INTERNATIONAL STANDARD

ISO/IEC 14496-3

Fourth edition
2009-09-01

AMENDMENT 7
2018-09

Information technology — Coding of audio-visual objects —

Part 3: **Audio**

AMENDMENT 7: SBR enhancements

Technologies de l'information — Codage des objets audiovisuels —

Partie 3: Codage audio

AMENDEMENT 7

Reference number
ISO/IEC 14496-3:2009/Amd.7:2018(E)

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

Published in Switzerland

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 of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html.

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

A list of all parts in the ISO/IEC 14496 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Information technology — Coding of audio-visual objects —

Part 3: Audio

AMENDMENT 7: SBR enhancements

1.5.1.2.6

Replace the first sentence with:

> The SBR Object contains the SBR-Tool and the MPEG-4 SBR Enhancements as signalled in the SBR extension element (bs_extension_id== EXTENSION_ID_ESBR) of Annex 8.A and can be combined with the audio object types indicated in Table 1.2. If a decoder detects and supports this SBR extension element, the decoder shall process the MPEG-4 SBR Enhancements.

1.5.2.2

In Table 1.4, update the row for Object Type SBR as follows:

SBR	fs = 24/48 kHz (in/out) (SBR tool)	3	2.5	1)
	fs = 24/48 kHz (in/out) (eSBR tool)	4.5	3	1)
	fs = 24/48 kHz (in/out) (Low Power SBR tool)	2	1.5	1)
	fs = 48/48 kHz (in/out) (Down Sampled SBR tool)	4.5	2.5	1)
	fs = 48/48 kHz (in/out) (Low Power Down Sampled SBR tool)	3	1.5	1)

1.5.2.3

Replace Table 1.11 with the following:

Level	Max. channels/ object	Max. AAC sampling rate, SBR not present [kHz]	Max. AAC sampling rate, SBR present [kHz]	Max. SBR sampling rate [kHz] (in/out)	Max. PCU	Max. RCU	Max. PCU Low power SBR	Max. RCU Low power SBR
1	NA	NA	NA	NA	NA	NA	NA	NA
2	2	48	24	24/48	11	11	7	8
3	2	48	48	48/48[a]	17	11	12	8
4	5	48	24/48[b]	48/48[a]	30	31	20	23
5	5	96	48	48/96	60	33	39	23
6	7	48	24/48[b]	48/48[a]	42	41	27	30
7	7	96	48	48/96	82	44	53	30

[a] For level 3, level 4 and level 6 decoders, it is mandatory to operate the SBR tool in downsampled mode if the sampling rate of the AAC core is higher than 24 kHz. Hence, if the SBR tool operates on a 48 kHz AAC signal, the internal sampling rate of the SBR tool will be 96 kHz, however, the output signal will be downsampled by the SBR tool to 48 kHz.

[b] For one or two channels the maximum AAC sampling rate, with SBR present, is 48 kHz. For more than two channels the maximum AAC sampling rate, with SBR present, is 24 kHz.

Replace Table 1.12 with the following:

Level[a]	Max. channels/ object	Max. AAC sampling rate, SBR not present [kHz]	Max. AAC sampling rate, SBR present [kHz]	Max. SBR sampling rate [kHz] (in/out)	Max. PCU	Max. RCU	Max. PCU HQ / LP SBR[e]	Max. RCU HQ / LP SBR[e]
1	NA	NA	NA	NA	NA	NA	NA	NA
2	2	48	24	24/48	11	11	9	10
3	2	48	24/48[c]	48/48[b]	17	11	15	10
4	5	48	24/48[d]	48/48[b]	30	31	20	23
5	5	96	48	48/96	60	33	39	23
6	7	48	24/48[d]	48/48[b]	42	41	27	30
7	7	96	48	48/96	82	44	53	30

[a] Level 2, 3, 4, 6 and 7 HE AAC v2 Profile decoders implement the baseline version of the parametric stereo tool. A level 5 decoder shall not be limited to the baseline version of the parametric stereo tool.

[b] For level 3, level 4 and level 6 decoders, it is mandatory to operate the SBR tool in downsampled mode if the sampling rate of the AAC core is higher than 24 kHz. Hence, if the SBR tool operates on a 48 kHz AAC signal, the internal sampling rate of the SBR tool will be 96 kHz, however, the output signal will be downsampled by the SBR tool to 48 kHz.

