INTERNATIONAL STANDARD

ISO/IEC 14496-4

Second edition
2004-12-15

AMENDMENT 22
2008-03-01

Information technology — Coding of audio-visual objects —

Part 4:
Conformance testing

AMENDMENT 22: AudioBIFS v3 conformance

Technologies de l'information — Codage des objets audiovisuels —

Partie 4: Essai de conformité

AMENDEMENT 22: Conformité AudioBIFS v3

Reference number
ISO/IEC 14496-4:2004/Amd.22:2008(E)

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

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.

Amendment 22 to ISO/IEC 14496-4:2004 was prepared by Joint Technical Committee 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 4:
Conformance testing

AMENDMENT 22: AudioBIFS v3 conformance

In subclause 4.4.3.3 Bitstreams, insert the following row into Table 6 after the row for AABper1-76:

Table 6 – Test Sequence Providers and Reason for Existence

ABv3_AAB01-05	Thomson	Tests AudioBIFS v3 node AdvancedAudioBuffer
ABv3_ACC01-05 ABv3_ACC06a-c ABv3_ACC07a-c ABv3_ACC09-13	FhG, Thomson, FT	Tests AudioBIFS v3 node AudioChannelConfig
ABv3_SS01a-b ABv3_SS02a-b ABv3_SS04-05	FhG, FT	Tests AudioBIFS v3 node SurroundingSound
ABv3_T3DA01-08	FhG	Tests AudioBIFS v3 node Transform3Daudio
ABv3_WS0101-03	FhG, Thomson,	Tests AudioBIFS v3 node WideSound
Abv3_Afxp_Ach01 Abv3_Afxp_Aco01 Abv3_Afxp_Aec01 Abv3_Afxp_Aeq01 Abv3_Afxp_Afi01 Abv3_Afxp_Afl01 Abv3_Afxp_Ana01 Abv3_Afxp_Are01 Abv3_Afxp_Asp01 Abv3_Afxp_Ast01 Abv3_Afxp_Avi01	Thomson	Tests AudioBIFS v3 AudioFXProto effects

In subclause 6.5.1 File name conventions, insert the following row into Table 29 in alphabetical order:

Table 29 – File name conventions

A.1 AudioBIFS v3	A.2 ABv3_<nodeAbbrev><coreSetup>	A.3 -- not applicable --

*In subclause 6.5.1 File name conventions, insert the following paragraph after "***_s<speedfac>** is a number refering to the decoder configuration with regard to the speed factor.":

<nodeAbbrev> is the abbreviation of one of the AudioBIFS v3 node names (see Table 6).

After subclause 6.11.5.4.4.2 Test scenes, insert the following subclauses:

6.12 AudioBIFS v3 Nodes

6.12.1 Introduction

This clause describes the conformance testing for the rendering and output of AudioBIFS v3 nodes, which are used for adding support for shaped sound sources (**WideSound** node), Ambisonics™ audio streams to two- and three-dimensional BIFS scenes. Furthermore they allow 3D sound in 2D visual scenes (**Transform3DAudio** node) and simplify the use of pre-defined audio effects (**AudioFXProto**s). With AudioBIFS v3 also a labelling mechanism was introduced that transports the channel configuration information through the AudioBIFS sub-graph and can be altered with the **AudioChannelConfig** node.

6.12.2 Composition Unit Inputs

The input audio streams used in the conformance testing of AudioBIFS v3 shall be outputs of an AAC decoder (AOT 2), and they are monophonic, stereophonic, 5.1 (surround) and Ambisonics™ sounds. They are explained below:

ABv3_CU01_2ch_AOT2_L1R2: Composition Unit Input AAC: white noise for one second on left, for one second on right channel. Duration: 8 seconds, sampling rate 44100 Hz, stereo.

ABv3_CU02_2ch_AOT2_L3R4: Composition Unit Input AAC: after 2 seconds silence white noise for one second on left, then for one second on right channel. Duration: 8 seconds, sampling rate 44100 Hz, stereo.

ABv3_CU03_2ch_AOT2_L5R6: Composition Unit Input AAC: after 4 seconds silence white noise for one second on left, then for one second on right channel. Duration: 8 seconds, sampling rate 44100 Hz, stereo.

ABv3_CU04_1ch_AOT2_M5: Composition Unit Input AAC: after 4 seconds silence white noise for one second. Duration: 8 seconds, sampling rate 44100 Hz, mono.

ABv3_CU05_1ch_AOT2_M6: Composition Unit Input AAC: after 5 seconds silence white noise for one second. Duration: 8 seconds, sampling rate 44100 Hz, mono.

