ANSI/CTA Standard

Loudness Standard for Over the Top Television and Online Video Distribution for Mobile and Fixed Devices

ANSI/CTA-2075

January 2020

Approved American National Standards
ANSI

Consumer Technology Association™

NOTICE

(Formulated under the cognizance of the CTA **R4 Video Systems Committee**.)

Published by
©CONSUMER TECHNOLOGY ASSOCIATION 2020
Technology & Standards Department
www.cta.tech

(This page left blank intentionally.)

FOREWORD

This standard was developed by the Consumer Technology Association's R4 Video Systems Committee.

TABLE OF CONTENTS

FIGURES

TABLES

Loudness Standard for Over the Top Television and Online Video Distribution for Mobile and Fixed Devices

1 INTRODUCTION

The rise in digital audio technology, with its significant improvements in dynamic range capability, brought program loudness challenges to content producers and consumers alike. Improved loudness measurement standards, coupled with industry guidelines and government regulation, all led to the successful management of loudness for traditional broadcast television throughout the world, whether delivered via Over the Air (OTA), cable, or satellite.

As consumers expand their video consumption to include new services provided by broadband internet connection, aka Over the Top (OTT) and Online Video Distribution (OVD), similar loudness management challenges arise. The Audio Engineering Society's AES71 Recommended Practice leverages successful techniques from the broadcast TV industry and combines them with other recommendations for content creation and distribution specific to OTT and OVD.

With OTT content now available at predictable loudness per AES71, the audio systems of mobile and fixed devices remain the final link in the delivery chain for producing consistent and enjoyable listener experiences. For example, listeners might still find that they cannot enjoy the sound in noisy environments. Importantly, the predictability and benefits of AES71 content can help CE designers optimize the listener experience.

Though beyond the scope of this standard, the predictability of AES71 content may also encourage designers to improve loudness matching when combining or switching compliant OTT content with other sources, such as system sounds, phone calls, the voice of a digital assistant, etc.

Experts familiar with AES71 and consumer electronics formed a group to create new guidelines for producing consistent and enjoyable listener experiences for OTT content on mobile and fixed devices. This standard is the result of their work.

1.1 Objectives of the Standard

Products that are compliant with this standard will improve the listening experience by:

- providing consistent loudness across different programs, service providers and advertising content,
- providing appropriate ranges for playback loudness for different devices and listening conditions,

- preventing excessive peak limiting or other processing that degrades the audio quality,
- preserving the artistic intent of wide content dynamic range (movies, drama, live music).

2 SCOPE

This standard applies to devices, whether mobile or fixed, that receive OTT and OVD services and create audio output either through a built-in loudspeaker or through an interface to an external transducer or another A/V device. It describes the management of loudness and dynamic range within a device, with or without the use of metadata, to optimize the listener experience considering:

a. the available resources of the device to achieve appropriate SPL,

b. environmental or physical conditions that affect the way a given loudness level and dynamic range are experienced,

c. user preferences that change the required loudness and dynamic range.

NOTE: This standard is intended to be used along with implementation information provided by codec suppliers.

3 REFERENCES

3.1 Normative References

The following standards contain provisions that, through reference in this text, constitute normative provisions of this standard. At the time of publication, the editions indicated were valid. All standards are subject to revision. Users of this standard are cautioned that newer editions of the referenced standards might or might not be compatible.

3.1.1 Normative Reference List

ATSC A/52, "Digital Audio Compression (AC-3, E-AC-3)," ATSC, 2018, https://www.atsc.org/standard/a522012-digital-audio-compression-ac-3-e-ac-3-standard-12172012/

ETSI TS 102 366 V1.4.1, "Digital Audio Compression (AC-3, E-AC-3) Standard, ETSI, 2017, https://www.etsi.org/deliver/etsi_ts/102300_102399/102366/01.04.01_60/ts_102366v010401p.pdf

ETSI TS 102 114 V1.4.1, "DTS Coherent Acoustics; Core and Extensions with Additional Profiles," ETSI, 2012, https://www.etsi.org/deliver/etsi_ts/102100_102199/102114/01.04.01_60/ts_102114v010401p.pdf

ETSI TS 103 190-1 V1.3.1, "Digital Audio Compression (AC-4) Standard; Part 1: Channel based coding." ETSI, 2018, https://www.etsi.org/deliver/etsi_ts/103100_103199/10319001/01.03.01_60/ts_10319001v010301p.pdf

ETSI TS 103 190-2 V1.2.1, "Digital Audio Compression (AC-4) Standard; Part 2: Immersive and personalized audio." ETSI, 2018, https://www.etsi.org/deliver/etsi_ts/103100_103199/10319002/01.02.01_60/ts_10319002v010201p.pdf

ETSI TS 103 491 V1.1.1, "DTS-UHD Audio Format; Delivery of Channels, Objects and Ambisonic Sound Fields." ETSI, 2017, https://www.etsi.org/deliver/etsi_ts/103400_103499/103491/01.01.01_60/ts_103491v010101p.pdf

ISO/IEC 14496-3:2019 Information technology – Coding of audio-visual objects -- Part 3: Audio, International Organization for Standardization (ISO), Geneva, Switzerland, https://www.iso.org/standard/76383.html

ISO/IEC 23003-4:2015 Information technology -- MPEG audio technologies -- Part 4: Dynamic Range Control, International Organization for Standardization (ISO), Geneva, Switzerland, https://www.iso.org/standard/66482.html

ISO/IEC 23008-3:2019 Information technology – High efficiency coding and media delivery in heterogeneous environments -- Part 3: 3D audio, International Organization for Standardization (ISO), Geneva, Switzerland, https://www.iso.org/standard/ 74430.html

3.1.2 Normative Reference Acquisition

ATSC Standards

- Advanced Television Systems Committee, 1776 K Street NW, Washington, DC 20006-2340, USA, Phone +1 202 872 9160, http://www.atsc.org

ETSI Standards

- ETSI, 650, Route des Lucioles, 06560 Valbonne - Sophia Antipolis, France, Phone: +33 (0)4 92 94 42 00, https://www.etsi.org

ISO Standards

- International Organization for Standardization, Chemin de Blandonnet 8, CP 401, 1214 Vernier, Geneva, Switzerland, Phone: +41 22 749 01 11, https://www.iso.org/store.html

3.2 Informative References

The following standards contain provisions that, through reference in this text, constitute informative provisions of this standard. At the time of publication, the editions indicated were valid. For dated references, only the edition cited applies. For undated references the latest edition of the referenced standards (including any amendments) applies. All standards are subject to revision. Users of this standard are cautioned that newer editions of the referenced standards might or might not be compatible.

3.2.1 Informative Reference List

AES71-2018, AES Recommended Practice - Loudness Guidelines for Over the Top Television and Online Video Distribution. Audio Engineering Society, Inc., 2018. http://www.aes.org/publications/standards/search.cfm?docID=107

AES-paper-6233, Preferred Listening Levels and Acceptance Windows for Dialog Reproduction in the Domestic Environment. Audio Engineering Society, Convention Paper 6233, October 2004. http://www.aes.org/e-lib/browse.cfm?elib=12890

ANSI/CTA-2034-A Standard Method of Measurement for In-Home Loudspeakers, February 2015, Washington DC, USA, https://www.cta.tech/Resources/Standards

ANSI S3.36-2012, "Manikin for Simulated In-Situ Airborne Acoustic Measurements," ANSI/ASA, 2012, https://webstore.ansi.org/standards/asa/ansiasas3362012

CTA-CEB11-C NTSC/ATSC Loudness Matching, Consumer Technology Association (CTA), January 2015, Washington DC, USA, https://www.cta.tech/Resources/Standards

CTA-CEB32.5 Recommended Practice for ATSC 3.0 Television Sets, Audio, June 2017, Washington DC, USA, https://www.cta.tech/Resources/Standards

EN 50332-1:2013, "Sound System Equipment: Headphones and Earphones Associated with Personal Music Players – Maximum Sound Pressure Level Measurement Methodology. Part 1: General method for one package equipment", CENELEC, 2013. https://tinyurl.com/y5yyfmhs

EN 50332-2:2013, "Sound System Equipment: Headphones and Earphones Associated with Personal Music Players – Maximum Sound Pressure Level Measurement Methodology. Part 2”, CENELEC, 2013. https://tinyurl.com/y4ucpsf7

EN 50332-3:2017, "Sound System Equipment: Headphones and Earphones Associated with Personal Music Players – Maximum Sound Pressure Level Measurement Methodology. Part 3: Measurement method for sound dose management”, CENELEC, 2017. https://tinyurl.com/y34y2f4s

EPA-600, “Speech levels in various noise environments,” U.S. Environmental Protection Agency, May 1977. https://nepis.epa.gov

IEC 60268-1, “Sound system equipment. Part 1: General” IEC, 1985, https://webstore.iec.ch/publication/1204

IEC 60268-7, “Sound system equipment. Part 7: Headphones and earphones” IEC, 2010, https://webstore.iec.ch/publication/1226

IEC 62368-1:2018, “Audio/video, information and communication technology equipment – Part 1L Safety requirements” IEC, 2018, https://webstore.iec.ch/publication/27412

ISO/IEC 23003-3:2012 Information technology -- MPEG audio technologies -- Part 3: Unified speech and audio coding, International Organization for Standardization (ISO), Geneva, Switzerland, https://www.iso.org/standard/57464.html

ITU-R BS.1770-4, “Recommendation ITU-R BS.1770-4. Algorithms to measure audio programme loudness and true-peak audio level,” ITU-R, 2015, https://www.itu.int/rec/R-REC-BS.1770-4-201510-I/en

ITU-T H.870, “Recommendation ITU-T H.870. Guidelines for safe listening devices/systems,” ITU-T, 2018, https://www.itu.int/rec/T-REC-H.870-201808-I/en

SMPTE ST 2095-1, “Calibration Reference Wideband Digital Pink Noise Signal,” SMPTE, 2015, https://ieeexplore.ieee.org/document/7395532

3.2.2 Informative Reference Acquisition

AES Standards

- Audio Engineering Society, Inc., 551 Fifth Ave., Suite 1225, New York, NY 10176, USA; Phone +1 212 661 8528; http://www.aes.org/publications/standards

ANSI Standards

- American National Standards Institute, 1899 L Street NW, Washington, DC, 20036, USA; Phone +1 202 293 8020; https://www.ansi.org

CENELEC Standards:

