INTERNATIONAL STANDARD

ISO/IEC 23003-4

First edition
2015-11-15

Information technology — MPEG audio technologies —

Part 4:
Dynamic Range Control

Technologies de l'information — Technologies audio MPEG —

Partie 4: Contrôle de gamme dynamique

Reference number
ISO/IEC 23003-4:2015(E)

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

Contents

Page

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*.

ISO/IEC 23003 consists of the following parts, under the general title *Information technology — MPEG audio technologies*:

— *Part 1: MPEG Surround*

— *Part 2: Spatial Audio Object Coding*

— *Part 3: Unified speech and audio coding*

— *Part 4: Dynamic Range Control*

Introduction

Consumer audio systems and devices are used in a large variety of configurations and acoustical environments. For many of these scenarios, the audio reproduction quality can be improved by appropriate control of content dynamics and loudness.

This part of ISO/IEC 23003 provides a universal dynamic range control tool that supports loudness normalization. The DRC tool offers a bitrate efficient representation of dynamically compressed versions of an audio signal. This is achieved by adding a low-bitrate DRC metadata stream to the audio signal. The DRC tool includes dedicated sections for clipping prevention, ducking, and for generating a fade-in and fade-out to supplement the main dynamic range compression functionality. The DRC effects available at the DRC decoder are generated at the DRC encoder side. At the DRC decoder side, the audio signal may be played back without applying the DRC tool, or an appropriate DRC tool effect is selected and applied based on the given playback scenario.

Information technology — MPEG audio technologies —

Part 4:
Dynamic Range Control

1 Scope

This part of ISO/IEC 23003 specifies technology for loudness and dynamic range control. This International Standard is applicable to most MPEG audio technologies. It offers flexible solutions to efficiently support the widespread demand for technologies such as loudness normalization and dynamic range compression for various playback scenarios.

2 Normative references

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

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

ISO/IEC 23001-8, *Information technology — MPEG systems technologies — Part 8: Coding-independent code points*

3 Terms, definitions and mnemonics

For the purposes of this document, the terms and definitions given in ISO/IEC 14496-12 and the following apply.

3.1 Terms

3.1.1
DRC sequence
series of DRC gain values that can be applied to one or more audio channels

3.1.2
DRC set
defined set of DRC sequences that produce a desired effect if applied to the audio signal

3.1.3
album
collection of audio recordings that are mastered in a consistent way. Traditionally, a collection of songs released on a Compact Disk belongs into this category, for example

3.2 Mnemonics

bslbf	bit string, left bit first, where “left” is the order in which bit strings are written in ISO/IEC 14496. Bit strings are written as a string of 1s and 0s within single quote marks, for example ‘1000 0001’. Blanks within a bit string are for ease of reading and have no significance
uimsbf	unsigned integer, most significant bit first
vlclbf	variable length code, left bit first, where “left” refers to the order in which the variable length codes are written
bit(n)	a bit string with n bits in the same format as bslbf
unsigned int(n)	an unsigned integer with n bits in the same format as uimsbf
signed int(n)	a signed integer with n bits, most significant bit first

4 Symbols (and abbreviated terms)

a_i	Filter coefficient
b	Band index of DRC filter bank (starting at 0)
b_i	Filter coefficient
deltaTmin	Smallest permitted DRC gain sample interval in units of the audio sample interval.
f_c	Cross-over frequency in Hz
$f_{c,norm}$	Cross-over frequency expressed as fraction of the audio sample rate.
$f_{c,norm,SB}(s)$	Cross-over frequency of audio decoder sub-band s expressed as fraction of the audio sample rate. The cross-over frequency is the upper band edge frequency of the sub-band.
f_s	Audio sample rate in Hz. If an audio decoder is present, it is the sample rate of the decoded time-domain audio signal.
N_{DRC}	Maximum permitted number of DRC samples per DRC frame. Identical to the number of intervals with a duration of *deltaTmin* per DRC frame.
N_{Codec}	Codec frame size in units of the audio sample interval $1/f_s$
M_{DRC}	DRC frame size in units of the audio sample interval $1/f_s$
π	Ratio of a circle’s circumference to its diameter
s	Audio decoder sub-band index (starting at 0)
TRUE/FALSE	Values of Boolean data type, which correspond to numerical 1 and 0, respectively.
z	Complex variable of the z-transform

