INTERNATIONAL STANDARD

ISO/IEC 23003-4

Second edition
2020-06

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 23003-4:2020(E)

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.

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) or the IEC list of patent declarations received (see http://patents.iec.ch).

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 Joint Technical Committee ISO/IEC JTC 1, *Information Technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia, and hypermedia*.

This second edition cancels and replaces the first edition (ISO 23003-4:2015), which has been technically revised. It also incorporates the Amendments ISO 23003-4:2015/Amd.1:2017 and ISO 23003-4:2015/Amd.2:2017. The main changes compared to the previous edition are as follows:

— Amendments to the previous edition that include enhancements, definitions of profiles and levels, reference software, and conformance are integrated.

A list of all parts in the ISO/IEC 23003 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.

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 document 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.

Loudness normalization is fully integrated with DRC and peak control to avoid clipping. A metadata-controlled equalization tool is provided to compensate for playback scenarios that impact the spectral balance, such as downmix or DRC. Furthermore, the DRC tool supports metadata-based loudness equalization to compensate the effect of playback level changes on the spectral balance.

The International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) draw attention to the fact that it is claimed that compliance with this document may involve the use of a patent.

ISO and IEC take no position concerning the evidence, validity and scope of these patent rights

The holders of these patent rights have assured ISO and IEC that they are willing to negotiate licences under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statements of the holders of these patent rights are registered with ISO and IEC. Information may be obtained from the patent database available at www.iso.org/patents.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights other than those in the patent database. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Information technology — MPEG audio technologies —

Part 4:
Dynamic range control

1 Scope

This document specifies technology for loudness and dynamic range control. It 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 are referred to in the text in such a way that some or all of their content constitutes requirements of this document. 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 14496-26:2010, *Information technology — Coding of audio-visual objects — Part 26: Audio Conformance*

ISO/IEC 23008-3:2019, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 3: 3D audio*

ISO/IEC 23091-3, *Information technology — Coding-independent code points — Part 3: Audio*

3 Terms, definitions and mnemonics

3.1 Terms and definitions

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

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

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

Note 1 to entry: Traditionally, a collection of songs released on a Compact Disk belongs into this category, for example.

3.1.4
conformance test bitstream
bitstream used for testing the conformance of MPEG-D DRC compliant audio decoders

3.1.5
conformance test case
conformance test category and a combination of one or more conformance test conditions for which a conformance test sequence is provided

3.1.6
conformance test condition
condition which applies to properties of a conformance test sequence in order to test a certain functionality of the MPEG-D DRC decoder

3.1.7
conformance test criteria
one or more conformance test tools and corresponding parameters applied to verify the conformance for a certain conformance test sequence

3.1.8
conformance test sequence
set of a conformance test bitstream, a decoder setting, an input audio file and a corresponding reference file

3.1.9
decoder input parameters
input parameters that are supplied to an MPEG-D DRC decoder in addition to a conformance test bitstream, a decoder interface bitstream and an input audio file

3.1.10
decoder setting
combination of a decoder interface bitstream and decoder input parameters that are supplied to an MPEG-D DRC decoder

3.1.11
input DRC set selection parameters
input parameter set for testing of a DRC gain decoder instance

Note 1 to entry: This parameter set is solely used for conformance testing in the context of the DRC gain decoder conformance test category (DrcGainDec).

3.1.12
reference audio file
decoded counterpart of a conformance test bitstream, a decoder setting and an input audio file

3.1.13
reference DRC set selection parameters
decoded counterpart of a conformance test bitstream and a decoder setting fed to the DRC set selection process

Note 1 to entry: This parameter set is an intermediate result of an MPEG-D DRC compliant decoder implementation solely used for conformance testing in the context of the DRC selection process test category (DrcSelProc).

3.1.14
reference file
reference audio file or reference DRC set selection parameters

3.2 Mnemonics

bslbf	bit string, left bit first, where "left" is the order in which bit strings are written in the ISO/IEC 14496 series NOTE 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.
byte_align()	number of bits to fill for byte alignment at the offset of *n* bits: byte_align(*n*) = 8 ceil (*n*/8) – *n*
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
mod	modulo operator: (x mod y) = x-y floor (x/y)
sizeof(x)	size operator that returns the bit size of a field x
TRUE/FALSE	values of Boolean data type, which correspond to numerical 1 and 0, respectively

4 Symbols

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

NOTE The cross-over frequency is the upper band edge frequency of the sub-band.

f_s audio sample rate in Hz

NOTE If an audio decoder is present, it is the sample rate of the decoded time-domain audio signal.