- CEN-CENELEC Management Centre, Rue de la Science 23, B-1040 Brussels, Belgium, phone + 32 2 550 08 11, https://www.cenelec.eu

CTA Standards and Bulletins:

- Consumer Technology Association, 1919 S. Eads St., Arlington, VA 22202, USA, Phone: +1 703 907 7600, https://www.cta.tech/Resources/Standards

EPA Documents

- Environmental Protection Agency, 1200 Pennsylvania Avenue, N.W., Washington, DC 20460, Phone +1 202 564 4700, http://www.epa.gov

IEC Standards

- International Electrotechnical Commission, Rue de Varembé 3, 1211 Geneva 20, Switzerland, Phone: +41 22 919 02 11, https://webstore.iec.ch

ISO Standards

- International Organization for Standardization, Chemin de Blandonnet 8, CP 401, 1214 Vernier, Geneva, Switzerland, Phone: +41 22 749 01 11, https://www.iso.org/store.html

ITU Standards

- International Telecommunication Union, Place des Nations 1211, Geneva 20, Switzerland, +41 22 730 5111; http://www.itu.int

SMPTE Standards

- Society of Motion Picture and Television Engineers, White Plains Plaza, 445 Hamilton Ave. STE 601, White Plains, NY 10601-1827, USA, Phone: +1 914 761 1100, https://www.smpte.org

4 COMPLIANCE NOTATION

This section defines compliance terms for use by this standard.

shall	This word indicates specific provisions that are to be followed strictly (no deviation is permitted).

shall not	This phrase indicates specific provisions that are absolutely prohibited.
should	This word indicates that a certain course of action is preferred but not required.
should not	This phrase means a certain possibility or course of action is undesirable but not prohibited.

5 DEFINITIONS

For the purposes of this standard, the following definitions apply:

Anchor-based loudness	A type of loudness measurement based on ITU-R BS.1770-4 applied to the content gated by activity of the anchor signal, which is commonly speech.
Content loudness	Loudness of a content asset after decoding when no loudness normalization or DRC is applied. It has a value of $L_{k,C}$.
Decoder device	A consumer device capable of decoding audio content associated with OTT services into a PCM format.
Dynamic Range Control	The process of continually adjusting audio signal level to reduce the level difference between loud and soft passages according to some desired objective.
Fixed device	A stationary decoder device.
Headroom	The margin between the highest level of the signal and the highest level that can be represented in, or conveyed by, the system. In the digital domain, for example, if the signal peak is at L_{TP} =-5 dBTP there is 5 dB of headroom.
Loudness	The subjective impression of a sound's acoustic strength. For purposes of this standard, loudness is an electrical measurement of audio content with a value of L_k such that various content having the same measured loudness will have the same subjective loudness.
Loudness request	A parameter with a value of $L_{k,R}$ to specify the desired output loudness of the audio decoder and metadata processing block.
Maximum SPL	The SPL of a transducer device produced by a test signal at the maximum volume setting according to Annex G.
Mobile device	A decoder device that is designed such that it can be taken along by the user and can operate almost anywhere.

Output loudness[1]	Expected loudness level with a value of $L_{k,out}$ at the output of the audio decoding and metadata processing block. The block may include gain and a peak limiter as shown in Figure 4. $L_{k,out}$ is equal to the loudness request value, if a request is present.
Over the air television	Linear television transmitted over-the-air by a licensed broadcast station intended for direct reception by the general public using an antenna.
Over the top television	The means to deliver video content via streaming, VOD, pay TV, IPTV and download via IP mechanisms. However, throughout this standard, over the top television should be understood to also include online video distribution.
Online video distribution	A service that offers video content by means of the Internet or other Internet Protocol (IP)-based transmission path.
Target layout	Input parameter to the decoder and metadata processing block that specifies the desired channel configuration of the output of that block.
Transducer device	Active physical transducer device connected to the decoder device for playback. It may be integrated into the decoder device or connected externally. An external transducer device may include a chain of external devices that actively delivers the audio signal to the physical transducer or transducer array.

6 SYMBOLS AND ABBREVIATIONS

For the purposes of this standard, the following symbols and abbreviations apply:

DRC	Dynamic Range Control or Dynamic Range Compressor
FS	Full Scale. The RMS level of a 997-Hz sine wave whose positive peak value reaches digital full scale is 0 dB FS
LKFS	K-weighted loudness level relative to Full Scale. A unit according to ITU-R BS.1770
LUFS	Loudness Units relative to Full Scale. A synonym for LKFS
OTA	Over The Air as defined in 5
OTT	Over The Top as defined in 5
OVD	Online Video Distribution as defined in 5
PLR	Peak to Loudness Ratio in dB. It has a value of L_{TP} - L_k

[1] CTA-CEB32.5 "Recommended Practice for ATSC 3.0 Television sets, Audio" defines this as Output Reference Level, and recommends it to be -24 LKFS or -31 LKFS.

RMS	Root Mean Square. The square root of the mean of the squares of sample amplitudes
SPL	Sound Pressure Level. The unit dB SPL refers to sound pressure level referenced to 20 µPa
G	Constant gain value in dB. To apply the gain, it is converted to the linear domain and the audio signal is scaled by the linear gain factor
L_k	Loudness level value measured according to Rec. ITU-R BS.1770 in units of LKFS
$L_{k,C}$	Loudness level value measured in LKFS of a content asset after decoding when no loudness normalization or DRC is applied
L_{TP}	True Peak level value of an audio signal, measured according to Rec. ITU-R BS.1770 in units of dBTP
$L_{k,R}$	Loudness request value in LKFS. A parameter to specify the desired output loudness $L_{k,out}$ of the audio decoder and metadata processing block
$L_{k,out}$	Output loudness level in LKFS as defined in 5
L_p	Sound pressure level value in dB SPL
$L_{p,max}$	Maximum SPL value in units of dB SPL as defined in 5

7 SYSTEM OVERVIEW

This section is informative.

This standard is applicable to devices that render both content from Over the Top Television and Online Video Distribution services that follow the practices addressed in the Audio Engineering Society Recommended Practice (RP) AES71, as well as other OTT content. The subject of this loudness and dynamic range control (DRC) standard encompasses the control of the audio decoder and metadata processing block within a mobile or fixed consumer device that supports OTT television. Metadata processing is commonly co-located with or integrated in the decoder. This standard identifies the key parameters needed to effectively enable metadata processing to achieve consistent and appropriate loudness and the application of DRC when necessary.

The AES71 RP is recognized in this standard so that content produced and distributed using its guidelines will be compatible with the various benefits provided by this standard. AES71 content playing back on a device compliant with this standard will offer an optimized listener experience regardless of the listening transducer device (micro-speaker, headphone, television, home theater) or environment (indoor or outdoor).

Figure 1 shows part of a device that processes OTT content. The device has an audio decoder and metadata processing block that processes the incoming OTT television content bitstream. This device is also referred to as a decoder device. The metadata processing is controlled by the input labeled loudness/DRC configuration. This input aggregates parameters that are derived from many other parameters in the parameter lookup block (grey). When processed by the system this enables the correct adjustment of loudness and dynamic range for the given playback scenario. The grey block, which is the group of parameters that combine to determine appropriate audio playback, encompasses the subject of this standard.

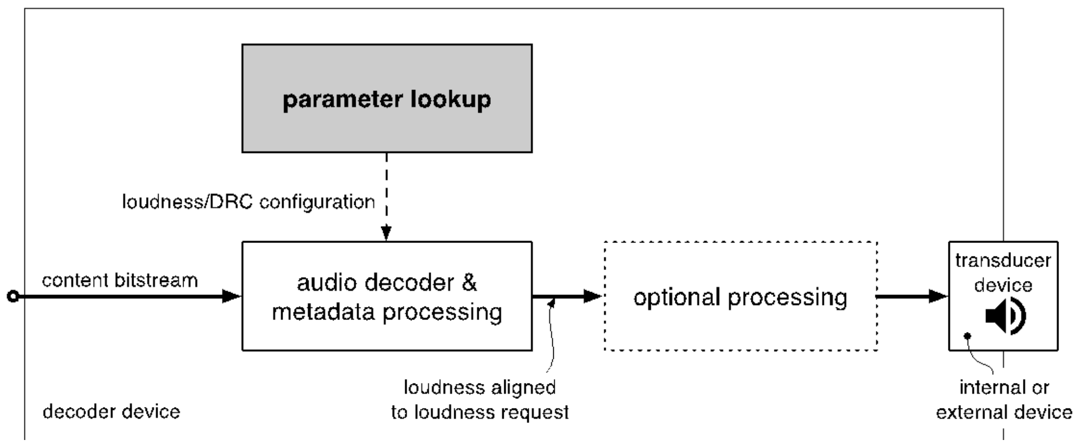

Figure 1 – High-level block diagram of conforming audio playback system in a consumer device.

The optional processing block shown provides additional audio processing that may be applied by the device. This block may be provided by the device designer is beyond the scope of this standard. This block may prove useful for control and loudness alignment of other audio sources along with the output of the audio decoder and metadata processing block.

Post-processing may also include loudspeaker protection and hearing protection which are out of scope for this standard. This post-processing may work independently from the system described here and may alter loudness and dynamics of the final output. The addition of other dynamics processing stages should be avoided, if possible, to minimize content degradations.

The transducer device is the active physical device connected to the system for playback. It can be internal (speaker(s) or micro-speaker(s)) or external (for example headphones, Bluetooth, Wi-Fi or HDMI connection) to the decoder device. When connected to the decoder device, a transducer device can have a limited SPL range in comparison to the full SPL range that a listener might enjoy. This can be caused by the characteristics of the transducer or its amplifier[2]. External devices can be connected by wire or wirelessly using standards that specify such connections for digital or analog interfaces. In general, the decoder's transducer device can be changed from one type to another by the user. The audio decoder and metadata processing block will adjust for such a change, if necessary. For example, when headphones are plugged into the decoder device, the audio playback switches from the internal loudspeaker to the headphones, then the headphones become the transducer device (see Figure 1). This also applies to an HDMI, Bluetooth or Wi-Fi connection, for example.

For some transducer devices that are connected to the decoder device by a digital interface, it may be required or preferred to output the unmodified content bitstreams of certain formats using a bitstream passthrough while other bitstream formats are processed as shown in Figure 1. In such a bitstream passthrough mode, the receiving device takes the role of the decoder device. The two devices can be set up to achieve consistent loudness across different content bitstream formats (see Annex C).