5 Technical overview

The technology described in this part of ISO/IEC 23003 is called DRC tool. It provides efficient control of dynamic range, loudness, and clipping based on metadata generated at the encoder. The decoder can choose to selectively apply the metadata to the audio signal to achieve a desired result. Metadata for dynamic range compression consists of encoded time-varying gain values that can be applied to the audio signal. Hence, the main blocks of the DRC tool include a DRC gain encoder, a DRC gain decoder, a DRC gain modification block, and a DRC gain application block. These blocks are exercised on a frame-by-frame basis during audio processing. Various DRC configurations can be conveyed in a separate bitstream element, such as configurations for a downmix or combined DRCs. The DRC set selection block decides based on the playback scenario and the applicable DRC configurations which DRC gains to apply to the audio signal. Moreover, the DRC tool supports loudness normalization based on loudness metadata.

A typical system for loudness and dynamic range control in the time domain is shown in Figure 1. A more complex system including downmixer and peak limiter is shown in Figure 2. The decoder part of the DRC tool is driven by metadata that efficiently represents the DRC gain samples and parameters for interpolation. The gain samples can be updated as fast as necessary to accurately represent gain changes down to at least 1 ms update intervals. In the following the decoder part of the DRC tool is referred to as "DRC decoder", which includes everything except the audio decoder and associated bitstream de-multiplexing.

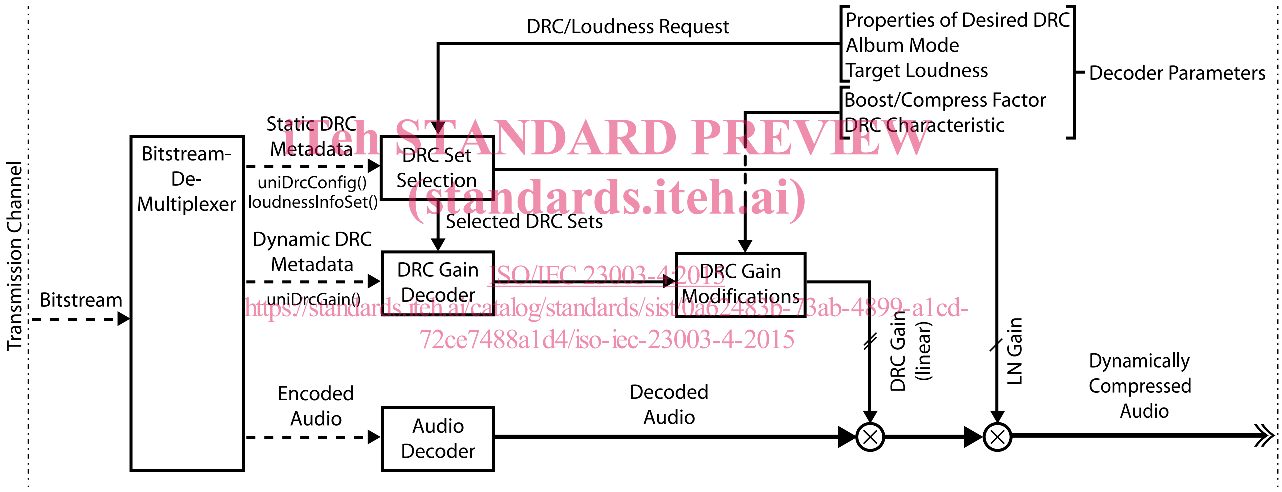

Figure 1 — Block diagram of a typical system with audio decoder and DRC tool modules to achieve loudness normalization (LN) and dynamic range control

