**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/WG11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 MPEG2014/N14462**

**April 2014, València, Spain**

|  |  |
| --- | --- |
| **Source** | Audio Subgroup |
| **Status** | Approved |
| **Title** | Active downmix control |

# Introduction

At the 108th MPEG meeting in Valéncia the MPEG-H 3D Audio work item progressed to Committee Draft (CD). This document outlines the current state of the downmix technology incorporated in the codec and presents a technology under consideration that provides more encoder control over the active downmix process in the MPEG-H 3D Audio decoder at minimum bitrate cost.

The MPEG Audio Subgroup invites interested broadcasters and other experts to evaluate the considered extended encoder control of the MPEG-H 3D Audio active downmix and assess and report whether it is considered a useful feature.

The MPEG Audio Subgroup kindly requests to provide any comments on the proposed active downmix control to national standardization bodies, such that their comments can be adequately taken into account in the ballot on the CD of MPEG-H.

# MPEG-H 3D Audio Decoder Downmix Features

The MPEG-H 3D Audio decoder contains a so-called *format converter* that flexibly renders audio channel content to arbitrary target loudspeaker setups. Format conversion is carried out by downmixing the transmitted audio channels to the target loudspeaker layout, applying a downmix matrix tailored to the transmission and reproduction channel configurations. The MPEG-H 3D Audio codec integrates a coding scheme to efficiently transmit downmix matrices, allowing for the simultaneous transmission of downmix matrices for multiple target setups in the same bitstream. If no downmix coefficients are transmitted in the bitstream for a target loudspeaker configuration, the MPEG-H 3D Audio format converter automatically generates a downmix matrix for mapping the transmitted audio channels to the reproduction setup.

Downmix coefficients (either transmitted or decoder generated) are applied in the MPEG-H 3D Audio decoder in an *active*, i.e. signal-adaptive, downmix process to avoid downmixing artifacts when coherent or partially coherent input signals are superimposed. The downmix algorithm makes use of two signal-adaptive processing steps to do so:

* Phase-alignment of coherent input signals that are added in the downmix.
* Energy-normalization of the downmix gains such that the input signal energy is preserved (also called *adaptive EQ*).

# Encoder Control over MPEG-H 3D Audio Active Downmix

Throughout the MPEG-H 3D Audio standardization, the active downmix in the MPEG-H format converter has been proven to deliver high-quality downmixes. However, broadcasters are used to the functionality of passive downmix implementations and request the possibility to optionally also allow for this legacy downmix processing, despite the downmix artifacts a passive downmix may introduce (uncontrolled signal boost and cancellation, coloration/comb filtering). Further, legacy content may have been mastered specifically for passive downmix compatibility.

## Current State of Active Downmix Control

To meet the broadcasters’ requirements, the MPEG Audio subgroup decided to add elements to the MPEG-H 3D Audio bistream that provide encoder control over the mode of operation of the MPEG-H downmix. The current MPEG-H codec specification allows a to optionally force the MPEG-H decoder to operate as a passive downmix by setting a flag in the MPEG-H 3D Audio bitstream.

The currently present signaling feature thus allows a broadcaster to decide between 2 downmix processing modes (active vs. passive downmix), e.g. depending on the production process of the content:

* Legacy content that has been created with restrictions imposed during production to ensure compatibility with passive downmixing may be reproduced using the passive downmix in the decoder.
* Content that has been produced without restrictions may be prone to passive downmix artifacts and thus the broadcaster may decide for active downmix processing for those items.

## Proposed Extended Active Downmix Control

While the binary active/passive switch gives a broadcaster basic control over the downmixing process in the MPEG-H 3D Audio decoder, it enforces a hard decision between the two currently available downmix options. However, the downmix algorithm in the MPEG-H 3D Audio decoder naturally allows for a less restrictive control of the downmix processing once the “active/passive” bitstream flag is replaced by a more fine-granular signaling in the bitstream.