8 TECHNICAL DETAILS

8.1 System Details and Input Parameters

A complete system diagram with the relevant blocks of a consumer device and applicable parameters is shown in Figure 2, which expands on Figure 1. The grey parameter lookup block shown encompasses the main strategy for solving the problems described in the scope of this standard. All the input parameters of that block are shown in Figure 2 and they are specified in Table 1. The input parameters shall only use the permitted values of Table 1, except for the user preferences. The settings of these parameters for some common use cases are explained below.

[2] In some regions, personal music players, headphones and earphones are manufactured to not exceed a fixed SPL limit according to CENELEC 50332-1 and 50332-2. It is recommended that such products adopt newer CENELEC 50332-3 regulations, when applicable, that do not require a fixed SPL limit, but a dose measurement instead (see also ITU-T H.870). This will result in improved preservation of the dynamics and more headroom for playback at a medium loudness request value of -24 LKFS in relatively noisy environments.

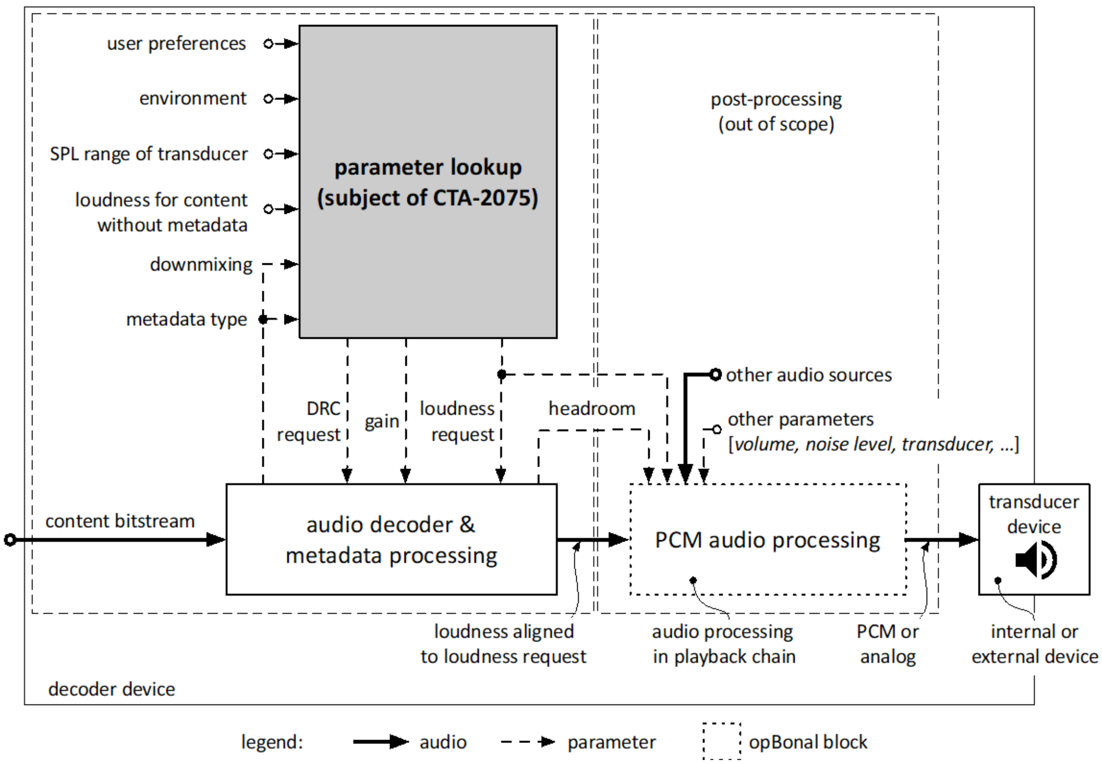

Figure 2 – Detailed block diagram of the audio playback system in a consumer device and parameter lookup block encompassing this standard.

Table 1 – Inputs to parameter lookup block for loudness and DRC control.

Input parameter	Permitted values[3]	Comment (informative)
user preferences	”max DRC”, ”late night”, ”DRC off”, “none” (see Table 2)	User preferences to adjust DRC. Other custom values are permitted.
environment	”ideal”, ”noisy”, ”unknown” (see Table 3)	Factors of the listener’s environment that may imply playback adjustments
SPL range of transducer device	”small”, ”medium”, ”large”, ”unknown” (see Table 4)	The range is mainly determined by the limited maximum SPL the transducer can deliver
loudness for content without metadata	$L_{k,C}$, “unknown”	For some specific content the loudness may be known
downmixing	“yes”, “no”	Indicates whether downmixing is active

[3] The labels in this column are provided to convey their meaning, but they are not necessarily intended as user-facing descriptors for user interfaces.

metadata type	MPEG-D DRC, AAC Metadata, (E)AC-3, AC-4, DTS-(U)HD, “none”	Derived from stream. The value of “none” may also be applied if loudness/DRC metadata of other formats is present but not used.

- The *user preferences* input supports DRC adjustments that may be provided through a user interface as listed in Table 2 (see also Annex D). A user preference is active, when it is different from “none”. In that case, it may cause an override of the parameters obtained from the parameter lookup. Other custom values are permitted.

Table 2 – Input parameters describing user preferences that may override the parameter lookup.

User preference[3]	Meaning
“max DRC”	Use the most aggressive DRC.
“late night”	An environment where the listener has the desire to reduce playback loudness. For example, to not disturb others.
“DRC off”	Turn DRC off according to 8.2.
“none”	No user preference is active, no override occurs.

- The *environment* input supports information about the playback scenario that may be provided by multiple sources as described in Table 3.

Table 3 – Input parameters describing the environment.

Environment[3]	Meaning
“ideal”	A quiet playback environment.
“noisy”	The listener is exposed to environmental noise.
“unknown”	The environment is unknown.

Examples:
  - a level-sensing device should be used to detect environmental conditions[4] to determine if the listening environment is noisy,

[4] For environmental condition detection, it is sufficient to sample the noise level at a low rate. The operating system of the decoder device should offer this service and send environmental data to the parameter look-up block. The determination of thresholds for what is a noisy environment and what is a quiet environment is left to

- a built-in car audio system might always assume a noisy environment.

- The *SPL range of the transducer device* is usually a known factor and the corresponding parameter shall be set appropriately (see Table 4). The characteristics of loudspeakers that are connected externally by the user are often unknown to the device and the "unknown" setting should apply in this case.

- A value for the "*loudness for content without metadata*" input may be known a priori in some specific cases, for example when it is known that all content of a specific streaming service is loudness normalized to that value.

- *Downmixing* may be necessary if the target layout deviates from the content channel configuration. The content channel configuration and *metadata type* is derived from the audio stream, typically as part of the audio decoding process. The target layout is determined by the transducer device's loudspeaker (or headphone) configuration, or the use of a virtualizer (a type of post-processing spatial renderer) known to the device and shall be signaled as such to the audio decoder when engaged.[5]

As previously noted, the optional post-processing block in Figure 2 is out of scope for this standard. It is shown to illustrate further loudness and dynamics processing that may be made available by the device. This type of post-processing may be engaged in cases where the transducer requires more aggressive DRC or increased loudness that exceed the instruction capability of the parameter processing block (see Annex B for more details).

8.2 Derivation of Control Parameters

Handling of a stream with a particular format by the decoder device is compliant with this standard if the applicable audio decoder is listed in this standard and if loudness and DRC metadata processing is supported for the metadata type(s) associated with that decoder as listed in 8.2.1 or if loudness and DRC processing for the stream without metadata is supported with that decoder as listed in 8.2.2.

To control the metadata processing, this standard includes the derivation of the loudness request value, the gain value, and the DRC request parameter by the

---

the device designer based on their level-sensing device, other device characteristics and judgement of what is appropriate.

[5]A virtualizer for stereo speaker or headphone playback can typically accept multichannel or immersive content without additional downmixing. A multichannel or immersive device's loudspeaker configuration may be changed by the user, therefore affecting the target layout and should be accounted for by the device. Both downmixing and virtualization can increase the peak level (L_{TP}).

parameter lookup block (see Figure 2) based on the inputs defined in Table 1.[6] The parameter lookup shall be done by performing the following steps in the order of occurrence:

a. Determine the SPL range of the transducer device.

b. Look up the loudness request value.

c. Determine the DRC request parameter and additional gain, if applicable.

When determining the SPL range of a given transducer device according to Table 4, the maximum SPL measurement according to Annex G can be considered. The maximum SPL measurement may be applied to assess if the system can reach a sufficient maximum SPL when output loudness normalization corresponding to the transducer category is applied. Table 4 includes suggested maximum SPL ranges for each transducer category.

Table 4 also specifies the loudness request values for all categories of transducer device range. The "unknown" case for the range applies when the decoder device has no information about the active transducer device SPL range.[7] If the metadata contains more than one type of loudness measurement, the selection of which data to use for the value of $L_{k,C}$ should be based on regional standards and recommendations as outlined in AES71-2018.[8] The audio decoder and metadata processing block shall dynamically adjust for changes of the transducer device, for example when headphones are plugged in.

Table 4 – Loudness request value for categories of transducer device SPL range.

SPL range of transducer	Suggested Maximum SPL, $L_{p,max}$	Loudness request value, $L_{k,R}$
"small"	below 75	-16
"medium"	between 70 and 90	-24
"large"	above 85	-31
"unknown"	not applicable	-24

[6] The loudness request determines the amount of content loudness normalization that is applied. Elevated loudness helps products with limited output SPL provide adequate loudness for the listener. DRC is intended to serve two primary purposes: make audio more intelligible under adverse listening conditions; and achieve more loudness when using products with limited output capability. Examples are limited SPL transducers, limited output power amplifiers, and limited output headphone drivers.

[7] The loudness request is elevated as the range decreases to achieve more consistent loudness across transducer devices with different ranges. This improves intelligibility and generates adequate loudness for transducers that have only low SPL output capability.

[8] There are two common approaches to measure loudness of OTT content. A regular ITU-R BS.1770 measurement includes program-independent gating. In contrast, anchor loudness is based on speech-gating to prevent parts of the audio signal that do not contain speech from contributing to the loudness measurement. Anchor loudness is expected to be consistent in professional productions according to best practices. Support for the loudness metadata depends on the metadata system (see Annex E) and it may vary by region (see AES71-2018).

