ATSC Proposed Standard: Revision of ATSC Recommended Practice A/85: Techniques for Establishing and Maintaining Audio Loudness for Digital Television

Doc. S42-032r12
9 June 2026

This document is a revision in process of A/85 as approved and published in 2013. It is in the final stages of balloting by the full ATSC membership. The ballot closes on 8 July 2026. Editorial changes may be made after approval and prior to publication. Readers are encouraged to download the final version of this document after publication, which is expected to be no later than 13 July 2026.

Advanced Television Systems Committee
1300 I Street, N.W., Suite 400E
Washington, D.C. 20005
202-872-9160

ATSC®, the Broadcast Standards Association®, is an international, non-profit organization developing voluntary standards and recommended practices for broadcast television and multimedia data distribution. ATSC member organizations represent the broadcast, professional equipment, motion picture, consumer electronics, computer, cable, satellite, and semiconductor industries. ATSC also develops implementation strategies and supports educational activities on ATSC standards. ATSC was formed in 1983 by the member organizations of the Joint Committee on Inter-society Coordination (JCIC): the Consumer Technology Association (CTA), the Institute of Electrical and Electronics Engineers (IEEE), the National Association of Broadcasters (NAB), the Internet & Television Association (NCTA), and the Society of Motion Picture and Television Engineers (SMPTE). For more information visit www.atsc.org.

Note: The user's attention is called to the possibility that compliance with this recommended practice may require use of an invention covered by patent rights. By publication of this document, no position is taken with respect to the validity of this claim or of any patent rights in connection therewith. One or more patent holders have, however, filed a statement regarding the terms on which such patent holder(s) may be willing to grant a license under these rights to individuals or entities desiring to obtain such a license. Details may be obtained from the ATSC Secretary and the patent holder.

Revision History

Version	**Date**
A/85:2009 approved	4 November 2009
A/85:2011 (successor document) approved	25 July 2011
A/85:2013 (successor document) approved Corrigendum No. 1 approved Hyperlinks to external .wav files fixed	12 March 2013 11 February 2021 16 May 2025
A/85:2026-XX (successor document) approved Clarified Loudness Measurement Guidelines Created Streaming Loudness Annex L Created One-Page Summary Table (Annex M)	<Date>

Table of Contents

Index of Figures and Tables

ATSC Recommended Practice: Techniques for Establishing and Maintaining Audio Loudness for Digital Television

1. SCOPE

This ATSC Recommended Practice (RP) provides guidance to creators and distributors of audio for high-definition (HD) or standard-definition (SD) television (TV) Content, as presented in both stereo (2.0-channel) and surround (5.1-channel) formats. It recommends production, distribution, and transmission guidelines to help provide the highest quality audio soundtracks to digital television (DTV) audiences.

This RP focuses on audio measurement and monitoring techniques for production and postproduction, insofar as they provide methods to effectively control Loudness for Content delivery or exchange. It recommends methods to effectively control TV audio Loudness, discusses audio metadata systems and their use, describes modern dynamic-range control, and provides recommendations for effective audio monitoring.

This RP specifies essential information on management of Content enabling consistent Loudness across transitions at Content boundaries between Long-form Content and Short-form Content[1]. Short-form Content includes commercial advertising (“spots”), public-service announcements (PSAs), promotional material (“promos”), political advertising, emergency information, news bulletins, and other discrete Content insertions into TV services that are otherwise composed primarily of Long-form Content.

This RP also includes several Annexes that provide additional background and supportive detail, including three that function as short, stand-alone “Quick Reference Guides” for use in three specific areas: 1) TV stations and MVPDs, 2) audio production and postproduction facilities, and 3) Streaming services.

1.1 Background and Introduction

Consumers do not expect large changes in audio Loudness between Long-form and Short-form TV Content, or between TV channels.

The NTSC analog television system used conventional audio dynamic-range processing at various points along the signal path to manage and provide consistent audio Loudness for broadcasts. This practice compensated for limitations in the dynamic range of analog equipment and controlled the various Loudness levels of audio received from suppliers. It also helped smooth the Loudness of Long-form to Short-form Content transitions. Though simple and effective, this practice permanently reduced dynamic range and changed the audio before it reached the audience. It modified the characteristics of the original sound, altering it from what the Content creator intended, to fit within the limitations of the analog system.

The AC-3 audio system defined in the ATSC 1.0 DTV standard [4] uses metadata to control Loudness and other audio parameters without permanently altering the dynamic range of the Content. The Content provider or DTV Operator encodes metadata along with the audio Content. The Dialog Normalization (dialnorm) metadata parameter causes receivers to adjust Loudness, and if properly used, can enable Content to have uniform Loudness without user knowledge or

[1] In other contexts, Long-form Content may be called “programs,” while Short-form Content may be referred to collectively as “interstitial(s),” or simply “commercials.”

intervention. It achieves results similar to a viewer using a remote control to set a comfortable volume between disparate TV Content segments, and TV channel-changing transitions. The dialnorm and other metadata parameters are integral to the AC-3 audio bit stream.

ATSC A/53, Part 5 [1] mandates the carriage of dialnorm and correctly set dialnorm values.

This RP is designed to enable the industry to establish and maintain proficiency in Loudness measurement, production monitoring, metadata usage, and contemporary dynamic-range practices.

This RP provides technical recommendations and information concerning:

- Loudness measurement using Recommendation ITU-R BS.1770 [3].
- Target Loudness for Content exchange without metadata.
- Setup of reference monitoring environments when producing for the expanded dynamic range of DTV, with consideration for multiple listening environments in the home.
- Methods to facilitate control of Loudness at Long-form to Short-form Content boundaries.
- Effective uses of audio metadata for production, distribution, and transmission of digital Content.
- Dynamic Range Control (DRC) within AC-3 audio, and contemporary conventional dynamic range control as an addition or alternative, including recommendations for Loudness and dynamics management at the boundaries between Long-form and Short-form Content.

Note that Content measured before publication of the original version of this document (ATSC A/85:2009), per the methods documented in a predecessor version, or as permitted by A/53, Part 5 [1], need not be measured again.

1.2 Organization

This document is organized as follows:

- **Section 1** – Outlines the scope of this document and provides a general introduction.
- **Section 2** – Lists references and applicable documents.
- **Section 3** – Provides a definition of terms, acronyms, and abbreviations for this document.
- **Section 4** – Explains the technical background of the AC-3 multichannel audio system.
- **Section 5** – Explains audio Loudness measurement based on techniques defined by ITU-R Recommendation BS.1770 [3].
- **Section 6** – Makes recommendations for exchange of Content without metadata.
- **Section 7** – Provide guidelines for the practical use of agile and fixed metadata within production, distribution, and transmission environments.
- **Section 8** – Describes methods to effectively control Loudness transitions between Short-form and Long-form Content.
- **Section 9** – Examines key issues relating to dynamic range management.
- **Section 10** – Specifies the setup of sound systems for DTV, including the alignment of control room monitor systems to a reference sound pressure level.
- **Annex A** – Program Loudness: Provides background on the BS.1770 [3] measurement algorithms for Loudness and True Peak.
- **Annex B** – Room Acoustics and Loudspeaker Placement: Discusses basic principles of audio control-room monitoring.

- **Annex C** – Room Correction: Examines issues relating to the interaction of sound from loudspeakers and the room.
- **Annex D** – Quick Reference Guide on Reference-Monitor Setup for Television Audio: provides guidance on setting the sound pressure level (SPL) reference for TV audio monitoring rooms.
- **Annex E** – Loudness Ranges: Examines the range of Loudness within which listeners will accept Loudness changes within and between Content items.
- **Annex F** – AC-3 Dynamic Range Control (DRC) Details.
- **Annex G** – AC-3 Metadata Parameters.
- **Annex H** – Quick Reference Guide for TV Station and MVPD Engineers on Loudness Management.
- **Annex I** – Quick Reference Guide for Audio Mixers and Editors Creating Content.
- **Annex J** – Requirements for Establishing and Maintaining Audio Loudness of Commercial Advertising in Digital Television.
- **Annex K** – Requirements for Establishing and Maintaining Audio Loudness of Commercial Advertising in Digital Television When Using Non-AC-3 Audio Codecs.
- **Annex L –** Guidelines for Establishing and Maintaining Audio Loudness of Internet Streaming Services When Using Metadata-based and Non-metadata-based Codecs.
- **Annex M –** Loudness and True Peak Quick Reference.

2. REFERENCES

All referenced documents are subject to revision. Users of this Recommended Practice are cautioned that newer editions might or might not be compatible.

2.1 Informative References

The following documents contain information that may be helpful in applying this Recommended Practice.

[1] ATSC A/53:2010, “ATSC Digital Television Standard, Part 5 – AC-3 Audio System Characteristics,” Advanced Television Systems Committee, Washington, DC, 6 July 2010.

[2] IEEE/ASTM SI 10-2002, “Use of the International Systems of Units (SI): The Modern Metric System,” Institute of Electrical and Electronics Engineers, New York, NY

[3] Recommendation ITU-R BS.1770-5, “Algorithms to measure audio programme loudness and true-peak audio level,” International Telecommunications Union, Geneva, Switzerland, 2012.

[4] ATSC A/52:2012, “Digital Audio Compression (AC-3, E-AC-3) Standard,” Advanced Television Systems Committee, Washington, DC 17 December 2012.

[5] IEC 60268-17 (1990), “Sound system equipment. Part 17: Standard volume indicators,” International Electrotechnical Commission, Geneva, Switzerland.

[6] SMPTE RP 200, “Relative and Absolute Sound Pressure Levels for Motion-Picture Multichannel Sound Systems,” Society of Motion Picture and Television Engineers, White Plains, NY.

[7] SMPTE 302M-2007, “Television – Mapping of AES3 Data into MPEG-2 Transport Stream,” Society of Motion Picture and Television Engineers, White Plains, NY, 2007.

[8] SMPTE 2020-1-2008, "Format of Audio Metadata and Description of the Asynchronous Serial Bitstream Transport"; SMPTE 2020-2-2008, "Vertical Ancillary Data Mapping of Audio Metadata – Method A"; SMPTE 2020-3-2008, "Vertical Ancillary Data Mapping of Audio Metadata – Method B," Society of Motion Picture and Television Engineers, White Plains, NY, 2008.

[9] Recommendation ITU-R BS.1534-3 (10/2015), "Method for the subjective assessment of intermediate quality level of coding systems," International Telecommunications Union, Geneva, Switzerland.

[10] Soulodre, G.A., Lavoie, M.C., and Norcross, S.G: "The Subjective Loudness of Typical Program Material," 115th Convention of the Audio Engineering Society, 2003.

[11] Soulodre, G.A.: "Evaluation of Objective Loudness Meters," 116th Convention of the Audio Engineering Society, Berlin, Germany, Preprint 6161.

[12] Recommendation ITU-R BS.775-4 (12/2022), "Multichannel stereophonic sound system with and without accompanying picture," International Telecommunications Union, Geneva, Switzerland.

[13] Soulodre, G.A., and Lavoie, M.C.: "Stereo and Multichannel Loudness Perception and Metering," 119th Convention of the Audio Engineering Society, San Francisco, Preprint 6618.

[14] AES TD-1001, "Multichannel sound systems and operations," Audio Engineering Society, New York, NY.

[15] Recommendation ITU-R BS.1771-1, "Requirements for loudness and true-peak indicating meters," International Telecommunications Union, Geneva, Switzerland.

[16] Eargle, J. M., "Audio Monitoring in Contemporary Post-Production Environments," SMPTE Journal, "Society of Motion Picture and Television Engineers, White Plains, NY, January 2005.

[17] Holman, Tomlinson, "Surround Sound Up and Running," Focal Press, 2nd edition, 2007.

[18] AES3-2009 (r2019), "Digital input-output interfacing - Serial transmission format for two-channel linearly represented digital audio data," Audio Engineering Society, New York, NY.

[19] "Model DP569 Dolby Digital Multichannel Audio Encoder User's Manual," Dolby Laboratories, Inc., San Francisco, CA.

[20] "Dolby DP570 Multichannel Audio Tool User's Manual," Dolby Laboratories, Inc., San Francisco, CA.

[21] AES17-2020, "AES standard method for digital audio engineering - Measurement of digital audio equipment," Audio Engineering Society, New York, NY.

[22] SMPTE RDD 6-2008, "Description and Guide to the Use of the Dolby E Audio Metadata Serial Bitstream," Society of Motion Picture and Television Engineers, White Plains, NY.

[23] SMPTE 337-2008, "Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface," Society of Motion Picture and Television Engineers, White Plains, NY.

[24] ATSC A/79, "Conversion of ATSC Signals for Distribution to NTSC Viewers," Advanced Television Systems Committee, Washington, D.C., 12 December 2008.

[25] CTA-CEB11-D, "NTSC/ATSC Loudness Matching," Consumer Technology Association, Arlington, VA, September 2019.

[26] ANSI C16.5-1942, "Standard Volume Indicator," American National Standards Institute, Washington, DC.

[27] IEC 60268-17, "Sound system equipment, Part 17: Standard volume indicators," International Electrotechnical Commission, Geneva, Switzerland, 1990.
[28] IEC 60268-10/IIa, IIb: 1991, "Sound system equipment, Part 10: Peak programme level meters," International Electrotechnical Commission, Geneva, Switzerland, 1991.
[29] ANSI S1.4-1983 (R 2001) with Amd.S1.4A-1995, "Specification for Sound Level Meters," American National Standards Institute, Washington, DC.
[30] Commercial Advertisement Loudness Mitigation (CALM) Act, Pub. L. 111-311, 47 U.S.C. § 621 (2011).
[31] AES Engineering Brief Number 38, "The Effect of Downmixing on Measured Loudness," Scott G. Norcross and Michel C. Lavoie, Audio Engineering Society, New York, NY, 2011.
[32] Riedmiller, J., S. Lyman, and C. Robinson: "Intelligent Program Loudness Measurement and Control: What Satisfies Listeners?," AES Pre-print, 115th Convention, October 2003, Audio Engineering Society, New York, NY.
[33] Schroeder, Manfred: "The Schroeder Frequency Revisited," Journal of the Acoustical Society of America, vol. 99, no. 5, pp. 3240–3241, Acoustical Society of America, Melville, NY.
[34] Report ITU-R BS.2054-4 (2014), "Audio levels and loudness," International Telecommunications Union, Geneva, Switzerland.
[35] AES TD-1009, "Improving Dialogue Intelligibility in Media," Audio Engineering Society, New York, NY, November 2025.
[36] EBU R 128 s2, "Loudness in Streaming," European Broadcasting Union, Geneva, Switzerland, November 2023.
[37] AES 71:2018 (R2023), "Recommended Practice Loudness Guidelines for Over the Top Television and Online Video Distribution," Audio Engineering Society, New York, NY, November 2025.
[38] ANSI/CTA-2075, "Loudness Standard for Over-the-Top Television and Online Video Distribution for Mobile and Fixed Devices," Consumer Technology Association, Arlington, VA, January 2020.

3. DEFINITION OF TERMS

With respect to definition of terms, abbreviations, and units, the practice of the Institute of Electrical and Electronics Engineers (IEEE) as outlined in the Institute's published standards [2] are used. Where an abbreviation is not covered by IEEE practice or industry practice differs from IEEE practice, the abbreviation in question will be described in Section 3.3 of this document.

3.1 Compliance Notation

This section defines compliance terms for use by this document:

vital – This word indicates a course of action to be followed strictly (no deviation is permitted).

should – This word indicates that a certain course of action is preferred but not necessarily required.

should not – This phrase means a certain possibility or course of action is undesirable but not prohibited.

3.2 Treatment of Syntactic Elements

This document contains symbolic references to syntactic elements used in the audio, video, and transport coding subsystems. These references are typographically distinguished by the use of a different font (e.g., restricted), may contain the underscore character (e.g., sequence_end_code) and may consist of character strings that are not English words (e.g., dynrng).

3.3 Acronyms and Abbreviations

The following acronyms and abbreviations are used within this document.

AAC – Advanced Audio Coding, a digital audio coding format described in MPEG-2 audio (ISO/IEC 13818-7) or MPEG-4 (ISO/IEC 14496-3) (NOTE: MPEG-4 defines loudness/DRC metadata that is not universally supported.)

AC-3 – A digital audio coding format used in ATSC 1.0, as described in ATSC A/52 [4]

AC-4 – A digital audio coding format as described in ATSC S/342 Part 2 and ETSI TS 103 190 Parts 1 and 2

AGC – Automatic Gain Control

dB – Decibel

dBFS – Decibels, relative to full-scale sine wave (per AES17 [21])

dBTP – Decibels, True Peak relative to full-scale

DBS – Direct Broadcast Satellite

DRC – Dynamic Range Control

DTS-UHD – A digital audio coding format described in ETSI TS 103 491

E-AC-3 – Enhanced AC-3, a digital audio coding format used in ATSC 1.0, as described in ATSC A/52 [4]

GPI – General Purpose Interface

MP2 – MPEG-1/2, Audio Layer II, a digital audio coding format described in ISO/IEC 11172-3 and ISO/IEC 13818-3

MP3 – MPEG-1/2), Audio Layer III, a digital audio coding format described in ISO/IEC 11172-3 and ISO/IEC 13818-3

MPEG-H Audio – A digital audio coding format described in ATSC A/342 Part 3 (officially MPEG-H 3D Audio) and ISO/IEC 23008-3

IP – Internet Protocol

ITU – International Telecommunication Union

LFE – Low Frequency Effects

LKFS – Loudness, K-weighted, relative to full-scale, measured with equipment that implements the algorithm specified by BS.1770 [3]. A unit of LKFS is equivalent to a decibel.

MPEG – Moving Pictures Expert Group

MVPD – Multichannel Video Programming Distributor (includes DBS service Operators, local cable system Operators, and cable multiple system Operators)

PPM – Peak Program Meter

SPL – Sound Pressure Level in decibels referenced to 20 $\mu N/m^2$

vMPVD – Virtual Multichannel Video Programming Distributor (e.g., MVPD operating over the Internet)

VU – Volume Unit [5]

xHE-AAC – A digital audio coding format using MPEG-D USAC Extended High-Efficiency AAC profile (ISO/IEC 23003-3) and MPEG-D DRC loudness control profile and dynamic range profile (ISO/IEC 23003-4)

3.4 Terms

The following terms are used within this document.