In this document we therefore propose to replace the active/passive downmix flag in the MPEG-H 3D Audio bistream by two optionally transmitted downmix setting parameters with very low increase in bitrate costs. This proposed extended signaling would

* allow for a more fine-granular encoder control over the decoder downmix,
* include the currently available options (fully active downmix, fully passive downmix),
* only require 5 additional bits per transmitted stream config if they are transmitted.

The proposed control parameters enable the broadcaster/content provider to optionally adapt the MPEG-H downmix to the properties of the content (e.g. resulting from the production process) or to his artistic intent. Together with the already present feature to transmit downmix matrices they give full encoder control over the MPEG-H downmix.

The transmission of the two proposed downmix setting parameters is optional. In case they are not sent or set to the default values, the downmix processing and downmix output signals are bitidentical to the ones produced by the MPEG-H 3D Audio reference model 2 (RM2). Implementing the proposed extended downmix control feature does not require changing the actual downmix processing, but merely modifies tuning constants and scales one intermediate gain value to control the effect strength of the active downmix features.

# Technical Description

This section describes the technical realization of the proposed feature to allow for the extended encoder control of the downmix mode of operation. For details on the changes to the CD text see the Annex at the end of this document.

## High-Level description

* Control parameters are signaled in the bitstream that control how the downmix in the MPEG-H decoder operates.
* Instead of the active/passive downmix switch that enforces a hard decision between the fully active downmix and the fully passive downmix, we propose controls that allow to ‘fade’ between the extreme settings (fully active ↔ fully passive), i.e. parameters that give more control than a hard active/passive switch.
* Two downmix setting parameters are proposed, see below for details:
	+ One for controlling the phase-alignment processing: phaseAlignStrength
	+ One for controlling the energy-preserving gain normalization (adaptive equalizer): adaptiveEqStrength
* The mode of operation of the two parameters in the actual downmix algorithm is straightforward. phaseAlignStrength only changes tuning constants, adaptiveEqStrength only scales the adaptive EQ gain towards 1.0 (=neutral, passive behavior).
* Default parameter settings lead to identical processing as defined in RM2 (has been tested for bit-identical output).
* Setting the parameters to phaseAlignStrength = adaptiveEqStrength = 0 results in a passive downmix (has been tested for bit-identical output compared to a simple passive downmix implementation).
* Gradual changes of the phaseAlignStrength, adaptiveEqStrength parameters reflect gradual changes of the downmix behavior between the extreme settings.

## Passive downmix feature

The passive downmix operation is selected by signaling phaseAlignStrength = adaptiveEqStrength = 0. Thereby, all signal-adaptive processing steps are deactivated – the downmix simply applies time-invariant downmix weights to all input signals and adds them up. With this setting the MPEG-H format converter downmix acts like a passive legacy downmix that does not adapt to the input signals’ properties.

## Control parameters description

Here there two control parameters are described. See the working draft changes in the Annex of this document for the implementation details.

### Phase-alignment strength

* A parameter that determines how much the phases of coherent input signals that are added in the downmix shall be aligned prior to adding the signals.
* Algorithmic description: phaseAlignStrength modifies the attraction curve f\_ICC(ICC) which maps inter-channel correlation coefficients (ICC) to attraction values f\_ICC that determine to which extent the input signal phases are aligned:
	+ Phases of signals for which f\_ICC is lower are phase-aligned less. For f\_ICC = 0.0 no phase-alignment is done.
	+ Figure 1 shows the resulting attraction curves f\_ICC(ICC, phaseAlignStrength). Note that the curve applied in the RM2 algorithm is marked red and is selected for the default setting of the phaseAlignStrength parameter.
* Parameter range: 0..7:
	+ 7 = maximum phase-alignment effect
	+ 3 = default setting (RM2 processing)
	+ 0 = signal adaptive phase-alignment switched off