M_{DRC} DRC frame size in units of the audio sample interval $1/f_s$

N_{DRC} maximum permitted number of DRC samples per DRC frame

NOTE 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 $1/f_s$

π ratio of a circle's circumference to its diameter

s audio decoder sub-band index (starting at 0)

z complex variable of the z-transform

5 Technical overview

The technology described in this document is called the "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. In addition to encoded time-varying gain values, the DRC gain decoder can also receive parametric DRC metadata for generation of time-varying gain values at the decoder. 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.

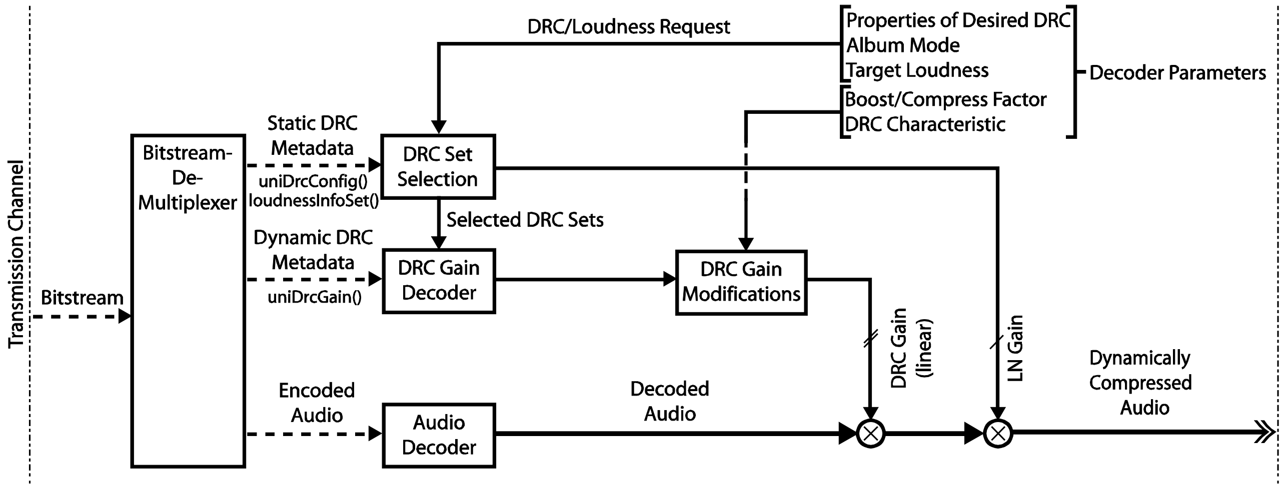

Figure 1 — Block diagram of a typical system with audio decoder and DRC tool modules to achieve loudness normalization (LN) and dynamic range control