Agile Metadata – Audio metadata values, including dialnorm, which can change at Content boundaries.

Anchor Element – The perceptual Loudness reference point or element around which other elements are balanced in producing the final mix of the Content, or that a reasonable viewer would focus on when setting the volume control.

BS.1770 – Recommendation ITU-R BS.1770 [3]. This specifies algorithms that provide numerical values indicative of the perceived Loudness and True Peak levels of the audio Content that is measured. Loudness meters and measurement tools that have implemented the BS.1770 algorithm will report Loudness in units of LKFS, and True Peak levels in units of dBTP.

Comfort Zone – the Comfort Zone is a range (+2.4 dB, -5.4 dB) of the change to audio Loudness that was found to be acceptable to a sample of listeners per [32]. The 0 dB point on the Comfort Zone scale is the average Target Loudness value or dialnorm of the channel.

Content –TV material distributed by an Operator.

dialnorm – An AC-3 metadata parameter, defined in ATSC A/52 [4], that is carried in the AC-3 bit stream. It is used to indicate how far the average Dialogue Level of the Content is below 0 LKFS. Valid values are unsigned integers within the range of 1 to 31. Loudness values outside this range cannot be expressed by dialnorm. The value of dialnorm is numerically equal to the absolute LKFS value of the Dialogue Level. The value of 0 is reserved.

Dialogue Gating – A process in which a signal gate is opened, allowing audio to pass and be measured when speech/dialogue is detected in the audio signal. This is used instead of the relative-level gate defined in BS.1770 when making dialogue-based Loudness measurements. It may also be called Speech Gating.

Dialogue Level – The Loudness, in LKFS units, of the Anchor Element[2]. When referring to an AC-3 encoder parameter, the term **Dialog Level** is used.

Dialogue Loudness – The Loudness, in LKFS units, of the dialogue element of a soundtrack.

Dolby E – Dolby E is a multichannel audio data-rate reduction technology designed for use in contribution and distribution, which also conveys Dolby E metadata.

Dolby E metadata – Metadata that is multiplexed into the Dolby E bitstream. Each metadata element is classified as either a Professional or Consumer type of metadata. Dolby E metadata is documented in SMPTE RDD 6 [22].

Downstream – A point in a distribution chain containing assembled Content where some Content boundaries may not be readily identifiable using automated methods.

DRC Profile – A collection of parameters that describe how Dynamic Range Control metadata is calculated.

File-based Scaling Device – A device used to apply an overall gain correction to audio Content stored as files.

[2] The term "Dialogue Level" is based on dialogue's widespread use as the Anchor Element for mixing of Content.

Framesync – Short term for "frame synchronizer."

Full-program Mix – The entire audio contents of a soundtrack, including all elements and audio channels (per BS.1770 [3], except LFE).

Integrated Loudness – The average Loudness in units of LKFS when measuring the entire duration of the Content (from start to finish).

Layback – A postproduction technique in which audio Content is rejoined and synchronized with its associated video Content after editing, mixing, or "sweetening."

Long-form Content – Episodic TV programming, movies, news, sporting events, or other non-Short-form Content. (Duration is typically greater than approximately three minutes, although shorter durations may apply for some Streaming TV services.)

Loudness – A perceptual quantity representing the magnitude (or "volume") of the physiological effect produced when a sound of a certain intensity stimulates the ear.

Low Frequency Effects – An optional single channel of limited bandwidth (below 120 Hz). It allows high sound pressure levels to be provided for low-frequency sounds and is typically intended to be fed to the subwoofer in multichannel audio playback systems.

Measured Loudness – The magnitude of an audio signal when measured with equipment that implements the algorithm specified by BS.1770 [3]. It is an approximation of perceived Loudness.

Mixing Level – An optional metadata parameter in the AC-3 bit stream that allows indication of the absolute sound pressure level calibration of the mixing studio that produced the Content.

Operator – A TV network, broadcast station, streaming service, DBS service, local cable system, cable multiple system operator (MSO) or other MVPD/vMVPD.

Programmatic Advertising – Automated downstream insertion of Short-form Content.

Short-form Content – Commercial advertising ("spots"), public-service announcements (PSAs), promotional material ("promos"), political advertising, emergency information, news bulletins, and other discrete insertions into or around Long-form Content. Also termed "interstitial" Content. (Duration is typically less than approximately three minutes.)

Speech Gating – See Dialogue Gating.

Streaming – General term for real-time delivery of IP-encoded media content via the Internet.

Target Loudness – A specified value for the Anchor Element (i.e., Dialogue Level), established to facilitate Content exchange from a Content supplier to an Operator.

True Peak – The maximum absolute level of the signal waveform in the continuous time domain, measured per BS.1770 [3]. Its units are dBTP (meaning decibels relative to nominal 100%, True Peak).

4. THE AC-3 MULTICHANNEL AUDIO SYSTEM

The ATSC AC-3 audio system is intended to deliver a reproduction of the original (unprocessed) Content at the output of the AC-3 decoder in a receiver, normalized to a uniform Loudness. It provides the capability to allow each listener the freedom to exert some control over the degree of dynamic range reduction, if any, that best suits their listening conditions. The dynamic-range processing part of the system is described in Section 9, but its operation is predicated on having properly normalized Content delivered to it.

The metadata parameter dialnorm is transmitted to the AC-3 decoder along with the encoded audio. The value of the dialnorm parameter indicates the Loudness of the Anchor Element of the Content. The dialnorm value of a very loud program might be 12, and of a soft one, 27. There is an attenuator at the output of the AC-3 decoder that applies appropriate attenuation to normalize all Content Loudness to -31 LKFS.

If the dialnorm parameter accurately reflects the overall Loudness of the Content, then listeners will be able to set their volume controls to their preferred listening levels and will not have to adjust the volume when the audio changes between Content types. If the system is used properly, the Loudness will also be consistent when changing TV channels.

Section 7 describes three methods of using audio metadata: Fixed, Preset, and Agile. Any one of these approaches will deliver consistent Loudness to the listeners; Operators are free to use the method that best suits their operational practices. Whichever approach is selected, the system's proper operation depends on the transmission of a value of dialnorm that correctly represents the loudness of the Anchor Element (typically, the Dialogue Level) of the Content, which, in turn, depends on accurate Loudness measurements of the transmitted audio Content.

5. LOUDNESS MEASUREMENT

Because Loudness is a subjective phenomenon, human hearing is the best judge of Loudness. When combined with a known mixing environment and a consistent monitoring level (such as that described in Section 10 of this document), experienced audio mixers using their sense of hearing can produce well-balanced sound with remarkably consistent Loudness. If all programs and commercials were produced at a consistent average Loudness, and if the Loudness of the Content were preserved through production, distribution, and delivery chains, listeners would not be subjected to annoying changes in Loudness within and between programs.

When measuring audio signals, there are two key parameters of interest: the True Peak level of the signal and its Loudness. The True Peak measurement enables the mixer to protect the program from clipping, and the Loudness measurement allows the mixer to protect the listener from annoying variations in Loudness. Although mixers balance mixes using their ears, an objective Loudness measurement helps to maintain consistent average Loudness within and between programs.

Practices have shifted from using VU and PPM metering to Loudness and True-Peak measurements using the ITU-R BS.1770 algorithms [3], with results reported in units of LKFS and dBTP.

This RP provides guidance that, if followed, will result in consistency in Loudness and avoidance of signal clipping. The specified measurement techniques are based on the Loudness and True-Peak measurements defined by ITU-R Recommendation BS.1770 [3]. The details of both measurements are in Annex A.

5.1 Overview of the BS.1770 Loudness Measurement Method

Loudness is measured by integrating or averaging the frequency- and channel-weighted power of the audio signals in all channels over the duration of the measurement. The general structure of the algorithm is shown in Figure 5.1.

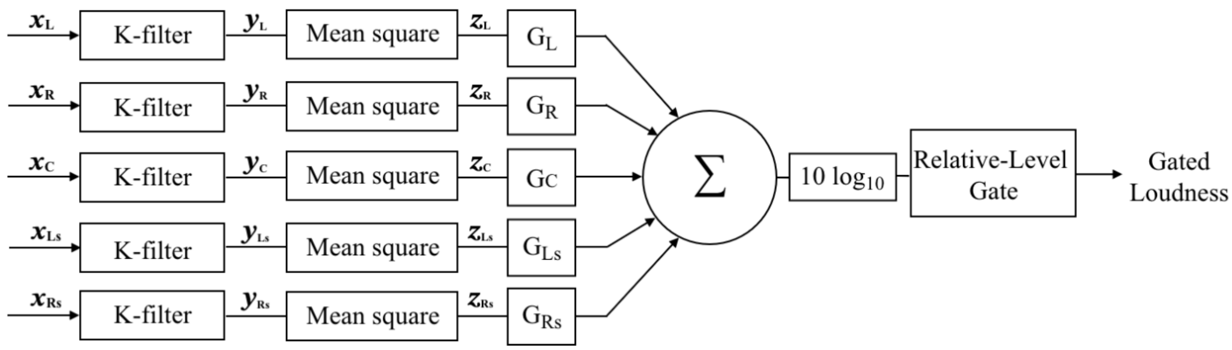

Figure 5.1 BS.1770 Loudness algorithm.

The BS.1770 method was validated in listening tests by comparing its results to the relative subjective Loudness of mono, stereo, and multichannel program material. Measured Loudness is reported as LKFS. A unit of LKFS is the same measure as a decibel. A -15 LKFS program can be made to match the Loudness of a quieter -22 LKFS program by attenuating it by 7 dB.

5.1.1 Understanding the Evolution and Essential Role of ITU-R BS.1770

Recommendation ITU-R BS.1770 [3] has been revised numerous times since its initial 2006 publication as BS.1770-0. Most notably, BS.1770-2, published in 2011, added a relative-level gate (as well as an absolute-level gate), making the algorithm more useful by excluding low-level signals or periods of silence that could skew the integrated loudness measurement. It is essential, however, that any BS.1770 version with a relative-level gate **not be used** as a substitute for the **dialogue**-based-gate that **is** necessary to measure the average loudness of Long-Form Content (see Section 5.2).

The BS.1770-3, -4, and -5 versions all have identical algorithms for Loudness and True-Peak measurement of stereo and 5.1 content. Therefore, any Loudness meter compliant with ITU-R BS.1770-3, BS.1770-4 or BS.1770-5 is usable to meet this document's recommendations on Loudness measurement of Short-form Content (see Section 5.2.5).

For Loudness measurements of Long-form Content, a **dialogue-gated**, BS.1770-1 meter—or, if unavailable, a BS.1770-2 or later meter with dialogue gating on and, ideally, relative-level gating off—should be used.[3] See Section 5.2.1.

True-Peak measurement in the original BS.1770-0 included optional elements that were removed in BS.1770-3 and subsequent versions to provide more consistent True-Peak measurements. Therefore, True-Peak measurements should ideally be performed using the True-Peak method defined in BS.1770-3 or later.

5.2 Overview of Making Loudness Measurements

The ATSC A/53 Digital Television Standard [1] mandates that the value of the dialnorm parameter present in the AC-3 (or E-AC-3) elementary stream indicates the Loudness of the encoded audio Content.

The goal for correctly managing Loudness is to measure and label the average Loudness of each piece of Content, regardless of the Content's dynamic range. The use of this measured

[3] Metering products may label their presets in different ways. Regardless of preset names, users should verify that the algorithms described here are used, in order to ensure accurate and consistently compliant measurements.

average Loudness value to normalize the Content helps eliminate annoying changes of Loudness from one piece of Content to another and is intended to enable listeners to enjoy programming with minimal adjustment of their volume controls.

The Anchor Element is the component of the sound mix that typical viewers focus on when setting their volume control. Dialogue forms the Anchor Element in the majority of media Content. During postproduction, it is usually possible to isolate the dialogue and to measure its Loudness accurately, regardless of the overall dynamic range or style of the finished program. If dialogue is **not** the Anchor Element of the Content (e.g., a music program), there is always some element the mix engineer uses as the Anchor Element around which the Loudness of all the other program elements is balanced. The Anchor Element must be measured and reported as the Loudness of Long-form Content.

The Anchor Element is typically dialogue for Long-form Content and is <u>always</u> Full-program Mix for Short-form Content. The loud and soft portions of a mix create the dynamics of the program around the Anchor Element's Loudness.

Measurement of dialogue Loudness for Long-Form content is especially critical with wide-dynamic-range Content (e.g., cinematic-style mix). If this type of content is normalized using Full-program-Mix Loudness, dialogue will typically be perceived lower in Loudness and not transition correctly to Short-form Content. Therefore, it is recommended to use measurement of dialogue Loudness for Long-form Content in general.

5.2.1 Additional Considerations

- When measuring Long-form Content, momentary or short-term Loudness measurements (as these terms are defined in BS.1771 [15]) **are not expected to remain** at one constant value. Loudness variations during portions of the Content that do not represent the Anchor Element may deviate from the target value. These variations are acceptable due to the intended dynamic range of properly mixed Content.
- Loudness Range (LRA) measurement [34] is often available in loudness-measurement tools. While LRA may be helpful to quantify dynamic range of both Full-program Mix and dialogue-based measured Content, its effectiveness in practice is uncertain. Therefore, LRA values **should not be enforced** in Content specifications.
- Similarly, while dialogue-intelligibility tools may be useful in Content creation, their resultant metrics **should not be enforced** in Content specifications. For more information on dialogue intelligibility, see [35].
- For cases in which it is impractical to measure the Anchor Element, see Section 5.2.4.
- Once finished Content moves downstream from production, subsequent Loudness measurements might not duplicate the measurements made during the production process. See Section 5.2.10.
- It should be noted that Content measured before the original effective date of this document, per the methods documented in a predecessor version or as permitted by ATSC A/53, Part 5 [1], need not be measured again.

5.2.2 Long-form Content Loudness Measurement During Postproduction

The average dialogue Loudness must be measured and reported as the Loudness for Long-form Content. This should be performed using BS.1770 (without the relative-level gate[4]) **along with a dialogue-gating algorithm** and measured over the entire duration (not act by act) of the Content's composite mix. (Versions of BS.1770 without relative-level gating can be referred to as BS.1770-1). This practice yields a dialogue-gated Loudness measurement that represents the Loudness of Long-form Content during postproduction and downstream. It is vital that this average dialogue Loudness value is reported as the Loudness of the Long-form Content.

For cases where it is impractical to measure the average dialogue Loudness, see Section 5.2.4.

5.2.3 Long-form Content Loudness Measurement During Real-Time Production (Live Event)

The intent of Loudness measurements made during a live event is to guide the mixer to produce the Content at an average predetermined or Target Loudness. The principle of measuring the Loudness of the Anchor Element of the Content also applies to live productions but is done in real time as the production progresses.

Some BS.1770 Loudness meters offer a sliding window or moving-average measurement mode that indicates the Loudness of a prior segment (typically 3 to 10 sec) of the program sound. Other implementations may use some kind of graphical representation of BS.1770-based Loudness to complement a simple Loudness figure. These meters may include a dialogue-gating algorithm as described in Section 5.2.1. Such modes should be used to help guide the mixer to the Target Loudness. Using these measurements as a guide to confirm Loudness is particularly useful when mixing in noisy locations, or when a consistent monitor level cannot be maintained.

Alternatively, if dialogue gating is not available on a meter, the mix engineer may acquire a valid Loudness measurement when speech, not whispered or shouted, is present and utilize this value as the Loudness of the anchor element.

A running dialogue-gated measurement of the live production from its start to any point thereafter may also provide an indication of the average Loudness of the anchor element of the Content.

5.2.4 Quality Control Measurement of Finished Long-Form Content Loudness

As in Section 5.2.1, during quality control, when only the finished composite mix is available, the average dialogue Loudness must be measured and reported as the Loudness for Long-form Content. This should be performed using BS.1770 (without the relative-level gate[5]) along with a speech-gating algorithm and measured over the entire duration (not act by act) of the Content's composite mix. (Versions of BS.1770 without relative-level gating can be referred to as BS.1770-1).

If measuring the average dialogue of the entire program as described above is not possible, a section of the Content that is representative of the Anchor Element (typically dialogue) should be isolated and measured and reported as Loudness of the Long-form Content.

[4] A relative-level gate was added to BS.1770-2 and subsequent versions. This is not a substitute for adding a speech-gating algorithm to BS.1770-1.

[5] A relative-level gate was added to BS.1770-2 and subsequent versions.

5.2.5 Measuring Loudness for Long-Form Content Containing Minimal or No Dialogue

In the case where the Content contains minimal dialogue, *any* component of dialogue (not whispered or shouted) should provide an anchor reference to which typical viewers will set their volume control. Any such dialogue component should be measured and reported as the Loudness of the Long-form Content. For automated processes, a minimum example value of 8% speech detection within the 5.1-channel soundtrack can be used, and this dialogue Loudness value can be effective for any stereo version, as well.

In the *complete* absence of dialogue, a Full-program-Mix measurement can be considered, however, regardless of the process used, **the goal is to normalize to a Loudness value that *perceptually* matches typical dialogue-based Content.** Report this as the Loudness of the Content.

Note that if wide-dynamic-range Content (e.g., cinematic-style mix) is normalized using Full-program-Mix Loudness, the Content typically will be perceived as softer than expected overall, and not transition correctly to Short-form Content.

Note also that *automated* processes measuring Content with a complete absence of dialogue may not yield the desired *perceptually* matched Loudness value. Therefore, it may be necessary to *alert Content recipients* to ensure no downstream Loudness normalization to the *specified* target value is performed on Content that has been *perceptually* matched. This process will ensure creative intent is preserved and reduces the need for listeners to readjust their volume control.

5.2.6 Measurement of Short-Form Content

A measurement of the long-term integrated (or average) Loudness of the Content's **Full-program Mix** over its full duration should be reported as the Loudness of all Short-form Content.

This measurement method provides an acceptable Loudness match between narrow dynamic range, Short-form Content and wide dynamic range, Long-form Content that is normalized using a dialogue-based Loudness measurement.

Those choosing to create and deliver wide-dynamic-range, Short-form Content (e.g., movie trailers) should note that the louder elements of this type of material will increase the Loudness measured with a Full-Program Mix measurement, and consequently reduce the dialog Loudness after normalization. This can cause an unacceptable match to Long-form material measured with an dialog-based method.