[c] If Parametric Stereo data is present the maximum AAC sampling rate is 24 kHz, if Parametric Stereo data is not present the maximum AAC sampling rate is 48 kHz.

[d] For one or two channels the maximum AAC sampling rate, with SBR present, is 48 kHz. For more than two channels the maximum AAC sampling rate, with SBR present, is 24 kHz.

[e] The PCU/RCU number are given for a decoder operating the LP SBR tool whenever applicable.

1.6.7.2

Replace Table 1.27 with the following:

Value of *n*	Additional delay[a]	Decoder operation mode
1	0	A) All operation modes not listed elsewhere in this table.
963	962	B1) HE-AAC or HE-AAC v2 decoder with SBR operated in dual-rate mode; decoding HE-AAC or HE-AAC v2 compressed audio.
482	481	B2) Same as B1), but with SBR operated in downsampled mode.
3011	3010	B3) Same as B1), but with SBR Enhancements enabled.
[a] The delay introduced by the post-processing is given in number of samples (per audio channel) at the output sample rate for the given decoder operation mode.		

4.5.2.8.1

Replace Table 4.112 with the following:

bs_extension_id	Meaning
0	reserved
1	reserved
2	EXTENSION_ID_PS[a]
3	EXTENSION_ID_ESBR[a]
[a] See subclause 8.A.	

After Table 4.119 add the following:

esbr_fill_bits Fill bits to enable correct parsing of the sbr_extension() element.

Annex 8.A

Replace the Annex title with:

Combination of the SBR tool with the parametric stereo tool and SBR Enhancements

8.A.1, paragraph 2

At the end of paragraph 2 add a new paragraph:

In addition, the harmonic transposers and SBR pre-processing as defined in ISO/IEC 23003-3 may be used in combination with the SBR tool as defined in 4.6.18. The bitstream element esbr_data() as defined in 8.A.2 conveys the information needed by these tools and is carried in the sbr_extension() container of the SBR bitstream.

8.A.2

Replace the first sentence with:

> The bitstream element ps_data() as defined in 8.4.2, and the bitstream element esbr_data() as defined in Table 8.A.2, are carried in the sbr_extension() container (see Table 8.A.1 below) provided by the SBR bitstream defined in 4.4.2.8. The semantics of the bs_extension_id field are given in Table 4.105 "bs_extension_id".

Replace Table 8.A.1 with the following:

Syntax	No. of bits	Mnemonic
sbr_extension(bs_extension_id, num_bits_left)		
{		
switch (bs_extension_id) {		
case EXTENSION_ID_PS:		
num_bits_left -= ps_data();		a
break;		
case EXTENSION_ID_ESBR:		
num_bits_left -= esbr_data();		b
break;		
default:		
bs_fill_bits;	num_bits_left	**bslbf**
num_bits_left = 0;		
break;		
}		
}		
a ps_data() returns the number of bits read.		
b esbr_data() returns the number of bits read.		

After Table 8.A.1 add a new table:

Table 8.A.2 — Syntax of esbr_data()

Syntax	No. of bits	Mnemonic
esbr_data(id_aac, bs_coupling)		
{		
num_bits_esbr = 1		
bs_sbr_preprocessing;	**1**	**uimsbf**
if (id_aac == ID_SCE) {		
num_bits_esbr += 1		
if (**sbrPatchingMode[0]** == 0) {	**1**	**uimsbf**
num_bits_esbr += 2		
sbrOversamplingFlag[0];	**1**	**uimsbf**
if (**sbrPitchInBinsFlag[0]**) {	**1**	**uimsbf**
num_bits_esbr += 7		
sbrPitchInBins[0];	**7**	**uimsbf**
} else		
sbrPitchInBins[0] = 0;		
} else {		
sbrOversamplingFlag[0] = 0;		
sbrPitchInBins[0] = 0;		
}		
} else if (id_aac == ID_CPE) {		
If (bs_coupling) {		
num_bits_esbr += 1		
if (**sbrPatchingMode[0,1]** == 0) {	**1**	**uimsbf**
num_bits_esbr += 2		
sbrOversamplingFlag[0,1];	**1**	**uimsbf**
if (**sbrPitchInBinsFlag[0,1]**) {	**1**	**uimsbf**
num_bits_esbr += 7		
sbrPitchInBins[0,1];	**7**	**uimsbf**
} else		
sbrPitchInBins[0,1] = 0;		
} else {		
sbrOversamplingFlag[0,1] = 0;		
sbrPitchInBins[0,1] = 0;		
}		
} else { /* bs_coupling == 0 */		
num_bits_esbr += 1		
if (**sbrPatchingMode[0]** == 0) {	**1**	**uimsbf**
num_bits_esbr += 2		
sbrOversamplingFlag[0];	**1**	**uimsbf**
if (**sbrPitchInBinsFlag[0]**) {	**1**	**uimsbf**
num_bits_esbr += 7		