It is up to the device designer to determine the transducer device category best suited for the product application and listeners. The lowest loudness request value that achieves adequate loudness for the listener will best preserve the content dynamics while minimizing the need for aggressive DRC profiles.

After the loudness request value has been determined, the applicable DRC request parameter and gain value are determined as specified in 8.2.1 for streams with metadata or in 8.2.2 for streams without metadata. Annex F illustrates the output loudness and loudness variations for several scenarios.

8.2.1 Streams with Metadata

For streams with metadata, the DRC request parameter and gain shall be determined as applicable using Table 5 to Table 12, where each Table is specific to an audio codec family and the associated metadata formats. The DRC request parameter and gain depend on the input parameters defined in Table 1 and the SPL range of the transducer as defined in Table 4.

Depending on the stream's metadata type, it may be necessary to apply a constant gain *g* to the decoded audio signal to achieve the requested loudness. Figure 3 to Figure 6 show the internal configuration of the audio decoder and metadata processing block for streams with various types of metadata (see also Annex E). If necessary, a multiplier is present to apply the appropriate gain. The applicable gain value shall be the value given in Table 7 to Table 12. Any necessary adjustment to the DRC metadata based on a gain scaling factor (see Annex A) shall be applied using the value given in the Tables, if applicable.

8.2.1.1 Control Parameters for MPEG Metadata Systems

MPEG metadata systems are controlled by the DRC request and loudness request as shown in Figure 3 with a peak limiter following the DRC processing. The control parameters shall be as defined in Table 5.

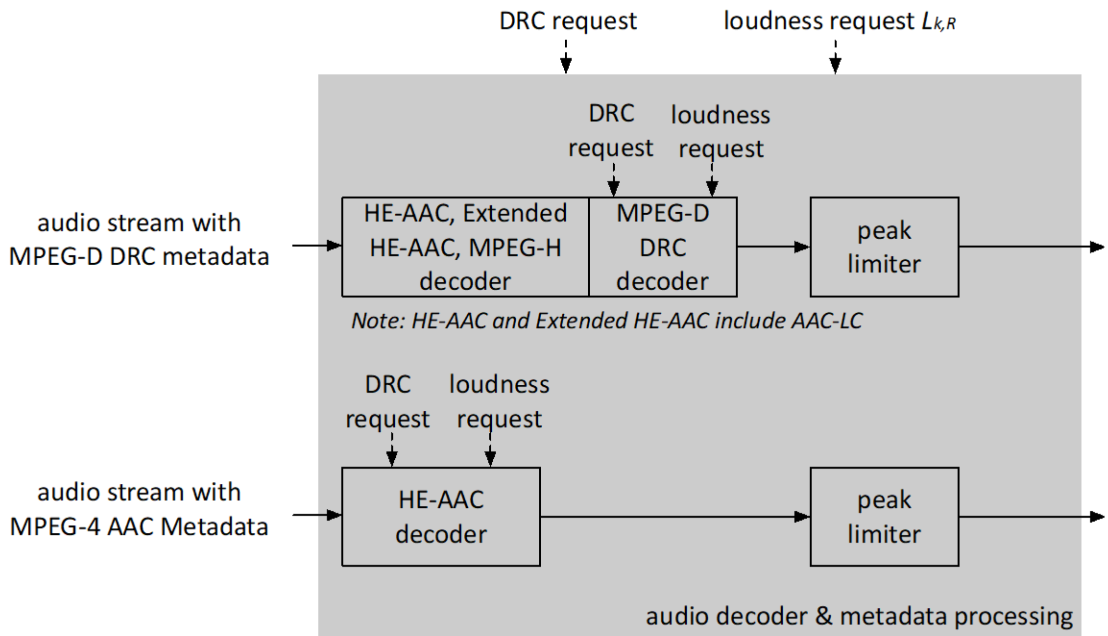

Figure 3 – Internal configuration of the audio decoder and metadata processing block of Figure 2 for MPEG metadata systems.

Table 5 – DRC request settings for MPEG metadata systems. The labels specifying the DRC request are defined in ISO/IEC 23003-4 and ISO/IEC 14496-3.

Environment	**Transducer SPL range**	**MPEG-D DRC**[a]	**MPEG-4 AAC Metadata**
		DRC request	**DRC request**
“ideal”, “unknown”	“large”	“general”[c]	“light compression”
	“medium”, “unknown”	“general”[c]	“light compression”[b]
“noisy”	“large”, “medium”, “small”, “unknown”	“noisy”	“heavy compression”
“ideal”, “unknown”	“small”	“limited”	“heavy compression”

[a] At the time of this writing, some platforms support MPEG-D DRC in MPEG-H 3D Audio, Extended HE-AAC and AAC-LC.
[b] If downmixing is active, the DRC request shall be “heavy compression”. Otherwise, an optional switch may be provided to the user to turn DRC off.
[c] A DRC request such as “general” specifies the desired characteristics of the output signal but may not always result in compression as explained in E.1.

The DRC request of Table 5 can be overridden based on user preferences as specified in Table 6. An override shall only occur if both transducer SPL range and environment have one of the values indicated. The “DRC off” settings are merely an example showing under which conditions an override occurs. Please note that the peak limiter will ensure that full scale is not exceeded even when DRC is disabled.

Table 6 – Override settings based on user preferences for MPEG metadata systems. The labels specifying the DRC request are defined in ISO/IEC 23003-4 and ISO/IEC 14496-3.

User preference	Transducer SPL range	Environment	MPEG-D DRC	MPEG-4 AAC Metadata
			DRC request	DRC request
“max DRC”	All	All	“noisy”	“heavy compression”
“late night”	“large”	“ideal”, “unknown”	“late night”	“light compression” [a]
	“medium”, “unknown”	“ideal”, “unknown”	“late night”	“heavy compression”
“DRC off”	“medium”, “large”, “unknown”	“ideal”, “unknown”	off	off

[a] If downmixing is active, the DRC request shall be “heavy compression”.

8.2.1.2 Control Parameters for Dolby Metadata Systems

Dolby AC-3 or E-AC-3 metadata systems shall be controlled by the DRC request and gain value as shown in Figure 4 with a peak limiter following the DRC processing and gain function. The control parameters shall be as specified in Table 7. The gain values are calculated as $g = L_{k,R} - L_{k,(E)AC3}$, where $L_{k,(E)AC3}$ is the specified decoder output loudness that corresponds with the particular DRC request. The DRC request shall use the labels in the DRC request column in Table 7 and Table 8 as defined in the standards referenced in E.3 and E.4.

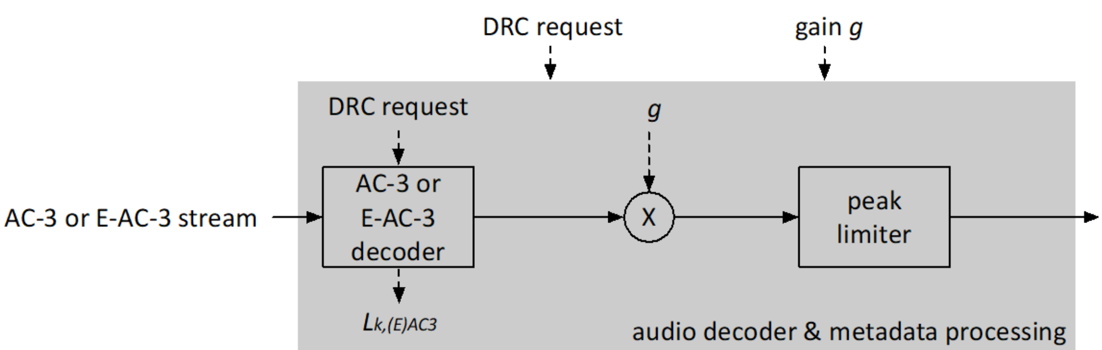

Figure 4 – Internal configuration of the audio decoder and metadata processing block of Figure 2 for Dolby AC-3 or E-AC-3 metadata systems.

The loudness request value and gain of Table 7 can be overridden based on user preferences as specified in Table 8. An override shall only occur if both transducer SPL range and environment have one of the values indicated. The “DRC off” settings are merely an example showing under which conditions an override occurs. Please note that the limiter will still ensure that full scale is not exceeded when the DRC is disabled.

Table 7 – DRC request settings and gain for AC-3 and E-AC-3 systems.

Environment	Transducer SPL range	AC-3, E-AC-3		
		DRC request	$L_{k,(E)AC3}$ **[LKFS]** [a]	**Gain** g
“ideal”, “unknown”	“large”	“Line mode”	-31	0
	“medium”, “unknown”	“RF mode”	-20	-4
“noisy”	“large”	“RF mode”	-20	-11
	“medium”, “unknown”	“RF mode”	-20	-4
All	“small”	“RF mode”	-20	+4

[a] For AC-3 and E-AC-3, the decoder responds to the DRC request by normalizing content to $L_{k,(E)AC3}$ and then gain g is applied to the decoder output if necessary. The value of $L_{k,(E)AC3}$ is provided for information.

Table 8 – Override settings based on user preferences for AC-3 and E-AC-3 systems.

User preference	Transducer SPL ranges where preference is applied	Environments where preference is applied	AC-3, E-AC-3		
			DRC request	$L_{k,(E)AC3}$ **[LKFS]** [a]	**Gain** g
“max DRC”	“large”	All	“Line mode” [b]	-31	0
	“medium”, “unknown”	All	“RF mode”	-20	-4
“late night”	“large”	“ideal”, “unknown”	“RF mode”	-20	-11

	"medium", "unknown"	"ideal", "unknown"	"RF mode"	-20	-4
"DRC off"[c]	"large"	"ideal", "unknown"	off	-31	0
	"medium", "unknown"	"ideal", "unknown"	off	-31	+7

[a] For AC-3 and E-AC-3, the decoder responds to the DRC request by normalizing content to $L_{k,(E)AC3}$ and then gain g is applied to the decoder output if necessary. The value of $L_{k,(E)AC3}$ is provided for information.
[b] For maximum DRC, the DRC gain scale factors should be set to 1, 1. (see Annex A)
[c] To turn DRC off, the DRC gain scale factors should be set to 0, 0. (see Annex A)

Dolby AC-4 metadata systems are controlled by the loudness request and gain value as shown in Figure 5. The control parameters shall be as specified in Table 9, where the decoder output loudness is calculated as $L_{k,AC4} = L_{k,R} - g$.

