Technical Specification Group Services and System Aspects
Meeting #16, Marco Island, Florida, 10-13 June 2002

TSGS#16(02)0207

ISO/IEC JTC 1/SC 29 N 4749

DATE: 2002-03-31

ISO/IEC JTC 1/SC 29

Coding of Audio, Picture, Multimedia and Hypermedia Information

Secretariat: Japan (JISC)

DOC. TYPE	Outgoing Liaison Statement
TITLE	Liaison Statement from SC 29/WG 11 to 3GPP on Advanced Text and 2D Graphics [SC 29/WG 11 N 4727, Attachment: WG 11 N 4705]
SOURCE	ISO/IEC JTC 1/SC 29/WG 11
PROJECT	
STATUS	In accordance with Resolution 3.9 taken at the 59th SC 29/WG 11 meeting, 2002-03-11/15, Jeju, Korea, Rep. of, the SC 29 Secretariat has forwarded this liaison statement to 3GPP. [Requested action: For SC 29's information]
ACTION ID	FYI
DUE DATE	
DISTRIBUTION	P, O and L Members of ISO/IEC JTC 1/SC 29 ; ISO/IEC JTC 1 Secretariat; ISO/IEC ITTF
ACCESS	Def

LEVEL				
ISSUE NO.		451		
FILE	NAME	29n4749c.htm	29n47491.doc	29n47492.doc
	SIZE (KB)		22.0	862.7
	PAGES		1	15

Secretariat, ISO/IEC JTC 1/SC 29
IPSJ/ITSCJ (Information Processing Society of Japan/Information Technology Standards Commission of Japan)*
Room 308-3, Kikai-Shinko-Kaikan Bldg., 3-5-8, Shiba-Koen, Minato-ku, Tokyo 105-0011 Japan
Telephone: 3-3431-2808; Facsimile: 3-3431-6493;
E-mail: ogura@itscj.ipsj.or.jp
*Standard Organization Accredited by JISC

Liaison Statement from WG 11 to 3GPP

See the file below:

- File Name: 29n47491.doc
- File format: MS Word for 97
- File size: 22,016 bytes
- Pages: 1

Attachment:
SC 29/WG 11 N 4705: Call for Proposals for MPEG-4 Systems Advanced Text and 2D Graphics

---

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

ISO/IEC/JTC1/SC29/WG11
MPEG2002/ N4727
March 2002, Jeju

Source : **Convenor**
Title : **Liaison to 3GPP**

In its meeting of March 11-15, 2002 in Jeju (Korea), WG 11 (MPEG) reviewed the need for a new work item on "Advanced Text and 2D Graphics".

On this issue, MPEG decided to issue a Call for Proposals, while realising that there is potential overlap with work items on "Timed Text" that already exist in 3GPP and that may be defined in W3C. As W3C, 3GPP and MPEG each approach the work item from different angles, MPEG believes that the work in these standardisation bodies can be complementary to a large extent.

MPEG would highly appreciate to jointly work with W3C and 3GPP on a solution for "Timed Text" so as to explore opportunities for achieving a solution with a level of commonality that is as high as possible and practical.

Attachment:
SC 29/WG 11 N 4705: Preliminary Call for Proposal on Advanced Text and 2D Graphics

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

ISO/IEC JTC1/SC29/WG11 N4705
Jeju – March 2002

Title: **Call for Proposals for MPEG-4 Systems Advanced Text and 2D Graphics**
Source: **Requirements**
Status: **Draft**

Call for Proposals for MPEG-4 Systems Advanced Text and 2D Graphics

Introduction

MPEG, a working group in ISO/IEC, has produced four important standards (MPEG-1, MPEG-2, MPEG-4 and MPEG-7) and is working on MPEG-21. Extension work is ongoing on the 'Intellectual Property Management and Protection' (IPMP) specification, with the goal to enhance interoperability in the consumption of protected content. The 'Multimedia Framework' MPEG-21 standard is in early development stages, with the Requirements phase well advanced and the first Call for Proposals issued in October 2000.

MPEG has identified the need for extensions of the Systems part of the MPEG-4 standard for which requirements are documented in Annex B of this document. This requirements mainly consist in :

- Text representation;
- 2D graphics representation.

This Call for Proposals invites submissions that fulfil some or all of these requirements.

Call for Proposals

All parties that believe they possess relevant technologies for MPEG-4 Systems Advanced Text and 2D Graphics are invited to submit proposals for consideration by MPEG. These parties do not necessarily have to be MPEG members. A full list explaining what proponents will be expected to submit can be found below.

Proposals are to be received by both Fernando Pereira and Olivier Avaro, either by postal mail or (preferred) electronically. Submissions shall be received by 15 July, 23:59 hours GMT. Proposals are to be formatted as Word97, HTML or PDF documents. If submitted by mail, they must be provided on a 1.44 MByte floppy disk or an ISO9660 CD-R..