5.2.7 Downmix Loudness Measurement and Monitoring

During production, the Loudness of surround programming should be measured in both its surround mix format and in its stereo (2.0-channel) downmix. This is necessary because of the high percentage of consumers experiencing the downmix of surround programming, and the possible Loudness disparity between the two formats.

The downmix Loudness measurement should be performed using the same downmix metadata coefficients as carried in downstream AC-3 bitstreams or used by discrete downmix equipment. This approach will identify any "Loudness build-up" in the stereo downmix.

Loudness build-up is typically caused by Content mixed in-phase to the three front channels of the 5.1-channel mix, which can result in downmix Loudness exceeding that of the surround mix by up to 3 dB. One method of avoiding this problem during production is to assign dialogue and lead-singer elements primarily or exclusively to the center channel only [31]. Alternatively, the surround mix may be adjusted in production/postproduction to mitigate Loudness build-up in the resultant downmix.

5.2.8 File-Based Measurement

The measurement methods described above also apply to Content that exists as digital files. In addition, file-based storage makes it practical to automate the Loudness measurements and to examine the dialnorm value (if any) that may have been assigned to the Content. The Content's Loudness may be adjusted, if necessary, to the desired or Target Loudness value by applying an overall gain correction[6], or the dialnorm value may be rewritten to match the measured Loudness of the Content.

5.2.9 Special Considerations for Measurement of Content Downstream

Once finished Content moves downstream from production (e.g., to MVPDs or broadcast station transmission facilities), it may not be possible or practical to measure or verify the Loudness that should correspond to the dialnorm value. Moreover, it might not be possible to distinguish between Long-form Content and Short-form Content, or to determine boundaries between Long-form and Short-form Content. Consequently, Loudness measurements made downstream may cross over Content-type boundaries.

Once finished Content moves downstream and determination of boundaries between Content segments is not practical, Loudness measurements may be made by direct use of formula (2) of BS.1770 [3]. Such measurements should not always be expected to precisely replicate measurements made as recommended in other sections of this Recommended Practice.

6. TARGET LOUDNESS AND TRUE PEAK LEVELS FOR CONTENT DELIVERY OR EXCHANGE

For delivery or exchange of Content without metadata (and where there is no prior arrangement by the parties regarding Loudness), the Target Loudness value should be -24 LKFS. Measurement tolerance of up to approximately ±2 dB around this value is anticipated, due to measurement variations. Content Loudness should **not** be targeted to the high or low side of this tolerance.

The True Peak level should be kept below -2 dBTP in order to provide headroom to avoid potential clipping due to downstream processing (such as audio coding used in delivery). Minor measurement tolerance of up to approximately ±0.5 dBTP around this value is anticipated, due to meter variations, and is acceptable. True Peak measurements should be performed prior to encoding, and it should be noted that the True Peak values may be different post-decode.

7. METADATA MANAGEMENT CONSIDERATIONS IMPACTING AUDIO LOUDNESS

An AC-3 encoder allows the setting of up to 28 metadata parameters concerning the characteristics of the accompanying audio in the bit stream (see Annex G). The parameters can be classified into three groups:

Informational metadata, which includes seven optional parameters that can be used to describe the encoded audio. These parameters have no effect on encoding or the decoded listening experience in the home.

Basic control metadata, which includes 19 parameters that determine the dynamic range compression, down-mixing, matrix decoding, and filtering used in certain operating modes of the professional encoder and consumer decoder. Optimizing the setting of these parameters for each program may enhance the listening experience under varying listening conditions and

[6] If the Content is stored in rate-reduced (coded) format, it may be necessary to decode the Content, alter its level, and then re-encode the Content.

with certain Content types. However, default values may be used without detriment to the listening experience.

Critical control metadata, which includes two parameters that are critical for proper encoding and decoding:

- **Channel mode** (acmod), which should be chosen correctly to engage proper channel formatting in the decoder to match the Content. Improper use of this parameter may alter a transmission and cause the loss of dialogue when encoding a 5.1 program. An example of such improper use is encoding a 5.1-channel soundtrack with Channel mode metadata set for 2.0 channels.
- **Dialog Level** (dialnorm), which ATSC A/53 [1] requires to be set correctly to prevent (potentially severe) Loudness variation during Content transitions on a channel, and when changing DTV channels. Incorrect dialnorm values can lead to variations in Loudness as large as 30 dB.

Apart from the parameters above, default values may be used for most of the other metadata parameters with acceptable results. Once mixers and producers become more familiar with these parameters by monitoring using available emulation systems, they can select values that further optimize the presentation of their program Content.

7.1 Importance of dialnorm

Carriage of and correct setting of the value of dialnorm is mandatory for DTV broadcasting in the United States (see ATSC A/53:2010, Section 5.5, “Dialog Level” [1]).

This RP identifies methods to ensure consistent DTV Loudness through the proper use of dialnorm metadata for all Content, and to thereby be compliant with A/53. Many of the principles for successful management of dialnorm may also apply to the management of other AC-3 metadata parameters.

As indicated in Section 6, measurement tolerance of up to approximately ±2 dB is anticipated, and this may lead to minor variations between the value of dialnorm and the actual program Loudness. Such minor variation is acceptable (due to the Comfort Zone – see Annex E); however, Operators should **not** intentionally operate at the high or low side of this tolerance.

7.2 Metadata Management Modes

The requirement for accurate dialnorm, channel mode (acmod), and other metadata can be met in three different ways, at the discretion of the operator:

- **Fixed metadata**: The AC-3 encoder Dialog Level is “fixed” to a single dialnorm value, and the Content Dialogue Levels are conformed to that setting.
- **Preset metadata**: AC-3 encoder “presets” are programmed, each with different dialnorm values and engaged via a “General Purpose Interface” (GPI) or other control interface.
- **Agile metadata**: The AC-3 encoder is configured to receive external metadata. An upstream “agile” dialnorm metadata system may be used to deliver changing dialnorm values to the encoder, corresponding to the changing Loudness at the Content boundaries.

When managed properly, all three methods provide a compliant and acceptable end result for the consumer. The majority of the discussion in this section of this RP focuses on the dialnorm parameter. Readers are encouraged to refer to Annex G and to research information on how the remaining metadata parameters may impact coding.

It is also possible for the Operator to apply a hybrid approach, choosing one of the methods for Loudness management and a different method for the remainder of the metadata: e.g., maintaining a fixed dialnorm value but switching channel mode as required.

7.3 Using Fixed dialnorm Metadata

The concept of fixed dialnorm is simply to “fix” the AC-3 encoder dialnorm setting to a single value within a network or broadcast system and to bring the Loudness of the encoder audio input signal into conformance with this setting. The Operator can choose any dialnorm value from 31 to 1; however, compliance with ATSC A/53 [1] requires the Operator to employ a value equal to the Loudness of all encoded audio Content. See Figure 7.1.

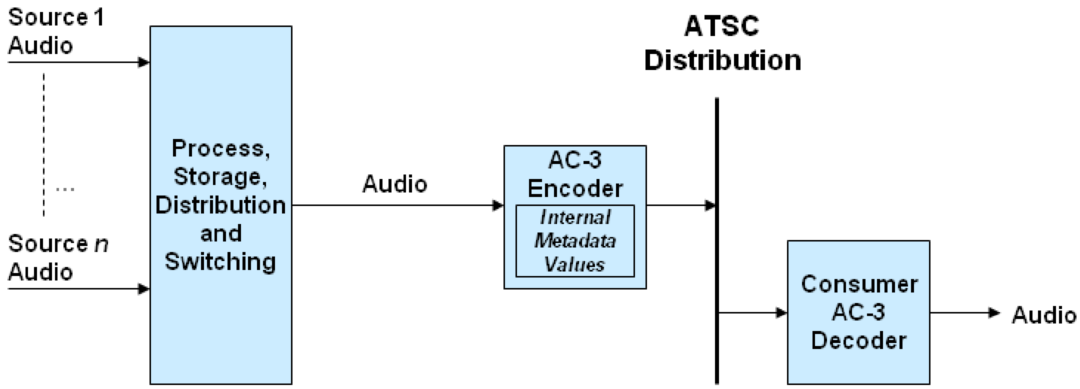

Figure 7.1 Fixed metadata concept.

7.3.1 Setting dialnorm by Long-Term Averaging Method

An Operator can achieve a first approximation of compliance with ATSC A/53 [1] by measuring the long-term average Loudness of the Operator’s audio output and setting the AC-3 encoder dialnorm parameter equal to this value. The averaging period should be chosen to include all types of Content. If the Dialogue Level of individual pieces of Content deviate significantly from that long-term average, however, the dialnorm parameter will not properly reflect the Dialogue Level of that Content. This situation should be addressed by the program originator or Operator and corrected (see Sections 7.3.2 and 7.3.3.). This method may not apply to Operators using Content with intentionally wide dynamic range.

7.3.2 Setting dialnorm for Production

A Content delivery specification should specify the Target Loudness for all Content. This establishes the anchor for layering of music and effects for the soundtrack. Content should be measured with a meter using the BS.1770 recommendation [3] to confirm the average Loudness of dialogue. The Content supplier should indicate the actual average Loudness with the deliverable. Cooperation between Content supplier and recipient is necessary to achieve successful Loudness management when implementing this practice.

7.3.3 Content Not Conforming to the Target Loudness

If the Operator needs to make use of Content not conforming to the established Target Loudness value, an offsetting gain or loss will need to be inserted to compensate. If the difference is unknown, it will be necessary to measure the Content Loudness before compensation is applied.

7.3.4 dialnorm and Loudness Quality Control

To ensure the proper match between dialnorm value and Loudness, the Operator should make use of Loudness metering during the quality-control process, and when necessary request redelivery or make compensating adjustments to ensure the Content meets the Target Loudness.

7.3.5 Emission dialnorm Setting for Compliance with A/53

An Operator receiving Content that is delivered at a Target Loudness, where there is no gain adjustment or processing after reception of the Content by the Operator, should set the value of dialnorm in the emission AC-3 encoder to match the originator's specified Target Loudness (often specified by contract, Content-delivery agreement, signal specification document, etc.). If any fixed gain or loss is applied downstream in the signal chain, the AC-3 encoder dialnorm value should be offset accordingly from the originator's Target Loudness.

For example, if the originator delivers audio with a Target Loudness of -24 LKFS, and no audio level gain or loss is incurred in the Operator's signal chain, dialnorm would be set at 24 in the AC-3 emission encoder[7]. However, if a gain of 3 dB is added in the signal chain, this would increase the Loudness of the Content to -21 LKFS, and dialnorm at the AC-3 emission encoder would therefore be set to 21. If instead an attenuation of 2 dB is introduced, this would reduce the Content's Loudness to -26 LKFS, and dialnorm would be set to 26.

If an automatic Loudness-control processor is utilized, the AC-3 encoder dialnorm value should be set to the average Dialogue Level set (and measured) at the output of the audio processor.

7.3.6 Fixed dialnorm Advantage

A fixed dialnorm system poses minimal risk to the Content. Fixed dialnorm has the advantage of simplicity, with no requirement for additional metadata equipment or data management. This approach can be used with every AC-3 encoder, and it is the only approach possible when using an encoder without metadata input or external GPI control.

7.4 Using Preset dialnorm Metadata

If the Operator needs to accommodate a small number of discrete changes to the dialnorm value or other metadata parameters, some AC-3 encoding systems can be configured to change between preset metadata values via external control (e.g., with a contact closure to a GPI). This method requires GPI external triggers for accurate preset signaling from the automation playlist or switcher. It is often used to switch between stereo and 5.1 encoding modes, even when the dialnorm remains fixed at a single value. See Figure 7.2.

[7] As noted in Section 3.4, dialnorm is expressed as an unsigned integer.

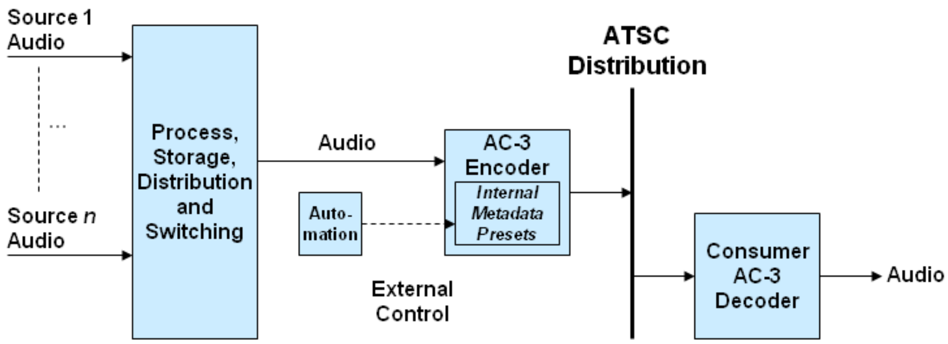

Figure 7.2 The preset metadata concept.

7.4.1 Implementation

The implementation of preset metadata is similar to "fixed" metadata. Predetermined preset values are loaded into the AC-3 encoder to accommodate known differences in Content Loudness. Compliance with ATSC A/53 [1] then requires that Content be delivered with Loudness matching one of the preset values, and that the automation system be programmed to change presets according to different Content Loudness values and channel modes.

7.4.2 AC-3 Framesync Requirement

Some AC-3 encoders reset and disrupt the audio bit stream output when a preset is changed. Depending on the ATSC encoder being used, this may result in an audible "glitch" on air. To avoid this potential problem, it may be necessary to provide an AC-3 framesync for the output of the AC-3 encoder to stabilize the AC-3 source for the ATSC encoder.

7.5 Using Agile dialnorm Metadata

An agile metadata system allows setting different dialnorm values for different Content that has different Loudness. This is accomplished by embedding the dialnorm parameter within the metadata bit stream accompanying the Content at an "upstream" location. The metadata is de-embedded just prior to the AC-3 encoder and connected to its external serial metadata input. The dialnorm setting changes appropriately on boundaries of the Content. See Figure 7.3.

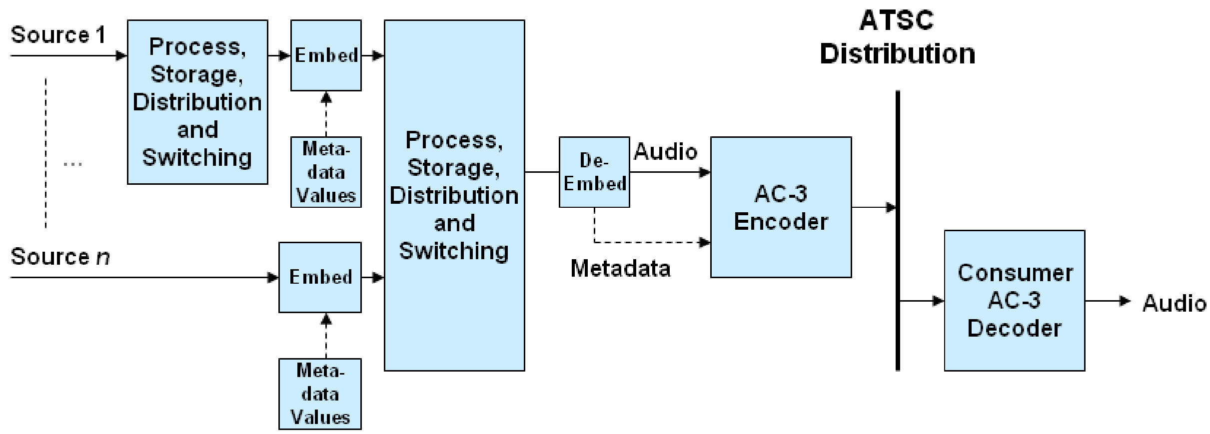

Figure 7.3 The agile metadata concept.

7.5.1 System Deployment

When the agile metadata approach is used by a network Operator, it will need to be employed throughout the plant of every broadcast station or MVPD head-end that receives Content from the network. This requires deployment of complex encoding and decoding equipment at all input, output, monitoring, and processing points in the distribution chain, from the metadata origin through to all AC-3 encoders. It is essential that the agile metadata reach the AC-3 encoder. Several approaches for agile metadata delivery and storage are available, and can be used separately or in combination. The AC-3 metadata (as detailed in Annex G) is a consumer subset of Dolby E metadata, as described in SMPTE RDD 6 [22]. This may be transported over serial data links, as vertical ancillary (VANC) data, or as data carried in compressed bitstreams. It may also be stored in file-based systems.

7.5.1.1 Dolby E Metadata Over Serial Link

Dolby E metadata in its baseband form may be carried via serial links. This approach may require a dedicated serial layer that remains carefully time-aligned to the audio and video signals.

7.5.1.2 Dolby E Metadata in VANC

Dolby E metadata can also be embedded within the VANC of standard- or high-definition serial digital systems and extracted downstream using the SMPTE 2020 [8] standard. This approach may require multiplexers and de-multiplexers, and requires support by video storage, encoding, processing, and distribution equipment with the ability to pass the VANC signal intact. (Some storage devices have limited or no VANC capability.)

7.5.1.3 Metadata and Codecs

Certain systems used for backhaul, distribution, and storage applications also have the ability to carry Dolby E metadata. These systems include the Dolby E compressed bitstream itself, and other proprietary formats[8], which require specialized audio encoders and decoders. They also may require equipment that can offset video timing to compensate for the encoding and decoding latency introduced. Most professional digital video equipment can be configured to pass these encoded signals through standard digital audio channels that comply with SMPTE 337 [23].

[8] E.g., Linear Acoustic *e-squared* format and the Ericsson *Quad Phase Aligned* format.

7.5.1.4 File-based Metadata

There are a large number of file-based techniques for storing Dolby E metadata—some of them standardized, and some of them proprietary. These are outside the scope of this document.

7.5.2 Production Technique – Live

In live production using an agile metadata approach, the TV production mix engineer selects a specific but arbitrary Loudness target for each program, with considerations for dynamic range, headroom, and the type and mood of the program. This parameter establishes the Loudness anchor for layering of music and effects for the soundtrack. Depending on the deliverable specification, the Target Loudness is either carried in the encoded signal by the value of the dialnorm parameter or communicated to the distributor for subsequent encoding.

