

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

**ISO/IEC JTC1/SC29/WG11
MPEG2007/N9519
October 2007, Shenzhen, CN**

**Source: Audio Subgroup
Status: Approved
Title: Call for Proposals on Unified Speech and Audio Coding**

1. Introduction

As mobile devices become multi-functional, and multiple devices converge into a single device, it is becoming prevalent for various types of content, including content that is a mix of speech and music, to be played on or streamed to mobile devices.

Hence, there is a strong market need for a codec that is able to provide consistent quality for mixed speech and music content and to do so with a quality that is better than codecs that are optimized for either speech content or music content.

Some examples of envisioned use cases for this technology are:

- Multi-media download to mobile devices
- User-generated content such as podcasts
- Digital radio
- Mobile TV
- Audio books

Therefore WG11 issues with this document a Call for Proposals for technology for coding of mixed speech and audio signals.

2. Technology covered in this call

WG11 is interested in technology that permits coding of signals having an arbitrary mix of speech and audio content, and that performs comparable to or better than the best coding technology that might be tailored specifically to coding of either speech or general audio content.

Minimum requirements for the technology are listed in ANNEX 1 of this document.

Note, however, that WG11 is under no obligation to proceed with standardization of the submitted technology.

3. Timetable and Procedures

3.1. Overview

A timetable for the Call for Proposals relative to specific MPEG meetings is given in the following table. Details of these meetings are available at <http://www.chiariglione.org/mpeg/meetings.htm>.

Meeting / Date	Action
82 nd MPEG Meeting	Issue Call for Proposals on Unified Speech and Audio Coding Issue Draft Evaluation Guidelines
83 rd MPEG Meeting	Issue final Evaluation Guidelines
21 Apr, 2008	Proponents must register intention to participate in Call
84 th MPEG Meeting	Issue Workplan for Evaluation
85 th MPEG Meeting	Call for Proposals responses due Selection of Reference Model 0 technology
86 th MPEG Meeting	Proponent of Reference Model 0 submits Working Draft text and Reference Software.

The following steps are envisioned for the standardization of the new technology:

- All proposals shall be prepared in accordance with the requirements set forth in this document and in the Workplan for Evaluation document.
- All proposals will be evaluated using the procedure described in the Evaluation Guidelines document. An important component of the evaluation process will be a subjective listening test to assess the quality of items coded by the proposed technology. Details of this subjective test will be given in the Workplan for Evaluation document.
- It is anticipated that the proposal with the highest Figure of Merit that also meets the requirements in Annex 1 will be selected as Reference Model 0, however all submitted information will be considered and selection of a proposal as Reference Model 0 will be by the consensus of the Audio Subgroup. Details on the Figure of Merit will be given in the Evaluation Guidelines document.
- It is expected that at the 86th MPEG meeting the proponents of the technology designated as Reference Model 0 will submit a detailed technical description, bitstream syntax and reference source code for the encoding and decoding process. Source code shall be in ANSI-C. A decoder compiled from the source code shall decode the proponent submitted bitstreams and produce the associated proponent submitted waveforms. MPEG standards do not specify a normative encoding process, hence encoder source code need only produce a bitstream that is consistent with the Reference Model bitstream syntax.
- Subsequent to the 86th MPEG meeting, a collaborative phase will improve upon the reference model using the MPEG core experiment process. Higher bitrates will be considered during the core experiment process.
- Prior to the conclusion of the standardization process WG11 will conduct a formal verification test and generate a report that characterizes the performance of the technology. Higher bitrates will be considered in the verification test.

3.2. Register

By 21 Apr 2008 (24:00 GMT) register an intention to participate in the Call by sending an email to the CFP Contact (see below). Email should indicate contact names and company.

3.3. Get Test Materials

Test items will be made available by 12 May 08. These test materials shall be used in proponent submission and are available at the CFP FTP site. Additional information on test materials and CFP FTP site will be provided in the Workplan for Evaluation.

3.4. Submit Coded Materials

By 26 May 08 (24:00 GMT) upload to the CFP FTP site the compressed data and decoded sound files associated with the submitted technology.

Compressed data shall be supplied at operating modes (i.e. bitrates for mono or stereo coding) listed in Table A-1. Compressed data can be in either MP4 file format or a proprietary format. Decoded sound files shall use WAV file format using 16 bit word lengths. Decoded waveforms must either be time-aligned with the original signals or proponents must indicate time delay (in samples) of decoded waveform relative to original waveform. Submitted technology shall not employ manual per-item tuning or multi-pass coding.

The Workplan for Evaluation document will provide details on test items and how and when to submit proponent materials.

3.5. Conduct Listening Test

The Workplan for Evaluation document will provide details on the listening test.

