INTERNATIONAL STANDARD

ISO/IEC 14496-3

Fourth edition
2009-09-01

AMENDMENT 5
2015-08-01

Information technology — Coding of audio-visual objects —

Part 3: **Audio**

AMENDMENT 5: Support for Dynamic Range Control, New Levels for ALS Simple Profile, and Audio Synchronization

Technologies de l'information — Codage des objets audiovisuels — Partie 3: Codage audio

AMENDEMENT 5: Aide pour le contrôle de plage dynamique, nouveaux niveaux pour profil simple ALS et synchronisation audio

Reference number
ISO/IEC 14496-3:2009/Amd.5:2015(E)

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.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

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. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

Information technology — Coding of audio-visual objects —

Part 3: Audio

AMENDMENT 5: Support for Dynamic Range Control, New Levels for ALS Simple Profile, and Audio Synchronization

1 Changes to the text of ISO/IEC 14496-3:2009

After 0.3.8.4, add:

0.3.9 Audio Synchronization Tool

The audio synchronization tool provides capability of synchronizing multiple contents in multiple devices. Synchronization is done by using audio features (fingerprint) extracted from the content. Neither common clock covering the multiple devices nor way to exchange time-stamps between the devices is required.

In the cover page of Part 3: Audio, replace:

This part of ISO/IEC 14496 contains twelve subparts:

with

This part of ISO/IEC 14496 contains thirteen subparts:

In the cover page of Part 3: Audio, add:

Subpart 13: Audio Synchronization

after

Subpart 12: Scalable lossless coding

In 1.3 Terms and Definitions, add:

1.3.z **Audio Sync**: Audio feature for synchronization

and increase the index-number of subsequent entries

In 1.5.1.1 Audio object type definition, amend Table 1.1 with the updates in the table below:

Object type ID	Audio object type	Gain control	[...]	Remark
0	Null			
[..]	[...]			
43	SAOC			
44	LD MPEG Surround			
45	SAOC-DE			
46	Audio Sync			
47 to 95	(reserved)			

After 1.5.1.2.40 add the following new subclauses:

1.5.1.2.41 Audio Sync object type

The Audio Sync object type conveys audio feature for multiple media stream synchronization (see ISO/IEC 14496-3 Subpart 13) in the MPEG-4 Audio framework.

In 1.5.2.1 (Profiles), Table 1.3 (Audio Profiles definition), add:

Object type ID	Audio object type	...
...	...	...
43	SAOC	
44	LD MPEG Surround	
45	SAOC-DE	
46	Audio Sync	

In 1.5.2.3 (Levels within the profiles), replace Table 1.13B and notes with:

— **Levels for the ALS Simple Profile**

Table 1.13B — Level for the ALS Simple Profile

Level	Max. number of channels	Max. sampling rate [kHz]	Max. word length [bit]	Max. number of samples per frame	Max. prediction order	Max. BS* stages	Max. MCC** stages
1	2	48	16	4096	15	3	1
2	2	48	24	4096	15	3	1
3	6	48	16	4096	15	3	1
4	6	48	24	4096	15	3	1

* BS: Block switching, ** MCC: Multi-channel coding

The BGMC tool and the RLS-LMS tool are not permitted. Floating-point audio data is not supported.

Insert the following new entries into Table 1.14 “audioProfileLevelIndication values” and adapt the “reserved for ISO use” range accordingly:

0x58	SAOC Dialogue Enhancement Profile	L1
0x59	SAOC Dialogue Enhancement Profile	L2
0x5A	ALS Simple Profile	L2
0x5B	ALS Simple Profile	L3
0x5C	ALS Simple Profile	L4
0x5D to 0x7F	reserved for ISO use	—

In 1.6.2.1 extend Table 1.15 "AudioSpecificConfig()"as follows:

Table 1.15 — Syntax of AudioSpecificConfig()

Syntax	No. of bits	Mnemonic
AudioSpecificConfig()		
{		
...		
switch (audioObjectType) {		
case 1:		
case 2:		
...		
case 43:		
saocPresentFlag = 1:		
saocPayloadEmbedding	**1**	**uimsbf**
SaocSpecificConfig():		
break;		
case 44:		
ldmpsPresentFlag = 1:		
ldsacPayloadEmbedding	**1**	**uimsbf**
LDSpatialSpecificConfig():		
break;		
case 45:		
saocDepresentFlag = 1:		
saocDePayloadEmbedding	**1**	**uimsbf**
SaocDeSpecificConfig():		
break;		
case 46:		
AudioSyncFeatureSpecificConfig():		
break;		
default:		
/* reserved */		
}		
}		

After 1.6.2.1.20 add the new subclause as follows:

1.6.2.1.21 AudioSyncFeatureSpecificConfig

Defined in ISO/IEC 14496-3 Subpart 13.