7.5.3 Production Technique – Non-Real-Time

With post-produced Content using the agile metadata approach, the final mix Loudness is determined either during program production or after it is complete. Depending on the deliverable specification, the Target Loudness is either carried in the encoded signal by the value of the dialnorm parameter or communicated to the distributor for subsequent encoding. Insertion of metadata in the deliverable can be accomplished through a dubbing process or by making use of the pre-read feature available in some video tape recorders. See Figure 7.4.

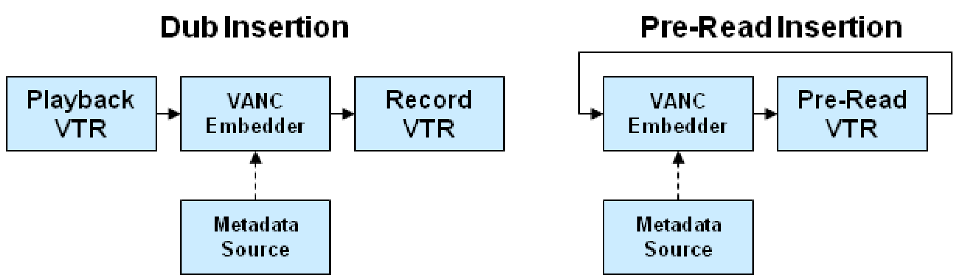

Figure 7.4 Metadata insertion options.

7.5.4 Production Monitoring

The soundtrack should be measured using the BS.1770 recommendation [3] to confirm that the average Loudness for the entire length of the production matches the chosen dialnorm value (see Section 5).

7.5.5 Semi-Agile Metadata

An Operator may use an agile metadata system but choose to simplify metadata authoring and insertion operations by specifying fixed Target Loudness values to be used by Content providers.

7.5.6 Impact of Metadata Loss on Content

A risk associated with the use of an agile metadata system is the potential for a severe discrepancy in Loudness between programs and between TV channels if metadata is lost. All AC-3 encoders with external metadata input provide a “reversion” feature to mitigate the impact of metadata loss. With this feature, the encoder can be configured to either retain the most recent metadata value or revert to an Operator-defined preset. While this feature can minimize the impact upon the consumer, the error in Loudness or other metadata parameters (such as channel mode) can still be significant. The reversion parameter should be chosen to minimize the impact of metadata loss on the presented Content.

7.5.7 Fixed-Agile Hybrid

In some instances, an Operator may choose to intentionally use the reversion feature to accommodate Content without metadata.

It is critical that Operators choose appropriate settings for all metadata parameters of the reversion preset, particularly ensuring that the Loudness of the distributed Content without metadata matches the pre-determined reversion dialnorm parameter. Reversion may also be used to protect against a loss of metadata, recognizing that the reversion metadata parameters may not exactly match that of the Content. In the event of a metadata loss, all Content being encoded under reversion will be subject to these parameters. It is especially critical that channel mode be set in a fashion to protect all Content under any circumstance. The inadvertent use of 2.0-channel mode with 5.1 Content will eliminate channels 3-6 of the encoded audio and put the Content at risk.

7.5.8 Advantages of Agile Metadata

An agile system presents the most flexibility for the Content provider without imposing creative limitations.

8. METHODS TO EFFECTIVELY CONTROL LOUDNESS TRANSITIONS BETWEEN LONG-FORM AND SHORT-FORM CONTENT

The ATSC 1.0 DTV audio system (AC-3), with its expanded dynamic range and techniques for managing Loudness, presents the possibility of undesirable Loudness changes at transitions to and from various pieces of Content, if not managed properly. This condition is known to annoy the audience by frequently forcing the listener to adjust the audio levels at transitions to maintain a comfortable volume. This condition can be alleviated when proper DTV Loudness management is applied.

AC-3 incorporates the necessary technology to mitigate variations in Loudness during Long-form to Short-form Content transitions. These techniques are described below:

8.1 Effective Solutions

Large Loudness variation during transitions can be effectively managed by ensuring dialnorm properly reflects the Loudness of all Content:

8.1.1 For Operators Using a Fixed dialnorm System

a) Ensure that all Content meets the Target Loudness and that long-term average Loudness matches the dialnorm value.
b) Employ a file-based scaling device to match long-term Loudness of non-conformant file-based Content to the Target Loudness.
c) Employ a real-time Loudness processing device to match the Loudness of non-conformant real-time Content to the Target Loudness.

See Section 7.3 for further details.

8.1.2 For Operators Using an Agile dialnorm System

a) Ensure that during program production, postproduction, or ingest, Content is measured (see Section 5.2) and labeled with the correct dialnorm value matching the actual Loudness of the specific Content.
b) Employ a file-based measurement and authoring device to set dialnorm to the average Loudness of the specific Content.

c) Employ a real-time processing device to match Content to a specific Loudness. Apply a dialnorm value, matching the Loudness of all Content processed by this device.

See Section 7.5 for further details.

8.2 Adverse Conditions

Notable conditions that may adversely impact transitions at Content boundaries are as follows:

- Long-form Content suppliers often increase dramatic impact by using program dynamics and manipulating Loudness to achieve a desired audience effect. This is sometimes done at the end of Long-form Content segments transitioning to Short-form Content.
- An extreme variation outside of the Comfort Zone (see Annex E) may cause a listener to adjust the volume to compensate for the large, temporary change in Loudness. When subsequent Short-form Content plays, the listener may need to readjust the volume yet again, to achieve an acceptable setting for the Short-form Content.

8.3 Summary Recommendations

Recommendations to lessen the negative impact of Loudness variation during Long-form to Short-form Content transitions include the following:

1) For fixed metadata systems, ensure properly targeted average Loudness of all Content.
2) For agile metadata systems, ensure that dialnorm authoring matches the measured Content Loudness for all Content elements.
3) Create awareness of Content suppliers concerning the potential use of excessive dynamic range and the possibility for listener complaints concerning transitions with large Loudness variations via the following methods:
   a) Document this condition in any Content delivery specifications an Operator requires of a supplier:
      i.) Describe negative impact to the audience created by mixing outside of the tolerance of the listener Comfort Zone (see Annex E) when going into and coming out of breaks.
      ii.) The Operator may rely on the Comfort Zone (see Annex E) for guidance on acceptable Loudness changes across the boundaries between Long-form and Short-form Content.
      iii.) Describe the expected Loudness of Short-form Content for the program supplier to create awareness of this situation and its potential for listener dissatisfaction and negative impact on the show.
   b) Ensure the use of the proper sound pressure level (SPL) in rooms used to mix and monitor Content soundtracks:
      i.) Refer Content suppliers to Section 10 of this RP endorsing the use of a suitable SPL monitoring level during Content postproduction. The selected SPL monitoring level should be appropriate for the size of the mixing room with consideration for the listening environment of the typical DTV audience. Lower SPL monitoring levels in postproduction monitoring environments yield louder mixes and more contained, appropriate dynamic range than louder mixing environments. A properly selected postproduction or playback monitoring environment is essential to establishing appropriate mixing levels for DTV.

ii.) Consider specifying a maximum True Peak value for the soundtrack. This practice constrains dynamic range by reducing headroom. It permits the audience to adjust the overall volume level with less risk of large Loudness variation.

4) The AC-3 DRC system should not be relied upon to control Long-form to Short-form Content Loudness variations The AC-3 DRC system should not be relied upon to control Long-form to Short-form Content Loudness variations.

8.4 Downstream Short-form Content Insertion

In the case of downstream insertion of Short-form Content, the Operator should ensure that the Loudness of the downstream insertion matches the dialnorm setting of the Content stream into which it is inserted.

If the originator's Content is decoded to baseband, the Loudness of the decoded audio should be measured, and the value of the re-encoder's AC-3 dialnorm value should be set to match the measured Loudness at the next stage of encoding. In this case either the Operator should modify the originator's Loudness to match the target value of the Operator's system, or the originator's Loudness value (as measured) should be used to set the dialnorm value in the next stage of AC-3 encoding. At this re-encoding stage it is critical that the other audio metadata parameters also be set appropriately and consistently.

9. DYNAMIC RANGE MANAGEMENT

The DTV audio system is capable of delivering very wide dynamic range (the range between the softest and loudest sounds). Content producers often take advantage of dynamic range as one of the methods to convey artistic intent.

However, there could be a conflict between the desire of the Content producer to deliver Content with wide dynamic range and the audience who cannot, or choose not to, enjoy the wider dynamic range. This could be caused by the inability of the viewer's equipment to reproduce the desired range of sounds, or the lack of an environment suitable to the enjoyment of the wide dynamic range. Thus, the goals of preserving the original dynamic range of the Content and satisfying viewers can often be at odds.

A goal of the AC-3 system is to provide Content producers with the greatest freedom and flexibility in the choice of DRC when producing Content. The AC-3 system conveys these DRC options to the viewer, where DRC system will interact with the viewer's input in a known and repeatable fashion.

There are several methods for controlling dynamic range. Some methods apply prior to audio encoding, some apply after decoding, and some span both domains:

1) One approach is traditional compression and/or limiting where gain control is applied to the audio prior to encoding.
2) Another approach is to use the AC-3 coding system, which generates gain control words during encoding but does not apply the gain control to the audio until after decoding, allowing users to optionally choose how much dynamic range they desire.

The primary difference between the two approaches is that the AC-3 approach (option 2 above) is "reversible," and the other approach is not. A hybrid of the two methods is also possible, allowing for some permanent and some "reversible" processing to be combined, in a balance determined by the broadcaster.

9.1 AC-3 Dynamic Range Control System ("Reversible")

To allow the Content to be enjoyed by the most viewers in the widest variety of listening environments, a mechanism to optionally restrict the dynamic range of the Content is designed into the DTV audio system. In simple terms, the AC-3 Dynamic Range Control (DRC) system can be considered a compressor/limiter that is split in half; the control signal is generated in the broadcast encoder, and the control signal can be applied in the consumer decoder. With this system, the audio itself is not modified from the original until it reaches the consumer's decoder, so the dynamic range control may be considered "reversible." Details of how DRC works are provided in Annex F.

Metadata is always transmitted as part of the coded audio bit stream. AC-3 encoders require certain metadata parameters to be specifically set. DRC is one of these parameters, and dialnorm, or the indicated Loudness parameter, is another. While separate from dialnorm, correct operation of the DRC system is very dependent on the proper measurement of Content and indication of the Loudness of that Content. See Section 5 of this document for measurement guidelines.

The DRC system modifies the dynamic range of the decoded Content by reducing the level of very loud portions of the Content to avoid annoying the viewer, and by raising the level of very quiet portions of the Content so that they are better adapted to the listening environment. If, however, the viewer has a home theatre or other listening environment that can reproduce the full dynamic range of the Content, the viewer can choose to turn off the DRC system, allowing that viewer to enjoy the Content with the full dynamic range as it was originally mixed.

When the DRC system is turned off by the viewer, the dynamic range of the Content will be identical to the input to the AC-3 encoder. Because the application of the DRC system is selectable by the viewer, the AC-3 system can provide multiple listening options for viewers, with the ability to accommodate different reproduction systems and listening environments.

9.1.1 Line Mode and RF Mode DRC

In the AC-3 audio encoder, two control signals are calculated by the DRC system:

1) One control signal is used to slightly reduce the dynamic range of the Content (suitable for directly connected audio components).
2) The other control signal is used for more aggressive dynamic range control, suitable either for equipment connected with analog RF-modulated signals (i.e., channel 3/4 antenna inputs), or for equipment that is otherwise incapable of handling the potential wide dynamic range of the original signal.

The two types of DRC control signals are calculated because there are two different scenarios where gain reduction/expansion may be desired.

One situation is where a gentle reduction of dynamic range is desired. This is default mode of many devices that that use "baseband" analog interconnections (often available on red and white "RCA" connector stereo audio outputs.) These connections are not a professional interface, but are often an acceptable consumer interface. This DRC control signal is called Line Mode in the encoder control interface in most professional products, because it usually refers to the line level analog audio interconnection process in the consumer's listening system. The name of this metadata parameter is dynrng.

A second, more aggressive DRC control signal is simultaneously calculated. It is designed to be used in products that could be interconnected by an analog RF-modulated signal. The dynamic range of these RF-modulated signals is usually less than signals using the line-level interconnection process. RF mode also adds 11 dB of signal boost for Loudness matching purposes with legacy analog systems. This mode is the default for MVPD set-top boxes as well as TVs with built-in DTV tuners and small built-in speakers. The name of this metadata parameter is compr. See CTA-CEB11 [25] for more information about Loudness matching recommendations for consumer products. (The recommendations in this document and in CTA-CEB11 were coordinated and are complementary.)

Either control signal will be selected as the default mode by the consumer equipment manufacturer, depending on the design and intended use of the product. For viewers who desire the full dynamic range of the original signal, these control signals could be ignored, resulting in the reproduction of the Content exactly as the Content producer created it. This method of using separate control and audio signals results in a "reversible" DRC system that is capable of simultaneously satisfying different viewers with different listening environments and situations.

9.1.2 Monitoring

There are benefits to using a system where the DRC "gain words" (see Annex F) are calculated in the encoder and applied at the decoder. One benefit of this type of system design is that it allows the compressor/limiter functions to be accurately previewed, or emulated, during production, well before the Content is encoded. Producers can check Content as it is being produced to hear how it will sound using pre-established dynamic range modes.

While the effects of this DRC system will be audible to mixers in a professional monitoring environment, the process should be considered in the context of the typical consumer in a typical home environment, where such gain reduction and expansion processes are usually not noticeable. In most situations, the effects of Line Mode DRC generally improve the portrayal of the Content by better fitting the audio within the reproduction capabilities of the viewer's equipment and listening space.

9.1.3 Relation to dialnorm

All DRC calculations are relative to and based on the indicated Loudness of Content as represented by the dialnorm metadata parameter. In other words, there is interaction between these two metadata parameters, in that the encoder needs to know how loud the Content is intended to be, so it can determine when the Content is either "too loud" or "too quiet." Dialnorm effectively sets this target. Therefore, it is very important to proper DRC operation that dialnorm accurately indicates the Loudness of the Content.

Dialnorm is also used to set the threshold of a somewhat hidden and inescapable overload protection process designed to protect downmixed audio from overloading consumer equipment. Overload protection has ballistics appropriate for preventing overload, but is far less than ideal for audio quality. It is good practice to avoid overload protection. This can be accomplished by ensuring that the dialnorm parameter accurately reflects the actual Loudness of the audio Content.

9.1.4 Professional Encoding

In AC-3 encoders, the gain reduction and expansion characteristics of the RF and Line DRC modes are determined by a group of DRC "profiles." These profiles describe many parameters, including the gain reduction (or "cut") range, gain expansion (or "boost") range, and attack and release times. Between these ranges is a linear range (or "null zone") where no gain reduction or expansion takes place. It is expected that the majority of professionally mixed Content will reside within the "null" range, where the Content will be delivered exactly as produced, with no additional (or continual) modification by the DRC system. Excursions beyond this null band might be used to convey a specific artistic intent. Note that the dialnorm parameter determines the position of this null band in the DRC system. It is essential that the dialnorm parameter accurately indicate the Loudness of the Content.

There are five profiles defined in the AC-3 encoder. The profiles are:

- Music Light
- Music Standard
- Film Light
- Film Standard
- Speech

The differences between Music and Film DRC choices may be subtle to a typical listener, but one or the other may be better applicable to certain types of Content. This can be best determined by monitoring with an appropriate emulator. The "Light" versions of the profiles have a much wider null area. Thus, gain reduction or expansion begins farther away from average program audio, resulting in less gain reduction or expansion than with the "Standard" version of the profile.

The Speech profile is, as the name would suggest, intended for programs that contain only speech (a "talk radio" format, for example.) This profile might introduce noticeable DRC artifacts on programs with music and effects. Please refer to Annex F for more information about the DRC profiles.

9.1.5 AC-3 DRC: Choosing "None"

There is also a choice called "None" that does not select any of the named DRC profiles. The selection of "None" (by the Operator) inhibits the generation of DRC control words.

"None" is an acceptable choice as long as the ramifications of **not** choosing a DRC profile are understood, as follows:

- The reversibility feature of the DRC system will not be available to the consumer.
- Selecting "None" prevents the viewer from selecting a DRC choice or enabling features such as "Late Night" or "Midnight," in some consumer equipment that use the RF DRC mode.
- Dynamic range should be controlled in another fashion by the Operator or by the program originator.
- There is a possibility that DTV sets with limited volume capacity will exceed their reproduction range.
- The RF mode's DRC control words are also used for protection limiting. Protection limiting prevents clipping in consumer decoders, which could result from an inappropriate dialnorm setting combined with very dynamic programming. Choosing "None" will not prevent the generation of protection-limiting DRC control words, which are very aggressive and not very artistic. The protection-limiting process has a very short attack time and a very long release time, and might cause objectionable audible artifacts.
- Systems using ATSC signals as a source for SD distribution (e.g., SD analog cable tiers) will not be able to use the RF mode DRC to establish an acceptable analog SD signal with reduced dynamic range. See ATSC Recommended Practice A/79 [24] for additional guidance.

In order for the AC-3 DRC system to be functional, a profile other than "None" should be used by Operators when appropriate.

9.2 Dynamic Range Processing With Metadata Interface ("Reversible")

Audio metadata processors have been developed that allow creation of different Line Mode and RF Mode gain values that can be stored in user-defined presets or profiles. Additional processing can be performed in the metadata domain, bringing the results much closer to traditional audio processing, but with the benefit of being "reversible." These technologies operate in tandem with an AC-3 encoder, measuring audio and calculating the gain values, then passing them to the encoder for insertion into the AC-3 bit stream. The application is similar to the original Line and RF features described above, but uses different profiles to satisfy different dynamic-range processing goals.

These alternative profiles allow for an adjustable ratio of permanent DRC processing versus metadata DRC processing. At one end of adjustment, permanent multiband audio processing is applied to the audio after it is first pre-conditioned by applied metadata to minimize the degree of permanent processing that might be required.

The resulting audio is AC-3 encoded with a fixed dialnorm value, and one of the traditional DRC profiles can be selected. At the other end of the adjustable ratio, audio that has been pre-conditioned by applied metadata is then analyzed, and DRC gain words for Line Mode and RF Mode are generated per Operator selection.