It is advisable that proponents attend the MPEG meeting, in Klagenfurt, Austria, where they will be asked to present their proposals. They are also invited to participate in the ensuing discussions at that meeting. Notably, the proposals will be discussed on the 20-21 July 2002 during a so-called 'MPEG Ad Hoc Group meeting', and during on the 22-26 July during the regular MPEG meeting. Please contact Olivier Avaro for details on attending the meeting (see address below).

After hearing the presentations, MPEG will evaluate the proposals against the requirements as set forth in Annex B of this Call. The evaluation leads to selecting relevant elements in the proposals and developing a specification based upon the responses to this Call. However, MPEG reserves the right to adopt none, one, or a combination of several proposals. All decisions will be made by consensus within the MPEG committee.

Proponents who are accredited MPEG delegates are responsible for the registration of their own proposals and their submission to the MPEG document repository in the usual manner. They shall upload their proposal before the submission deadline (16 July 2002 at 23:59 GMT) ***and*** send a notification email to both Fernando Pereira and Olivier Avaro. Proposals submitted by proponents who are not accredited MPEG delegates, will be made available to MPEG by the recipients to the call.

Proposals will be made available to accredited MPEG delegates through its password protected document repository. Proposals will also be disclosed to other Proponents at the meeting. MPEG will not disclose proposals to non-MPEG delegates but will not prevent such disclosure by the proponents themselves. The disclosure of proposals to all proponents, after the submission deadline but before the meeting, is highly encouraged. This will allow an informed discussion at the meeting. The email addresses of proponents who are not accredited MPEG delegates will be provided to all proponents to facilitate the exchange of proposals.

Further information (including documents mentioned above and this Call) can be obtained from the MPEG home page at http://www.cselt.it/mpeg. The following can also provide information:

Fernando Pereira
Chairman MPEG Requirements Group
Instituto Superior Técnico - Instituto de Telecomunicações, Portugal
Tel: + 351 21 8418460
Email: Fernando.Pereira@lx.it.pt

Olivier Avaro
Chairman MPEG Systems Group
France Telecom R&D, France
Tel: +33 1 45 29 40 97
Email: Olivier.avaro@francetelecom.com

Note: Please do not send confidential information or proposals to Fernando Pereira or Olivier Avaro.

Requirements for Responding to the CfP

Also, any submission shall include ***all*** of the following:

1. A completed Registration Form as found in Annex C

2. An assessment (e.g., a table) of which requirements are met and how (see Annex B).
   Note: Proponents are not expected to provide a solution with the objective of meeting as many of the requirements as possible. However, as an aid to the evaluation process, it is very important that proponents identify those requirements that their submission aims to address;

3. A description of how the proposed technology will interact with the other elements of the MPEG-4 Systems Architecture as presently described in the Annex A of this document;

4. Evidence of an implementation (e.g. an accompanying demonstrator) including a detailed documented description of the system, and any other relevant information;

Also, any submission ***may*** include:

1. Any other additional information relevant to help the evaluation of the submission.

Source code and IPR

Proponents are advised that, upon acceptance by MPEG for further evaluation, MPEG may require that software support for certain parts of any proposed technology be made available in source code. This code is then to be included in the MPEG-4 standard, in the form of Reference Software. Commitment to provide such software is required by the July 2002. Actual availability in that case is required as soon as a technology is adopted. If Proponents feel that any aspects of their technology should not be made available in source code, then they should clearly state which aspects and why.

Furthermore proponents are advised that this Call is being made under the auspices of ISO/IEC, and as such, subject to the ISO/IEC Intellectual Property Rights Policy as approved by the ISO and IEC councils in 1994 (see attachment iso_iec_ipr_policy.doc).

Timetable and Logistics

15 July, 2002, 23:59 hours GMT	Proposals to be received by both Fernando Pereira and Olivier Avaro either by postal mail or (preferred) electronically. See above for addresses In addition, proponents who are accredited MPEG delegates are responsible for the registration of their own proposals and their submission to the MPEG document repository in the usual manner. Proposals submitted by proponents who are not accredited MPEG delegates, and therefore cannot do this, will be registered by the above individuals. Proposals are to be formatted as Word97, HTML or PDF documents. If submitted by mail, they must be provided on a 1.4MB floppy disk or an ISO9660 CD-R
20-21 July 2002	**MPEG Ad-Hoc Meeting for evaluation of proposals** Proponents are invited to attend to present their proposals and participate in discussions
July 2002	**MPEG Meeting (Klagenfurt, Austria)** Proponents are invited to attend, further present their proposal(s), and participate in relevant discussions
July 2002	Commitment to provide Reference Software
December 2002	Delivery of Reference Software