Figure 1: Attraction curves as a function of ICC and phaseAlignStrength

### Adaptive EQ strength

* + A parameter that determines how strong the downmix gains are modified (normalized) to preserve the energy of the input signals in the downmix output
	+ Algorithmic description: The adaptiveEqStrength parameter determines to which degree the signal-adaptive energy-preservation gain is scaled towards the neutral gain factor, i.e. towards the constant 1.0. adaptiveEqStrength = 1 leaves the adaptive gain untouched (=RM2 behavior), whereas adaptiveEqStrength = 0 results in a constant, thus signal-independent gain of 1.0.
	+ Parameter range: 0..7:
		- 7 = full energy preservation within gain limits (-10dB to +8dB compensation gain, i.e. RM2 processing)
		- 0 = signal adaptive gain normalization switched off

# Conclusion

In this contribution we present the addition of two optional parameters to the MPEG-H 3D Audio static configuration data that allows encoder control of how the downmix in the MPEG-H decoder should operate. The parameters allow a broadcaster/content provider to signal a downmix setting that is best suited for the transmitted audio content or that optimally reflects his artistic intent for decoder downmixes. The combination of the proposed signaling parameters and the already present feature to transmit downmix matrices gives full encoder control over the MPEG-H downmix, including the possibility to enforce a passive, i.e. legacy downmix processing in the decoder.

The parameters allow to individually set the effect strength of the two signal-adaptive processing principles of the MPEG-H format converter downmix - the signal-adaptive phase-alignment as well as the signal-adaptive energy-preservation equalizer. The control parameter ranges cover a fully passive downmix processing as well as the RM2 active downmix processing.

No algorithmic changes are applied to the downmix processing itself. The transmitted control parameters simply modify tuning parameters of the downmix algorithm that have been set in RM2 and scale an intermediate gain value towards 1.0.

Updates to the Working Draft 2 text (WD2, defining RM2) are given in the Annex of this document. A reference software implementation is readily available within a short editing period.

The MPEG Audio Subgroup kindly requests interested audio experts and/or industry representatives to assess the presented technology and comment to national standardization bodies on the usefulness of the proposed extended encoder control over the MPEG-H 3D Audio downmix.

Annex –Changes to Working Draft 2 (WD2) Text

The proposed changes to the WD2 text are:

* Addition of the elements phaseAlignStrength and adaptiveEqStrength to control the active downmix behavior
* Change of name of config extension ID to reflect that not only downmix matrices are signaled in the config extension

Note that the WD2 text presents the state of the MPEG-H 3D Audio work item text *before* addition of the binary active/passive flag to the bitstream.

The changes are listed in detail in the following:

*In Table 13, Syntax of mpegh3daConfigExtension(), replace*

|  |
| --- |
|  case ID\_CONFIG\_EXT\_DMX\_MATRIX: |
|  DownmixMatrixSet() |

*by*

|  |
| --- |
|  case ID\_CONFIG\_EXT\_DOWNMIX: |
|  downmixConfig() |

*In subclause 5.2.2, Decoder configuration, replace*

5.2.2.4 Syntax of downmix matrix elements

Table 15 — Syntax of DownmixMatrixSet()

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| DownmixMatrixSet() |  |  |
| { |  |  |
|  **numDmxMatrices** | **5** | **uimsbf** |
|  for (k=0; k<numDmxMatrices; ++k) { |  |  |
|  **downmixId;** | **6** | **uimsbf** |
|  **CICPspeakerLayoutIdx;** | **6** | **uimsbf** |
|  **DmxMatrixLenBits** = escapedValue(8,8,12); | **8..28** |  |
|  DownmixMatrix(inputConfig(UsacChannelConfig()),  inputCount(UsacChannelConfig()),  outputConfig(CICPspeakerLayoutIdx),  outputCount(CICPspeakerLayoutIdx) ); | **DmxMatrixLenBits** |  |
|  } |  |  |
| } |  |  |

*by*

5.2.2.4 Syntax of downmix configuration

Table 15 — Syntax of downmixConfig()