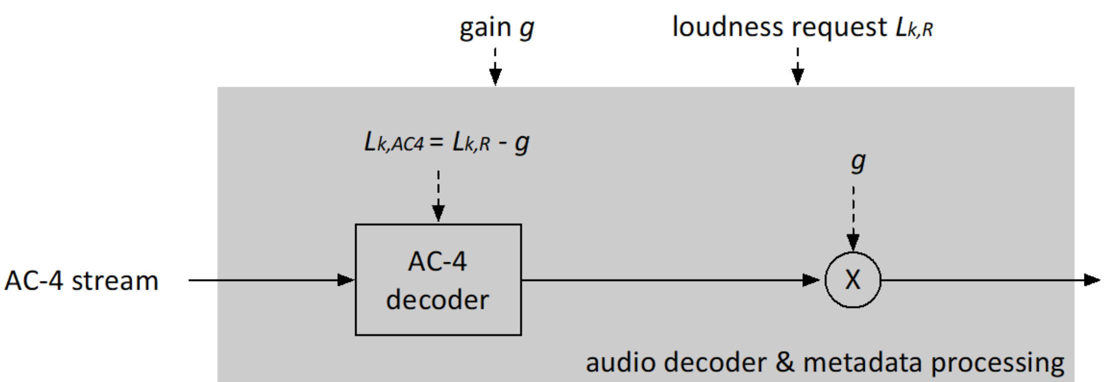

Figure 5 – Internal configuration of the audio decoder and metadata processing block of Figure 2 for Dolby AC-4 metadata systems.

Table 9 – Loudness and DRC configuration for AC-4 systems.

Environment	**Transducer SPL range**	**Loudness/DRC Configuration**	
		$L_{k,AC4}$ [LKFS] [a]	**Gain g [dB]**
"ideal", "unknown"	"large"	-31	0
	"medium", "unknown"	-24	0
"noisy"	"large"	-24	-7

	“medium”, “unknown”	-16	-8
All	“small”	-16	0

[a] The AC-4 decoder uses the loudness value $L_{k,AC4}$ to normalize content to the specified decoder output level and then gain g is applied to the decoder output if necessary. The loudness value is computed as $L_{k,AC4} = L_{k,R}$ - g. The decoder internally selects the appropriate DRC mode and the loudness correction gains for downmixes to be applied based on the loudness value $L_{k,AC4}$, specified output channel configuration and metadata in the bitstream.

The loudness request value and gain of Table 9 can be overridden based on user preferences as specified in Table 10. An override shall only occur if both, transducer SPL range and environment, have one of the values indicated. The “DRC off” settings are merely an example showing under which conditions an override occurs. Please note that the limiter will still ensure that full scale is not exceeded when the DRC is disabled.

Table 10 – Override setting based on user preferences for AC-4 systems.

User preference	**Transducer SPL range**	**Environment**	**Loudness/DRC Configuration**	
			$L_{k,AC4}$ [LKFS] [b]	**Gain *g* [dB]**
“max DRC” [a]	“large”	All	-31	0
	“medium”, “unknown”	All	-24	0
“late night”	“large”	“ideal”, “unknown”	-31	0
	“medium”, “unknown”	“ideal”, “unknown”	-24	0
“DRC off” [a]	“large”	“ideal”, “unknown”	-31	0
	“medium”, “unknown”	“ideal”, “unknown”	-24	0

[a] DRC can be disabled in the AC-4 decoder and the internal peak limiter will ensure that full scale is not exceeded.
[b] The AC-4 decoder uses the loudness value $L_{k,AC4}$ to normalize content to the specified decoder output level and then gain g is applied to the decoder output if necessary. The loudness value is computed as $L_{k,AC4} = L_{k,R}$ - g. The decoder internally selects the appropriate DRC mode and the loudness correction gains for downmixes to be applied based on the loudness value $L_{k,AC4}$, specified output channel configuration and metadata in the bitstream.

8.2.1.3 Control Parameters for DTS Metadata Systems

DTS-HD has no DRC operation. All audio output is normalized to -31 LKFS (except when downmixing, see Annex E.5). DRC, gain adjustments, and peak limiting functions all need to be provided after the decoder. The gain value is computed as $g = L_{k,R} - L_{k,DTS\_HD}$, based on the decoder output loudness normalization at L_{k,DTS_HD} = -31 LKFS.

DTS-UHD is controlled by the DRC request and loudness request as shown in Figure 6. A peak limiter is included in the decoder. The control parameters shall be as specified in Table 11. The DRC request shall use the labels in the DRC request column in Table 11 and Table 12 as defined in the standards referenced in section E.6.

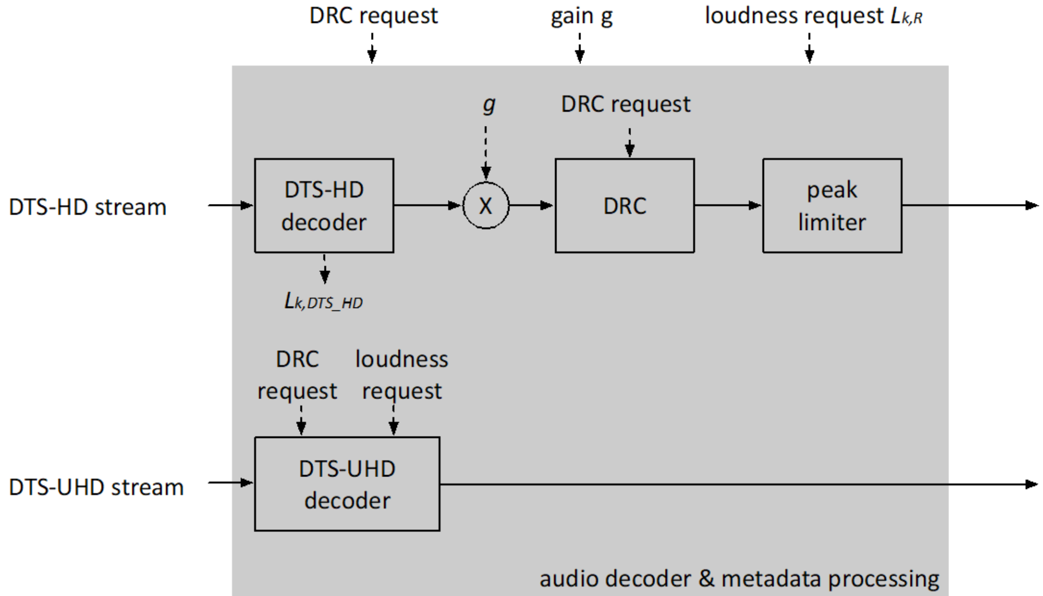

Figure 6 – Internal configuration of the audio decoder and metadata processing block of Figure 2 for DTS metadata systems.

Table 11 – DRC request settings and gain for DTS metadata systems.

Environment	**Transducer SPL range**	**DTS-HD**	**DTS-UHD**
		Gain *g* [a]	**DRC request**
"ideal", "unknown"	"large"	$L_{k,R}$ +31	off
	"medium", "unknown"	$L_{k,R}$ +31	"low", gain scaling of 50% (see also Annex A)
"noisy"	"large", "medium", "unknown"	$L_{k,R}$ +31	"high"
"ideal", "noisy", "unknown"	"small"	$L_{k,R}$ +31	"high"

[a] The DTS-HD decoder applies a protective gain offset when downmixing. The offset is based on bitstream downmix metadata and the specified output channel configuration, and is available to the system in order to enable gain compensation. The compensating gain offset is added to the gain, *g*, value.

The loudness request value and DRC request of Table 11 can be overridden based on user preferences as specified in Table 12. An override shall only occur if both, transducer SPL range and environment, have one of the values indicated. The “DRC off” settings are merely an example showing under which conditions an override occurs. Please note that the limiter will still ensure that full scale is not exceeded when the DRC is disabled.

Table 12 – Override settings based on user preferences for DTS metadata systems.

User preference	**Transducer SPL ranges where preference is applied**	**Environments where preference is applied**	**DTS-HD Loudness/DRC Configuration**	**DTS-UHD Loudness/DRC Configuration**
			Gain *g* [a]	**DRC request**
“max DRC”	All	All	$L_{k,R}$ +31	“high”
“late night”	“large”, “medium”, “unknown”	“ideal”, “unknown”	$L_{k,R}$ +31	“medium”
“DRC off”	“large”, “medium”, “unknown”	“ideal”, “unknown”	$L_{k,R}$ +31	off

[a] The DTS-HD decoder applies a protective gain offset when downmixing. The offset is based on bitstream downmix metadata and the specified output channel configuration, and is available to the system in order to enable gain compensation. The compensating gain offset is added to the gain, *g*, value.

8.2.2 Streams without Metadata

For streams without metadata, device-supplied DRC should be applied as indicated in Table 13 and Figure 7. For a consistent user experience between streams with and without metadata, this DRC should achieve similar compression characteristics as metadata-based DRCs. It can take advantage of the available information of the parameter lookup block (see Table 1) to control the compression characteristics similarly to the metadata-based DRCs.

The gain value applied to streams without metadata and known loudness shall be $g = L_{k,R} - L_{k,C}$. As indicated in Figure 2, the system or media player API should support the configuration of the content loudness $L_{k,C}$ for content without metadata. $L_{k,C}$ can then be configured accordingly, for example, when OTT content of a service is known to conform to AES71 but does not provide metadata.

The gain value applied to streams without metadata and with unknown loudness should be derived by the following method:

- assume that the loudness of the stream conforms to AES71, which is $L_{k,C}$ = -23 LKFS for Europe and $L_{k,C}$ = -24 LKFS for other regions. This results in a gain of $g = L_{k,R} - L_{k,C}$. Please note that the output loudness will deviate from the loudness request when the stream loudness does not conform to AES71.