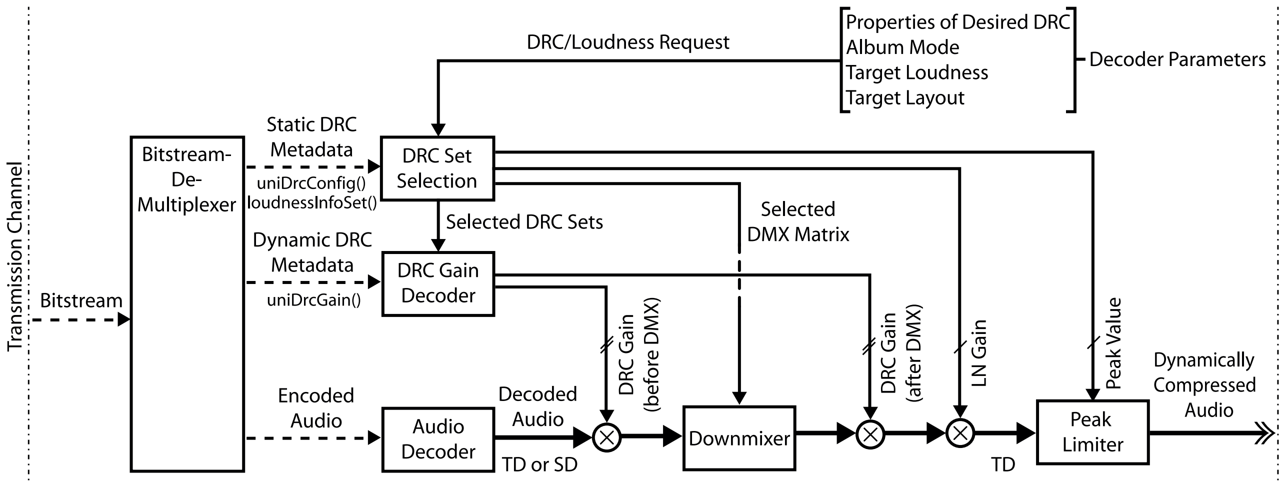

Figure 2 — Block diagram of a more complex system including downmixer and peak limiter (TD = time-domain, SD = subband-domain)

6 DRC decoder

6.1 DRC decoder configuration

6.1.1 Overview

The DRC configuration information can be received in-stream using the static payloads uniDrcConfig() and loudnessInfoSet() described below, or it can be delivered by a higher layer, such as 14496-12 (see Table 1). The basic decoding process of the static information is virtually the same. The difference consists mainly in a few syntax changes and reduced field sizes to increase the bit rate efficiency of the in-stream configuration. The syntax of the in-stream static payload is given in 7.3. The associated metadata encoding is given inA.6. The static DRC payload is evaluated once at the beginning of the decoding process and it is monitored subsequently. For static DRC payload changes during playback see 6.8.

Table 1 — Overview of configuration (setup) and separate metadata track in ISO/IEC 14496-12

	Sample Entry Code	*Setup (in sample entry)*	*Track reference*	*Sample format*
Audio Track	As specified for the audio codec in use (unchanged)	DRCInstructions box using negative values for drcLocation	'adrc' referring to the metadata tracks carrying gain values	As specified for the audio codec in use (unchanged)
Metadata Track	'unid'	(none)	(none)	Each sample is a uniDrcGain() payload

The static payload is divided into five logical blocks:

— channelLayout();

— downmixInstructions();

— drcCoefficientsBasic(), drcCoefficientsUniDrc();

— drcInstructionsBasic(), drcInstructionUniDrc();

— loudnessInfo().

Except for the channelLayout(), multiple instances of a logical block can appear. The DRC decoder combines the information of the matching instances of up to five logical blocks for a given playback scenario. Matching instances are found by matching several identifiers (labels) contained in the blocks.

From the static payload the decoder can also extract information about the effect of a particular DRC and various associated loudness information, if present. If multiple DRCs are available, this information can be used to select a particular DRC based on target criteria for dynamics and loudness (see 6.3)

uniDrcConfig() contains all blocks except for the loudnessInfo() blocks which are bundled in loudnessInfoSet(). The last part of the uniDrcConfig() payload can include future extension payloads. In the event that a *uniDrcConfigExtType* value is received that is not equal to UNIDRCCONFEXT_TERM, the DRC tool parser must read and discard the bits (otherBit) of the extension payload. Similarly, the last part of the loudnessInfoSet() payload can include future extension payloads. In the event that a *loudnessInfoSetExtType* value is received that is not equal to UNIDRCLOUDEXT_TERM, the DRC tool parser must read and discard the bits (otherBit) of the extension payload.

The top level fields of uniDrcConfig() include the audio sample rate, which is a fundamental parameter for the decoding process (if not present, the audio sample rate is inherited from the employed audio codec). Moreover, the top level fields of uniDrcConfig() include the number of instances of each of the logical blocks, except for the channelLayout() block which appears only once. The top level fields of loudnessInfoSet() only include the number of loudnessInfo() blocks. The five logical blocks are described in the following.