Traditional settings such as attack, release, threshold, AGC range, gate, and freeze can be adjusted very much like a traditional (i.e., non-metadata-enabled) audio processor, but the results are instead included in the AC-3 bit stream along with the encoded audio. This allows for processing that may be more aggressive than the original five profiles. The original audio is delivered to the consumer, and the DRC values will be applied to the audio by default, but can also be bypassed by the viewer if desired.

9.3 Dynamic Range Processing Without Metadata Interface ("Irreversible")

In analog AM radio broadcasting, automatic gain control (AGC) systems were created to ensure proper modulation of a broadcast station's carrier. Incorrect modulation could lessen coverage if the average level was too low and could cause interference if the peak level was too high. This was made more difficult by FM systems, which employ pre-emphasis to increase the level of high frequencies prior to transmission to minimize noise. Essentially these AGC systems (often termed "modulation controllers") were then controlling the peak-to-average ratio of program audio and thus controlling dynamic range. Unlike the metadata-based approaches described above, gain changes are immediately applied to the input signal prior to transmission to consumers, and the results are therefore irreversible.

The earliest AGC systems were wideband, using a single level detector controlling a single gain element to control the entire audio bandwidth. In the early days of broadcasting, this may have been an acceptable practice because audio bandwidth was restricted, and better audio consistency was achieved by skilled Operators. However, as audio bandwidth increased and Operator intervention decreased, additional techniques had to be developed to maintain control and minimize audible side-effects of the AGC process.

A possible artifact of wideband AGC is that the level of one part of the audio spectrum can be affected by another unrelated part of the audio spectrum that has more power. This can be illustrated by imagining an audio signal comprised of a low frequency bass drum and a higher frequency flute. A simple wideband AGC would vary the control signal—and thus the level of all frequencies—in response to the higher energy of the bass drum. This level change might be appropriate for the bass drum, but would cause inappropriate fluctuation in the level of the flute signal, an effect commonly referred to as “pumping” or “breathing.” The development of multiband processing systems significantly improved this issue since the input signal is divided into two or more frequency bands, and a separate AGC is applied to each of the bands. Multiband systems allow more control with less interaction between different parts of the audio spectrum, but can also change the spectral balance of the applied signal.

Additional techniques have evolved to minimize the side effects such as “pumping,” overload due to pre-emphasis, and increasing background noise. These include multi-stage processing, look-ahead processing, clipping, and gating. Multi-stage processing involves placing two or more AGC sections of differing speeds in series, allowing peaks to be controlled separately from control of average level. Look-ahead processing separates detection and adjustment portions of the AGC by a delay so that gain changes are applied at precisely the instant required. Clipping is a process where the signal peaks are truncated and may result in audible distortion. Gating prevents very low-level signals and noise from being unnecessarily increased by the AGC, and can be used along with thresholds of other stages to create a no-processing null zone. Taken together, all of these techniques go far beyond the original goals of simply protecting against under or over-modulation, and in many cases have been used to produce a sound that is very different from the original.

The goal of modern DTV dynamic-range control systems is different than its analog predecessors. The elimination of the need for pre-emphasis dramatically reduces the degree of processing necessary. Newer devices may also employ more sophisticated detection schemes based upon or compatible with BS.1770 [3]. If the desire is to simply manage Loudness with minimal impact on program Content, experience has shown that multiple stages of light processing can be very effective while still maintaining much of the program’s audio integrity. Additionally, Long-form to Short-form Content boundary issues can be improved with the use of irreversible dynamic range processing, not part of the AC-3 system[9].

9.4 Consumer Experience

Viewers may interact with the DRC process within AC-3 bit streams in several ways. Generally, the MVPD set-top box and DTVs with integrated tuner/decoders will default to a specific DRC mode. Usually this will be RF Mode, as these devices either contain RF modulators or feed TVs directly that may only support minimal dynamic range. Often the cost and flexibility of the equipment will determine how many other choices the viewer might have.

Home theatre A/V receivers will often have the most DRC listening choices, but the menus used to navigate these choices may lack consistency and “interoperability” with other products. In this context, interoperability means that terminology used to describe options may differ or conflict with terminology used in other devices. For example, the concept of ‘wide,” “standard” (or “normal”), and “narrow” are commonly used. In this context, “wide” refers to no DRC being engaged, resulting in the original, wide dynamic range being reproduced. “Standard” engages the profile selected for line mode during encoding. “Narrow” engages the profile selected for RF mode during encoding. As noted earlier, RF mode adds 11 dB of signal boost for Loudness-matching purposes with legacy analog systems.

10. AUDIO MONITORING SETUP

This section documents the recommended setup of sound systems for DTV operations, including the alignment of control-room monitor systems to a reference sound pressure level corresponding

[9] Manufacturers of devices designed for this purpose should be consulted for more information.

to a particular electrical audio-signal level, and the equalization of such sound systems. Monitoring conditions for a variety of spaces used for DTV production and postproduction are shown in Table 10.1.

This section is not intended to replace SMPTE RP 200 [6] for cinema sound monitoring in larger spaces.

This section documents the recommended original setup conditions for monitoring, as well as measurements and adjustments that may be undertaken from time to time for quality control of monitoring for spaces used for DTV production and quality control. This section is intended for planners, design engineers, installation engineers, maintenance engineers, and users of such facilities.

A "Quick Reference Guide" for setting monitor levels is provided in Annex D.

> Note: The test signals referenced in this document are from the TMH Test Disc Series distributed by The Hollywood Edge, and are copyrighted. Free use of the copyright is granted to television and video studios, networks, and postproduction houses for alignment purposes. All other rights reserved. The TMH Test Disc Series makes available many other test signals for electrical and acoustical testing.

10.1 Background

As noted in Section 1.1, consumers do not expect large changes in audio Loudness from Long-form to Short-form Content, and from channel to channel. Two level-related features are built into the AC-3 system specified in the ATSC standard: dialogue normalization and dynamic range control. If these are set and used correctly, they go a long way towards achieving the goals of maintaining the artistic integrity of mixes while delivering a dynamic range suitable for the conditions of the consumer. The use of a standardized monitor layout, standard electrical and acoustical reference levels, and reference spectrum response by all Long-form and Short-form Content providers makes for a common basis in monitoring. Along with the features built into the AC-3 system, practicing the recommendations in this section of the RP should help satisfy the mutual needs of producers and consumers.

A universal observation is that the same sound pressure level is perceived as louder in small rooms, such as control rooms, than in large rooms, such as cinemas. The reference sound pressure level in this document has been tested for interchangeability with SMPTE RP 200 [6] employed in large spaces.

The use of reference conditions for monitoring has been shown to improve interchangeability of Long-form and Short-form Content mixes, and thus is desirable.

10.2 Characteristics of Rooms and Spaces

Five categories of sound control rooms and postproduction spaces are defined. They are listed in Table 10.1, which shows the characteristics for each type.

Table 10.1 Categories of Audio Control Rooms used in Television Production

Category	Characteristics
I	Principal audio monitoring control rooms with specialized acoustics and sound systems. Channel range up to 5.1 (3 front/2 surround/0.1 low frequency enhancement). Well isolated from other operations. Widest frequency and dynamic ranges equal to best home cinemas properly aligned. This type of room may be used for quality control at the network level, for example checking program material for conformance to

	delivery requirements when a question arises at ingest stations. Sound monitor quality dominates over production requirements in this category of room. Broadcast organizations might be expected to have only a small number of such rooms.
II	Audio-mostly production spaces with equipment needs and placement supplanting absolute audio monitoring conditions, although audio monitoring is still expected to be good. Channel number equal to highest number used for material originating in the room. Good isolation from other operations. This type of room may be used for program origination, with its output occasionally subject to check in a Category I room. Low-frequency range and headroom may be somewhat restricted compared to a Category I room.
III	Audio editing spaces, premix, and prelay rooms, and other spaces the output of which is typically expected to be integrated into programs in a Category II room or better. If used for final mixing, apply the level and equalization recommended practice herein.
IV	Trucks and booths for program mixing. These spaces have special considerations due to their small room volume, high background noise level, high level of early reflections, and communication needs in a production environment.
V	Headphone monitoring systems recommendations. Used for ingest stations in crowded environments, quality control in machine rooms, and the like.

10.3 Installation

Sound systems should be installed insofar as practical to the recommended practices in reference documents [12] [14] [16] [17] in Type I, II, and IV spaces. Type III spaces should also fulfill the requirements of the recommended practices if they are to be used for final mixing.

Attention is drawn to the requirement for equal-distance spacing of the loudspeakers from the primary mix monitoring location, or, if that is impractical, the need to use time delay to make the time of arrival at the mix location constant among the loudspeaker channels. This is necessary due to the very great precision of human hearing in locating phantom images appearing between adjacent pairs of loudspeakers. This requirement applies especially across the front channels, and if the surround loudspeakers are closer than the front loudspeakers to the listening location, they should be delayed.

The reference documents show the expected anechoic frequency response of monitor loudspeakers and systems. In-situ measurements of loudspeakers in control rooms, however, show strong deviations from the anechoic response of the loudspeakers, in particular due to room boundary loading conditions at very low frequencies, with standing wave modal effects through the range typically from about 80 Hz to 500 Hz. For this reason, room equalization is highly desirable to the point of necessity for higher quality spaces. A description of the concerns for equalizing rooms is given as Annex C.

It has been observed that the widest deviation in octave-to-octave balance of TV audio mixes occurs at the frequency extremes, i.e., below 100 Hz and above 8 kHz. This probably occurs due to the varying responses of monitor systems in these ranges. For this reason, the operational room response curve given in Figure 10.1 applies to monitoring in Type I spaces. Type II and IV spaces may have somewhat restricted low-frequency range and headroom compared to Type I spaces.

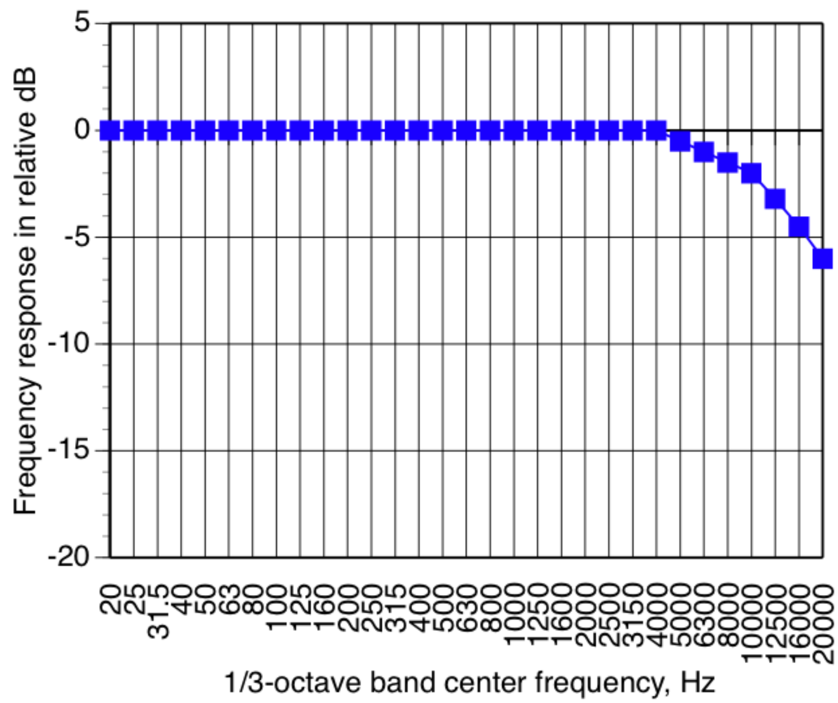

Figure 10.1 Operational room electro-acoustic response curve.

Figure 10.1 is a quasi-steady-state measurement to be made with small, low-diffraction microphone(s) at and around the principal listening position for each main channel (Left, Right, Center, Left Surround, and Right Surround) in turn. Further information is given in Annex C.

Attention is also drawn to the level alignment of the LFE or 0.1 channel. Some confusion occurs due to its requirement to have 10 dB of in-band gain compared to the main channels. This does not mean that it should measure 10 dB greater in sound pressure level than the main channels when calibrating. Because the LFE channel's bandwidth is more limited than that of the main channels, its level playing a pink noise source of the correct spectrum and electrical level will be approximately 4 dB greater than any one of the main channels.

10.4 Reference Level Calibration

The procedure for calibrating reference sound pressure level varies in level and method across the various categories, with the objective of making programs interchangeable across a range of listening conditions. The steps below should be performed as follows:

Step 1. Perform electrical alignment of the system under calibration using the following test:

- Signal 1. The 440 Hz sine wave tone at -20 dBFS recorded on the left channel of the stereo file linked below. Import this into a digital audio workstation and duplicate it for each channel.
  https://www.atsc.org/wp-content/uploads/2020/04/440Hz_left_Ch-20dB.wav
- Use this tone to align the output meters of the equipment in use for -20 dBFS on digital meters. Use a meter with adequate resolution. It is best if the playback device and the console input, channel, and output meters, can all be set to precisely unity gain.

- On analog meters, this level is meant to be set to 0 VU (ANSI C16.5 [26], IEC 60268-17 [27]), to 4 on the scale of 7 of the BBC peak meter (IEC 60268-10/IIa, IIb [28], although note that this meter displays 4 dB/step, so that the maximum calibrated value corresponds to -8 dBFS; this problem is ameliorated by the fact that the PPM has an attack time of 80 percent reading in 10 ms, much slower than a True-Peak meter), and to corresponding values for other variant meters.
- See Section 10.5.1 for more information on this test signal.

Step 2. (Skip to Step 3 if no sound level meter is available.) Perform acoustical level alignment of the system under calibration using the following test:

- Signal 2. 500 Hz – 2 kHz band limited pink noise at -20 dBFS recorded on the left channel of the stereo file linked below. Import it into a digital audio workstation and copy it to each of the Left, Right, Center, Left Surround, and Right Surround channels, one at a time. Maintain the unity gain structure established with the sine wave tone. See Section 10.5.3 for more information on this test signal.
  https://www.atsc.org/wp-content/uploads/2025/06/MidRngPinkNoise_-20dB.wav
- Ignore console or other electrical meters, as they will read differently on noise than on tone, since the peak level of this stochastic (i.e., random) noise source is higher than its RMS level, which in turn is higher than its average level. On peak meters, the range will be around 10 dB higher than the sine-wave tone, whereas on a VU meter (which is an average-responding RMS-calibrated instrument per IEEE C16.5-1954[10]) the level will read about 1 dB low.
- Apply this signal to each channel in turn and adjust the appropriate controls affecting the monitor level only for the standard acoustical level (see below). In any one given installation, the items that could affect monitor acoustical level are the monitor output level control on the console, which is suggested to be set at a marked standard value, any room/loudspeaker equalizer level controls, and any power-amplifier gain controls or powered-loudspeaker input sensitivity controls. Ordinary issues of headroom and signal-to-noise ratio through a chain of multiple units also apply.

The standard acoustical level should be measured using a sound-level meter, preferably meeting the standard ANSI S1.4 [29] Type 2, using slow reading (1s integration time) and C-weighting. It should be measured at the position of the center of the head of a listener seated at the main mixing location, with the meter oriented towards the channel under calibration. Note that the body of the person making the measurement should be to one side of the meter, not behind it relative to the source, where the reflection from the body might affect the reading.

The standard acoustical reference level for each category and for various room volumes within the categories is given in Table 10.2.

[10] The Standard Volume Indicator was first standardized in an American Standard originally published in 1942 [26], revised in 1954, rewritten as an IEEE/ANSI standard in 1991, and withdrawn in 1999. The ANSI designation C16.5 no longer exists.

Table 10.2 Reference Sound Pressure Level

Categories	Room Volume in Cubic Feet	SPL in dB re: 20 μN/m^2
I, II	> 20,000	85*
	10,000 < 19,999	82
	5,000 < 9,999	80
	1,500 < 4,999	78
	< 1,499	76
III	Depends on room usage. For editing purposes, may be controlled by the editor for use with the material at hand. For final program mixing, follow the recommendations for categories I, II above.	
IV	< 1,500	76
V		Use 2 cc coupler and set 440 Hz level to 76 dB.
* Per SMPTE RP 200 [6]		

Step 3. Dub (duplicate) the original speech file downloaded (see link below) using unity gain into the center or mono channel of the playback device in use. In the case where only two-channel stereo monitoring exists, dub the original speech into left and right channels with an attenuation of -3 dB into each of the channels, and confirm that they are in sync to the sample. Ensure that the signal path is the same as was used in step 1 above, so that this playback encounters the same gain conditions as the tone of step 1 and the noise of step 2. Play the speech, which is recorded with a measured Loudness of -24 LKFS, and observe that its level is normal. If no sound level meter is available and you have skipped here from Step 2, confirm that you have performed Step 1 if you have such equipment and adjust this track for the most comfortable listening level.

Note that this speech recording has been adjusted in level so that it measures -24 LKFS to standard BS.1770 [3].

https://www.atsc.org/wp-content/uploads/2025/06/Speech_left_Ch-20dB.wav

Step 4. If there is a separate subwoofer, feed the LFE channel electrically with the signal linked below, and set the level of the subwoofer for +4 dB re: the main channels, when measuring with a C-weighted, slow-reading sound level meter.

https://www.atsc.org/wp-content/uploads/2025/06/LFPinkNoise_-20dB.wav

10.5 Test Signal Details

10.5.1 Definition of 0 dBFS

Uncertainty in measurement of the level of noise with electrical meters has been described above in Section 10.2. The definition of level in digital audio systems is given in AES17 [21], where a full-scale sine wave is defined as 0 dBFS; this definition applies in this document. However, many software programs indicate level on virtual meters by means of a conventional RMS calculation, leading to a full-scale sine wave reading -3.01 dBFS, which is incorrect in the context of this document.

10.5.2 440 Hz Sine Wave Tone

A sine wave tone is the simplest and most widely employed test signal in audio. Typical uses include signal tracing and level setting. A sine wave has two properties useful for testing: First, the level is constant so it can be measured with great accuracy, and second, a mid-frequency sine-wave tone does not stress the limits of any system, so it typically passes through unprocessed.