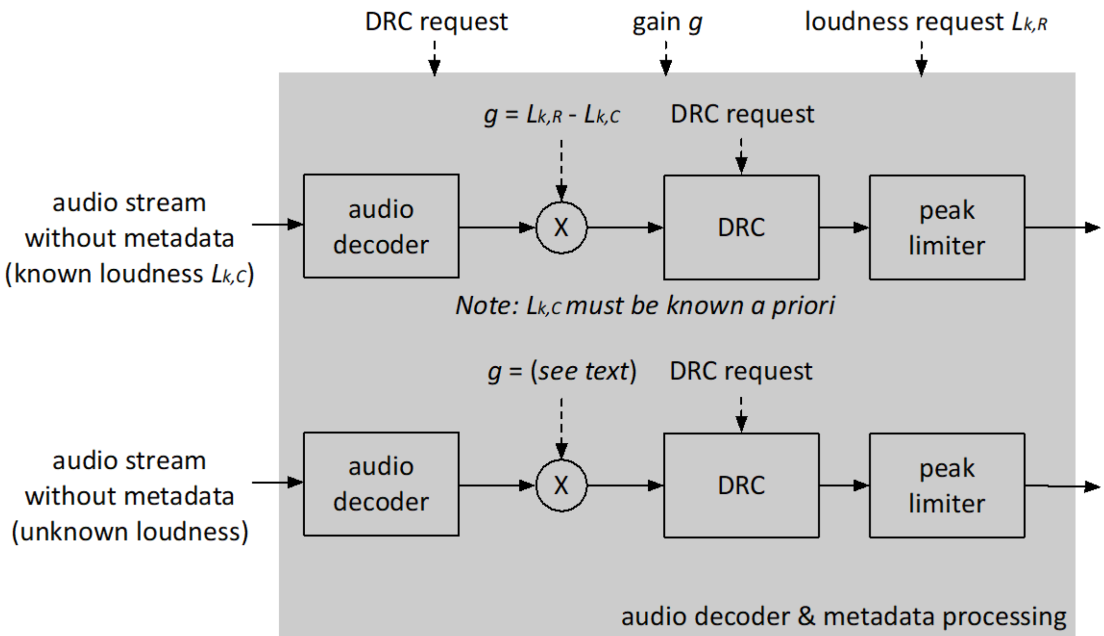

Figure 7 – Internal configuration of the audio decoder and metadata processing block of Figure 2 for bitstreams without metadata.

Table 13 – DRC request settings for bitstreams without metadata.

Environment and user preferences	Transducer SPL range	No metadata, known loudness $L_{k,C}$		No metadata, unknown loudness
		DRC request	Gain g	DRC request
All	All	optional	$L_{k,R} - L_{k,C}$	See text

The loudness request value and DRC request of Table 13 can be overridden based on user preferences as specified in Table 14. An override shall only occur if both, transducer SPL range and environment, have one of the values indicated. The "DRC off" settings are merely an example showing under which conditions an override occurs. Please note that the limiter will still ensure that full scale is not exceeded when the DRC is disabled.

Table 14 – Override settings based on user preferences for bitstreams without metadata.

User preference	Transducer SPL ranges where preference is applied	Environments where preference is applied	No metadata, known loudness $L_{k,c}$	No metadata, unknown loudness
			DRC request	DRC request
"max DRC"	All	All	*aggressive*	*aggressive*
"late night"	"large", "medium", "unknown"	"ideal", "unknown"	*appropriate for "late night"*	*appropriate for "late night"*
"DRC off"	"large", "medium", "unknown"	"ideal", "unknown"	*off*	*off*

8.3 Peak Limiter

In all cases, a sample peak limiter shall be used to prevent clipping (see also Figure 3 to Figure 7). Table 15 indicates which specific systems provide a peak limiter option as part of the audio decoder and metadata processing block that can be used for this purpose. However, it may be preferred to place a peak limiter outside of the audio decoder and metadata processing block to prevent potential clipping caused by additional signal processing or the combining of other signals downstream. If a peak limiter operates downstream, it is recommended to disable the peak limiter of the audio decoder and metadata processing block, if possible, to reduce potential limiter artifacts.

The choices of the peak limiter placement in the playback chain depends mostly on the overall system design. For example, designs that allow placing the limiter after the volume control and before the DAC result in less necessary limiting when the volume control setting is lowered. This way, potential limiter artifacts can be reduced. Other examples are designs that use a PCM audio representation with fixed-point or floating-point format. Because floating-point format practically does not saturate, a limiter is not needed until the signal reaches the DAC, which usually has saturation. For integer PCM formats, well-known methods have been established to ensure that the signal stays within the provided integer value range for each processing stage and a peak limiter may be needed to maintain sufficient signal-to-noise ratio.

Table 15 – Availability of a peak limiter in specific metadata systems.

System	Peak limiter option present
MPEG-D DRC	Yes
MPEG-4 AAC Metadata	No
AC-3, E-AC-3	Yes (required for “noise” and “small range” only”)
AC-4	Yes (limiter always present)
DTS-HD	No
DTS-UHD	Yes

ANNEX A
(INFORMATIVE)
OPTIONAL DRC GAIN ADJUSTMENTS

Metadata processing may support modifications of the dynamic DRC gain that is applied to the audio signal to achieve a compression effect. This is commonly performed with support of two gain scaling factors. To modify the DRC gain, one factor is applied to negative DRC gains in dB. The other factor is applied to positive DRC gains in dB. The supported range of the factors is usually [0.0, 1.0] so that the compression effect can be gradually reduced. For example, when a factor of 0.0 is applied, DRC gains are reduced to 0.0 dB. Table A.1 lists the metadata types that support this kind of gain modification.

For example, in a use case for DRC gain adjustments for noisy environments, the DRC gain modification permits an adjustment of the compression needed to suit the actual environmental noise level. Another potential use case is the adjustment of compression to best fit the transducer range of a playback device. For transducers that can produce greater SPL, less compression may be sufficient.

Table A.1 – Gain modification support.

Metadata type	Support for gain modifications
MPEG-D DRC	Yes (*compress* and *boost* factor)
MPEG-4 AAC Metadata	Yes (“light compression” only)
AC-3, E-AC-3	Yes (“Line Mode” only)
AC-4	Yes
DTS-HD	No
DTS-UHD	Yes

ANNEX B
(INFORMATIVE)
OPTIONAL POST-PROCESSING

For the optional post-processing shown in Figure 2, available information including metadata may be useful. Table B.1 lists some of this information and its potential use case.

Table B.1 – Useful information for post-processing block.

Information	Example use cases
Loudness request value	When system or other audio sources are mixed with the primary audio source, their loudness should be adjusted based on the loudness request value.
DRC request	The DRC applied may impact the type and tuning of post-processing.
Headroom	For some metadata systems the headroom at the processor output is a known value. Headroom may be useful for proper adjustment of any signal processing downstream.
Noise level	The environmental noise level may be used to adjust the volume.
SPL range of transducer	The SPL range of different transducer devices may impact the type and tuning of post-processing.

The optional post-processing block may include additional loudness and dynamics processing. This may be used in cases when the transducer device requires more aggressive DRC or increased loudness. By taking into account the information listed in Table B.1, unnecessary cascading of a postprocessor DRC with a DRC in the metadata processor should be avoided.

The loudness of streams without metadata and with unknown loudness can unpredictably deviate from the assumptions in 8.2.2 and lead to inconsistent levels. Avoiding these problems while preserving artistic intent are reasons why the use of metadata according to 8.2.1 is highly recommended to ensure predictable output loudness.

For those non-ideal streams where metadata has not been provided and the loudness of the streams is unknown, non-metadata-based loudness control algorithms are typically used. Techniques such as real-time Automatic Loudness Control (ALC) may be able to mitigate unexpected loudness jumps. If employed, care should be taken in the design of the ALC algorithm to optimize it for the control of average program loudness and to minimize common processing side effects such as pumping, signal distortion, sound coloration, noise modulation or other artifacts which can negatively impact the content.

A possible improvement could be to design the ALC algorithm to use an initial gain setting based on data stored about previous measurements of the average program loudness level for different services. This data might be collected by the device itself or obtained by other, external facilities. A priori measured data about the stream, such as content loudness ($L_{k,C}$), loudness range, and peak-to-loudness ratio (PLR) can also be helpful to derive appropriate ALC parameters.

More examples that may benefit from implementing optional post-processing are listed below.

a) The use of Surround/immersive upmixers, (used to adapt stereo, 5.1, or 7.1 content for stereo speakers or headphones or their virtualizers)
b) Speaker or headphone spatial virtualization
c) Bass management (crossover filters plus subwoofer output)
d) Room correction and/or user adjustable equalization
e) Room simulation (reflections and reverberation)
f) Mixing/ducking of the OTT content when combined with other device audio
g) Digital signal scaling plus analog gain compensation for headroom management while maintaining playback loudness, particularly useful for a), b), and d)
h) Equalization and protection for internal loudspeaker(s)
i) Hearing protection according to regulations such as EN 50332-3 or IEC 62368-1
j) Final peak limiting prior to DAC to protect additional audio and post-processing

B.1 COMBINING OTT AUDIO WITH OTHER DEVICE SOURCES

The post-processing block may combine OTT audio with other sources, such as system sounds, phone calls, the voice of a digital assistant, etc. The known output loudness (loudness request value) of the audio decoding and metadata processing block is important for guiding the loudness adjustment of other sources with the OTT audio source to achieve overall uniform loudness for the listener. Most of the information available to the grey block in Figure 2 may be useful to the post-processing block.

ANNEX C
(INFORMATIVE)
BITSTREAM OUTPUT

The decoder device (source device) may have an optional digital output, for example over HDMI, S/PDIF, Bluetooth, or WiFi, that supports the output of a coded audio bitstream. The bitstream output can be of two types:

- Bitstream passthrough:
  The coded audio bitstream extracted from the OTT program by the decoder device (source device) is passed to the digital output. Devices compliant with this standard should not alter, add, or remove loudness metadata or DRC metadata in a bitstream.

- Transcoded bitstream:
  A transcoded bitstream is a means to provide a bitstream format that is compatible with certain sink devices or digital interfaces that do not support specific codecs used in OTT delivery. For example, some soundbars with S/PDIF interface require certain bitstream formats for multi-channel audio. In such cases, it may be advantageous to transcode the OTT audio format to a compatible codec format.

Some digital interfaces also support PCM formats in addition to bitstream formats. If PCM is supported by the interface and by the sink device for all supported OTT content formats without compromising the spatial reproduction, for example due to required downmixing, it is not necessary to use a bitstream format for output. In that case, consistent and optimal loudness control is achieved by applying the processing shown in Figure 2 to generate the PCM output.

However, since the digital interface and the sink device may not support PCM for every input format, for example in some cases of multi-channel audio, it may be necessary to output (or pass through) an audio bitstream. In that case, the sink device decodes the audio bitstream and should produce the same loudness as it would when receiving the loudness-adjusted PCM signal as above. When switching between sources that require bitstream output and others, loudness jumps may occur if the system is misconfigured.

