
**Information technology — Multimedia
application format (MPEG-A) —**

**Part 12:
Interactive music application format**

*Technologies de l'information — Format pour application multimédia
(MPEG-A) —*

Partie 12: Format d'application musicale interactive



Reference number
ISO/IEC 23000-12:2010(E)

AVA Documents
Document Sharing Hub
© ISO/IEC 2010

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2010

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword	iv
Introduction.....	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Overview.....	2
4.1 General	2
4.2 File contents and organization	2
4.3 Usage model	2
5 Supported components of Interactive Music AF.....	3
6 File Structure	4
6.1 Table for boxes	4
6.2 File structure for a single type file	5
6.3 File structure for a multiple type file	6
6.4 Backward compatibility to legacy music player	6
6.5 Hierarchical structure of audio tracks.....	7
6.6 Preset Information.....	9
6.7 Interactivity Rules.....	13
6.8 Timed text data	19
6.9 Metadata	20
7 Brand Identification.....	22
Annex A (informative) Mode Usage in Interactive music AF player	23
Annex B (informative) Personal-mix Usage Example	26

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 23000-12 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

ISO/IEC 23000 consists of the following parts, under the general title *Information technology — Multimedia application format (MPEG-A)*:

- *Part 1: Purpose for multimedia application formats* [Technical Report]
- *Part 2: MPEG music player application format*
- *Part 3: MPEG photo player application format*
- *Part 4: Musical slide show application format*
- *Part 5: Media streaming application format*
- *Part 6: Professional archival application format*
- *Part 7: Open access application format*
- *Part 8: Portable video application format*
- *Part 9: Digital Multimedia Broadcasting application format*
- *Part 10: Video surveillance application format*
- *Part 11: Stereoscopic video application format*
- *Part 12: Interactive music application format*

Introduction

The market for interactive music service which provides users with an experience of recomposing the music to their taste is getting formed and matured. Hence, a standardized form needs to be specified to guarantee the interoperability of the interactive music contents in order to broaden and to boost the interactive music service market.

Interactive music application format (IM AF) defines a file format designed for interactive music service. It specifies how to combine the multiple audio tracks before conventional mixing process, with associated information for a presentation in a well-defined format that facilitates storage, interchange, management, editing, and presentation of interactive music contents in interoperable ways.

Information technology — Multimedia application format (MPEG-A) —

Part 12: Interactive music application format

1 Scope

This part of ISO/IEC 23000 specifies a file format designed for interactive music services. It integrates the multiple audio tracks with appropriate additional information for enabling users to experience various preset mixes and to make their own mixes complying with interactivity rules imposed by the producer.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 10918-1:1994, *Information technology — Digital compression and coding of continuous-tone still images: Requirements and guidelines*

ISO/IEC 11172-3:1993, *Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 3: Audio*

ISO/IEC 14496-3:2009, *Information technology — Coding of audio-visual objects — Part 3: Audio*

ISO/IEC 14496-12:2008, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

ISO/IEC 15938-5:2003, *Information technology — Multimedia content description interface — Part 5: Multimedia description schemes*

ISO/IEC 23003-2, *Information technology — MPEG audio technologies — Part 2: Spatial Audio Object Coding (SAOC)*

3GPP TS 26.245, Version 6.1.0, *Transparent end-to-end Packet switched Streaming Service (PSS) — Timed text format*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

3.1

track

collection of related samples in an IM AF file

NOTE For media data, a track corresponds to a sequence of images or sampled audio and an audio track is either mono or stereo. For hint tracks, a track corresponds to a streaming channel.

4 Overview

4.1 General

This clause provides necessary information on the content, the organization and the playback of an IM AF file.

4.2 File contents and organization

Creating an IM AF file involves formatting different types of media data, especially multiple audio tracks with interactivity data and storing them into an ISO-Based Media File Format (ISO-BMFF).