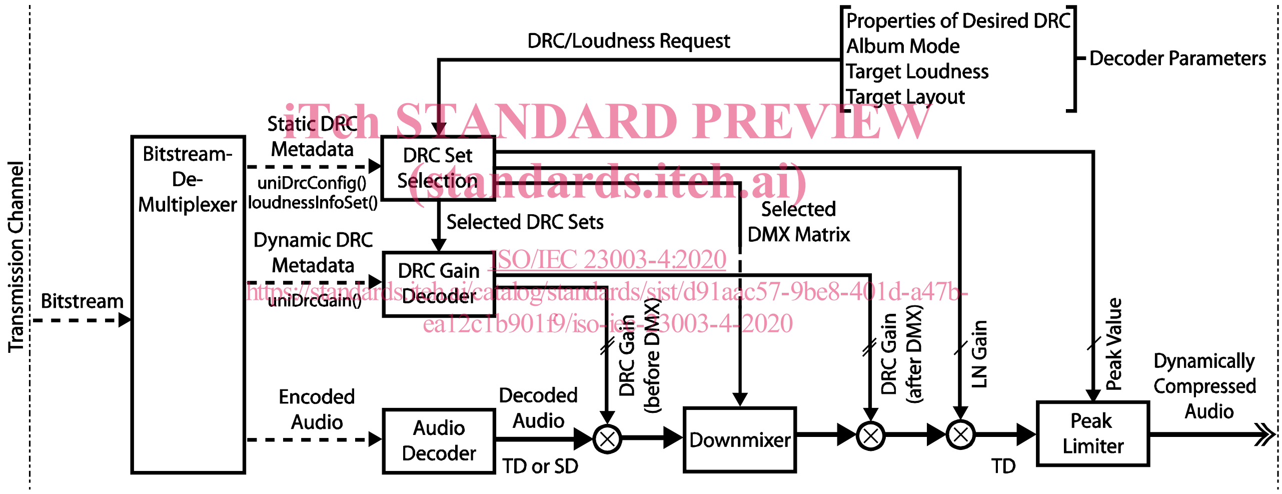

Figure 2 — Block diagram of a more complex system including downmixer and peak limiter (TD = time-domain, SD = subband-domain)

The DRC tool provides support for loudness equalization, sometimes called “loudness compensation”, that can be applied to compensate for the effect of the playback level on the spectral balance. For this purpose, time-varying loudness information can be recovered from DRC gain sequences to dynamically control the compensation module. While the compensation module is out of scope, the interface describes in which frequency ranges the loudness information should be applied.

A flexible tool for generic metadata-controlled equalization is provided. The tool can be used to reach the desired spectral balance of the reproduced audio signal depending on a wide variety of playback scenarios, such as downmix, DRC, or playback room size. It can operate in the sub-band domain of an audio decoder and in the time domain.

The DRC tool is specified in Clause 6. The tool may be subject to profiles and levels that shall be in accordance with Annex I. The bitstream field decoding of the DRC tool shall be in accordance with Annex A. If an interface for external parameter control of the DRC tool is used, it shall conform to Annex B.

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 in ISO/IEC 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 in A.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 6.12.

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 several logical blocks:

— channelLayout();

— downmixInstructions(), downmixInstructionsV1();

— drcCoefficientsBasic(), drcCoefficientsUniDrc(), drcCoefficientsUniDrcV1();

— drcInstructionsBasic(), drcInstructionUniDrc(), drcInstructionUniDrcV1();

— loudnessInfo(), loudnessInfoV1();

— drcCoefficientsParametricDrc();

— parametricDrcInstructions();

— loudEqInstructions();

— eqCoefficients();

— eqInstructions().

Except for the channelLayout(), drcCoefficientsParametricDrc(), and eqCoefficients(), multiple instances of a logical block can appear. The DRC decoder combines the information of the matching instances of the 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 shall 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 shall read and discard the bits (*otherBit*) of the extension payload. Each extension payload type in uniDrcConfig() or loudnessInfoSet() shall not appear more than once in the bitstream if not stated otherwise. An extension payload of type UNIDRCCONFEXT_V1 shall preceed an extension payload of type UNIDRCCONFEXT_PARAM_DRC in the bitstream if both payloads are present. For ISO/IEC 14496-12, configuration extension payloads are provided according to Table 76.

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 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 base channel count represents the total number of base content channels. The base channel count value shall serve as the value of baseChannelCount for parsing the downmixInstructions(), downmixInstructionsV1(), drcInstructionsUniDrc(), drcInstructionsUniDrcV1() and eqInstructions() payloads as specified in Clause 7.

6.1.2.2 downmixInstructions() and downmixInstructionsV1()

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. In contrast to downmixInstructions(), the downmixInstructionsV1() payload includes an offset for all downmix coefficients and the coefficient decoding does not depend on the LFE channel assignment. The downmixInstructions() box for ISO/IEC 14496-12 contains the corresponding metadata of either one of the in-stream payloads as indicated by the *version* parameter of the box.

6.1.2.3 drcCoefficientsBasic(), drcCoefficientsUniDrc(), and drcCoefficientsUniDrcV1()

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 document. 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 shall increase