INTERNATIONAL ORGANISATION FOR STANDARDISATION

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

ISO/IEC JTC1/SC29/WG11
MPEG2001/N3793
January 2001

Title: Call for Proposals for New Tools for Audio Coding
Author: Audio Sub group
Status: Approved

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, at the 53rd MPEG meeting, in Beijing, MPEG issued a Call for Evidence Justifying the Testing of Audio Coding Technology (N3641). Evidence submitted in response to the Call was examined at the 55th MPEG meeting, and it was determined that there was technology that might improve upon the MPEG-4 standard.

Therefore WG11 issues with this document a call for proposals of new audio coding technology.

Technology covered in this call

WG11 is interested in technology that

1.1. improves compression efficiency of audio signals or speech signals by means of bandwidth extension, and that is forward and backward compatible with existing MPEG-4 technology;
1.2. improves compression efficiency of high-quality audio signals by means of parametric coding. It is very desirable that this technology builds upon the existing MPEG-4 HILN or other MPEG-4 technology.

All proposals of technology have to fulfil the requirements that are defined in ANNEX I, which is to supply a technical description and evidence of the performance of the proposed technology.

The following steps are planned for the standardisation of the new technology:

1. All proposals have to fulfil the requirements that are defined in ANNEX I.
2. In the case of multiple proposals, a comparative test will determine verification model 1.
3. There will be a Collaborative phase to improve upon the verification model using the core experiment procedure.
4. Formal verification test.

Conditions for the specific technologies in the call follow.

BW extension

Procedure

Bandwidth extension is a tool that is not yet available within MPEG-4. This call asks for technology that addresses bandwidth extension of one or both of:

1. general audio signals, to extend the capabilities currently provided by MPEG-4 general audio coders.
2. speech signals, to extend the capabilities currently provided by MPEG-4 speech coders.

A single technology that addresses both of these signals is preferred. This technology shall be both forward and backward compatible with existing MPEG-4 technology. In other words, an MPEG-4 decoder can decode an enhanced stream and a new technology decoder can decode an MPEG-4 stream. There are two possible configurations for the enhanced stream: MPEG-4 AAC streams can carry the enhancement information in the DataStreamElement, while all MPEG-4 decoders can accept a MPEG-4 Elementary Stream and a second Elementary Stream containing the enhancement information.

Acceptance criteria

In the formal verification test it is determined whether the developed technology provides a significant increase in coding efficiency in comparison to MPEG-4. The developed technology shall satisfy both of the following two criteria. (The MUSHRA test methodology will be used).

The target bit-rate for the proposed coder is approximately 24 kbit/s per channel for general audio, and the target bit-rate is approximately 8 kbit/s per channel for speech.

1. With the developed technology operating at the target bit-rate and MPEG-4 operating at 25% higher bit-rate, the developed technology shall have a mean score that is comparable to or better than the mean score of MPEG-4.
2. With both coders operating at the target bit-rate, none of the items shall be worse in a statistical sense for the developed technology.

As a result of optimising, the quality of the compatible part might be less than that of a regular encoder. The quality of the core coder shall be compared to MPEG-4 operating at a bit-rate 25% lower than the target bit-rate.

Parametric coding

Procedure

The MPEG-4 standard already provides a parametric coding scheme for coding of general audio signals for low bit-rates (HILN). This call asks for technology that addresses parametric coding of general audio signals for the higher quality range, to extend the capabilities currently provided by HILN. Whenever possible this technology should build upon the existing MPEG-4 HILN or other MPEG-4 technology.

Acceptance criteria

In the formal verification test it is determined whether the developed technology provides a significant increase in coding efficiency in comparison to MPEG-4. The developed technology shall satisfy both of the following two criteria. (The MUSHRA test methodology will be used).

The target bit-rate for the proposed coder is approximately 24 kbit/s per channel.