Mechanisms to avoid loudness jumps should be based on the SPL range of the transducer device according to Table 4, which determines the output loudness. The output loudness value should be matched by the sink device in order to achieve the same loudness normalization as the decoder device (source device). Note that the transducer is shared by both devices, therefore both devices should have the same transducer SPL range information.

Transcoding may be necessary for digital interfaces or devices that have limited or no support for PCM. For example, Bluetooth interfaces may not support PCM or loudness and DRC metadata. In that case, the audio decoder and metadata processing block of Figure 2 is applied in the decoder device (source device) before re-encoding the audio signal.

ANNEX D
(INFORMATIVE)
DRC OPERATION

The DRC settings specified in 8.2 are designed to provide a good listening experience in various scenarios considering content, transducers, and environment. This is achieved by applying dynamic range compression when necessary. There are provisions to use a minimum of compression when possible, such as:

- A compressor may have a "dead band", meaning that it does not compress unless the audio signal level is outside of the dead band range. Content that was already compressed will be less affected by such a DRC.
- Depending on the content, an MPEG-D DRC encoder can decide how much compression should be applied, if at all. Even if a DRC request is set on the decoder, no compression may be applied.
- In a quiet environment, the amount of applied compression mainly depends on the SPL range of the transducer. A large SPL range results in a loudness request of -31 with the lowest or no compression.

Refer to 8.2.1 for specifics on setting DRC and loudness request values for each audio decoder.

D.1 OPTIONS FOR USER CONTROL OF DRC

Some users may still prefer to override the DRC settings of the automatic process. To support this use case, user input options to modify the DRC settings are specified in 8.2 and may be implemented. The corresponding settings can be set by the user for example as part of a device setup procedure or other user interface options. It is specifically recommended to provide a user option to turn DRC off according to 8.2, for cases where the user wants to play back content with its fullest dynamic range.

D.2 IMPLEMENTATION CONSIDERATIONS

Some products may be able to support a request to turn DRC off except when downmixing, e.g. the common case of downmixing 5.1 programs to stereo. This is usually acceptable. However, depending on the class of product or market requirements, it may be important to offer a DRC off capability whether downmixing or not.

Some audio decoders are able to provide downmixed outputs without applying DRC, but the output levels are scaled down several dB to preserve peak headroom. In other cases, downmixing may be applied after the decoder has produced the complete multichannel PCM output. In this case either the downmixer or the decoder may apply the necessary gain scaling for downmix protection. As long as the scaling value is known to the system, it can be compensated to maintain loudness normalization.

Analog gain compensation typically occurs in the volume control stage following the DAC, utilizing part of the control's total range isolated from the user for this purpose. Annex B mentions additional examples of signal processing where analog gain compensation may be beneficial.

ANNEX E
(INFORMATIVE)
METADATA SYSTEMS

This Annex provides an overview of basic concepts and practices for the metadata systems that are covered in this standard.

E.1 MPEG-D DRC

MPEG-D DRC [ISO/IEC 23003-4] specifies a loudness and DRC metadata tool that can be combined with various audio codecs. The USAC [ISO/IEC 23003-3] codec, which is part of Extended HE-AAC, and the MPEG-H 3D Audio [ISO/IEC 23008-3] codec are specified with mandatory MPEG-D DRC support and the bitstreams of these formats always contain the corresponding metadata. At the time of this writing, some platforms support MPEG-D DRC in MPEG-H 3D Audio and Extended HE-AAC which includes AAC-LC.

Minimum requirements for MPEG-D DRC are defined in two Profiles, the Loudness Control Profile and the Dynamic Range Control Profile of ISO/IEC 23003-4. Decoder processing should support the Dynamic Range Control Profile in systems that implement CTA-2075. The Dynamic Range Control Profile requires the presence of loudness and DRC metadata and the corresponding decoder processing capabilities. Table E.1 lists the mandatory DRC metadata of the DRC Profile in detail. The term “DRC effect” is used to label each mode of DRC according to its use case. The supported loudness $L_{k,out}$ indicates what maximum loudness level is supported by a particular DRC. A DRC mode may have no compression effect. For example, if the original content has a small PLR, it may not be necessary to reduce the dynamic range to achieve a given loudness value.

Table E.1 – Mandatory Dynamic Range Control bitstream Profiles for MPEG-D DRC [ISO/IEC 23003-4].

DRC effect	Maximum supported loudness $L_{k,out}$
Late night (“night”)	-24
Noisy Environment (“noisy”)	-16
Limited Playback range (“limited”)	-16
General compression (“general”)	-24

MPEG-D DRC does not mandate the use of specific DRC characteristics for metadata generation. Therefore, technology upgrades are possible when DRC improvements are made available. DRCs can be tuned specifically for different content types, playback environments, downmix, and output loudness. In addition to the DRCs mandated by the DRC Profiles, more DRCs may be used to achieve finer granularity. For example, DRCs that apply to a specific downmix only may be included.

Optional support for additional loudness metadata is provided that includes a peak level value. Metadata for loudness and peak level of a downmix is supported indicating how the loudness was measured. Examples are speech gating or gating based on the application of BS.1770-4.

Loudness requests and DRC requests can be applied independently – any associated processing is disabled if no request exists

E.2 MPEG-4 AAC METADATA

The family of AAC codecs based on ISO/IEC 14496-3 supports the carriage of loudness and DRC metadata according to ISO/IEC 14496-3:2009, Section 4.5.2.7 and ISO/IEC 14496-3:2009/Amd.4:2013, Section 4.5.2.14.

The Programme Reference Level indicates the content loudness $L_{k,C}$ of the transmitted audio content. During decoding, this parameter is used to match the decoder output level to a given loudness request $L_{k,R}$.

The dynamic range of the decoded audio signal can be adapted to different listening scenarios based on dynamic DRC gains embedded in the AAC bitstream. Dependent on the specific implementation, the application of those gains can be controlled by a decoder interface.

MPEG-4 AAC Metadata provides two types of DRC gains: one for ‘light compression’ and one for ‘heavy’ compression. At the encoder, those gains are inserted for each audio frame according to the configured compression characteristic.

At the time of writing, decoder support of MPEG-4 AAC Metadata is integrated on the following platforms: Android, macOS, iOS, and others.

E.3 AC-3, E-AC-3

The AC-3/E-AC-3 codecs defined in ATSC A/52 include the mandatory carriage of loudness and DRC metadata (in the form of gain values) where all decoders support the application of the metadata.

The metadata parameter *dialnorm* is used to indicate the content loudness, $L_{k,C}$, of the encoded audio and is applied during the decoding process to normalize the audio according to the selected operating mode. AC-3 natively supports two operating modes, where E-AC-3 supports three. The operating mode defines the decoder output loudness, $L_{k,(E)AC3}$ as follows:

- **Line mode** the output level is normalized to -31 LKFS;
- **RF mode** where the output level is normalized to -20 LKFS,

The value of *dialnorm* defines the center of the null-band of the DRC profile that is used to calculate the DRC metadata during the encoding process.

The DRC system in AC-3/E-AC-3 is a reversible process where separate DRC gain coefficients, *dynrng* for line-mode and *compr* for RF-mode, are calculated in the encoder and carried in the bitstream. Five different DRC profiles (Music Light, Music Standard, Film Light, Film Standard, Speech) which are explicitly defined in ETSI TS 102 366 and are selected during encoding. This gives content creators flexible control of how their content is decoded. The DRC metadata is applied during the decoding, depending on the decoder operating mode active in the decoder device. It can also be selected to not apply DRC and output the decoded audio program with unaltered content dynamic range. To account for possible overload during downmixing, the encoder calculates the worst-case overload for each audio block and includes this in the DRC coefficients.

E.4 AC-4

ETSI TS 103 190-1 V1.3.1 (2018-02) defines a parameter L_{out} as "Reference output level for the DRC processor, supplied by the system". This parameter corresponds to the $L_{k,AC4}$ value in this standard and is an input parameter to the decoder. The parameter L_{out} can have any value and is not limited to the discrete values specified in this standard. Based on the value of $L_{k,AC4}$, the decoder selects the corresponding DRC Decoder Mode from the bitstream. This can be one of the predefined DRC Decoder Modes "Home Theater", "Flat Panel", or "Portable". The predefined "Portable" DRC Decoder Mode provides two variants, one for speaker ("Portable Speaker") and one for headphones ("Portable Headphones"), which can be externally signaled depending on whether playback is over speakers or headphones with limited output SPL capability (i.e. small dynamic range).

Alternatively, a "custom DRC decoder mode" can be used. There are four optional custom DRC decoder modes available for each presentation in the AC-4 stream. The custom DRC decoder mode specifies the range of output reference levels for which it is valid and because of their higher DRC decoder mode ID they have priority over the predefined DRC decoder modes.

Table E.2 – AC-4 DRC Decoder Modes and mapping of the predefined modes to values of $L_{k,AC4}$.

AC-4 DRC Decoder Mode ID	Name	Range of $L_{k,AC4}$
0	"Home Theatre"	-31 LKFS ≤ $L_{k,AC\text{-}4}$ ≤ -27 LKFS
1	"Flat panel TV"	-26 LKFS ≤ $L_{k,AC\text{-}4}$ ≤ -17 LKFS
2	"Portable – Speakers"	-16 LKFS ≤ $L_{k,AC\text{-}4}$ ≤ 0 LKFS

AC-4 DRC Decoder Mode ID	Name	Range of $L_{k,AC4}$
3	“Portable – Headphones”	-16 LKFS ≤ $L_{k,AC-4}$ ≤ 0 LKFS
4 … 7	Custom defined	As per drc_output_level_from and drc_output_level_to.

For each DRC Decoder Mode, specific DRC data can either be transmitted as DRC gains, or as a DRC curve that controls the DRC tool in the decoder. The DRC tool then generates DRC gains to be applied. Independent of gains or curve, the DRC tool also scales the signal to match the output reference level L_{out}.

AC-4 uses downmix correction gains to compensate for a loudness change during the downmix

The AC-4 decoder allows DRC to be enabled or disabled, and DRC gains that are either transmitted or generated by the DRC tool to be scaled in the decoder, based on user or system input.

Every Dolby AC-4 decoder implementation is equipped with a final limiter to ensure that full scale is not exceeded.