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| downmixConfig () |  |  |
| { |  |  |
|  **downmixConfigType**; if (downmixConfigType == 0 || downmixConfigType == 2) { **phaseAlignStrength;** **adaptiveEqStrength;** } else if (downmixConfigType == 1 || downmixConfigType == 2) { | **2****3****3** | **uimsbf****uimsbf****uimsbf** |
|  **numDmxMatrices** | **5** | **uimsbf** |
|  for (k=0; k<numDmxMatrices; ++k) { |  |  |
|  **downmixId;** | **6** | **uimsbf** |
|  **CICPspeakerLayoutIdx;** | **6** | **uimsbf** |
|  **DmxMatrixLenBits** = escapedValue(8,8,12); | **8..28** |  |
|  DownmixMatrix(inputConfig(UsacChannelConfig()),  inputCount(UsacChannelConfig()),  outputConfig(CICPspeakerLayoutIdx),  outputCount(CICPspeakerLayoutIdx) ); | **DmxMatrixLenBits** |  |
|  } }} |  |  |

*In* *Table 45, Value of usacConfigExtType, replace*

|  |  |
| --- | --- |
| ID\_CONFIG\_EXT\_DMX\_MATRIX | 1 |

*by*

|  |  |
| --- | --- |
| ID\_CONFIG\_EXT\_DOWNMIX | 1 |

*In subclause 5.4, Configuration Element Descriptions, replace*

5.4.2 Downmix matrix

5.4.2.1 General

Downmix matrix coefficients may be transmitted by the encoder to enable control over the format conversion process at the decoder. Transmission is facilitated by means of a ConfigExtension of Type ID\_CONFIG\_EXT\_DMX\_MATRIX. Each downmix matrix signals its associated target speaker layout that determines the matrix dimensions and identifies which kind of downmix matrix operation the transmitted coefficients are suitable for.

*by*

5.4.2 Downmix configuration

5.4.2.1 General

Downmix matrix coefficients and/or active downmix setting parameters may be transmitted by the encoder to enable control over the format conversion process at the decoder. Transmission is facilitated by means of a ConfigExtension of Type ID\_CONFIG\_EXT\_DOWNMIX. The ConfigExtension may contain downmix matrices as well as active downmix setting parameters.

If downmix matrices are transmitted, each downmix matrix signals its associated target speaker layout that determines the matrix dimensions and identifies which kind of downmix matrix operation the transmitted coefficients are suitable for.

*In subclause 5.4.2.2, Data Elements and Variables, add the following element/variable descriptions:*

**downmixConfigType** This parameter allows to signal whether active downmix control parameters (value 0) or downmix matrices (value 1) or both (value 2) are transmitted. Value 3 is reserved.

**phaseAlignStrength** Strength of the phase-alignment in the active downmix. Valid parameter range: 0..7. The value 3 signals the default setting, 0 deactivates the phase-alignment.

**adaptiveEqStrength** Strength of the downmix gain normalization in the active downmix. Valid parameter range: 0..7. . The value 7 signals the default setting (=maximum preservation of signal energy).

*In subclause 9.3.2.3, Phase-alignment matrix formulation, replace*

The  values are mapped to an attraction measure matrix with elements

 ,

*by*

The  values are mapped to an attraction measure matrix with elements

 ,

where *PasMax, PasCurveSlope, PasCurveShift* are derived from Table 57,

Table 57 — phase attraction mapping curve parameters. phaseAlignStrength shall be set to 3 if no other value has been transmitted in the bitstream.

|  |  |  |  |
| --- | --- | --- | --- |
| phaseAlignStrength | *PasMax* | *PasCurveSlope* | *PasCurveShift* |
| 0 | 0 | 0 | 0 |
| 1 | 0.0714 | 0.1701 | -0.0891 |
| 2 | 0.1548 | 0.3771 | -0.1896 |
| 3 | 0.25 | 0.625 | -0.3 |
| 4 | 0.3571 | 0.9184 | -0.4184 |
| 5 | 0.4762 | 1.2623 | -0.5427 |
| 6 | 0.6071 | 1.6624 | -0.6707 |
| 7 | 0.75 | 2.125 | -0.8 |

*At the end of the same subclause replace*

the final phase-aligning mixing matrix elements follow as

.

*by*

the final phase-aligning mixing matrix elements follow as

,

where *AES =* adaptiveEqStrength/7. adaptiveEqStrength shall be set to the value 7, if no other value has been transmitted in the bitstream.