1. With the developed technology operating at the target bit-rate and MPEG-4 operating at 25% higher bit-rate, the developed technology shall have a mean score that is comparable to or better than the mean score of MPEG-4.
2. With both coders operating at the target bit-rate, none of the items shall be worse in a statistical sense for the developed technology.

Timetable and Procedures

Register:

Register by 15 May, 2001 an intention to participate in the Call. 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 that will be proposed (e.g. bandwidth extension for general audio, bandwidth extension for speech, or parametric coding for general audio).

Submit Coded Materials:

Submit by 1st June, 2001, the following: the bitstreams, decoders and decoded sound files (*.wav) associated with the proposed algorithms. . Decoders shall be delivered as executables on x86 Linux or Win32 platforms. Proponents that have already submitted these materials in response to the Audio Call for Evidence (N3641) do not need to submit them again.

Submit Documentation:

Submit as contributions to the July MPEG meeting:

- A description of the technology having sufficient detail to permit technical discussions.
- Evidence of the performance of the technology (according to the guidelines of ANNEX I)

All proponents need to submit a description. Proponents that have already submitted evidence in response to the Audio Call for Evidence (N3641) are not required to submit this again. Proponents that are MPEG members shall register these documents as contributions to the July MPEG meeting and send title and author information to Schuyler Quackenbush prior to the time of the close of the contribution registry. Proponents that are not MPEG members shall email the documentation to Schuyler Quackenbush prior to 27 June, 2001, so that he can register them as contributions. The proposer's documents should be written in Microsoft Word.

Participate:

Attend the July MPEG meeting (details on meeting location and date will be communicated via email). It is strongly urged that experts familiar with the proposed technology attend in order to allow discussions on details of the proposals.

Comparative Tests:

At the July MPEG meeting timetables and procedures will be defined for conducting three tests that will compare the proposed technology, one test for each of bandwidth extension of general audio signals (at approximately 24 kb/s), bandwidth extension of speech signals (at approximately 8 kb/s, if sufficient evidence is provided) and parametric coding of general audio signals (at approximately 24 kb/s). The MUSHRA test methodology will be used and the timetable for submissions of proponent materials for the test will be determined at the July MPEG meeting.

Core Experiments:

The best technology, as identified by the comparative tests, will be Verification Model 1 and be the basis for subsequent core experiments. Proponents whose technology is selected as Verification Model 1 and all proponents participating in the core experiment process shall supply a detailed description of their technology. Core experiments will be conducted according to Core Experiment Methodology for MPEG-4 Audio, N1748.

In core experiments, the performance of equivalent MPEG-4 blocks will be checked. In the case of comparable performance, the existing MPEG-4 technology will have preference.

Verification Tests

The performance of the new technology will be measured via a formal subjective test, to be carried out prior to the Committee Draft stage of the standardization process. The acceptance criteria (enumerated above) must be met in order for the technology to progress in the standardization process.

ANNEX I

All proponents shall supply a description of the technology having sufficient detail to permit technical discussions