The development plan is currently as follows:

July 2002	Adoption of *Working Draft*
October 2002	Adoption of *Working Draft v2.0*
December 2002	Adoption of *Committee Draft*
March 2003	Adoption of *Final Committee Draft*
July 2003	Adoption of *Final Draft International Standard*

References and Supporting Information

Annexes to this document:

A: Background information
B: MPEG-4 Advanced Text and 2D Graphics Requirements
D: Registration form

Attached in the archive file containing the CfP are the following documents:

- Overview of the MPEG-4 Standard (ISO/IEC JTC1 SC29 WG11 N4668)
- ISO/IEC Intellectual Property Policy (iso_iec_ipr_policy.doc)

The following referenced documents can be found on MPEG's website: www.cselt.it/mpeg

- MPEG Systems FAQ (ISO/IEC JTC1 SC29 WG11 N4718)

Annex A: Background Information

MPEG-4 "Version 1" Systems Architecture

1 Overview

ISO/IEC 14496 specifies a system for the communication of interactive audio-visual scenes. This specification includes the following elements:

1. the coded representation of natural or synthetic, two-dimensional (2D) or three-dimensional (3D) objects that can be manifested audibly and/or visually (audio-visual objects) (specified in part 1,2 and 3 of ISO/IEC 14496);
2. the coded representation of the spatio-temporal positioning of audio-visual objects as well as their behavior in response to interaction (scene description, specified in this part of ISO/IEC 14496);
3. the coded representation of information related to the management of data streams (synchronization, identification, description and association of stream content, specified in this part of ISO/IEC 14496); and
4. a generic interface to the data stream delivery layer functionality (specified in part 6 of ISO/IEC 14496).

The overal operation of a system communicating audio-visual scenes can be paraphrased as follows:

At the sending terminal, the audio-visual scene information is compressed, supplemented with synchronization information and passed to a delivery layer that multiplexes it into one or more coded binary streams that are transmitted or stored. At the receiving terminal, these streams are demultiplexed and decompressed. The audio-visual objects are composed according to the scene description and synchronization information and presented to the end user. The end user may have the option to interact with this presentation. Interaction information can be processed locally or transmitted back to the sending terminal. ISO/IEC 14496 defines the syntax and semantics of the bitstreams that convey such scene information, as well as the details of their decoding processes.

This part of ISO/IEC 14496 specifies the following tools:

- a terminal model for time and buffer management;
- a coded representation of interactive audio-visual scene description information (Binary Format for Scenes – BIFS);
- a coded representation of metadata for the identification, description and logical dependencies of the elementary streams (Object descriptors and other Descriptors);
- a coded representation of descriptive audio-visual content information (object content information – OCI);
- an interface to intellectual property management and protection (IPMP) systems;
- a coded representation of synchronization information (sync layer – SL); and
- a multiplexed representation of individual elementary streams in a single stream (FlexMux).

These various elements are described functionally in this subclause and specified in the normative clauses that follow.

2 Architecture

The information representation specified in ISO/IEC 14496-1 describes the means to create an interactive audio-visual scene in terms of coded audio-visual information and associated scene description information. The entity that composes and sends, or receives and presents such a coded representation of an interactive audio-visual scene is generically referred to as an "audio-visual terminal" or just "terminal". This terminal may correspond to a standalone application or be part of an application system.