ABv3_CU06_1ch_AOT2_M0-3: Composition Unit Input AAC: 3 seconds white noise starting from the beginning. Duration: 3 seconds, sampling rate 44100 Hz, mono.

ABv3_CU07_6ch_AOT2_surround: Composition Unit Input AAC: 0.5 seconds silence, then for 0.2 seconds an 880Hz sinus tone on center channel, 0.3s silence, 0.5s 440Hz sin on front left channel, 0.5s 440Hz sin on front right channel, 0.4s 220Hz sin on left surround channel, 0.1s silence, 0.4s 220Hz sin on right surround channel, 0.1s silence, 1s white noise on LFE channel. Duration: 4 seconds, sampling rate 44100 Hz, 5.1 channels.

ABv3_CU08_1ch_AOT2_one: Composition Unit Input AAC: male voice ‘one’. Duration: 1 second, sampling rate 44100 Hz, mono.

ABv3_CU09_1ch_AOT2_two: Composition Unit Input AAC: male voice ‘two’. Duration: 1 second, sampling rate 44100 Hz, mono.

ABv3_CU10_1ch_AOT2_three: Composition Unit Input AAC: male voice ‘three’. Duration: 1 second, sampling rate 44100 Hz, mono.

ABv3_CU11_1ch_AOT2_four: Composition Unit Input AAC: male voice ‘four’. Duration: 1 second, sampling rate 44100 Hz, mono.

ABv3_CU12_1ch_AOT2_five: Composition Unit Input AAC: male voice ‘five’. Duration: 1 second, sampling rate 44100 Hz, mono.

ABv3_CU13_1ch_AOT2_six: Composition Unit Input AAC: male voice ‘six’. Duration: 1 second, sampling rate 44100 Hz, mono.

ABv3_CU14_2ch_AOT2_LCR_Binaural: Composition Unit Input AAC: male voices ‘left, center, right’, repeated 1 time. Duration: 10 seconds, sampling rate 44100 Hz, 2-channel binaural.

ABv3_CU15_2ch_AOT2_applause: Composition Unit Input AAC: applause. Duration: ~12 seconds, sampling rate 44100 Hz, 2-channel stereo.

ABv3_CU16_1ch_AOT2_applause_mono: Composition Unit Input AAC: applause. Duration: ~12 seconds, sampling rate 44100 Hz, 1-channel mono.

ABv3_CU17_2ch_AOT2_orchestra: Composition Unit Input AAC: orchestra. Duration: ~12 seconds, sampling rate 44100 Hz, 2-channel stereo.

ABv3_CU18_1ch_AOT2_orchestra_mono: Composition Unit Input AAC: orchestra. Duration: ~12 seconds, sampling rate 44100 Hz, 1-channel mono.

ABv3_CU19_1ch_AOT2_W: Composition Unit Input AAC: the omni-directional channel (0th order component W) of a 3D ambisonic recording of an ambient sound field: a complex, mostly static, outdoor scene; birds singing all around, splashes caused by stones thrown in the water first on the left then on the right, a child and adults speaking on the front, slightly on the left. In the test sequences, it has to be decoded and played synchronously with the following audio streams (ABv3_CU20_… to ABv3_CU22_...). Duration: ~14.3 seconds, sampling rate 44100 Hz, 1-channel.

ABv3_CU20_2ch_AOT2_XY: Composition Unit Input AAC: the horizontal, bidirectional (1st order) channels X and Y of the same ambisonic recording as ABv3_CU19_1ch_AOT2_W. Duration: ~14.3 seconds, sampling rate 44100 Hz, 2-channels.

ABv3_CU21_1ch_AOT2_Z: Composition Unit Input AAC: the vertical bidirectional (1st order) channel Z of the same ambisonic recording as ABv3_CU19_1ch_AOT2_W. Duration: ~14.3 seconds, sampling rate 44100 Hz, 1-channel.

ABv3_CU22_2ch_AOT2_UV: Composition Unit Input AAC: the horizontal, 2nd order channels U and V of the same ambisonic recording as ABv3_CU19_1ch_AOT2_W. Duration: ~14.3 seconds, sampling rate 44100 Hz, 2-channels.

ABv3_CU23_1ch_AOT2_ambITU_C: Composition Unit Input AAC: the center channel of a 5.0 playback version of the recording described for ABv3_CU19_1ch_AOT2_W that could be played over a standard ITU loudspeaker arrangement. Duration: ~14.3 seconds, sampling rate 44100 Hz, 1 channel.