The level of the tone linked below is -20 dB relative to Full Scale (dBFS).

https://www.atsc.org/wp-content/uploads/2025/06/440Hz_left_Ch-20dB.wav

The 440 Hz frequency was chosen for three reasons: 1) It is in the flattest portion of the equalization frequency response curve of the BS.1770 [3] specification; 2) It is musically relevant as reference "A above middle C" (also called "A4") on the musical scale[11], and it is not harmonically related to the various sampling frequencies in use, so that all code values within its range are exercised. Use of a tone at a precise tuning frequency helps by being able to identify audibly when there is an error in playback sample rate compared to recorded sample rate.

During the development of this test signal, it was found that producing a test sine wave at precisely -20 dBFS, then adding ±1 LSB of triangular-probability-density function, white-noise dither (as required for a distortion-free test signal), triggered some meters to read one division higher than -20 dBFS, which could result in up to a 2 dB error in units found. Therefore, the standard level was adjusted downwards in amplitude by one bit, and the dither added, so that the peak level of the signal is precisely -20 dBFS including dither, and it reads so consistently across test meters.

10.5.3 Band-Limited Pink Noise

Band-limited pink noise, provided in the linked file below, is preferred over full-band noise for several reasons. At low frequencies, below the Schroeder frequency12 in rooms, standing waves have a strong influence on level, and including such frequencies in the band of measurement adds uncertainty. Also, low-frequency energy Content in a stochastic signal leads to greater level variations vs. time. At higher frequencies, uncertainty in microphone response vs. angle, room response curve, and room absorption lead to less accuracy. On the other hand, narrow-band noise or tones are too greatly influenced by loudspeaker response and room acoustics. Therefore, a band of two octaves centered on 1 kHz has been determined as the most useful for the main channels, while one centered on 40 Hz has been determined as the most useful for the LFE channel.

https://www.atsc.org/wp-content/uploads/2025/06/MidRngPinkNoise_-20dB-1.wav

[11] While other frequencies have been (and occasionally still are) used, A440 is now commonly used as a reference frequency to calibrate acoustic equipment and to tune musical instruments. It is standardized as such in ISO 16:1975.

[12] See [33].

Annex A: Content Loudness and True Peak Measurement

A.1 INTRODUCTION

This Annex provides background to the BS.1770 Loudness and True Peak measurement algorithms. A detailed description of the measurement algorithms can be found in [3].

A.2 CONTENT LOUDNESS

The perceived Loudness of an audio signal can be considered in a number of ways. Short-term changes in Loudness are a reflection of the continuously changing nature and dynamics of the audio Content. Examples include moment-to-moment changes within Content that convey changes in mood or context. Short-term Loudness changes can also occur at the transition from one Content element to another. Some of these changes can be inherent given that each Content element may have been created for different purposes. They may also occur if the Content elements were created following different production practices.

A listener's impression of the Loudness of a piece of Content (a TV program, an advertisement, etc.) is not formed by the loud or soft portions of the entire Content element, but by the component that forms the "anchor point" of the entire mix. The loud and soft portions of a mix create the dynamics of the Content; the Anchor Element determines its overall or average subjective Loudness to the listener. No matter what the dynamic range of the Content, matching the Loudness of the Anchor Elements by applying simple gain offsets to bring each piece of Content to the same overall Loudness will allow listeners to enjoy the stream of Content without having to readjust their volume controls.

In 2001 the Radiocommunication sector of the International Telecommunication Union (ITU-R) initiated a study to identify an objective measure of the perceived Loudness of typical broadcast material. The study resulted in the adoption of a method to calculate the long-term integrated or average Loudness of an audio signal, which is reflected in Recommendation ITU-R BS.1770 [3]. This measurement is intended for mono, stereo and multichannel signals. A separate ITU-R Recommendation, BS. 1771 [15], defines the measurement of short-term and momentary Loudness.

The first phase of the ITU-R study was to develop a subjective test method to examine the perceived or subjective Loudness of typical monophonic material. A three-member panel made up of ITU-R Working Party 6P SRG3 members selected 48 test sequences consisting of a wide range of broadcast material and a reference sequence consisting of English female speech was chosen to establish a target Loudness level. The playback level of the reference sequence was set to 60 dB SPL, A-weighted, slow. In addition to the use of a single common reference sequence, test subjects were encouraged to compare all test sequences to one another. This test approach is similar to [9], where the subject is able to compare various test items to one other within a trial. Each of the test sequences was repeated at two amplitude levels, thus creating a total of 96 monophonic audio sequences for the subjects to match. A total of 97 subjects participated at five different test sites. The results of the subjective tests [10] formed the basis for evaluating the performance of various proposed Loudness algorithms and meters.

Seven proponents submitted ten monophonic Loudness meters/algorithms for evaluation. In addition to the meters, two additional Loudness algorithms were submitted by the evaluation lab to serve as a performance baseline. These two algorithms were simple mean-square calculations: The first used a simple frequency-weighting filter while the second was unweighted. The simple weighting filter was a revised low-frequency B-weighting curve (RLB). A comparison of the submitted Loudness meters/algorithms with the subjective database revealed that the simple frequency-weighted mean-square algorithm performed the best [11].

To verify the performance of this algorithm, a second subjective database was created from formal subjective tests conducted at one of the test sites. This test had 20 subjects provide a perceived Loudness rating to 96 monophonic audio sequences and used the same subjective testing methodology as the first round of tests. An analysis of the results from this second subjective database confirmed the performance of the simple frequency-weighted averaging algorithm.

In an effort to extend the algorithm to multichannel audio signals, a third round of tests was carried out using 144 audio sequences (48 monophonic, 48 stereo, and 48 multichannel), with 20 subjects participating. The test used the same subjective methodology as the previous two tests. The reference consisted of English female speech with stereo ambience and low-level background music. The loudspeakers were configured as described in Recommendation ITU-R BS.775 [12]. The results of this third subjective test [13] resulted in the design of the multichannel Loudness algorithm shown in Figure A.1.

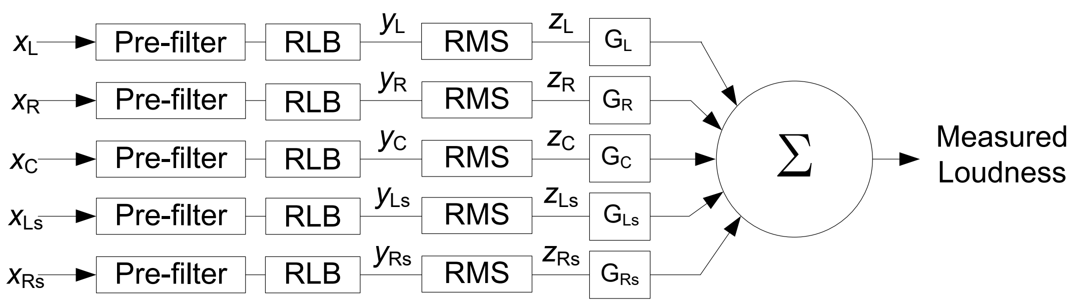

Figure A.1 Block diagram of multichannel Loudness algorithm.

The Loudness of each channel is measured separately, given a channel-weighting gain (GL, etc.) and summed to give the total resulting Loudness value. In addition to the RLB weighting curve, shown in Figure A.2, a pre-filter is added to account for the effects of the human head, and its frequency response is shown in Figure A.3. The combination of the two filters is known as a K-weighting filter.

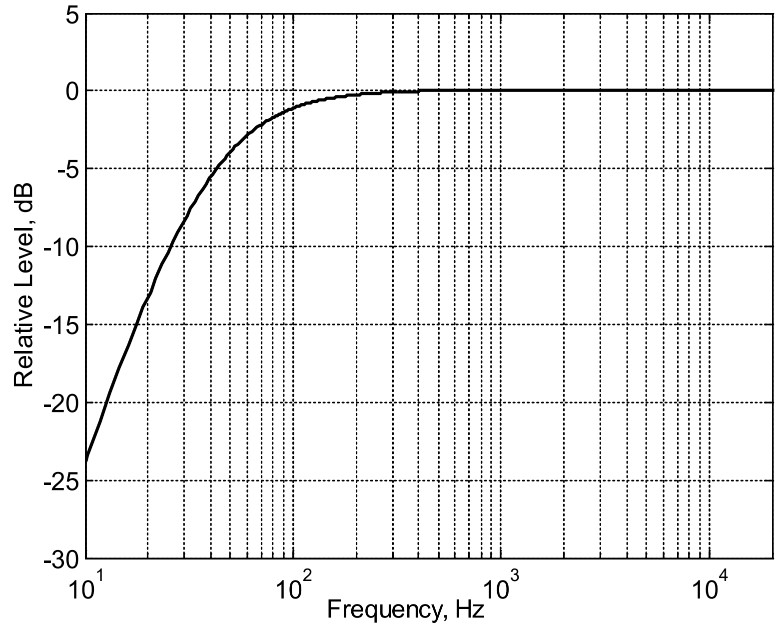

Figure A.2 The RLB weighting curve.

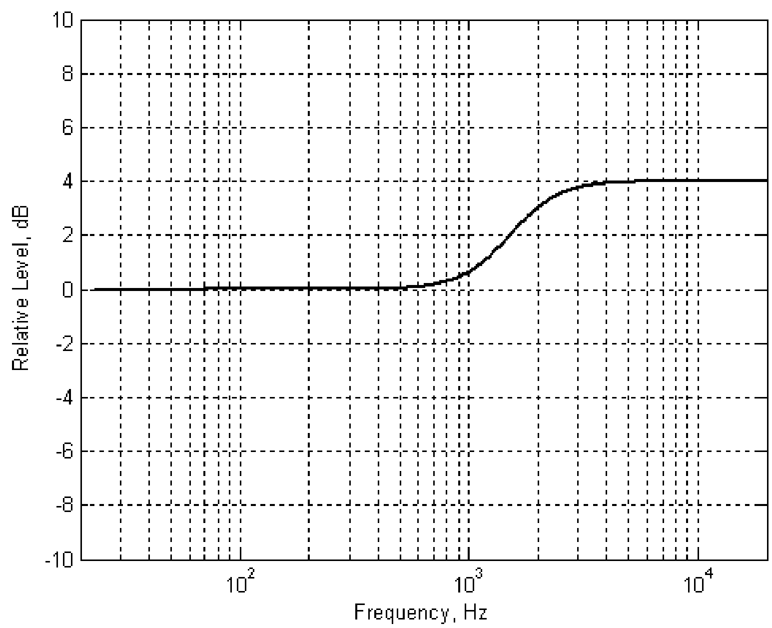

Figure A.3 Pre-filter response used to account for the acoustics effects of the human head.

The performance of the algorithm can be illustrated by comparing the subjective impressions of Loudness against measured values. This is shown in Figure A.4, where the measured Loudness for each of the 240 audio sequences from the three subjective tests is plotted against the subjective Loudness. Perfect agreement would result in the points falling on the diagonal line shown in the diagram. In this graph the correlation of subjective assessments to measured values is 0.977.

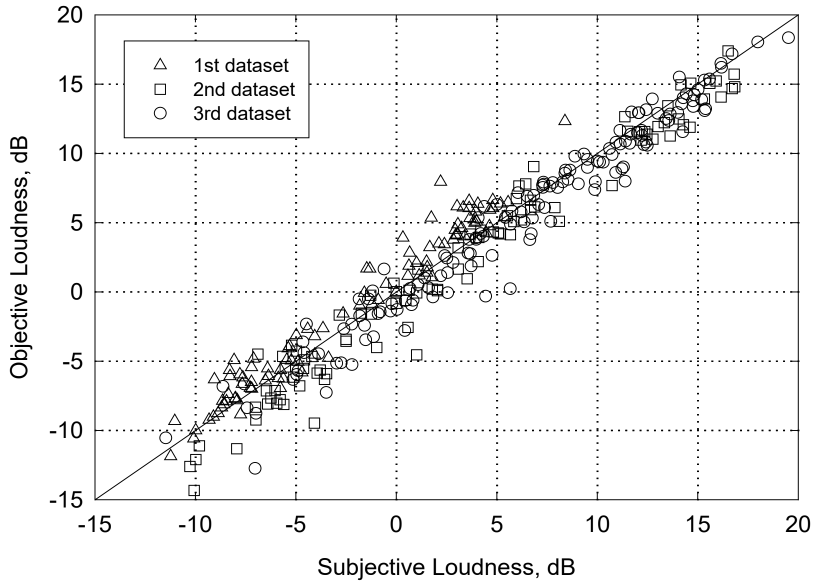

Figure A.4 Combined results for all three datasets (r = 0.977).

A subsequent refinement was the addition of two level-driven “gating” mechanisms. The basic BS.1770 Loudness measurement is an integrated or averaged power measurement, made over the entire duration of the Content being measured. The “gated” measurement algorithm works in a similar fashion but calculates the weighted average power in 400 ms blocks of time, updating the average power of this interval every 100 ms. Repeatable measurements, especially over relatively short pieces of Content depend on starting and stopping the measurement period predictably. A low-level, absolute “gate” was added to exclude any Content below -70 LKFS, effectively freezing the integration period and eliminating any measurement differences caused by an Operator pushing a start/stop button at different times, or by extended periods of silence in a Content segment affecting the measured loudness. This greatly simplifies the task of measuring the Loudness of the dialogue or Anchor Element when the audio is not continuous.

Since the remaining sound intensities above -70 LKFS contribute to the overall Loudness measurement, quiet sounds like environmental noises are included in the calculations, reducing the measured overall Loudness. Perceptually, however, a relatively quiet period during a program typically doesn’t substantially affect the listener’s impression of the overall Loudness of the program. A second “relative” gate was added to overcome this effect.

The threshold of the relative gate is set 10 dB below the absolute gated measurement. The relative gate removes any blocks whose Loudness is below this threshold.

Real time measurements can be made by “binning” all the blocked Loudness measurements (rather like a histogram) and recalculating the relative-gated Loudness every 100 ms, by ignoring any bins whose Loudness is more than 10 dB below the absolute-gated measurement.

Users should be aware that a long-term, integrated, relative-gated measurement of wide dynamic range Content will indicate that the Content is louder than an anchor-based measurement of the same Content. This is because the relative-gated measurement tends to focus on the loudest parts of the mix, and ignore the rest.

The result is that a relative-gated measurement will tend to underestimate the Loudness of the Anchor Element of Content with a wide dynamic range, when compared to an anchor-based measurement of the same Content. Normalization will, therefore, tend to push the Anchor Element (dialogue) Loudness measurement down, relative to the Loudness of the Content normalized to a relative-gated measurement.

A.3 TRUE PEAK

Modern digital audio systems have greatly simplified the manipulation and distribution of audio signals. Peak metering of these signals typically takes the form of displaying the maximum absolute sample value over a given measurement period. This narrow focus on peak sample values has made it easy to overlook the underlying continuous waveform. This can lead to unanticipated audio overloads, inconsistent peak readings, and other hidden problems. True Peak level measurements provide a more accurate description of the audio signal, which can be useful in preventing these problems.

Figure A.5 illustrates the potential for inaccurate readings when using typical sample-peak meters. In this diagram the maximum level of the continuous waveform is higher than the maximum sample value. This under-reporting of the audio level is small at low frequencies but can be significant at higher frequencies or with signals containing sharp transients. Inconsistencies in sample-peak readings can result when the samples do not fall on the same locations in the audio signal. In the diagram below, a small phase shift between the sampling clock and the audio signal can result in different sample-peak readings.

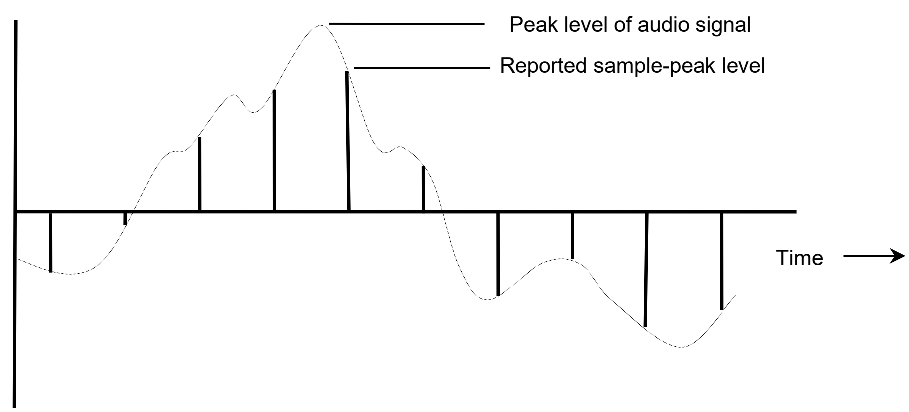

Figure A.5 Continuous-signal peak level versus sample-peak level.

Annex 2 of BS.1770 [3] describes an algorithm for estimating the True Peak level of a linear PCM audio signal. A simplified description of the algorithm is shown in Figure A.6.

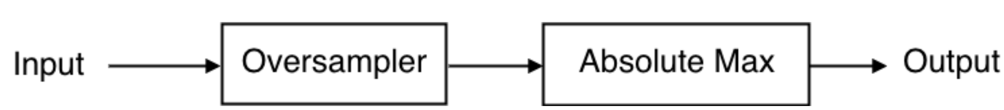

Figure A.6 Basic structure of the BS.1770 True Peak estimation algorithm.

Incoming signals are oversampled by four times to obtain a more accurate representation of the audio signal. This translates to 192 kHz oversampling for signals originally sampled at 48 kHz.

Annex B: Room Acoustics and Loudspeaker Placement

While a full discussion of control room acoustics and loudspeaker placement are beyond the scope of this RP, the quality of control room monitoring can be dramatically improved by following certain basic principles.

- *Control Low Frequency Room Modes Using Irregular Dimensions and Effective Low-Frequency Absorption*

Every room has "modes" determined by the geometry that cause sound to resonate at certain frequencies. Making walls non-parallel does not prevent this. In small rooms used for control monitoring, the resonant lowest frequency is within the audible range and the peaks of resonance are widely spaced. Left uncontrolled, this results in large irregularities in the frequency response, which varies from channel to channel, no matter how good the loudspeakers. Simple equalization is of limited usefulness in resolving this issue because the peaks of resonance are very narrow. Mixing under such conditions is a challenge, as individual notes in a music mix can jump out as sounding far too loud, while adjacent notes do not.