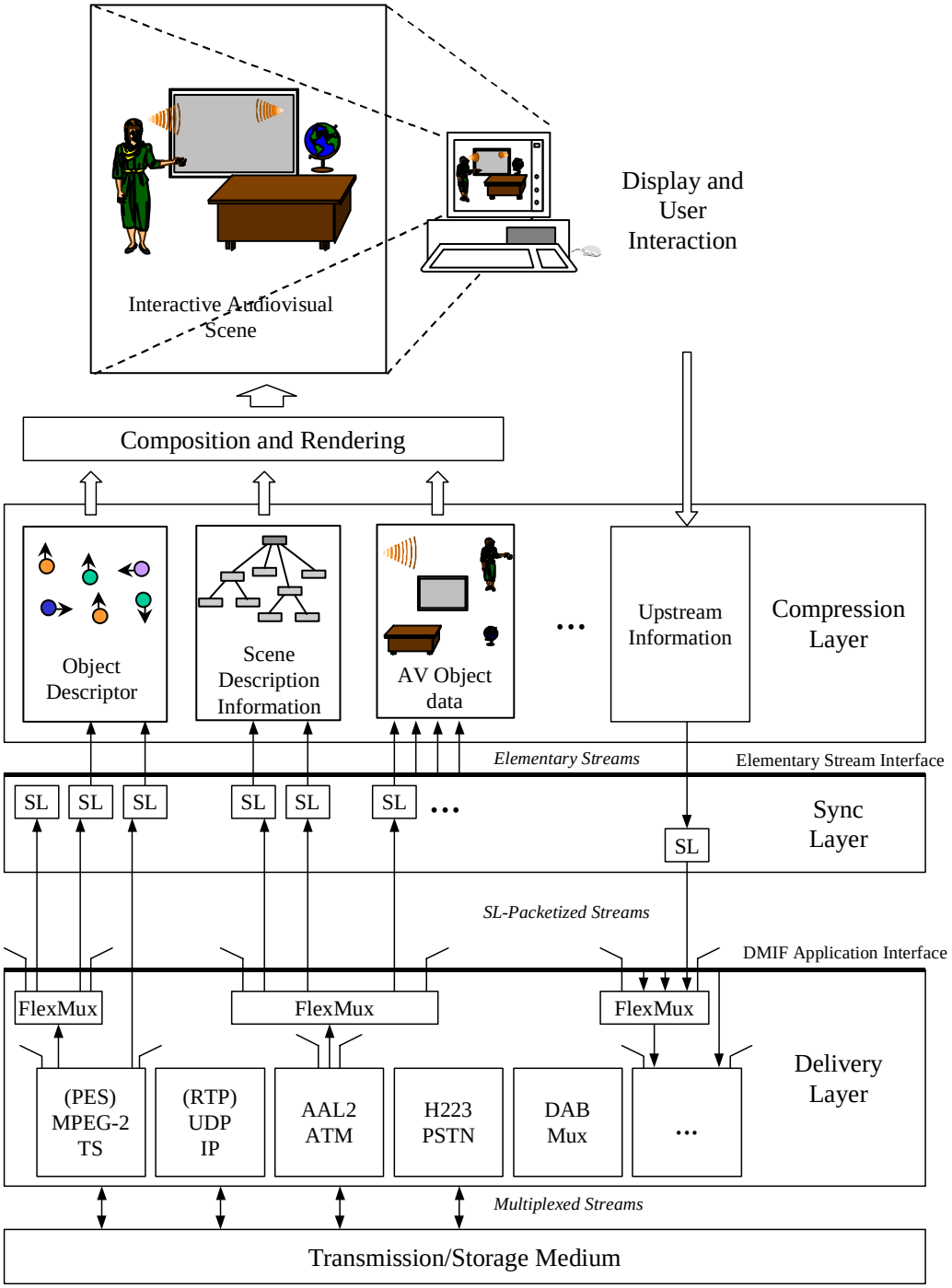

Figure 1 - The ISO/IEC 14496 terminal architecture

The basic operations performed by such a receiver terminal are as follows. Information that allows access to content complying with ISO/IEC 14496 is provided as initial session set up information to the terminal. Part 6 of ISO/IEC 14496 defines the procedures for establishing such session contexts as well as the interface to the delivery layer that generically abstracts the storage or transport medium. The initial set up information allows, in a recursive manner, to locate one or more elementary streams that are part of the coded content representation. Some of these elementary streams may be grouped together using the multiplexing tool described in ISO/IEC 14496-1.

Elementary streams contain the coded representation of either audio or visual data or scene description information. Elementary streams may as well themselves convey information to identify streams, to describe logical dependencies between streams, or to describe information related to the content of the streams. Each elementary stream contains only one type of data.

Elementary streams are decoded using their respective stream-specific decoders. The audio-visual objects are composed according to the scene description information and presented by the terminal's presentation device(s). All these processes are synchronized according to the Systems Decoder Model (SDM) using the synchronization information provided at the synchronization layer.

These basic operations are depicted in Figure 1, and are described in more detail below.

3 Terminal Model: Systems Decoder Model

The systems decoder model provides an abstract view of the behavior of a terminal complying with ISO/IEC 14496-1. Its purpose is to enable a sending terminal to predict how the receiving terminal will behave in terms of buffer management and synchronization when reconstructing the audio-visual information that comprises the presentation. The systems decoder model includes a systems timing model and a systems buffer model which are described briefly in the following subclauses.

3.1 Timing Model

The timing model defines the mechanisms through which a receiving terminal establishes a notion of time that enables it to process time-dependent events. This model also allows the receiving terminal to establish mechanisms to maintain synchronization both across and within particular audio-visual objects as well as with user interaction events. In order to facilitate these functions at the receiving terminal, the timing model requires that the transmitted data streams contain implicit or explicit timing information. Two sets of timing information are defined in ISO/IEC 14496-1: clock references and time stamps. The former convey the sending terminal's time base to the receiving terminal, while the latter convey a notion of relative time for specific events such as the desired decoding or composition time for portions of the encoded audio-visual information.