E.5 DTS-HD

The DTS-HD codec does not support DRC metadata. It does support content loudness metadata in the form of “DIALNORM.” Based on DIALNORM metadata, the decoder applies gain reduction as needed to normalize all content output at -31 LKFS.

The DTS-HD decoder applies a protective gain offset when downmixing. The offset is based on bitstream downmix metadata and the specified output channel configuration and is available to the system in order to enable gain compensation to maintain loudness normalization.

Refer to ETSI TS 102 114 for DIALNORM parameter values.

E.6 DTS-UHD

The DTS-UHD codec supports DRC and loudness metadata. The decoder can output any reasonable loudness level, e.g. from -31 to -16 LKFS, with or without downmixing, and will apply DRC and/or peak limiting, referenced to the output loudness, to achieve the result. DRC and loudness requests are API-controlled.

To minimize peak limiting during decoding or downmixing, a reduced loudness request value may be used in conjunction with analog gain compensation to maintain loudness normalization for the listener.

The DRC module present in the UHD decoder includes nine predefined DRC profiles. Each tier, "Low", "Medium" and "High" has three variants, -A, -B, and -C.

"Low", "Medium" and "High" describe the general amount of compression that will be applied, with "A", "B", and "C" defining the general symmetry between the amount of gain boost vs. cut:

A: less aggressive attenuation to loud content

B: less aggressive boost to quiet content

C: equal amount of attenuation and boost

The percentages of gain boost and cut can be adjusted within each profile, and specific slow/fast attack and release times are associated with each profile.

Low-C, Medium-C and High-C are the default presets in the decoder. The product designer may override these selections and choose different profiles for Low, Medium and High based on specific needs of the product.

Refer to ETSI TS 103 491 for details on the Loudness and DRC parameters.

ANNEX F
(INFORMATIVE)
OVERALL LOUDNESS AND LOUDNESS VARIATIONS WITHIN CONTENT

At the output of the audio decoder and metadata processing block of Figure 2, it is expected that the overall loudness will be close to the requested loudness. Loudness variations will be reduced if a DRC is applied.

As an example, Figure F.1 illustrates the overall loudness and perceived loudness range of a stream using MPEG-D DRC metadata. The reduced range is apparent when DRC is engaged. For the “noisy” DRC setting, the shifted grey arrows illustrate a level shift due to an optional post-processing step to increase the loudness, dependent on the environmental noise level.

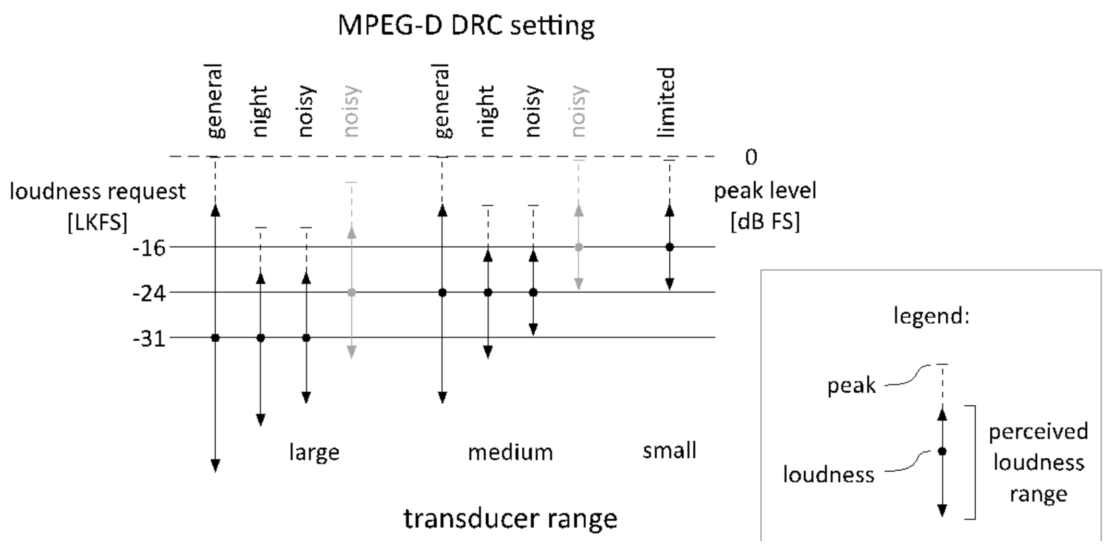

Figure F.1 – Illustration of loudness and range after MPEG-D DRC metadata is applied.

ANNEX G
(INFORMATIVE)
MAXIMUM SPL MEASUREMENT

The maximum SPL of the transducer device is the SPL of a test signal at the maximum volume setting. It should be considered when assigning one of the three SPL range categories per Table 4. This ensures that the system can reach a sufficient maximum SPL when the output loudness normalization corresponding to the specific category is applied. Suggested ranges for the maximum SPL per category are provided by Table 4 for guidance.

The suggested maximum SPL measurement procedure is based on the acoustical pressure $p_{DRP}(t)$ at the drum reference point (DRP) of one of a manikin's ears. The manikin is located at the anticipated listener position when the test signal is radiated from the transducer device. If the transducers are not located in a symmetrical arrangement relative to the median plane of the manikin, the measurement should be done at the ear that is exposed to the highest sound pressure. The acoustical environment for the measurement should mimic a typical use case scenario for the device under test.

For the measurement, the decoder and metadata processing block reproduces the test signal. Any post-processing should only include signal processing that is always applied when the transducer device is active, such as loudspeaker protection or equalization. The volume control should be set to the maximum volume.

A sound source with known SPL is used for calibration. The maximum SPL $L_{p,max}$ of the transducer device is computed based on $p_{DRP}(t)$ by applying the inverse diffuse field response of the manikin and an A-weighting filter to the recorded test signal. The calibrated result is equivalent with an A-weighted loudness measurement in the diffuse field without manikin. A suitable diffuse field response is specified by ANSI S3.36, for example.

G.1 RECOMMENDED TEST SIGNAL

The test signal should simulate the spectral energy distribution and dynamics of typical OTT content. Band-limited pink noise is commonly used for such purposes and provides a reasonable approximation of the spectral distribution. Such pink noise can be generated, for example, as specified in one of the following standards:

- SMPTE ST 2095-1
- IEC 60268-1, Section 7, "Simulated Programme Signal"
- ANSI/CTA-2034-A

The channel layout of the test signal is identical to that of the transducer device. Each channel signal contains pink noise at the same level and it is not correlated with any other channel signal. The loudness level of the PCM test signal including all channels should be -24 LKFS.

G.2 LOUDSPEAKER MEASUREMENT

During the measurement, all loudspeakers of the transducer device radiate the test signal at the same time.

G.3 HEAD-MOUNTED TRANSDUCER MEASUREMENT

Head-mounted transducers, such as headphones, are attached to a manikin's head for the measurement. The measurement of $p_{DRP}(t)$ can be done in accordance with IEC 60268-7, section 8.5.2., for example.

G.4 EXAMPLE RESULTS

Figure G.1 shows example results of maximum SPL measurements of four transducer devices that span all three transducer SPL range categories. The SPL range boundaries according to Table 4 are shown as grey bars. Based on the measurement, the transducer devices are categorized according to Table G.1 and the corresponding loudness request values are assigned.

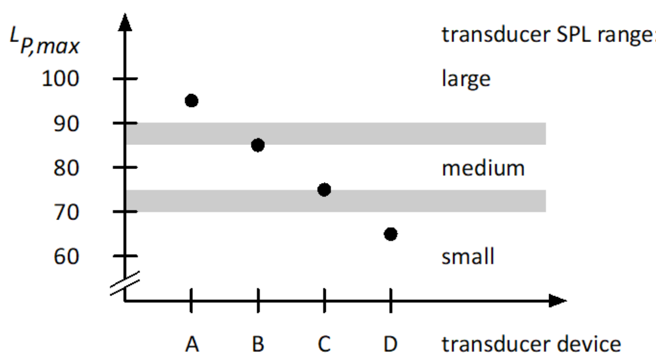

Figure G.1 – Example results of maximum SPL measurements for four transducer devices.

Table G.1 – Example transducer device category assignment.

Transducer device	Transducer SPL range	Loudness request value
A	"large"	-31
B, C	"medium"	-24
D	"small"	-16

Figure G.2 shows the SPL that would be produced by each transducer device when the assigned loudness request value would be applied during the maximum SPL measurement. It illustrates that the approximated desired listening level in quiet can be reached by the devices. In noise, device A is still able to exceed the desired listening level, even though the loudness request has the lowest value of -31. Desired listening levels are in the range of measurements reported in AES-paper-6233 and EPA-600 for TVs and home theaters.

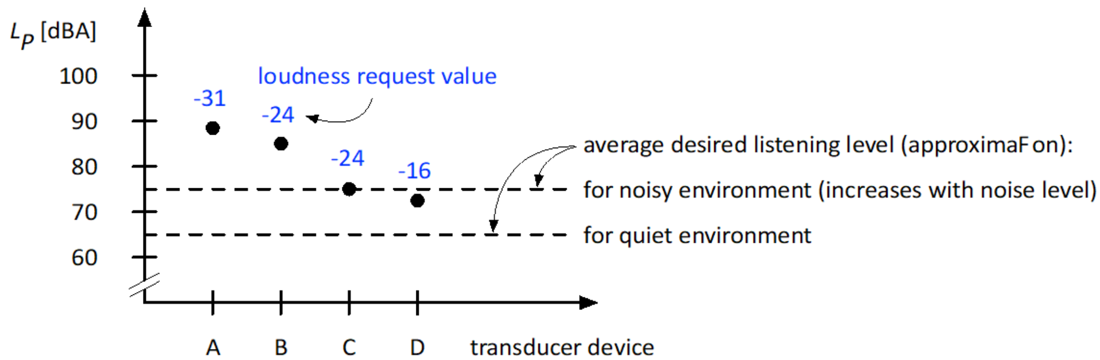

Figure G.2 – Expected SPL of the transducer devices in Figure G.1 when applying the loudness request value as indicated.

Consumer Technology Association Document Improvement Proposal

If in the review or use of this document a potential change is made evident for safety, health or technical reasons, please email your reason/rationale for the recommended change to standards@CTA.tech.

Consumer Technology Association
Technology & Standards Department
1919 S Eads Street, Arlington, VA 22202
FAX: (703) 907-7693 standards@CTA.tech

Consumer
Technology
Association™