

ISO/IEC JTC 1/SC 29

Secretariat: JISC

Voting begins
on: **2015-07-21**

Voting terminates
on: **2015-09-21**

Information technology — Coding of audio-visual objects —

Part 11: Scene description and application engine

*Technologies de l'information — Codage des objets audiovisuels —
Partie 11: Description de scène et moteur d'application*

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.

IN ADDITION TO THEIR EVALUATION AS BEING ACCEPTABLE FOR INDUSTRIAL, TECHNOLOGICAL, COMMERCIAL AND USER PURPOSES, DRAFT INTERNATIONAL STANDARDS MAY ON OCCASION HAVE TO BE CONSIDERED IN THE LIGHT OF THEIR POTENTIAL TO BECOME STANDARDS TO WHICH REFERENCE MAY BE MADE IN NATIONAL REGULATIONS.



Reference number
ISO/IEC FDIS 14496-11:2015(E)

© ISO/IEC 2015



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2015, Published in Switzerland

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

Foreword.....	vi
0 Introduction	viii
0.1 Scene Description	viii
0.2 Extensible MPEG-4 Textual Format	x
0.3 MPEG-J	x
1 Scope	1
2 Normative references	1
3 Additional reference	2
4 Terms and definitions	2
5 Abbreviations and Symbols	7
6 Conventions	7
7 MPEG-4 Systems Node Semantics	8
7.1 Scene Description	8
7.2 Node Semantics	24
7.3 Informative: Differences Between MPEG-4 Scripts and ECMA Scripts	181
7.4 Informative: FlexTime behavior	182
7.5 Informative: Implementation of MaterialKey node	183
7.6 Informative: Example implementation of spatial audio processing (perceptual approach).....	184
7.7 Informative: MPEG-4 Audio TTS application with Facial Animation	188
7.8 Informative: 3D Mesh Coding in BIFS scenes	188
7.9 Profiles	189
7.10 Metric information for resident fonts	220
7.11 Font metrics for SANS SERIF font (Albany)	221
7.12 Font metrics for SERIF font (Thorndale)	227
7.13 Font metrics for TYPEWRITER font (Cumberland)	234
8 BIFS	242
8.1 Introduction	242
8.2 Decoding tables, data structures and associated functions	242
8.3 Quantization	247
8.4 Compensation process	257
8.5 BIFS Configuration	258
8.6 BIFS Command Syntax	262
8.7 BIFS Scene	274
8.8 BIFS-Anim	305
8.9 Interpolator compression	310
8.10 Definition of bodySceneGraph nodes	349
8.11 Adaptive Arithmetic Decoder for BIFS-Anim	357
8.12 Informative : Adaptive Arithmetic Encoder for BIFS-Anim	359
8.13 View Dependent Object Scalability	360
9 The Extensible MPEG-4 Textual Format.....	381
9.1 Introduction	381
9.2 XMT-A Format	381
9.3 XMT-Ω Format	433
9.4 XMT-C Modules	478
9.5 XMT Schemas.....	486
9.6 Informative: XMT/X3D Compatibility.....	486
9.7 Informative: The usage of XMT-A BitWrapper element in authoring side	487
10 MPEG-J	500

10.1	Architecture	500
10.2	MPEG-J Session	502
10.3	Delivery of MPEG-J Data.....	503
10.4	MPEG-J API List	506
10.5	Informative: Starting the Java Virtual Machine	512
10.6	Informative: Examples of MPEG-J API usage	513
Annex A	(normative) Curve-based animators.....	522
Annex B	(normative) Procedural textures algorithms	525
Annex C	(informative) Text Processing in BIFS.....	530
Annex D	(informative) Patent statements.....	532
Bibliography	533

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC 14496-11 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information Technology*, Subcommittee SC 29, *Coding of Audio, Picture, Multimedia and Hypermedia Information*.

This second edition cancels and replaces the first edition, which has been technically revised.

ISO/IEC 14496 consists of the following parts, under the general title *Information technology — Coding of audio-visual objects*:

- *Part 1: Systems*
- *Part 2: Visual*
- *Part 3: Audio*
- *Part 4: Conformance testing*
- *Part 5: Reference software*
- *Part 6: Delivery Multimedia Integration Framework (DMIF)*
- *Part 7: Optimized reference software for coding of audio-visual objects* [Technical Report]
- *Part 8: Carriage of ISO/IEC 14496 contents over IP networks*
- *Part 9: Reference hardware description* [Technical Report]
- *Part 10: Advanced Video Coding*
- *Part 11: Scene description and application engine*
- *Part 12: ISO base media file format*
- *Part 13: Intellectual Property Management and Protection (IPMP) extensions*
- *Part 14: MP4 file format*