ABv3_CU24_2ch_AOT2_ambITU_LR: Composition Unit Input AAC: the front left and front right channels the 5.0 playback version described above. Duration: ~14.3 seconds, sampling rate 44100 Hz, 2 channels.

ABv3_CU25_2ch_AOT2_ambITU_SLSR: Composition Unit Input AAC: the surround left and surround right channels the 5.0 playback version described above. Duration: ~14.3 seconds, sampling rate 44100 Hz, 2 channels.

Abv3_CU26_2ch_AOT2_seawash_cl.media: Composition Unit Input AAC: center channel and left channel ‘seawash’. Duration: ~20 seconds, sampling rate 48000 Hz, 2 channels.

Abv3_CU27_2ch_AOT2_seawash_rs.media: Composition Unit Input AAC: right channel and surround channel ‘seawash’. Duration: ~20 seconds, sampling rate 48000 Hz, 2 channels.

Abv3_AudioNaturalReverb_ImpulseResponse.wav: Original impulse response for binary comparison for test scene Abv3_aFXP_aNa01.

6.12.3 Compositor Output

The output of the audio compositor will be investigated for conformance, and shall be a single output, N channel (depending on the spatialization and reproduction method used) PCM audio stream. The input audio streams are at 16 bit signed integer sample format, and the processing defined by the Advanced Audio BIFS nodes in the scene will be carried out at an accuracy of at least 16 bits.

Because of the non-normative nature of implementing many of the AudioBIFS features, no sample-wise comparison is done to the output sound from the compositor. Some of the features can be evaluated in a static situation (no dynamic changes, such as sound source or viewpoint movements, in the 3D environment) by measuring certain parameters of the compositor's output. Some functionalities, on the other hand, require testing in a dynamic situation where only subjective evaluation can be used (the user is listening to the sound compositor output, and watching the visual compositor output if visual components are present).

In addition to objective and subjective testing a third on one is needed for AudioBIFS v3, the parameter printout. Therefore the printout of certain parameters (like the absolute position of an audio source or the current channel configuration of the audio data) can be compared with a given reference. Note, that the parameter printout does not print the *contents* of a node’s field, but the status information like channel configuration that has to be passed through the audio scene graph along with the audio data or through the BIFS scene graph like the transform hierarchy. The parameter printout should be ASCII text with the format

```
parameterName value_1 … value_n
```

whereby value should be in the format of the corresponding field and logged in the top-level sound nodes. In case of a SFFloat/MFFloat field type the 1.16 float format should be used.

6.12.4 Conformance Tests for AudioBIFS v3 Nodes

The following chapters contain the detailed description of the conformance tests for the nodes **AdvancedAudioBuffer**, **AudioChannelConfig**, **SurroundingSound**, **Transform3Daudio**, **WideSound** and **AudioFXProto** effects.

6.12.4.1 Testing of AdvancedAudioBuffer Node

The **AdvancedAudioBuffer** node provides an interface for stored sound. This node has corrected functionality and enhanced reload mechanism compared to the **AudioBuffer** node, e.g to accumulate snippets of sound in the **AdvancedAudioBuffer**. These snippets can be accessed directly or as the full accumulated content.

6.12.4.1.1 BIFS components needed in the conformance testing

For testing the load and playback mechanism of the **AdvancedAudioBuffer** node a minimal set of BIFS nodes is needed: Besides the root node and one grouping node (**Group**, **OrderedGroup**) one top-level AudioBIFS node (**Sound**, **Sound2D**) as well as a node that connects the AudioBIFS sub-graph with the decoder (**AudioSource**) is required.

6.12.4.1.2 Conformance testing procedure

Conformance testing of the **AdvancedAudioBuffer** node requires the player to support the parameter printout of **AdvancedAudioBuffer** node's fields. For subjective testing the listener has to check if playing back the scene has the described effect.

The basic behaviour of this node is determined by the **loadMode** field. For each of these five modes a test scene shall be used to check the functionalities. The read/write access to the different content blocks shall be verified with the corresponding scene by a printout of the index of the content block. Additional subjective listening tests shall be performed.

Test Scenes:

ABv3_AAB01 This scene is used for testing the compatibility mode (**LoadMode**=0) (compatible to **AudioBuffer**, see functionality and semantics in the node definition) of the **AdvancedAudioBuffer** node by loading ABv3_CU8_1ch_AOT2_one with the help of the **AudioSource** node into the internal buffer of the **AdvancedAudioBuffer** node. After 10 seconds (t=10s) the ‘one’ will be repeated five times, which can be tested subjectively.