In 1.6.2.2.1 extend Table 1.17 "Audio Object Types" as follows:

Table 1.17 — Audio Object Types

Object type ID	Audio object type	Definition of elementary stream payloads and detailed syntax	Mapping of audio payloads to access units and elementary streams
0	NULL		
...			
43	SAOC	ISO/IEC 23003-2	
44	LD MPEG Surround	ISO/IEC 23003-2	
45	SAOC-DE	ISO/IEC 23003-2:2010/Amd.3	
46	Audio Sync	ISO/IEC 14496-3 Subpart 13	

In Table 4.57 add:

Table 4.57 — Syntax of extension_payload()

Syntax	No. of bits	Mnemonic
extension_payload(cnt)		
{		
extension_type;	**4**	**uimsbf**
align = 4;		
switch(extension_type) {		
case EXT_DYNAMIC_RANGE:		
return dynamic_range_info();		
case EXT_UNI_DRC:		
return uniDrc();		
case EXT_SAC_DATA:		
return sac_extension_data(cnt);		
case EXT_SAOC_DATA:		
return saoc_extension_data(cnt);		
case EXT_LDSAC_DATA:		
return ldsac_extension_data(cnt);		
case EXT_SBR_DATA:		
return sbr_extension_data(id_aac, 0);		Note 1
case EXT_SBR_DATA_CRC:		
return sbr_extension_data(id_aac, 1);		Note 1
case EXT_SAOC_DE_DATA:		
return saoc_de_extension_data(cnt);		
case EXT_DATA_LENGTH:		
...		

In Table 4.121 add:

Table 4.121 — Values of the extension_type field

1. Symbol	2. Value of extension_type	3. Purpose
EXT_FILL	‘0000’	bitstream payload filler
EXT_FILL_DATA	‘0001’	bitstream payload data as filler
EXT_DATA_ELEMENT	’0010‘	data element
EXT_DATA_LENGTH	‘0011’	container with explicit length for extension_payload()
EXT_UNI_DRC	’0100‘	Unified dynamic range control
EXT_LDSAC_DATA	‘1001’	LD MPEG Surround
EXT_SAOC_DATA	‘1010’	SAOC
EXT_DYNAMIC_RANGE	‘1011’	dynamic range control
EXT_SAC_DATA	‘1100’	MPEG Surround
EXT_SBR_DATA	‘1101’	SBR enhancement
EXT_SBR_DATA_CRC	‘1110’	SBR enhancement with CRC
EXT_SAOC_DE_DATA	‘1111’	SAOC-DE
-	all other values	Reserved: These values can be used for a further extension of the syntax in a compatible way.
Note: Extension payloads of the type EXT_FILL or EXT_FILL_DATA have to be added to the bitstream payload if the total bits for all audio data together with all additional data are lower than the minimum allowed number of bits in this frame necessary to reach the target bitrate. Those extension payloads are avoided under normal conditions and free bits are used to fill up the bit reservoir. Those extension payloads are written only if the bit reservoir is full.		

In 4.5.14.1.1 Data elements, replace:

Table AMD4.7 - Definition of downmix procedure

stereo_downmix_mode	downmix procedure
0	Lo/Ro
1	Lt/Rt

with:

Table AMD4.7 - Definition of downmix procedure

stereo_downmix_mode	downmix procedure
0	Lo/Ro
1	Lo/Ro or Lt/Rt

In 4.5.2.14.2 “Decoding Process”, rename the headline of 4.5.2.14.2.1

4.5.2.14.2.1 Downmixing from 5.1 to Stereo

as

4.5.2.14.2.1 Downmixing from 5.1 to Stereo/Mono

Immediately after this headline add a new subclause headline:

4.5.2.14.2.1.1 Downmixing to Stereo

In 4.5.14.2.1.1 Downmixing to stereo, replace:

if **stereo_downmix_mode** is 0,

$$L' = L + C \times b + Ls \times a + LFE \times c$$

$$R' = R + C \times b + Rs \times a + LFE \times c$$

else if **stereo_downmix_mode** is 1,

$$L' = L + C \times b - (Ls + Rs) \times a + LFE \times c$$

$$R' = R + C \times b + (Ls + Rs) \times a + LFE \times c$$

where **surround_mix_level**, “a” and **center_mix_level**, “b” are shown as “Multiplication factor” in Table AMD4.8. C, L, R, Ls, Rs are the source signals and L’ and R’ are the derived stereo signals. LFE channels should be omitted from the mixdown (i.e. c is equal to zero) if **ext_downmixing_lfe_level_status** is “0”. If **ext_downmixing_lfe_level_status** is “1”, the LFE mix level “c” shall be derived as shown in Table AMD4.9.

with:

if **stereo_downmix_mode** is 0,

$$Lo = L + C \times b + Ls \times a + LFE \times c$$

$$Ro = R + C \times b + Rs \times a + LFE \times c$$

else if **stereo_downmix_mode** is 1,

$$Lo = L + C \times b + Ls \times a + LFE \times c$$

$$Ro = R + C \times b + Rs \times a + LFE \times c$$

or

$$Lt = L + C \times b - (Ls + Rs) \times a + LFE \times c$$

$$Rt = R + C \times b + (Ls + Rs) \times a + LFE \times c$$

where **surround_mix_level**, “a” and **center_mix_level**, “b” are shown as “Multiplication factor” in Table AMD4.8. C, L, R, Ls, Rs are the source signals and Lo/Ro or Lt/Rt are the derived stereo signals.

If stereo_downmix_mode is “0”, the decoder should apply a downmix by obtaining Lo and Ro. If stereo_downmix_mode is “1”, the decoder may obtain Lt and Rt as an alternative to Lo and Ro.

LFE channels should be omitted from the mixdown (i.e. c is equal to zero) if **ext_downmixing_lfe_level_status** is “0”. If **ext_downmixing_lfe_level_status** is “1”, the LFE mix level “c” shall be derived as shown in Table AMD4.9.

Further, after Table AMD4.9, insert the following subclause:

4.5.2.14.2.1.2 Downmixing to Mono

$$M' = L + R + 2 \times C \times b + (Ls + Rs) \times a + 2 \times LFE \times c$$

where **surround_mix_level**, “a” and **center_mix_level**, “b” are shown as “Multiplication factor” in Table AMD4.8. C, L, R, Ls, Rs are the source signals and M’ is the derived mono signal. LFE channels should be omitted from the mixdown (i.e. c is equal to zero) if **ext_downmixing_lfe_level_status** is “0”. If **ext_downmixing_lfe_level_status** is “1”, the LFE mix level “c” shall be derived as shown in Table AMD4.9.

In 4.5.2.14.2.5 after “ Table AMD4.12: Default values after synchronization” add:

In addition the “actual compression value” shall be set to 1.0 (0 dB).

Add new section 4.5.2.16 immediately before 4.5.3 with the following text:

4.5.2.16 Unified Dynamic Range Control

The DRC tool specified in ISO/IEC 23003-4 is supported. The corresponding data is carried in an extension payload with the type EXT_UNI_DRC. The DRC tool is operated in regular delay mode and the DRC frame size has the same duration as the AAC frame size.

The time resolution of the DRC tool is specified by *deltaTmin* in units of the audio sample interval. It is calculated as specified in 23003-4. Specific values are provided here as examples based on the following formula:

$$\text{deltaT}\min = 2^M .$$

The applicable exponent M is found by looking up the audio sample rate range that fulfils:

$$f_{s,min} \le f_s < f_{s,max}$$

Table — AMD5.1 — Lookup table for the exponent *M*

fs,min [Hz]	*fs*,max [Hz]	*M*
8 000	16 000	3
16 000	32 000	4
32 000	64 000	5
64 000	128 000	6

Given the codec frame size N_{Codec}, the DRC frame size in units of DRC samples at a rate of *deltaTmin* is:

$$N_{DRC} = N_{Codec} 2^{-M} .$$

For AAC, the DRC tool of 23003-4 offers mandatory decoding capability of up to four DRC subbands using the time-domain DRC filter bank. Optionally, more DRC subbands can be supported by replacing the time-domain DRC filter bank by a uniform 64-band QMF analysis and synthesis filter bank, such as the one defined for HE-AAC. DRC sets that contain more than four DRC subbands must contain gain sequences that are all aligned with the QMF domain.

For HE-AAC and HE-AACv2 decoders the DRC gains are applied to the sub-bands of the QMF domain immediately before the synthesis filter bank.

The drcLocation parameter shall be encoded according to Table AMD5.2.