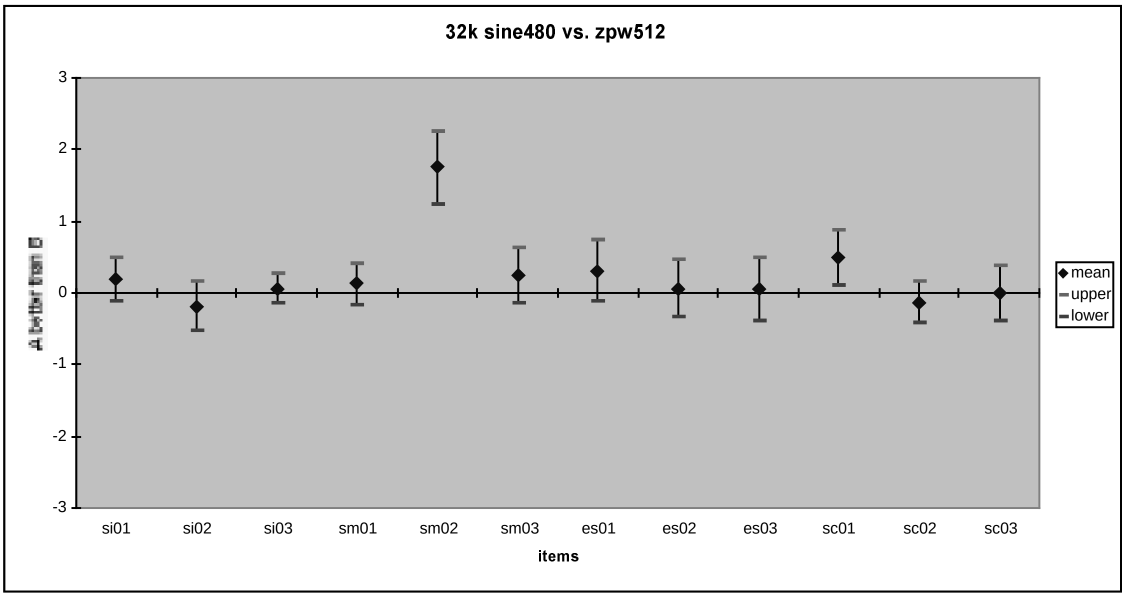
For the demonstration of evidence, the same quality assessment methodology shall be used as was successfully developed and employed for the MPEG-4 Core Experiment process (N1748). This is described in the following guidelines:

- 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.
- 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 randomised 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 or 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.
- 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 results of the listening tests are to be given by the average scores and the 95 % confidence interval. All listener responses should be considered in the analysis of variance; responses to the trials of A/B and B/A should not be averaged before analysis. An example listening test result is given below.



The following test material will be used in presenting the evidence:

- speech signals [es*, j*]
- single instruments (monophonic, i.e. one note sounding at a time) [si*]
- simple sound mixtures (material with. several notes sounding at a time) [sm*]
- 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:

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

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
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).

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.

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

ANNEX II

Reference Quality MPEG-4 Material

The following table lists companies able to provide “reference quality” encoders or encoded material.

Company	Contact	Email	Technology
Philips	Ralf Funken	ralf.funken@philips.com	Narrowband CELP coding
Philips	Ralf Funken	ralf.funken@philips.com	Wideband CELP coding
NEC	Toshiyuki Nomura	t-nomura@ccm.cl.nec.co.jp	Narrowband CELP coding
FhG	Bernhard Grill	grl@iis.fhg.de	General Audio Coding

Procedure for access to original, MPEG-4 and Proponent materials.

An FTP site shall be used to provide access to original test signals, MPEG-4 and proponent encoded signals (i.e. bitstreams, decoders and decoded signals. The bitstream supplier shall declare, for each supplied bitstream, the average bit rate and the algorithmic delay for running the encoder and decoder over a constant rate channel at the declared average rate. Whenever possible MPEG-4 bitstreams shall be supplied in MPEG-4 file format. For all MPEG-4 and Proponent executables, the computer platform shall be either x86 Linux or x86 Win32. MPEG-4 CELP decoders employing postfiltering shall be allowed. The output sampling rate of the decoder is unrestricted. Decoded waveforms shall be in *.wav format

NEC will put the 8 kb/s MPEG-4 narrowband CELP encoded materials on the FTP site before 2400hrs GMT 1 June 2001.

The decoder executables, bitstreams and decoded materials shall be put on an FTP site prior to 24:00 hrs GMT 1 June, 2001. The FTP address, user name and password will be announced on the mpeg-audio-call email list 2 weeks after the close of the 56th MPEG meeting.
If a proponent does not wish to place it’s decoder on the FTP site, the proponent should inform Schuyler Quackenbush. In this case the proponent will sent its decoder to a neutral party, who will

- check that supplied proponent bitstreams decode to the corresponding *.wav files
- determine average bitrate of the proponent bitstreams.