3.2 Buffer Model

The buffer model enables the sending terminal to monitor and control the buffer resources that are needed to decode each elementary stream in a presentation. The required buffer resources are conveyed to the receiving terminal by means of descriptors at the beginning of the presentation. The terminal can then decide whether or not it is capable of handling this particular presentation. The buffer model allows the sending terminal to specify when information may be removed from these buffers and enables it to schedule data transmission so that the appropriate buffers at the receiving terminal do not overflow or underflow.

4 Multiplexing of Streams: The Delivery Layer

The term delivery layer is used as a generic abstraction of any existing transport protocol stack that may be used to transmit and/or store content complying with ISO/IEC 14496. The functionality of this layer is not within the scope of ISO/IEC 14496-1, and only the interface to this layer is considered. This interface is the DMIF Application Interface (DAI) specified in ISO/IEC 14496-6. The DAI defines not only an interface for the delivery of streaming data, but also for signaling information required for session and channel set up as well as tear down. A wide variety of delivery mechanisms exist below this interface, with some of them indicated in Figure 1. These mechanisms serve for transmission as well as storage of streaming data, i.e., a file is considered to be a particular instance of a delivery layer. For applications where the desired transport facility does not fully address the needs of a service according to the specifications in ISO/IEC 14496, a simple multiplexing tool (FlexMux) with low delay and low overhead is defined in ISO/IEC 14496-1.

5 Synchronization of Streams: The Sync Layer

Elementary streams are the basic abstraction for any streaming data source. Elementary streams are conveyed as sync layer-packetized (SL-packetized) streams at the DMIF Application Interface. This packetized representation additionally provides timing and synchronization information, as well as fragmentation and random access information. The sync layer (SL) extracts this timing information to enable synchronized decoding and, subsequently, composition of the elementary stream data.

6 The Compression Layer

The compression layer receives data in its encoded format and performs the necessary operations to decode this data. The decoded information is then used by the terminal's composition, rendering and presentation subsystems.

6.1 Object Description Framework

The purpose of the object description framework is to identify and describe elementary streams and to associate them appropriately to an audio-visual scene description. Object descriptors serve to gain access to ISO/IEC 14496 content. Object content information and the interface to intellectual property management and protection systems are also part of this framework.

An object descriptor is a collection of one or more elementary stream descriptors that provide the configuration and other information for the streams that relate to either an audio-visual object or a scene description. Object descriptors are themselves conveyed in elementary streams. Each object descriptor is assigned an identifier (object descriptor ID), which is unique within a defined name scope. This identifier is used to associate audio-visual objects in the scene description with a particular object descriptor, and thus the elementary streams related to that particular object.

Elementary stream descriptors include information about the source of the stream data, in form of a unique numeric identifier (the elementary stream ID) or a URL pointing to a remote source for the stream. Elementary stream descriptors also include information about the encoding format, configuration information for the decoding process and the sync layer packetization, as well as quality of service requirements for the transmission of the stream and intellectual property identification. Dependencies between streams can also be signaled within the elementary stream descriptors. This functionality may be used, for example, in scalable audio or visual object representations to indicate the logical dependency of a stream containing enhancement information, to a stream containing the base information. It can also be used to describe alternative representations for the same content (e.g. the same speech content in various languages).

6.1.1 Intellectual Property Management and Protection

The intellectual property management and protection (IPMP) framework for ISO/IEC 14496 content consists of a normative interface that permits an ISO/IEC 14496 terminal to host one or more IPMP Systems. The IPMP interface consists of IPMP elementary streams and IPMP descriptors. IPMP descriptors are carried as part of an object descriptor stream. IPMP elementary streams carry time variant IPMP information that can be associated to multiple object descriptors.

The IPMP System itself is a non-normative component that provides intellectual property management and protection functions for the terminal. The IPMP System uses the information carried by the IPMP elementary streams and descriptors to make protected ISO/IEC 14496 content available to the terminal. An application may choose not to use an IPMP System, thereby offering no management and protection features.

6.1.2 Object Content Information

Object content information (OCI) descriptors convey descriptive information about audio-visual objects. The main content descriptors are: content classification descriptors, keyword descriptors, rating descriptors, language descriptors, textual descriptors, and descriptors about the creation of the content. OCI descriptors can be included directly in the related object descriptor or elementary stream descriptor or, if it is time variant, it may be carried in an elementary stream by itself. An OCI stream is organized in a sequence of small, synchronized entities called events that contain a set of OCI descriptors. OCI streams can be associated to multiple object descriptors.

6.2 Scene Description Streams

Scene description addresses the organization of audio-visual objects in a scene, in terms of both spatial and temporal attributes. This information allows the composition and rendering of individual audio-visual objects after the respective decoders have reconstructed the streaming data for them. For visual data, ISO/IEC 14496-1 does not mandate particular composition algorithms. Hence, visual composition is implementation dependent. For audio data, the composition process is defined in a normative manner in ISO/IEC 14496-1 (MPEG-4 Systems) and ISO/IEC 14496-3 (MPEG-4 Audio).