6.1.2 Description of logical blocks

6.1.2.1 channelLayout()

The channelLayout() block includes the channel count of the audio signal in the base layout. It may also include the base layout unless it is specified elsewhere. For use cases where the base audio signal represents objects or other audio content, the channel count represents the total number of base content channels.

6.1.2.2 downmixInstructions()

This block includes a unique non-zero downmix identifier (downmixId) that can be used externally to refer to this downmix. The targetChannelCount specifies the number of channels after downmixing to the target layout. It may also contain downmix coefficients, unless they are specified elsewhere. For use cases where the base audio signal represents objects or other audio content, the downmixId can be used to refer to a specific target channel configuration of a present rendering engine.

6.1.2.3 drcCoefficientsBasic(), drcCoefficientsUniDrc()

A drcCoefficients block describes all available DRC gain sequences in one location. The block can have the basic format or the uniDrc format. The basic format, drcCoefficientsBasic(), contains a subset of information included in drcCoefficientsUniDrc() that can be used to describe DRCs other than the ones specified in this standard. drcCoefficientsUniDrc() contains for each sequence several indicators on how it is encoded, the time resolution, time alignment, the number of DRC sub-bands and corresponding crossover frequencies and DRC characteristics. The crossover frequencies must increase with increasing band index. Alternatively, explicit indices in a decoder sub-band domain can be specified for the assignment of DRC sub-bands. The sub-band indices must also increase with increasing band index. If the DRC gains are applied in the time-domain by using the multi-band DRC filter bank specified in 6.4.11, explicit index signalling is not allowed. The index of the DRC characteristic indicates which compression characteristic was used to produce the gain sequence. The DRC location describes where these gain sequences can be found in the bitstream. The DRC gain sequences in that location are inherently enumerated according to their order of appearance starting with 1.

The DRC location field encoding depends on the audio codec. A codec specification may include this specification, and use values 1 – 4 to refer to codec-specific locations as indicated in Table 1. For example, for AAC (ISO/IEC 14496-3), the codec-specific values of the DRC location field are encoded as shown in Table 3.

Table 2 — Encoding of drcLocation for in-stream payload

drcLocation n	Payload
0	*Reserved*
1	Location 1 (Codec-specific use)
2	Location 2 (Codec-specific use)
3	Location 3 (Codec-specific use)
4	Location 4 (Codec-specific use)
$n > 4$	*reserved*

Table 3 — Codec-specific encoding of drcLocation for MPEG-4 Audio

drcLocation n	Payload
1	uniDrc() (defined in Clause 7)
2	dyn_rng_sgn[i] / dyn_rng_ctl[i] in dynamic_range_info() (defined in ISO/IEC 14496-3:2009 subpart 4)
3	compression_value in MPEG4_ancillary_data() (defined in ISO/IEC 14496-3:2009/AMD 4:2013)
4	*reserved*

The DRC frame size can optionally be specified. It must be provided if the DRC frame size deviates from the default size specified in 6.4.2. If not specified, the default frame size is used.

The in-stream drcCoefficient syntax is given in Table 42 and Table 44. The syntax for the corresponding block for ISO/IEC 14496-12 (ISO base media file format) is shown in Table 43 and Table 45. The corresponding blocks carry essentially the same information. Values that are identically included in both blocks are coded the same way except for drcLocation.

In ISO base media file format (see ISO/IEC 14496-12), for each codec that can be carried in MP4 files and that also carries DRC information, there is a specific definition of how the location is coded, using the DRC_location field (see Table 4). A negative value of DRC_location indicates that a DRC payload is in an associated meta-data track. That track is the n-th linked via a track reference of type 'adrc' (audio DRC) from the audio track, where n = abs(DRC_location), and the sample-entry type in the meta-data track indicates in which format the coefficients are stored. Table 3 defines the specific entries of the drcLocation field for AAC. Some example use cases are discussed in C.10.

If the uniDrc() payload is stored in a separate track in the ISO base media file format (ISO/IEC 14496-12), then the track is a metadata track with the sample entry identifier 'unid' (uniDrc), with no required boxes added to the sample entry. The time synchronization with the linked audio track is the same as if the payload was in-stream.