An IM AF file is composed of:

- multiple audio tracks which represent the several parts (e.g. instruments and/or voices) of a song (i.e. a musical piece);
- groups of audio tracks which define a hierarchical structure of audio tracks (e.g. all guitars of a song can be gathered in a same group);
- preset data which is pre-defined mixing information on multiple audio tracks (e.g. karaoke and rhythmic version);
- rules which introduce specific data related to user's interaction (e.g. definition of action allowed concerning audio tracks/groups selection and audio volume control);
- additional media can be used to enrich the user's interaction space (e.g. timed text synchronized with audio tracks which can represent the lyrics of a song, images related to the song, to the music album and/or to the artist);
- metadata can be used in the same goal (e.g. data used to describe the song, the music album and/or the artist).

4.3 Usage model

By means of an interactive music player, users can listen and handle audio tracks that compose an IM AF file (and eventually the other media and metadata). Two possible mix modes are defined for interaction and playback:

- the Preset-mix mode;
- the User-mix mode.

In Preset-mix mode, user selects one preset among the presets stored in the IM AF file then audio tracks are mixed according to preset parameters associated with selected preset.

In User-mix mode, user selects/unselects the audio tracks/groups and/or controls the volume of each of them. Each user's interaction is analyzed in order to determine compatibility with rules defined by the producer and/or the artist. This compatibility is verified by means of a rule analyzer (see Figure 1). A detailed description of this checking is provided in Annex A.

Figure 1 shows the block diagram for both the Preset-mix mode and the User-mix mode. The preset selection, the audio track/group selection and the volume change are the possible user's interaction.

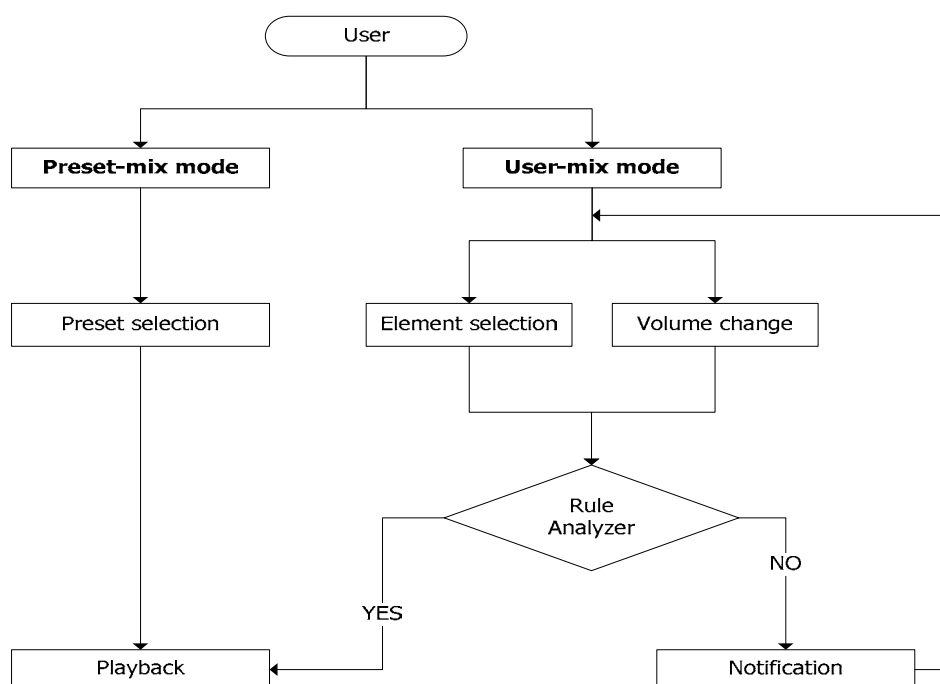


Figure 1 — The illustration of IM AF usage mode

5 Supported components of Interactive Music AF

In Table 1, all the supported components of IM AF are listed with the specification reference of the respective component. Note that IM AF does not mandate all the syntax and semantics of the 'Specification' column of Table 1. Instead, it does mandate the component specification with some restrictions on usages, which are defined in the specification of 'Restriction' column of the table.