The scene description is represented using a parametric approach (BIFS - Binary Format for Scenes). The description consists of an encoded hierarchy (tree) of nodes with attributes and other information (including event sources and targets). Leaf nodes in this tree correspond to elementary audio-visual data, whereas

intermediate nodes group this material to form audio-visual objects, and perform grouping, transformation, and other such operations on audio-visual objects (scene description nodes). The scene description can evolve over time by using scene description updates.

In order to facilitate active user involvement with the presented audio-visual information, ISO/IEC 14496-1 provides support for user and object interactions. Interactivity mechanisms are integrated with the scene description information, in the form of linked event sources and targets (routes) as well as sensors (special nodes that can trigger events based on specific conditions). These event sources and targets are part of scene description nodes, and thus allow close coupling of dynamic and interactive behavior with the specific scene at hand. ISO/IEC 14496-1, however, does not specify a particular user interface or a mechanism that maps user actions (e.g., keyboard key presses or mouse movements) to such events.

Such an interactive environment may not need an upstream channel, but ISO/IEC 14496 also provides means for client-server interactive sessions with the ability to set up upstream elementary streams and associate them to specific downstream elementary streams.

6.3 Audio-visual Streams

The coded representations of audio and visual information are described in ISO/IEC 14496-3 and ISO/IEC 14496-2, respectively. The reconstructed audio-visual data are made available to the composition process for potential use during the scene rendering.

6.4 Upchannel Streams

Downchannel elementary streams may require upchannel information to be transmitted from the receiving terminal to the sending terminal (e.g., to allow for client-server interactivity). Figure 1 indicates the flowpath for an elementary stream from the receiving terminal to the sending terminal. The content of upchannel streams is specified in the same part of the specification that defines the content of the downstream data. For example, upchannel control streams for video downchannel elementary streams are defined in ISO/IEC 14496-2.

MPEG-4 Version 2 Systems Architecture

1 Overview

Version 2 of the MPEG-4 systems extends this framework to support more sophisticated functionality such as 3D spatialization of the acoustic response, Hierarchical 3D Mesh Coding and Java support. Version 2 also specifies a file format to store MPEG-4 content. The work involving scene graphs is addressed in Advanced BIFS section, the Java enabled functionality is addressed in MPEG-J section, file format is addressed in MPEG-4 File Format, and the changes required to the ESM to accommodate new functionalities like SMPTE Camera positioning descriptor is specified in the ESM sections.

Version 2 builds on version 1 of MPEG-4 systems. All version 2 systems decoding version 2 content shall also decode version 1 content. This document is an amendment to the ISO/IEC 14496-1.

2 Advanced BIFS

The BIFS part of this document is an amendment to version 1 Scene Description (See, ISO/IEC 14496-1 clause 9). This first amendment to BIFS (Binary Format for Scenes) is referred to as Advanced BIFS, and it consists of a set of nodes that bring new functionalities to the BIFS scene descirption. These functionalities are:

1. 3D-spatialization of acoustic response of the environment, and advanced sound source directivity modeling in interactive 3-D scenes. This is provided by a set of nodes referred to as Advanced Audio BIFS.
2. Body animation of either a default body model present at the decoder or of a downloadable body model.

3. Chroma keying via the MaterialKey node.
4. Hierarchical 3-D mesh coding enabled by a Hierarchical3Dmesh node that is used to attach 3-D mesh streams specified in the Visual part of the FPDAM.
5. Application signalling via the ServerCommand node that enables sending back-channel messages in BIFS scenes.

3 MPEG-J

The MPEG-J is a programmatic system (as opposed to the parametric system offered by MPEG-4 Version 1) which specifies API for interoperation of MPEG-4 media players with Java code. By combining MPEG-4 media and safe executable code, content creators may embed complex control and data processing mechanisms with their media data to intelligently manage the operation of the audio-visual session. A block diagram of the MPEG-J player in an MPEG-4 system player environment is shown in Figure 1. The lower half of this drawing depicts the parametric MPEG-4 System player also referred to as the Presentation Engine. The MPEG-J subsystem controlling the Presentation Engine, also referred to as the Application Engine, is depicted in the upper half of Figure 1.

The Java application is delivered as a separate elementary stream to the MPEG-4 terminal. There it will be directed to the MPEG-J run time environment, from where the MPEG-J program will have access to the various components and data of the MPEG-4 player, in addition to .the basic packages of the language (java.lang, java.io, java.util). MPEG-J specifically does *not* support downloadable decoders.