Table 4 — Encoding of drcLocation for ISO/IEC 14496-12

drcLocation n	Payload
$n < 0$	*DRC payload located in |n|-th linked meta-data track*
0	*reserved*
1	Location 1 (Codec-specific use)
2	Location 2 (Codec-specific use)
3	Location 3 (Codec-specific use)
4	Location 4 (Codec-specific use)
$n > 4$	*reserved*

6.1.2.4 drcInstructionsBasic(), drcInstructionsUniDrc()

A drcInstructions block includes information about one specific DRC set that can be applied to achieve a desired effect. This block can have the basic format or the uniDrc format. The basic format, drcInstructionsBasic(), contains a subset of information included in drcInstructionsUniDrc() that can be used to describe DRCs other than the ones specified in this standard. The information included in drcInstructionsUniDrc() consists mainly of pre-defined description elements such as the DRC set effect and the DRC gain sequences that are applied. The drcSetEffect field contains several effect bits as listed in Table A.32. Multiple bits can be set unless otherwise noted. Note that if no effect bit is set at all, the DRC set is ignored in the DRC set selection (see 6.3). Each drcInstructions block carries a unique non-zero identifier drcSetId. A downmixId is included to indicate if this DRC set applies to a certain downmix with this identifier. A downmixId of zero indicates that the DRC set is applied to the base layout. A downmixId of 0x7F indicates that the DRC set can be applied before or after the downmix.

Since such a DRC can be applied to any downmix, it has only one channel group including all channels. If a “Ducking” bit is set in the drcSetEffect field, the DRC set is applied before any downmix specified by the downmix ID, i.e. the DRC set is always applied to the base layout and the downmix is generated thereafter. The downmixId 0x7F is not permitted for a ducking DRC set. In all other cases, the DRC set is applied to the channel configuration indicated by the downmix ID.

A second DRC set may be specified for certain configurations. These configurations include cases where, e.g. one DRC set is used for dynamic range compression and the other for clipping prevention (“Clipping” bit is set); or, e.g. one DRC set is applied before and the other after the downmix. In those cases, the second DRC set contains a non-zero field *dependsOnDrcSet* that has the value of the *drcSetId* of the first DRC set it depends on. The declared DRC set effects of the second DRC set do not take into account the effects of the first DRC set. If the first DRC set is not designed to be used without combining it with another DRC set, the *noIndependentUse* flag must be set to 1. In that case, the DRC set can only be used in combination with another DRC set as indicated by the *dependsOnDrcSet* field of the other set that is combined with it.

Usually, each audio channel is assigned to a DRC gain sequence. A collection of channels assigned to the same DRC gain sequence is called “channel group”. The assignment of a DRC gain sequence to a channel group is done in the order of first appearance of the sequence index when iterating through all channels (see also Table 14). A DRC gain sequence index *bsSequenceIndex* == 0 indicates that the assigned channel will be passed through by the DRC tool without processing unless otherwise noted. Note that therefore *bsSequenceIndex* is effectively 1-based, wereas the corresponding indices (*sequenceIndex*) for processing are zero-based.

If subsequent channels are assigned the same sequence index, the field *repeatSequenceCount* indicates how many channels will have the same sequence not including the first.

The *drcLocation* field is used in the same way as the *drcLocation* field in the drcCoefficients (see 6.1.2.3). Certain entries of the drcLocation field allow adding drcInstructions information to gain sequences defined elsewhere. Some use cases are discussed in C.10.

The field *limiterPeakTarget* declares the peak target level used by the encoder-side DRC, if applicable. For example, if a limiter is used to generate the DRC gain sequence, it is configured to control the audio sample magnitude to not exceed this peak target level. *limiterPeakTarget* is represented in dBFS and encoded according to Table A.27.

If *limiterPeakTarget* is present, and the only drcSetEffect is “clipping prevention”, the gain sequence is to be shifted by the negative sum of *loudnessNormalizationGainDb* and *limiterPeakTarget* if the negative sum is greater than 0. Afterwards, the gain sequence is saturated at the threshold of 0 dB so that only negative gains (dB) occur. With this mechanism it is possible to send gains for clipping prevention in expectation of a high *loudnessNormalizationGainDb*. If *loudnessNormalizationGainDb* is lower than expected, the gains are applied only as far as needed, and the dynamic range can be kept as high as possible.