3.6. Submit Documentation

Submit as contributions to the 85th MPEG meeting:

- A description of the technology having sufficient detail to permit technical discussions. This can be at a rather high level, and should not contain any proprietary information or company trade secrets.
- High-level analysis of the complexity of the decoder, for example : CPU load (as measured on an x86 platform) and memory requirement (RAM and ROM). CPU load is calculated by $((\text{decoder execution time})/(\text{duration of decoded signal}))$ in %.
- Bitrate of coded material for each operating mode (i.e. target bitrate for mono or stereo signals). This shall be reported as
 - Average bitrate for an entire coded item.
 - The size of the decoder input buffer needed when operating over a constant rate channel at that rate, as in a real-time transmission system.
 - Maximum, minimum and average bitrate over all sub-segments of the coded item where the sub-segment length is on the order of 100 ms.
- Theoretical algorithmic delay of encoding and decoding system, expressed in samples for a given sampling rate. This should neglect issues of limited processing power and transmission channel bandwidth.

Proponents that are MPEG members shall register these documents as contributions to the 85th MPEG meeting and send title and author information to the CfP Contact prior to the time of the close of the contribution registry. Proponents that are not MPEG members shall email the documentation to the CfP Contact no later than 11 Jul 08 (24:00 GMT), so that he can register them as contributions. The proponent's documents shall be written in Microsoft Word.

3.7. *Participate*

Attend the 85th MPEG meeting (details on meeting location and date will be communicated via email). Proponent experts familiar with the technology are strongly urged to attend this meeting so that all aspects of the proposal can be discussed. In the absence of such experts, it may be impossible for MPEG to have an informed discussion about the features and performance of a technology.

3.8. *Evaluation and selection of technology*

At the 85th MPEG meeting and associated AHG meetings (to be held on the Saturday and Sunday prior to the MPEG meeting at the MPEG meeting location), the proposed technologies will be evaluated using the process set forth in the Evaluation Guidelines document. Information considered will be listening test results and submitted proponent documents.

The CfP contact will submit as a contribution to the 85th MPEG meeting a unified test report that pools the results of all individual test sites. This data will be used to determine which submissions meet the Requirements and to calculate the Figure of Merit for each submission. Details on these calculations will be given the Evaluation Guidelines document.

The proponent whose technology is selected as Reference Model 0 will be required to demonstrate that the submitted coded materials can be obtained from the original test items using their encoder and decoder executables.

4. Core Experiments

The best-performing technology, as identified using the Evaluation Guidelines, will be Reference Model 0 and be the basis for subsequent core experiments. Proponents whose technology is selected as Reference Model 0 and all proponents participating in the subsequent core experiment process shall supply a detailed description of their technology at the MPEG meeting following the inclusion of the technology into the Reference Model. Core experiments will be conducted according to Core Experiment Methodology for MPEG-4 Audio, document N7140.

5. Verification Test

The performance of the new technology will be measured prior to final balloting in the standardization process. The Requirements in Annex I must be met in order for the technology to progress in the standardization process.

6. Call for Proposals Contact

To register for this Call or for any other questions concerning the Call, contact:
Schuyler Quackenbush

Audio Subgroup Chair
Email srq@audioresearchlabs.com
Phone +1 908 490 0700

7. Call for Proposals FTP Site

Information on the location of test material and the location for proponents to upload coded material will be available in the Workplan for Evaluation.

8. References

The following informational documents on MPEG-4 may be accessed at:
<http://www.chiariglione.org/mpeg/meetings.htm> or directly from ISO at
<http://www.iso.org>

1. N9254, Framework for Exploration of Speech and Audio Coding
2. N7140, Revised core experiment methodology for MPEG-4 audio
3. ISO/IEC 14496-3:2005 MPEG Audio Third Edition
4. AMR-WB+ source code is available at :
<http://www.3gpp.org/ftp/Specs/html-info/26-series.htm>
<http://www.3gpp.org/ftp/Specs/html-info/26304.htm>

ANNEX 1 - Requirements

We define the following terms:

- VC is Virtual Codec
- NT is new technology.
- VC performance is the better score of the two state-of-the-art codecs (HE-AAC v2 and AMR-WB+) on a per item basis (i.e. pooling all listeners responses for that item).

The Requirements for the work on unified coding of Speech and Audio are:

- That the performance of NT shall not be worse than the performance of VC when both are operated at the same bitrate.
- Although the performance of NT is most important at low bitrates, e.g. below 24kb/s/channel, it shall not be worse than the performance of current MPEG technology when both are operated at higher bitrates.

The New Technology shall be evaluated at the operating modes (bitrate and number of channels) listed in Table A-1. The Evaluation Guidelines document will define a process for determining whether a submitted technology meets the Requirements and, in order to select one submission from amongst those that meet the Requirements, define a Figure of Merit for determining the technology with the best performance.

Table A-1: Operating Modes.

Bitrate	Stereo/Mono
64 kbps	Stereo
32 kbps	Stereo
24 kbps	Both
20 kbps	Both
16 kbps	Both
12 kbps	Mono

The complexity of the submitted technology will be taken into consideration as part of the Reference Model 0 selection process.

The Audio Subgroup envisions that the final performance of a possible work item shall be that it fulfills the requirements stated above and that NT has significantly better performance than VC at the lower bitrate range.