IM AF file includes various kinds of audio stream such as MP3, AAC, SAOC and PCM. Also it includes JPEG for image, 3GPP Timed Text for text and MPEG-7 MDS for metadata. In this specification, ISO-BMFF is used as the base file format structure.

Table 1 — Supported components of IM AF

Type	Component Name	Abbreviation	Specification	Restriction
File Format	ISO Base Media File Format	ISO-BMFF	ISO/IEC 14496-12:2008	Defined in sub-clause 6.1
Audio	MPEG-1 Audio Layer III	MP3	ISO/IEC 11172-3:1993	None
	MPEG-4 Audio AAC profile	AAC	ISO/IEC 14496-3:2005	None
	MPEG-D SAOC Baseline profile	SAOC	ISO/IEC 23003-2:200x	None
	PCM	PCM	-	None
Image	JPEG Image	JPEG	ISO/IEC 10918-1	Baseline coding
Text	3GPP Timed Text	3GPP TT	3GPP TS 26.245	None
Metadata	MPEG-7 Multimedia Description Scheme	MDS	ISO/IEC 15938-5:2003	Defined in sub-clause 6.9

Note that audio shall be done in accordance with MP3 (ISO/IEC 11172-3), AAC (ISO/IEC 14496-3) and SAOC (ISO/IEC 23003-2). And requirements of JPEG (ISO/IEC 10918-1) apply for image in IM AF.

6 File Structure

6.1 Table for boxes

The file format of IM AF can be represented in two kinds of structures for single IM AF content and multiple IM AF contents. The file format structure of this specification is derived from the ISO-BMFF standard. Table 2 shows the structure of the boxes and their description. The mandatory boxes are marked with an asterisk (*).

Unless specifically mentioned in this specification, the requirements of ISO-BMFF (ISO/IEC 14496-12) apply for the listed boxes.

Table 2 — Table for boxes of IM AF

*	ftyp					file type and compatibility
*	moov					container for all the metadata
		mvhd				movie header, overall declarations
		trak				container for an individual track or stream
*			tkhd			track header, overall information about the track
			tref			track reference container
			edts			edit list container
				elst		an edit list
*			mdia			container for the media information in a track
*				mdhd		media header, overall information about the media
*				hdlr		handler, declares the media (handler) type "soun" for audio data "text" for timed text data "hint" for protocol hint track
*				minf		media information container
					smhd	sound media header, overall information (sound track only)
					hmhd	hint media header, overall information (hint track only)
					nmhd	Null media header, overall information (some tracks only)
*					dinf	data information box, container
*					dref	data reference box, declares source(s) of media data in track
*					stbl	sample table box, container for the time/space map
*					stsd	sample descriptions (codec types, initialization etc.)
*					stts	(decoding) time-to-sample
*					stsc	sample-to-chunk, partial data-offset information
					stsz	sample sizes (framing)
					stz2	compact sample sizes (framing)
*					stco	chunk offset, partial data-offset information
					co64	64-bit chunk offset
		grco				container for the groups
			grup			group box, describes the structure (hierarchy)
*		prco				container for the presets
*			prst			preset box, container for the preset information
		ruco				container for rules
			rusc			selection rule box, container for a selection rule
			rumx			mixing rule box, container for a mixing rule
	mdat					media data container
	free					free space
	skip					free space
	meta					Metadata
*		hdlr				handler, declares the metadata (handler) type
		dinf				data information box, container
			dref			data reference box, declares source(s) of metadata items
		iloc				item location
		iinf				item information
		xml				XML container
		bxml				binary XML container
		pitm				primary item reference

6.2 File structure for a single type file

Figure 2 illustrates a single type file structure of IM AF which is containing single movie presentation with associated data. This type of file structure mainly consists of ftyp, moov and mdat boxes. The moov box describes the presentation of the scene in which more than one trak box are contained. The trak box contains the presentation description for one media. A media in each trak box can be individual audio, image and text as shown in Figure 2. The trak box supports time information for the synchronization with other media of other trak box. The mdat box contains the real contents which are described in the trak box or the trak box may import the content by URL without depositing the content in the mdat box.