A few categories of APIs have been defined for different scopes. For Scene graph API the objective is to provide access to the scene graph: to inspect the graph, to alter nodes and their fields, and to add and remove nodes within the graph. The Resource Manager API is used for regulation of performance: it provides a centralized facility for managing resources. The Terminal Capability API is used when program execution is contingent upon the terminal configuration and its capabilities, both static (that do not change during execution) and dynamic. Media Decoders API allow the control of the decoders that are present in the terminal. The Network API provides a way to interact with the network, being compliant to the MPEG-4 DMIF Application Interface. Complex applications and enhanced interactivity are possible with these basic packages.

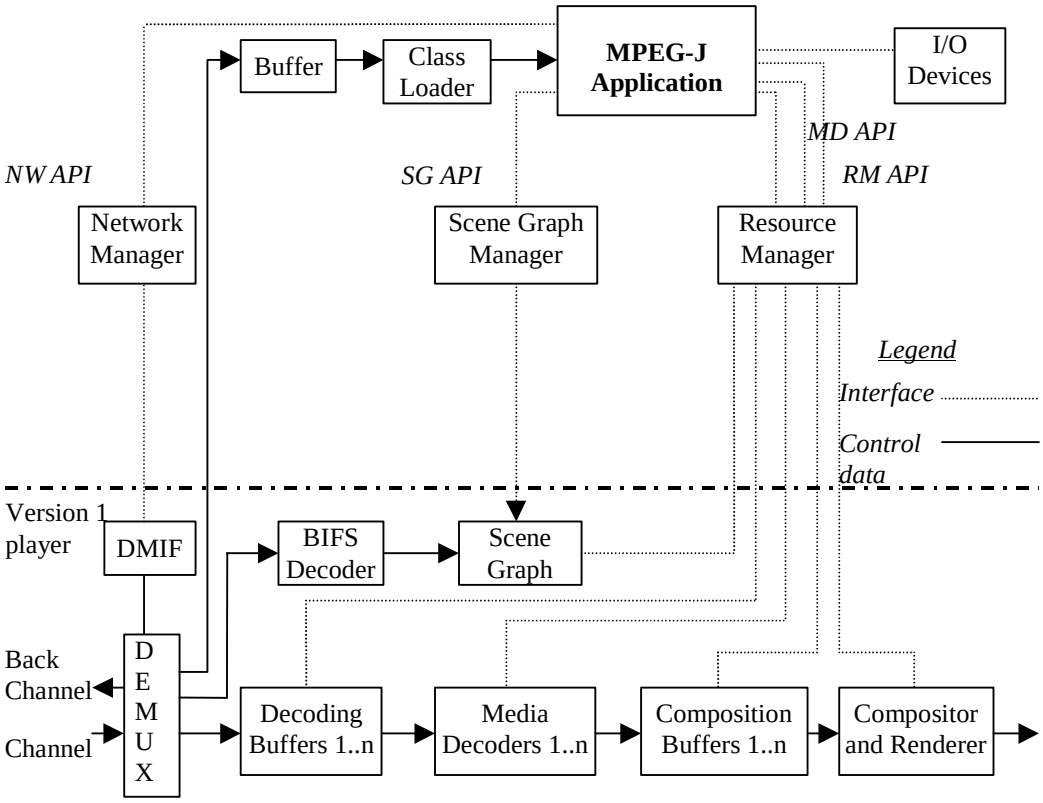

Figure 1 Location of interfaces in the architecture of an MPEG-J enabled MPEG-4 System

4 MPEG-4 File Format

The MP4 file format is designed to contain the media information of an MPEG-4 presentation in a flexible, extensible format, which facilitates interchange, management, editing, and presentation of the media. This presentation may be 'local' to the system containing the presentation, or may be via a network or other stream delivery mechanism (a TransMux). The file format is designed to be independent of any particular TransMux while enabling efficient support for TransMuxes in general. The design is based on the QuickTime® format from Apple Computer Inc.

The MP4 file format is composed of object-oriented structures called 'atoms'. Each atom is identified by a unique tag and a length. Most atoms describe a hierarchy of metadata giving information such as index points, durations, and pointers to the media data. This collection of atoms is contained in an atom called the 'movie atom'. The media data itself is located elsewhere; it can be in the MP4 file, contained in one or more 'mdat' or media data atoms, or located outside the MP4 file and referenced via URL's.

The file format is a streamable format, as opposed to a streaming format. That is, the file format does not define an on-the-wire protocol, and is never actually streamed over a transmission medium. Instead, metadata in the file known as 'hint tracks' provide instructions, telling a server application how to deliver the media data over a particular TransMux. There can be multiple hint tracks for one presentation, describing how to deliver over various TransMuxes. In this way, the file format facilitates streaming without ever being streamed directly.