If *gainScalingPresent* == 1, the gain scaling coefficients must be applied to the channel group. If *gainOffsetPresent* == 1, the gain offset value must be applied to the channel group as shown in Table 16. Similarly, if *duckingScalingPresent* == 1, the scaling factor must be applied to the associated ducking gain sequence for that channel group.

The in-stream drcInstructions syntax is given in Table 46 and Table 48. The syntax for the corresponding block for ISO/IEC 14496-12 is shown in Table 47 and Table 49. The corresponding blocks carry essentially the same information. Values that are identically included in both blocks are coded the same way except for drcLocation. Further information on the coding of drcLocation is defined in 6.1.2.3.

6.1.2.5 loudnessInfo()

A loudnessInfo() block includes loudness and peak information. A downmix identifier and DRC set identifier indicate which configuration the information applies to. Hence, this block can be associated with the audio signal without DRC and without downmix, or with any specific DRC and/or downmix applied. If a DRC with a dependent DRC set is applied, the loudness information describes the output of the combined DRCs. A loudnessInfo() block can either represent an individual content item or the entire album. Typically, all content items of an album include identical album loudnessInfo() blocks.

If *downmixId* is zero, then loudnessInfo() applies to the base layout. If the *drcSetId* is zero, then loudnessInfo() applies to the audio signal without DRC processing.

The fields *samplePeakLevel* and *truePeakLevel* represent the level of the maximum sample magnitude in dBFS and the true peak in dBTP, respectively, of the associated audio content before or after audio encoding as defined in Reference [4]. The *measurementSystem* field includes standardized systems and others (see Table A.37). System 3 is defined as ITU-R BS.1770-3 with pre-processing. The pre-processing is a high-pass filter that models the typical limited frequency response of portable device loudspeakers. System 4 is defined as “User”. It means that the corresponding *methodValue* reflects a (subjective) user preference. System 5 is defined as “Expert/Panel”. It means that the corresponding *methodValue* represents a (subjective) expert or panel preference.

The *methodDefinition* field according to Table A.36 specifies how the *methodValue* is derived. The mixing level is compatible with “mixlevel” in ATSC A/52.[1] It indicates the absolute acoustic sound pressure level of an individual channel during the final audio mixing session. The peak mixing level is the acoustic level of a sine wave in a single channel whose peaks reach 100 percent in the PCM representation. The absolute SPL value is typically measured by means of pink noise with an RMS value of −20 or −30 dB with respect to the peak RMS sine wave level. The value of mixing level is not typically used within the DRC tool, but may be used by other parts of the audio reproduction system.

The room type field is compatible with “roomtyp” in ATSC A/52.[1] It indicates the type and calibration of the mixing room used for the final audio mixing session. The value of *roomtyp* is not typically used by the DRC tool, but may be used by other parts of the audio reproduction system.

The loudnessInfoSet() payload contains all loudnessInfo() blocks. The in-stream syntax of loudnessInfoSet() is given in Table 37. For the ISO base media file format the slightly different syntax of “LoudnessBox” is used as defined in ISO/IEC 14496-12.

6.1.3 Derivation of peak and loudness values

The loudnessInfo() blocks provide optional values that describe loudness and peak. Several DRC decoder processes depend on these values, hence, when the loudness information is partially or entirely absent, fallback values are used as shown in Table 5. For peak values, a default value is to be used. Some other values can be drawn from the loudnessInfo() block of the base layout.

Table 5 — Default and fallback values of loudnessInfo

Value	Default	1st fallback: use value from loudnessInfo() of base layout with same DRCsetId	2nd fallback: use value from loudnessInfo() of the base layout without DRC
truePeakLevel	0.0	No	No
samplePeakLevel	0.0	No	No
programLoudness	Undefined	Yes	Yes
anchorLoudness	Undefined	Yes	Yes
loudnessRange	Undefined	No	No
Maximum loudness range	Undefined	No	No
Maximum momentary loudness	Undefined	No	No
Maximum short-term loudness	Undefined	No	No
Short-term loudness	Undefined	No	No
Mixing level	Undefined	Yes	Yes
Room type	Undefined	Yes	Yes

The *signalPeakLevel* of a DRC set is determined as specified in Table 6, where peak related metadata entries are selected dependent on their availability and dependent on the drcSetId, and the requested downmixId. If no explicit peak information is available, *signalPeakLevel* is estimated from downmix coefficients and others. The estimates based on downmix coefficients hold for passive downmixers and might hold for specific active downmixers.