ABv3_ AAB02 This scene is used for testing the reload mode (**LoadMode**=1) by loading ABv3_CU08_1ch_AOT2_one with the help of the AudioSource node into the internal buffer of the AdvancedAudioBuffer node. The clip will be repeated five times with loop=enabled from t=2s until t=7s. After 10 seconds the content will be replaced be clip ABv3_CU08_1ch_AOT2_two, which will be repeated five times from t=12 until t=17. Clip ABv3_CU08_1ch_AOT2_three will be loaded after 20 seconds and repeated five times from t=22 until t=27. The playback should be tested subjectively.

ABv3_ AAB03 This scene is used for testing the accumulate mode (**LoadMode**=2) by loading ABv3_CU8_1ch_AOT2_one at t=0 with the help of the **AudioSource** node into the internal buffer of the **AdvancedAudioBuffer** node. At t=3s ABv3_CU08_1ch_AOT2_two will be loaded into the buffer and at t=6s ABv3_CU08_1ch_AOT2_three will be loaded into the buffer. At t=10s all three clips will be played contiuous (‘one, two, three’) two times while loop is enabled. The loop mechanism will be tested by setting loop to false at t=19s and start replay at t=20s for 6 seconds. Only one block (‘one, two, three’) should be heard due to the disabled loop mode. The playback should be tested subjectively.

ABv3_ AAB04 This scene is used for testing the continuous accumulate mode (**LoadMode**=3) by loading ABv3_CU08_1ch_AOT2_one, ABv3_CU08_1ch_AOT2_two and ABv3_CU08_1ch_AOT2_three at t=0, t=3s and t=6s into the internal buffer of the **AdvancedAudioBuffer** node with the help of **AudioSource**. At t=10s all three clips will be played contiuous (‘one, two, three’) two times while loop is enabled. At t=18s a fourth clip (ABv3_CU08_1ch_AOT2_four) will be appended and with the restriction of **length** = 3 seconds, the first clip shall be deleted. At t=20s all three clips will be played contiuously (‘two, three, four’) two times. This sequence will be repeated with appending ABv3_CU08_1ch_AOT2_five and playing back (‘three, four, five’) at t=30s and appending ABv3_CU08_1ch_AOT2_six and playing back (‘four, five, six’) at t=40s. The playback should be tested subjectively.

ABv3_AAB05 This scene is used for testing the accumulate mode with limited number of buffer blocks of the **AdvancedAudioBuffer** node. In this mode the number of accumulaed blocks will be set instead of a maximum length. First a sequence ‘one, two, three' will be loaded and will be played back two times at t=10s. Then ‘four’ will be loaded into the buffer by replacing ‘one’, played back two times at t=20s. At t=27s the block in the middle (‘three’) will be deleted and at t=29 ‘five’ will be appended. This block (‘two, four, five’) will be played back two times at t=31s.

At t=40s the latest block (‘five’) is individually addressed and shall be played 3 times while loop is still enabled. At t=43s the block in the middle (‘four’) will be addressed individually and played back 3 times. At t=46s the first block (‘two’) will be addressed individually and played back 3 times until t=49s. At t=48,5s **loop** will be set to false. The actual active block will be played until its end.

At t=50s, t=55s and t=60s clip ‘four’, ‘five’ and ‘two’ will be played only one time.

The playback should be tested subjectively.

6.12.4.2 Testing of AudioChannelConfig Node

This node is used to label the audio data in the audio subtree to supply the audio presenter with the required information for multi-channel or soundfield signals. The node has the standard audio node interfaces, but no signal processing capability. The samples are passed through and get new channel configuration information.

6.12.4.2.1 BIFS components needed in the conformance testing

For testing the labelling mechanism of the **AudioChannelConfig** node a minimal set of BIFS nodes is needed: Besides the root node and one grouping node (**Group**, **OrderedGroup**) one top-level AudioBIFS node that has a **spatialize** field (**Sound**, **Sound2D**, **DirectiveSound**) as well as a node that connects the AudioBIFS sub-graph with the decoder (**AudioSource**) is required.

6.12.4.2.2 Conformance testing procedure

Conformance testing of the **AudioChannelConfig** node requires the player to support the parameter printout of the channel configuration that reaches the top-level sound nodes. As the way this information is transported through the AudioBIFS sub-graph is implementation-dependent the channel configuration should be printed in the form of the **AudioChannelConfig** node's fields.