In rooms with few large objects, choose the three dimensions to be unequal, avoiding simple ratios between dimensions by at least ±5 percent. Always provide low frequency absorption to dampen the room response at resonant frequencies. Low frequencies can only be absorbed by high-quality materials thicker than 2", at a minimum, or 2" material covering a deeper air cavity. Carpet on floors or walls is of virtually no benefit. Suspended ceilings with high-performance absorptive tile with a 4" layer of backing absorbing material can provide effective low-frequency absorption to dampen room modes in one plane. Deep corner absorption, located at either wall-wall or wall-ceiling boundaries, can also be effective.

- *Distribute High Quality Absorption and Diffusive Materials Randomly and Avoid Hard Parallel Surfaces*

Too much room reverberation makes it difficult to hear the details of a mix. In addition, "ringing" effects, where a series of mid to high frequencies resonate, occur if two hard surfaces are parallel, especially when other room surfaces are absorptive.

To control both effects, distribute 2" high quality absorption randomly on walls and ceilings to eliminate parallel surfaces where opposite areas are reflective. Minimize the use of wall carpet and other thin absorption, which are only useful at very high frequencies. Where surfaces cannot be made absorptive, as in the case of windows, tilting the inner glazing layer can eliminate the parallel surface. Diffusive elements are also useful in producing a space with reasonable but controlled reverberation.

- *Position Loudspeakers and Absorption to Prevent Discrete Reflections*

A flat hard surface behind the mix position creates "comb filter" effects from the main monitor loudspeakers. The nature of the comb filter changes in frequency as the mix engineer moves forward and back in the room. A similar effect, at higher frequencies, occurs when the loudspeaker direct sound reflects off a large console face. The effect on the mix engineer can mask real comb filter problems in the mix.

Discrete rear-wall reflections should be prevented by a combination of absorption and diffusion on that surface. Consider locating the loudspeakers to prevent console-face reflections from reaching the ears of the mix engineer.

- *Control Mid-Bass Resonance Through Placement and Equalization*

A perfectly "flat" loudspeaker measured anechoically can sound very different in the lower mid-range in installations, depending on its room placement, due to reinforcement and canceling caused by reflections from adjacent reflective surfaces near the loudspeaker.

This effect can be mitigated by avoiding placing loudspeakers equidistant from two walls, by bringing the loudspeakers further from the reflective surfaces or by providing mid-range absorption on the adjacent walls. Since the effects are rather broad, equalization can be very useful in flattening the response.

- *Choose Loudspeakers with Flat Response and Broad Dispersion*

The above measure cannot correct for loudspeakers with an irregular frequency response, or where the response varies dramatically with listening angle. Such loudspeakers cannot be trusted to represent the frequency balance of the final mix.

Choose loudspeakers that sound consistent as you move from on-axis to a position at the fringe of the listening area. Look for evidence that the loudspeaker frequency response is flat across the audio range. Consider the use of a sub-woofer to extend the low-frequency response.

See the references "Audio Monitoring in Contemporary Post-Production Environments" [16], and "Surround Sound: Up and Running" [17] for further information.

Annex C: Room Correction

Room correction is needed in monitoring environments in order to address numerous problems that arise from the interaction of sound from loudspeakers and the room. The negative effects of small room acoustics produce a number of audible artifacts including imaging distortion due to unwanted reflections and frequency response anomalies that are particularly problematic in the low frequency range.

Traditionally, room equalization has been performed with analyzer/equalizer systems that use a single microphone measurement of pink noise and parametric or graphic equalizers. There are severe limitations in the performance of such systems because:

1) A single room measurement cannot provide enough information about the low frequency performance in the listening area.
2) Pink noise can only measure the magnitude response and does not have any time response information.
3) Parametric or graphic equalizers do not provide enough resolution even with 30 bands.
4) The IIR filters used in such equalizers can suffer from phase anomalies particularly as they become narrower.

Effective room correction needs to:

- Capture time domain information so that the effects of reflections can be properly accounted for.
- Capture frequency domain information with sufficiently high resolution in the low frequencies to address the typical problems found in small rooms.
- Combine multiple measurements from the listening area to account for low frequency variations due to standing waves.
- Reduce modal ringing.

The requirements for effective room correction can be met by using FIR filters for the equalization solution. However, standard textbook FIR approaches are not sufficient. Frequency-weighting methods are typically employed in the FIR filter design in order to distribute the filter power non-linearly with frequency, thus allocating more correction power to the lower frequencies. Furthermore, simple spatial averaging of multiple measurements is also not sufficient. The spatial distribution of acoustical problems is not uniform and some locations exhibit larger problems than others. A non-linear spatial weighting method also needs to be employed when combining the measurements.

Annex D: Quick Reference on Reference-Monitor Setup for Television Audio

D.1 PRODUCTION AND POSTPRODUCTION FACILITY MONITOR SETUP

1. Copy the first test tone, 1 kHz sine wave at -20 dBFS, to a digital audio workstation, and then copy it into each of the channels to be employed, such as 5.1.
2. Play the file, setting the output level controls of the workstation to unity gain, the console input to unity gain, and the console master level control to unity gain. You may keep the monitor level control low for the time being.
3. Set the level of each channel in turn for the sine-wave test level for the meter in use, such as -20 dBFS for digital meters, or 0 VU for VU meters.
4. Copy the second test tone, band-limited pink noise at -20 dBFS, to each channel of the digital audio workstation.
5. Play one channel at a time.
6. Put a sound level meter at the position of the center of the head of the normal position for the Operator, pointing it at the loudspeaker channel in use. Keep your body perpendicular to the meter, off to one side, as the level can be affected by a strong reflection off your body.
7. Put the master monitor level control at a standardized, repeatable setting that you will use in mixing.
8. Set the individual channel level controls, such as power amplifier gain controls, or gain controls on powered loudspeakers for the Sound Pressure Level measured C-weighted and slow-reading shown in Table D.1.

Table D.1 Reference Sound Pressure Level

Categories	Room Volume in Cubic Feet	SPL in dB re 20 μN/m^2
I, II (Mix rooms)	> 20,000	85*
	10,000 < 19,999	82
	5,000 < 9,999	80
	1,500 < 4,999	78
	< 1,499	76
III (Edit rooms sometimes used for mixing)	Depends on room usage. For editing purposes, may be controlled by the editor for use with the material at hand. For final program mixing, follow the recommendations for categories I, II above.	
IV (Booths, vans)	< 1,500	76
V (Headphones)	Use 2 cc. Coupler and set 400 Hz level to 76 dB.	
* Per SMPTE RP 200 [6]		

Note that Table D.1 reproduces Table 10.2 Reference Sound Pressure Level for the reader's convenience.

First test tone, sine wave:
https://www.atsc.org/wp-content/uploads/2020/04/440Hz_left_Ch-20dB.wav
Second test tone, band-limited pink noise:
https://www.atsc.org/wp-content/uploads/2025/06/MidRngPinkNoise_-20dB.wav

D.2 REALLY QUICK REFERENCE GUIDE FOR MONITOR SETUP

If you can't use the band-limited pink noise to calibrate the monitor levels, the speech sample referred to in Section 10.4 may be used to set the monitor level appropriately. Make sure that the signal path from the device playing back the speech sample to the input to the monitor system is set to unity gain. Edit the mono source file into the single channel for mono playback, into left and right at –3 dB each for stereo playback, and into the center channel of 5.1-channel and other multichannel systems. Play back the speech sample, and adjust the master monitor gain to put the speech at your most comfortable listening level. Since the Loudness of the speech sample is -24 LKFS, programs mixed so that the anchor program material such as dialogue matches this level will have approximately the same Loudness.

https://www.atsc.org/wp-content/uploads/2025/06/Speech_left_Ch-20dB.wav

Annex E: Loudness Ranges

The information in this annex is not to be confused with Loudness Range (LRA) [34].

The Comfort Zone is the range of Loudness within which a listener will accept Loudness changes within and between Content items. A subjective experiment was undertaken to determine this range, and the other "Loudness tolerance" points (see [32]). The results of the experiment are also shown in Figure E.1.

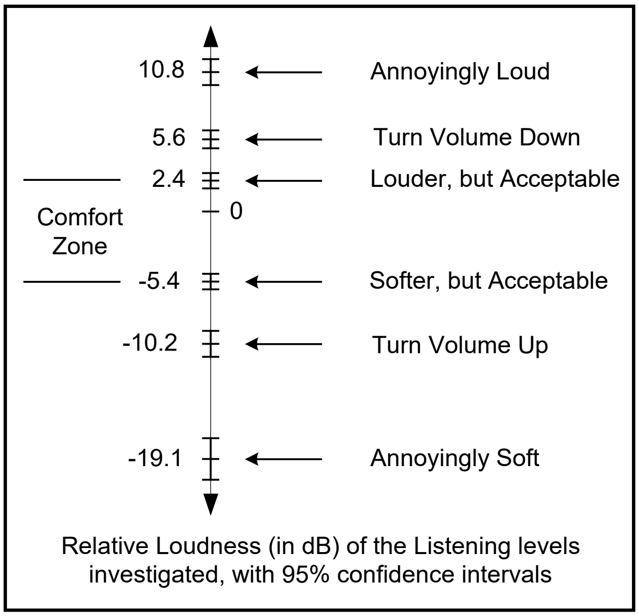

Figure E.1 Critical Loudness levels.

The experiment mimicked the transitions between Long-form and Short-form Content within a channel, and changes between channels. Subjects were placed in a typical listening or viewing situation, and asked to switch between samples of Long-form and Short-form Content. They were presented with a total of five different paired, monophonic Reference and Test items, reproduced by a single loudspeaker in front of them.

Subjects were instructed to adjust the master playback level until the Reference item was reproduced at what they considered to be a "comfortable volume," the zero point in the diagram. The experimenter then asked them to set the Test volume control ("volume" is a more familiar term than Loudness to most listeners) to one of the six points shown in the diagram. They could switch at will between the reference and test items as often as they wanted. Once they had settled on the requested Loudness difference, the offset was recorded. The questions were asked in random order, and the order of presentation of the pairs of reference and test items was randomized between subjects. The Reference item and Test items were taken from another experiment that provided a library of constant subjective-Loudness Content.

Since the Reference and Test items were known to be equally loud, the gain offset that the subjects applied to the test item in response to the questions from the experimenter were a direct measurement of the subjects' comfort zone and the other critical-Loudness levels investigated.

It is interesting to note that a gain increase of two to three dB in level is enough to move the subjective Loudness of a typical program out of the listener's Comfort Zone, and toward the point at which they would like to turn the volume down. There is more latitude available on the softer side of the "comfortable volume" point (shown here as "0").

The ambient noise level in the listening room used for the tests was quite low; similar to that in a rural living room on a tranquil evening. Since the "Annoyingly Soft" point can reasonably be expected to fall somewhere above the ambient noise level in the listening environment, the figure of –19.1 dB probably depends strongly on the ambient noise level. The other points are far enough above the ambient that their relative distribution should not be affected.

Annex F: AC-3 Dynamic Range Control (DRC) Details

F.1 DRC OVERVIEW

The AC-3 system is NOT intended to mimic the audio processing used in the current analog TV system. It is instead intended to provide (1) a consistent dialogue Loudness between programs, and (2) to allow individual listeners to reduce the dynamic range excursions around this common Loudness level if they so desire.

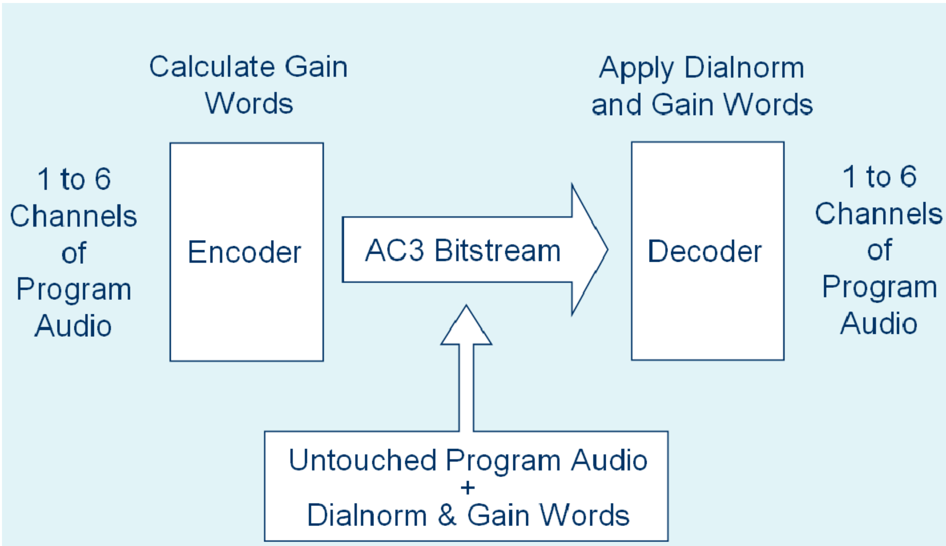

Figure F.1 AC-3 DRC basic concept.

The basic concept of the dynamic range control part of the AC-3 system is to do all the calculation required to reduce the dynamic range of the program to fit the selected dynamic range "profile" in the AC-3 encoder, then to send the original (wide dynamic range) version of the program (Content) to the AC-3 decoder along with the DRC gain words that describe the gain changes necessary to reduce the dynamic range of the program to fit the selected profile. (See Figure F.1.) The listener then has the option of applying the gain words, or not, depending on their listening requirements. One listener may choose not to apply the gain words, and will hear the original dynamics of the program, while another may choose to listen to the reduced dynamic range. In either case, the Loudness of the dialogue, or "anchor point" of all programs will be normalized, regardless of whether the DRC data is applied or not.

The choice of compression "profile" is up to the program producer, as this allows them to make an artistic choice of how the program dynamics are reduced. The AC-3 system can provide independent profiles for two modes (Line mode and RF mode) as explained later.

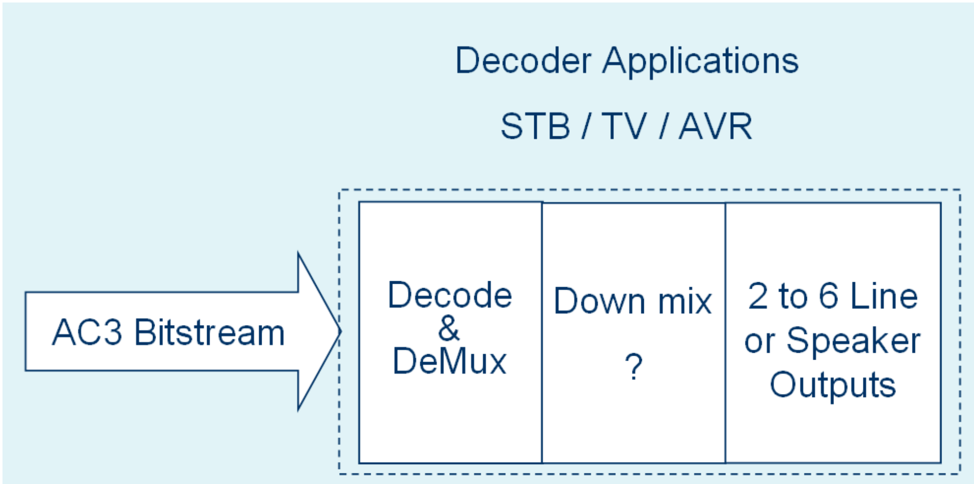

Figure F.2 AC-3 decoder requirements (Line Mode).

Wherever decoders are used, there are two main “modes” required. The first, shown in Figure F.2, is as a source of line level signals used to provide the audio for a home theatre or other “hi-fi” application. In this case, usually only a light compression is required, if any is needed at all.

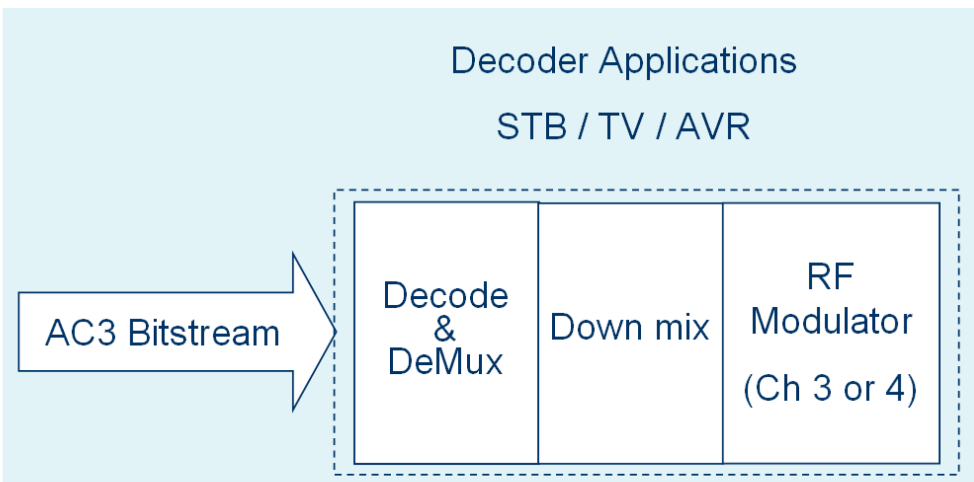

Figure F.3 AC-3 decoder requirements (RF Mode).

The second mode (Figure F.3) is needed to drive an RF modulator. In this case, the program will have to be downmixed, and the Loudness boosted to provide a reasonable match to the other analog off air signals used by the TV set. The system will also have to provide more dynamic range reduction than in the previous case, due to the boosted program signal level, matching the smaller dynamic range of typical existing processed TV programs, and possibly because of the listening conditions.

F.2 AC-3 DRC METADATA PLACEMENT IN BITSTREAMS

The AC-3 system supplies gain instructions for both types of dynamic range reduction, and also provides limiting in case the level build up inherent in downmixing is enough to produce clipping.

The gain words for the so called Line mode shown in Figure F.2 are called “dynrng” words. The gain words for the so called RF mode shown in Figure F.3 are called “compr” words.

The AC-3 bitstream carries both dynrng and compr gain words (as shown in Figure F.4), as well as the limiting gain words when required. AC-3 decoders can apply either RF or Line mode dynamic range control, depending on the listener’s preference (though some CE manufacturers may limit the choice in specific devices).