The metadata in the file, combined with the flexible storage of media data, allows the MP4 format to support streaming, editing, local playback, and interchange of content, thereby satisfying the requirements for the MPEG4 Intermedia format.

Annex B: MPEG-4 Advanced Text and 2D Graphics Requirements

Text Composition Requirements

The current MPEG-4 specification is recognized as being insufficient to satisfy rudimentary text composition requirements. Solutions proposed for inclusion in the specification should address the following requirements.

Note – The purpose of this call is not to call for fonts that would be resident in the terminal.

1 Requirements

1.1 Precision

1.1.1 Requirement

The MPEG-4 Standard shall allow the precise layout and formatting of text in the terminal composition.

1.1.2 Specification

a) The MPEG-4 specification shall allow font metrics to be known at the decoder.

b) MPEG-4 text encoding shall provide support for spatial scalability of fonts.

1.2 Internationalization

1.2.1 Requirement

The MPEG-4 Standard shall allow the specification of text and its composition utilizing international character sets.

~~**4.3Compatibility**~~

~~***Requirement***~~

~~**Solutions adopted for MPEG-4 text object should attempt to reuse/extend existing MPEG-4 mechanisms, but they must not invalidate existing bit streams.**~~

~~4.4~~1.3 Flexibility

1.3.1 Requirement

MPEG-4 text should allow signaling and transmission of a variety of fonts and font attributes (e.g. bold, italics, underline, bold italic, color, etc.).

2 Evaluation Criteria

2.1 Compatibility

Solutions adopted for MPEG-4 text should attempt to reuse/extend existing MPEG-4 mechanisms, and they must not invalidate existing bit streams.

2.2 Complexity

Solutions should not be too (computationally, implementation) complex.

2.3 Quality

Solutions should be technically sound. Solutions that can be generalized to 3D are preferred.

2.4 Efficiency

Solutions should be efficient in memory and bandwidth.

Streaming Text Requirements

1 Requirements

1.1 Requirement

MPEG-4 shall support the streaming of text in a compressed format.

2 Evaluation Criteria

2.1 Compatibility

Solutions based on technology that is compatible with existing or developing standards are preferred.

Requirements for Advanced2D Graphics

1 Requirements

1.1 Requirement

MPEG-4 shall support the following list of 2D graphics features:
It shall be possible to :

- allow the use of a viewport in conjunction with Layer2D, with all appropriate options to deal with possible differences of aspect ratio between the viewport and the enclosing Layer2D
- use a 2x3 matrix to specify a 2D transformation

- draw circular, parabolic and elliptical arcs
- use line properties line-cap and line-join whose properties can be respectively butt, round, square or mitter, round, bevel

- fill and texture-fill curves, even when moveTo types are used
- texture the (wide) outline of shapes (including text)
- use gradient and patterns as textures
- texture text
- outline text
- have both text fill and text outline

- define the shape of each character in a font using a set of Shape nodes
- easily combine individual text objects with different decorations to build a continuous text
- render text on a baseline that follows any path/curve
- specify the following properties of text:
  - alignment-baseline: it specifies which baseline of this element has to be aligned with the baseline of its parent element. It can take the following values: baseline, top, before-edge, text-top, text-before-edge, middle, bottom …
  - baseline-shift: it allows for having subscripted and superscripted text with an arbitratry shifting.
  - kerning: it specifies whether the inter-glyph spacing is based on the kerning-table of the font.
  - letter-spacing: it specifies the space between two letters, possibly after addition of a kerning space.
  - word-spacing: it specifies the spacing between two words, possibly after addition of the letter space.
  - text-decoration: a text can either be underlined, be overlined, stroke through, or blink.
  - Unicode-bidi: This properties allows for handling text that mixes characters that have to be rendered from left to right and characters that have to be rendered from right to left (e.g. a text that mixes hebrew and english).

- Font-stretch: This property indicates the desired amount of condensing or expansion in the glyphs used to render the text.
- Font-variant : This property indicates whether the text is to be rendered using the normal glyphs for lowercase characters or using small-caps glyphs for lowercase characters.
- Font-weight: This property refers to the boldness or lightness of the glyphs used to render the text, relative to other fonts in the same font family. It can take the following values: bold, bolder, lighter, 100, 200, 300 … to 900.

2 Evaluation Criteria

2.1 Compatibility

a) Solutions based on technology that is compatible with existing or developing standards are preferred.
b) Solutions should attempt to reuse/extend existing MPEG-4 mechanisms, and they must not invalidate existing bit streams.

Annex C: Registration form

Company:	
Contact name:	
Address:	
Phone number:	
Fax:	
Email:	
Title of submission:	
Abstract:	
Time requested for presentation (subject to review by organisers):	
Demo intended:	yes / no
Equipment and other support required for demo:	

Name:

Date:

Signatures :