6.12.4.2.2.1 Unique generalChannelFormat

In this subclause the testing of the correct composition of one or more audio signal of the same generalChannelFormat is described, i.e. all sources in the scene are channel-oriented presets (or subsets thereof), parametric channel oriented or Ambisonics™ oriented.

6.12.4.2.2.1.1 Pass through

Test scene for parameter printout and subjective testing:

Abv3_ACC01

This scene is used for testing whether the channel configuration of the audio elementary stream is correctly passed through the AudioBIFS tree. For that purpose a 5.1-channel standard multi-channel configuration sound source will be used. The audio compositor shall recognize the channel configurations of the elementary stream and map the channels to the appropriate speaker(s). The 5.1 configuration shall be recognizable in the *parameter printout* of the channel configuration from the top level node.

For *subjective testing* the CU input used for this scene result in a distinct scheme: On every channel white noise is being played for one second, starting with the front left channel, followed by the front right, surround left, surround right, center channel and finally the LFE channel (5s<t<6s).

ABv3_CU07_6ch_AOT2_surround is used as input sounds for this scene.

6.12.4.2.2.1.2 ChannelPreset

Abv3_ACC02

This scene is used for testing whether an alternative channel configuration stream will be passed correctly through the AudioBIFS tree. Therefore **generalChannelFormat** will be set to the ‘ChannelPreset’ mode. A binaural recorded sequence will be used. After start-up the clip will be marked as ‘normal’ stereo. At t=12s the clip will be played again and marked as ‘binaural’ stereo. The configuration shall be recognizable in the *parameter printout* of the channel configuration from the top level node.

For *subjective testing* the properties of the audio presenter must be taken into account. The presenter might have a crosstalk-canceller for loudspeaker playback or a HRTF cross-convolution for headphone playback. In this case the clip will be heard 'spatialized' false during the first 12 seconds and for t>=12s the clip will be heard correctly due to the binaural marked content which should lead to the crosstalk-canceller being enabled for loudspeaker listening or disabled HRTF cross-convolution in the headphone listening case.

ABv3_CU08_1ch_AOT2_LCR_Binaural is used as input sounds for this scene.

6.12.4.2.2.1.3 ChannelPresetSubset

Test scene for parameter printout and subjective testing:

ABv3_ACC03 — This scene is used for testing the combination of two stereo and two mono sound sources to a 5.1-channel sound source. The audio compositor shall recognize the channel configurations of the sound sources and map the channels to the appropriate speaker(s). It shall be recognizable in the *parameter printout* of the channel configuration that each top-level sound node addresses an individual subset of a 5.1 configuration (the values of **fixedPreset** are 000110b=6 for the front channels, 011000b=24 for the surround channels, 000001b=1 for the center channel and 100000b=32 for the LFE channel).

For *subjective testing* the CU inputs used for this scene result in a distinct scheme: On every channel white noise is being played for one second, starting with the front left channel, followed by the front right, surround left, surround right, center channel and finally the LFE channel (5s<t<6s).

ABv3_CU01_2ch_AOT2_LR12, ABv3_CU02_2ch_AOT2_LR34, ABv3_CU04_1ch_AOT2_M5 and ABv3_CU05_1ch_AOT2_M6 are used as input sounds for this scene.

ABv3_ACC04 — This scene is used for testing the combination of two mono sources to one stereo source. The audio compositor shall recognize the channel configurations of the sound sources and map the channels to the appropriate speaker. It shall be recognizable in the *parameter printout* of the channel configuration that each top-level sound node addresses an individual channel (the values of **fixedPreset** are 01b=1 for the left channel and 10b=2 for the right channel).

For *subjective testing* the CU inputs used for this scene result in a distinct scheme: On the left channel white noise is being played for three seconds, followed 1 second silence and then one second white noise on the right channel.

ABv2_CU06_1ch_AOT2_M0-3 and ABv3_CU04_1ch_AOT2_M5 are used as input sounds for this scene.

Abv3_ACC05 — This scene is used for testing whether a dynamic changed channel configuration stream (relabelled channels) will be passed correctly through the AudioBIFS subtree. Therefore **generalChannelFormat** will be set to the 'ChannelPresetSubset' mode. A stereo signal ('applause') will be used. After start-up the clip will be marked as L,R signal. At t=6s the clip will be marked as LS, RS. The configuration shall be recognizable in the *parameter printout* of the channel configuration from the top level node.

For *subjective testing* the clip should be recognized from the left and right loudspeakers after start-up. After 6 seconds the clip should be heard from the left and right surround loudspeakers.

Abv3_Cu15_2ch_Aot2_Applause is used as input sounds for this scene.