- *Part 15: Advanced Video Coding (AVC) file format*
- *Part 16: Animation Framework eXtension (AFX)*
- *Part 17: Streaming text format*
- *Part 18: Font compression and streaming*
- *Part 19: Synthesized texture stream*
- *Part 20: Lightweight Application Scene Representation (LSeR) and Simple Aggregation Format (SAF)*
- *Part 21: MPEG-J GFX*

Introduction

1.1 Scene Description

1.1.1 Overview

ISO/IEC 14496 addresses the coding of audio-visual objects of various types: natural video and audio objects as well as textures, text, 2- and 3-dimensional graphics, and also synthetic music and sound effects. To reconstruct a multimedia scene at the terminal, it is hence not sufficient to transmit the raw audio-visual data to a receiving terminal. Additional information is needed in order to combine this audio-visual data at the terminal and construct and present to the end-user a meaningful multimedia scene. This information, called scene description, determines the placement of audio-visual objects in space and time and is transmitted together with the coded objects as illustrated in Figure 1. Note that the scene description only describes the structure of the scene. The action of assembling these objects in the same representation space is called composition. The action of transforming these audio-visual objects from a common representation space to a specific presentation device (i.e. speakers and a viewing window) is called rendering.

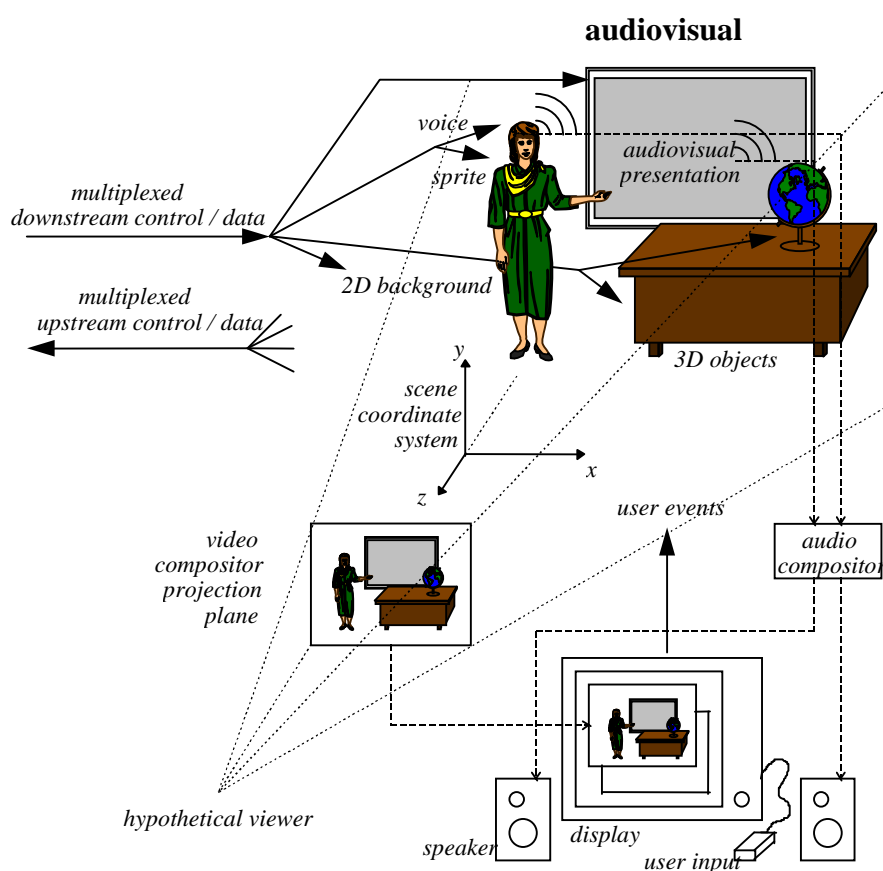


Figure 1 — An example of an object-based multimedia scene

Independent coding of different objects may achieve higher compression, and also brings the ability to manipulate content at the terminal. The behaviors of objects and their response to user inputs can thus also be represented in the scene description.

The scene description framework used in this part of ISO/IEC 14496 is based largely on ISO/IEC 14772-1:1998 (Virtual Reality Modeling Language – VRML).

1.1.2 Composition and Rendering

ISO/IEC 14496-11 defines the syntax and semantics of bitstreams that describe the spatio-temporal relationships of audio-visual objects. For visual data, particular composition algorithms are not mandated since they are implementation-dependent; for audio data, subclause 7.1.1.2.13 and the semantics of the AudioBIFS nodes normatively define the composition process. The manner in which the composed scene is presented to the user is not specified for audio or visual data. The scene description representation is termed “Binary Format for Scenes” (BIFS).