 % confidence interval. An example listening test result is given below:



- A minimum of two test sites must be used, one of which is a company that is independent from the proposer. Each site will report its test results separately.
- The following test material will be used in presenting the evidence:
    1. speech signals [es*, j*]
    2. single instruments (monophonic, i.e. one note sounding at a time) [si*]
    3. simple sound mixtures (material with. several notes sounding at a time) [sm*]
    4. complex sound mixtures [sc*]

For coders claiming to address mono or stereo general audio signals, the following test set shall be used:

Test Item	Description
es01	vocal (Suzan Vega)
es02	German speech
es03	English speech
si01	harpsichord
si02	castanets
si03	pitch pipe
sm01	bagpipes
sm02	glockenspiel
sm03	plucked strings
sc01	trumpet solo and orchestra
sc02	orchestral piece
sc03	contemporary pop music

This material is available at 48 kHz sampling rates. Proposers can create other sampling rates by using the **ResampAudio** sample rate conversion tool:

HYPERLINKhttp://www.tnt.uni-hannover.de/soft/audio/packages/afsp/

For coders claiming to address multi-channel general audio signals, the following test set shall be used:

No.	Name	Description
1	pitch_pipe	Pitch Pipe
2	harpsichord	Harpsichord
3	triangle	Triangle
4	cast_pan1	Castanets panned across the front, noise in surround
5	elliot1	Female and male speech in a restaurant, chamber music
6	mancini	Orchestra - strings, cymbals, drums, horns
7	station_master1	Male voice with steam-locomotive effects
8	clarinet_theatre	Clarinet in centre front, theatre foyer ambience, rain on windows in surround
9	thalheim1	Piano front left, sax in front right, female voice in centre
10	glock	Glockenspiel and timpani

This material is available at 48 kHz sampling rates. Proposers can create other sampling rates by using the **ResampAudio** sample rate conversion tool.

For speech coders the test material is restricted to speech-dominated material. The following test set shall be used:

Test Signal	Signal Type
es01	English/German Speech
es02	English/German Speech
es03	English/German Speech
es04	English/German Speech
es05	English/German Speech
es06	English/German Speech
es07	English/German Speech
es08	English/German Speech
es09	English/German Speech
mp4_08	English multiple Speaker
js01	Japanese Speech
js02	Japanese Speech
js03	Japanese Speech
js04	Japanese Speech
js05	Japanese Speech
js06	Japanese Speech

js07	Japanese Speech
jb02	Japanese Speech with Background Noise
jm01	Japanese Speech, Multiple Speakers
jp01	Japanese Speech, Sentence Pair

This material is available in both 8 kHz and 16 kHz sampling rates.

At least one test site must evaluate English/German test signals using English or European language speakers. Similarly, at least one test site must evaluate Japanese test signals using Japanese language speakers. For each test site, results must be reported in two segments:

1) English/German test signals as evaluated by English or European language speakers, or Japanese test signals as evaluated by Japanese language speakers.

2) All test signals as evaluated by all listeners (at that test site).

3 Test Areas

The coding conditions for MPEG-4 coders that are to be used for comparing coding efficiency are given in this section. The proposer does not have to test all coding conditions of the MPEG-4 coders, although multiple coding conditions may provide more compelling evidence.

3.1 Low, Medium, and High Bitrate Coding Efficiency

The audio part of the MPEG-4 standard provides a toolbox containing tools and algorithms covering a wide range of bit rates. The coding conditions for audio coding and for speech coding are listed in the following two tables:

number of channels	bit rate per channel
1/2/5	64 kb/s
1/2	48kb/s
1/2	32kb/s
1/2	24kb/s
1/2	16 kb/s
1/2	8 kb/s

Coding Conditions for the Audio Coding Efficiency Test

signal sampling rate	number of channels	bit rate
16	1	24 kb/s
16	1	16 kb/s
8	1	12 kb/s
8	1	6 kb/s
8	1	2 kb/s

Coding Conditions for the Speech Coding Efficiency Test

3.2 2.3 Scalable Coding

Bitrate scalability bitstreams consists of multi-layer bitstreams, for example, a base layer bitstream and multiple enhancement layer bitstreams. The coding conditions for MPEG-4 coders that are to be used for comparing coding efficiency and bitrate scalability are given below.

Base layer rate and number of channels	Number of enhancement layers	Enhancement layer rates and number of channels
24 kb/s, mono	2	16 kb/s, stereo
64 kb/s, stereo	4	8 kb/s, stereo

3.3 Robustness in Error Prone Environments

To support communication over noisy channels, MPEG-4 has technology that provides both unequal rate forward error correction and also bitstream formats that are resilient to bit errors. The coding conditions for MPEG-4 coders that are to be used for comparing coding efficiency and error robustness are given below.

Bitrate and number of channels	Channel error conditions
96 kb/s, stereo	Critical and Very Critical
16 kb/s, mono	Critical and Very Critical

The error conditions of this test are described in the table below. Bursty error sequences are used, as might be found in a typical wireless mobile transmission channel. The error conditions are defined as follows:

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

Error sequences were generated using the Gilbert Model (a 2-state Markov Model). Bit errors occur only within the error burst, during which the bit error rate is 50 %. The probability of making a transition from a burst interval to a clear channel interval and back is:

Probability of BAD to GOOD (P_BADtoGOOD) = 1.0 / AverageBurstLength (in bits)

Probability of GOOD to BAD = AverageBER * P_BADtoGOOD * (0.5 - AverageBER)

Software and parameter sets for generating these channel error conditions are available on request.

It is assumed that AudioSpecificConfig() of the MPEG-4 coder is transmitted through an error-free control channel.

4 Further Information

For information about MPEG-4 technology and any questions related to test conditions, software and test sequences please contact:

Dr Schuyler Quackenbush
Chairman, MPEG Audio Subgroup
AT&T Laboratories, Room E133
180 Park Avenue
Florham Park, NJ, 07932, USA
Phone: ++1 973 360 8551
FAX: ++1 973 360 7111
Email: srq@research.att.com.

5 References

The following informational documents on MPEG-4 may be accessed through following link: **http://www.cselt.it/mpeg**.

1. N3444, 52nd MPEG meeting, 2000, MPEG-4 Overview Document,
2. N2724, 51st MPEG meeting, MPEG-4 Applications Document
3. N1419, 37th MPEG meeting, Report on the Formal Subjective Listening Tests of MPEG-2 NBC multichannel audio coding.
4. N1420, 37th MPEG meeting, Overview of the Report on the Formal Subjective Listening Tests of MPEG-2 NBC multichannel audio coding
5. N2006, 42nd MPEG meeting, Report on the MPEG-2 AAC Stereo Verification Tests
6. N2276, 44th MPEG meeting, Report on the MPEG-4 audio NADIB verification tests
7. N2424, 45th MPEG meeting, MPEG-4 Audio verification test results: Speech Codecs
8. N2425, 45th MPEG meeting, MPEG-4 Audio verification test results: Audio on Internet
9. N3075, 50th MPEG meeting, Report on MPEG-4 Version 2 Audio Verification Test