Besides this type of file structure includes grco, prco and ruco boxes. The grco box contains zero or more grup boxes and each grup box describes the hierarchy structure of a single group which is composed of one or more audio tracks and/or groups. The prco box contains one or more prst boxes which describe the predefined mixing information. The ruco box contains zero or more rusc boxes and/or rumx boxes which describe the interactivity rules related to selection and/or mixing of audio tracks contained in the file.

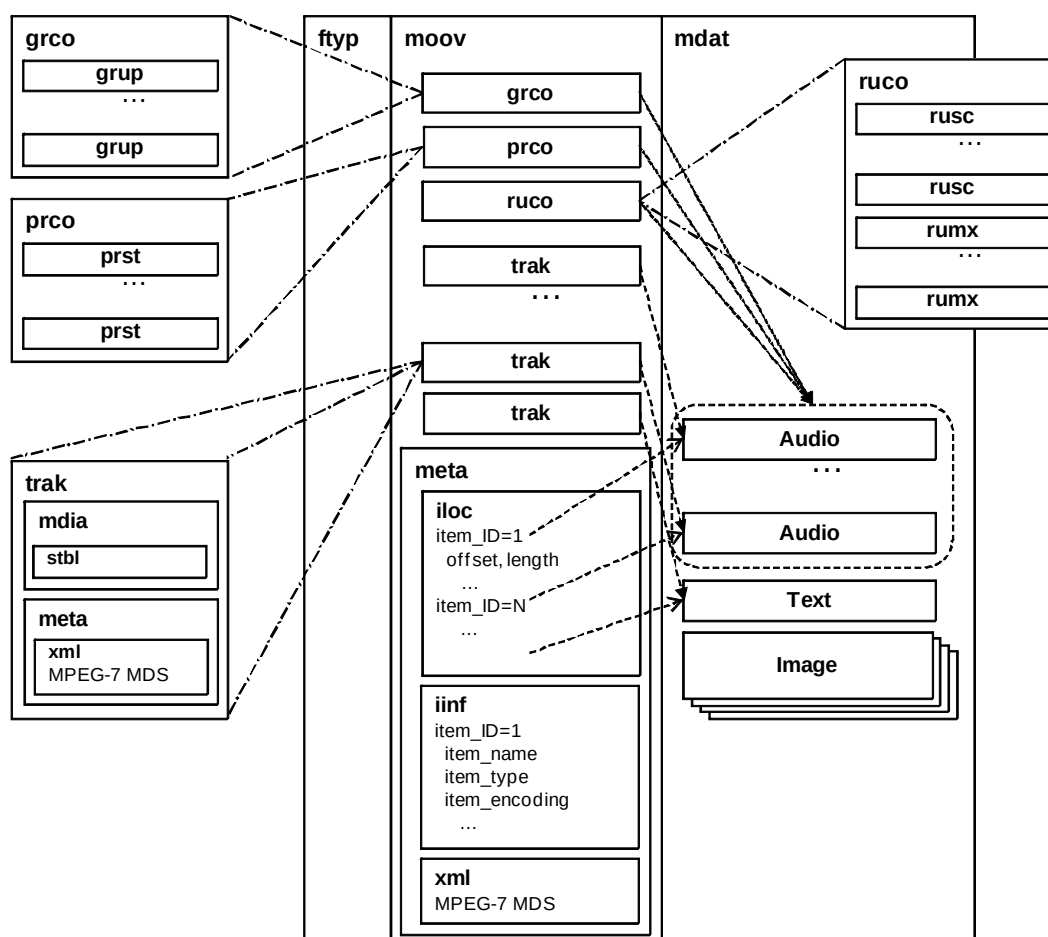


Figure 2 — A single type file structure with meta box