NOTE *bs_sbr_preprocessing* is defined in ISO/IEC 23003-3:2012, 6.2.12. *sbrPatchingMode[ch], sbrOversamplingFlag[ch], sbrPitchInBinsFlag[ch]* and *sbrPitchInBins[ch]* are defined in ISO/IEC 23003-3:2012, 7.5.

Table 8.A.2 *(continued)*

Syntax	No. of bits	Mnemonic
sbrPitchInBins[0];	**7**	**uimsbf**
} else		
sbrPitchInBins[0] = 0;		
} else {		
sbrOversamplingFlag[0] = 0;		
sbrPitchInBins[0] = 0;		
}		
num_bits_esbr += 1		
if (**sbrPatchingMode[1]** == 0) {	**1**	**uimsbf**
num_bits_esbr += 2		
sbrOversamplingFlag[1];	**1**	**uimsbf**
if (**sbrPitchInBinsFlag[1]**) {	**1**	**uimsbf**
num_bits_esbr += 7		
sbrPitchInBins[1];	**7**	**uimsbf**
} else		
sbrPitchInBins[1] = 0;		
} else {		
sbrOversamplingFlag[1] = 0;		
sbrPitchInBins[1] = 0;		
}		
}		
}		
I		
if (num_bits_esbr < 6)		
num_fill_bits = 6 - num_bits_esbr		
else		
num_fill_bits = 0		
esbr_fill_bits	num_fill_bits	**bslbf**
}		

NOTE *bs_sbr_preprocessing* is defined in ISO/IEC 23003-3:2012, 6.2.12. *sbrPatchingMode[ch], sbrOversamplingFlag[ch], sbrPitchInBinsFlag[ch]* and *sbrPitchInBins[ch]* are defined in ISO/IEC 23003-3:2012, 7.5.

8.A.3

Replace the subclause heading with:

8.A.3 Decoding process for the PS tool

8.A.4

At the end of 8.A.4 add the following new subclause:

8.A.5 Decoding process for the SBR Enhancements

Semantics and decoding process for the SBR Enhancements are defined in ISO/IEC 23003-3:2012, 7.5.

8.C.7

At the end of 8.C.7 add the following new subclauses:

8.C.8 Recommendations for encoding settings of SBR Enhancements

8.C.8.1 General

The following clause provides recommendations for usage of the SBR Enhancements.

In order to improve the subjective quality for audio content with harmonic frequency structure and strong tonal characteristics, in particular at low bitrates, activation of the SBR Enhancements is recommended (sbr_extension(EXTENSION_ID_ESBR, ...)). The values of the corresponding bitstream element (i.e. esbr_data()), controlling these tools, may be determined in the encoder by applying a signal dependent classification mechanism.

8.C.8.2 SBR Enhancements

8.C.8.2.1 Harmonic transposer

Generally, the usage of the harmonic patching method (**sbrPatchingMode** == 0) is preferable for coding music signals at very low bitrates, where the core codec may be considerably limited in audio bandwidth. This is especially true if these signals include a pronounced harmonic structure. Contrarily, the usage of the regular SBR patching method is preferred for speech and mixed signals, since it provides a better preservation of the temporal structure in speech.

8.C.8.2.2 Pre-processing

In order to improve the performance of the MPEG-4 SBR transposer, a pre-processing step can be activated (**bs_sbr_preprocessing** == 1) that strives to avoid the introduction of spectral discontinuities of the signal going in to the subsequent envelope adjuster. The operation of the tool is beneficial for signal types where the coarse spectral envelope of the low band signal being used for high frequency reconstruction displays large variations in level.

8.C.8.2.3 Signal adaptive frequency domain oversampling

In order to improve the transient response of the harmonic SBR patching, signal adaptive frequency domain oversampling can be applied (**sbrOversamplingFlag** == 1). Since signal adaptive frequency domain oversampling increases the computational complexity of the transposer, but only brings benefits for frames which contain transients, the use of this tool is controlled by the bitstream element, which is transmitted once per frame and per independent SBR channel.