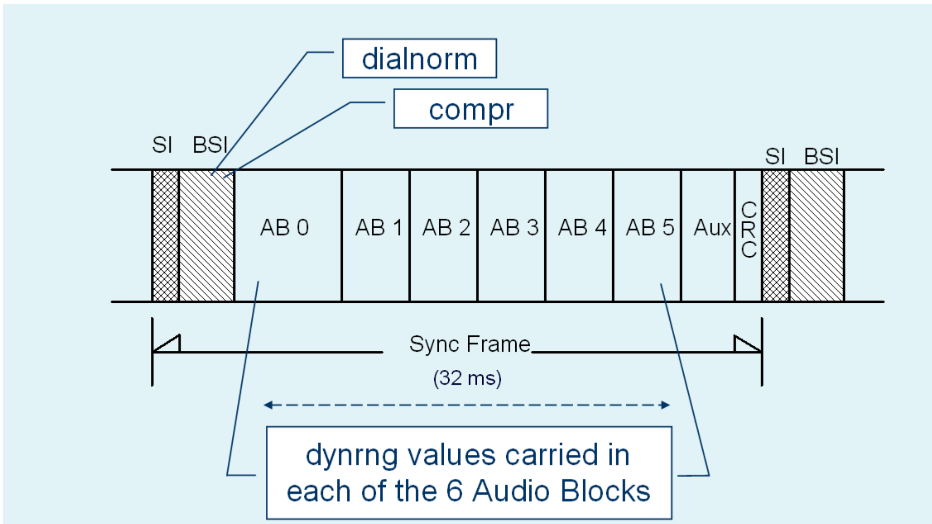

Figure F.4 AC-3 DRC metadata placement in bitstream.

All the dynamic range control words, including any limiting gain instructions are generated in the AC-3 encoder, sent to the AC-3 decoder and applied in the AC-3 decoder. The transmitted information includes the dialnorm data, which is used to normalize the dialogue or “anchor point” Loudness of each program as a whole.

The dialnorm and compr gain words are sent to the AC-3 decoder once every 32 ms, while the dynrng gain words are carried at 6 times this rate, or each ~ 5.3 ms.

F.2.1 Calculation of AC-3 Gain Words

As shown in Figure F.5, the first task is to calculate the normalized Loudness and peak levels of the program material (further detail shown in Figure F.6). acmod tells the encoder how many channels the program has.

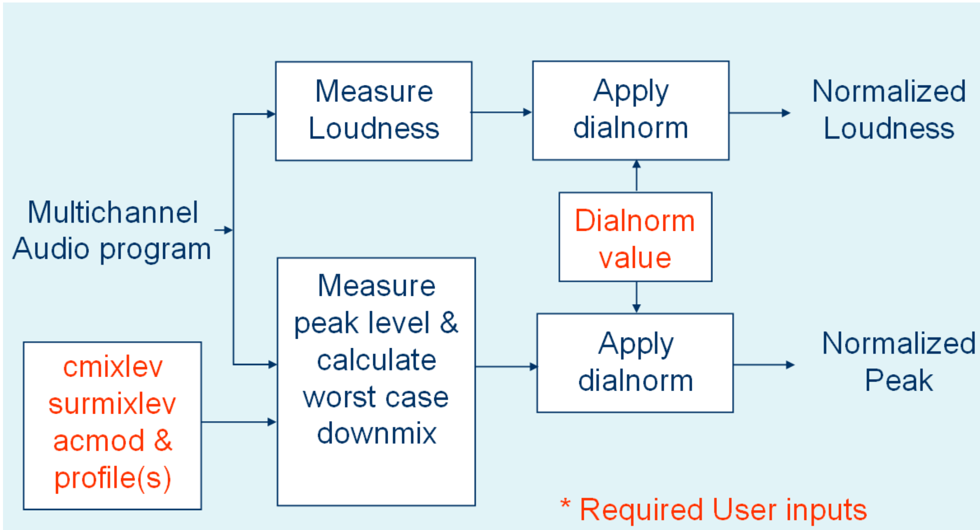

Figure F.5 Calculation of AC-3 gain words by the encoder (Part 1).

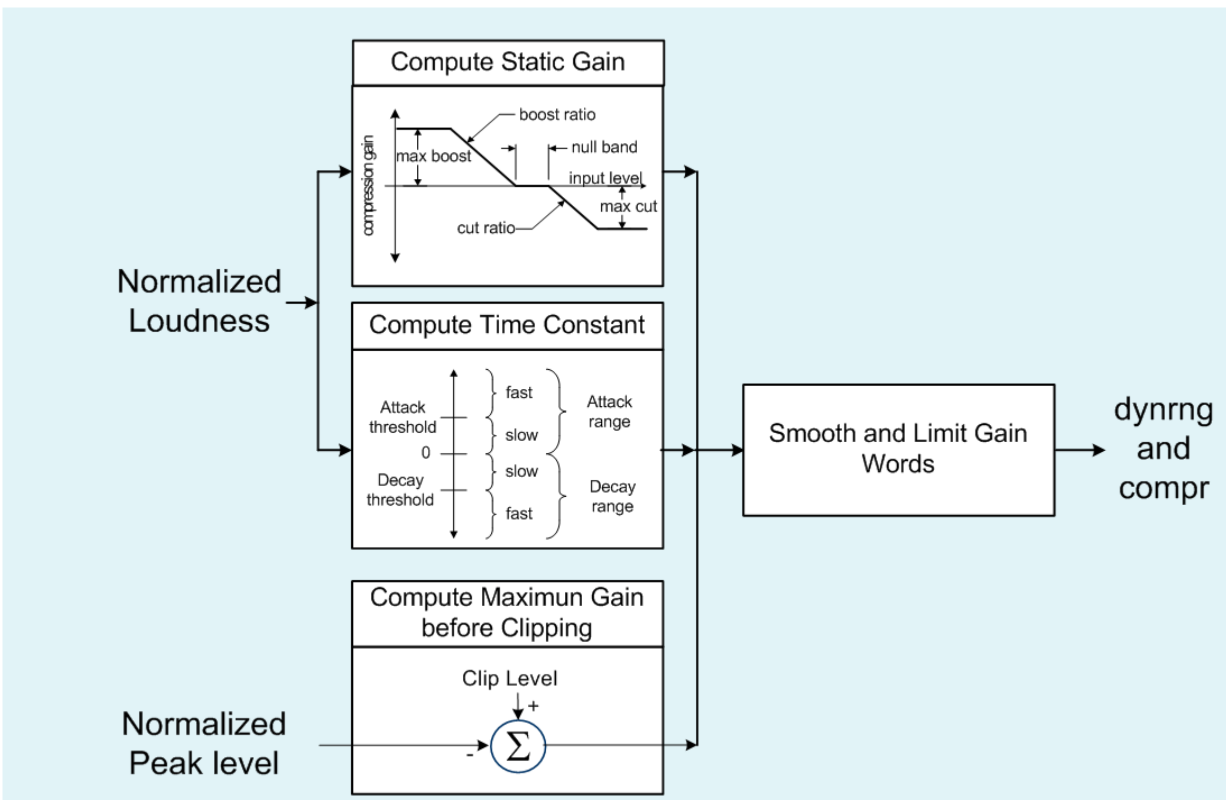

Figure F.6 Calculation of AC-3 gain words by the encoder (Part 2).

For the dynrng gain words, the Loudness is calculated for each 256 samples (Audio Blocks represent 512 samples – 5.3 ms – of audio, but because of the "overlap add" structure of the blocks, the Loudness of each audio block is represented by the average of two 256 sample measurements). Loudness values are thus produced every ~ 5.3 ms.

The Loudness of each channel is calculated individually, then summed to get the overall Loudness of the program. (The Loudness weighting curve is shown in Figure F.8).

The peak level calculation has to include the center and surround channel downmixing coefficients (cmixlev and surmixlev) selected (ideally) during the production process. The output of the calculation is the maximum peak level of the worst case downmix condition.

Loudness and peak values are normalized by the overall program dialnorm value before being passed to the subsequent calculation stages.

The Loudness and peak level data used for the "compr" gain word calculations represent these values over an entire AC-3 sync frame (32 ms).

The audio decoder includes an overlap and add, thus even the coarsely timed "compr" gain words are applied smoothly.

The dynamic range control portion of the AC-3 system has to deal with all types of program material, so it includes a method of adjusting the attack and decay time constants that are determined automatically depending upon the program material..

The audio data after Loudness normalization is mapped onto the selected DRC profile to generate a static gain value. It is also used to select one of four time constants, depending on whether the input Loudness is larger (attack) or smaller (decay) than the smoothed Loudness, and by how much.

If the input Loudness is larger than the smoothed Loudness, and the amount is greater than the attack threshold, then the "fast attack" time constant is selected. This is generally a very fast time constant, designed to provide very quick convergence of the compressor for very loud events.

If the input Loudness is larger than the smoothed Loudness, but not by more than the attack threshold, then the "slow attack" time constant is selected. This is generally a moderately fast time constant, designed to provide reasonably quick convergence of the dynamic range controller for moderately loud events

If the input Loudness is smaller than the smoothed Loudness, but not by more than the decay threshold, then the "slow decay" time constant is selected. This is generally a very slow time constant, designed to provide slow and therefore inaudible release of the compressor during soft passages.

If the input Loudness is smaller than the smoothed Loudness, and the amount is greater than the decay threshold, then the "fast decay" time constant is selected. This is generally a moderate time constant, designed to provide a faster release at the end of very loud events.

The other principal part of the DRC data generation process is the peak limiting function. The clip level is known (0 dBFS) thus the difference between that and the normalized peak level of the program is the maximum allowable gain before clipping. If the static gain words call for more gain, the smooth and limit gain block limits the static gain words to the maximum allowable gain. The time constant just computed ensures that the gain changes will not be abrupt enough to cause objectionable artifacts, and are appropriate to the program material.

Note that there are two parallel processes going on; one running at the audio block rate (~ 5.3 ms) generating dynrng gain words, and the other running at the sync frame rate (32 ms) generating the compr gain words.

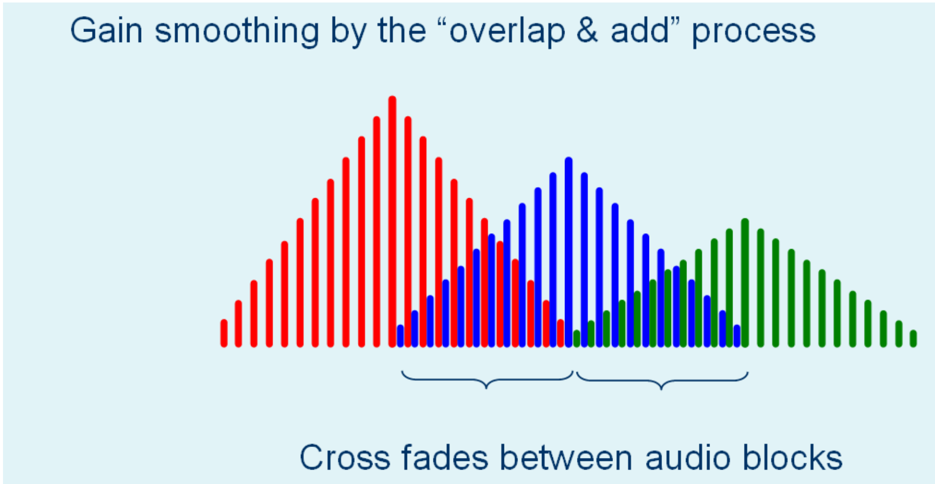

Figure F.7 Gain smoothing.

This is a very simple idea of how the "overlap and add" transitions between audio blocks happen, and the effect they have on the gain steps (quantized into 0.2 dB steps for dynrng, and 0.4 dB steps for compr) applied by the dynamic range control system.

The triangular shape shown in Figure F.7 is an approximation (for easy drawing) of the actual windowing function applied to the data.

Dynrng words apply to an entire group of 512 samples, but each group of 512 samples overlaps the adjacent group by half, so the effect when the sample values are added is a crossfade between groups, which smooths out any steps between them.

The compr words apply to an entire sync frame, but the sync frames also use the overlap add process, smoothing out the compr gain steps as well.

F.2.2 The DRC Loudness Weighting and Input-Output Curves

The DRC weighting curve emphasizes low frequencies more than the B-weighting curve, which is normally used to weight "moderate level" (i.e., close to TV listening levels) sounds. (See Figure F.8.)

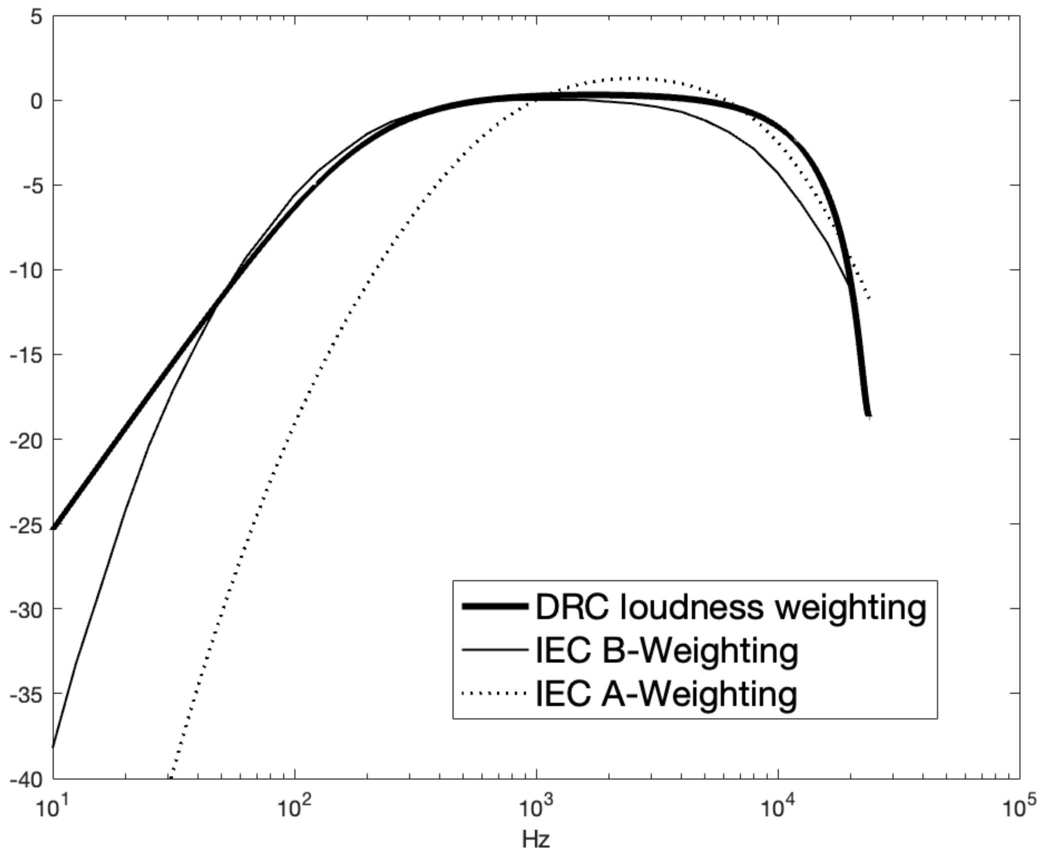

Figure F.8 DRC Loudness weighting curve.

The total Loudness of a multichannel source is computed as the sum of the weighted RMS power levels in each channel. Both the weighting and Loudness summation predate the ITU Loudness measurement, which can be seen as a refinement of this method.

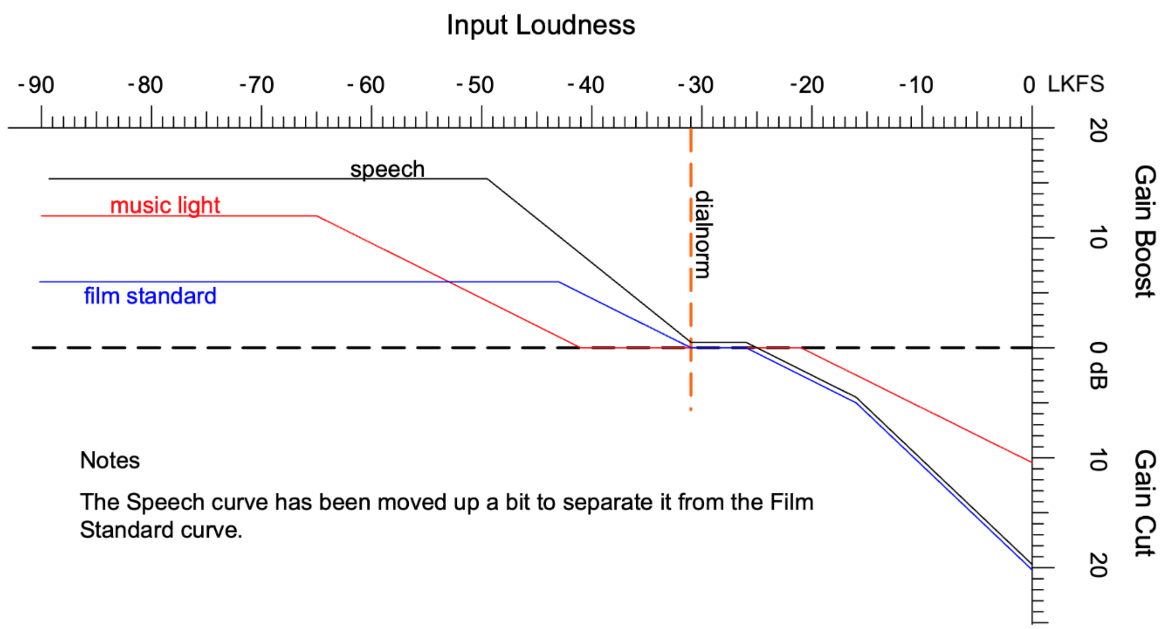

Figure F.9 DRC profiles input/output characteristics.

The Null zones of Film Standard and Speech are not symmetric around the dialnorm value because with typical program material and the degree of dynamic range reduction desired, the relatively rapid boost zone attack, combined with a slower decay time tends to leave the program in the asymmetric null zone. (See Figure F.9.)

F.2.3 DRC Encoder Parameters for Setting Metadata

This illustrates some of the front panel menu choices provided by the Dolby 569 AC-3 Encoder. There is a PC based remote control program for the Dolby 569 that presents all the parameters on a single screen. The Dolby 570 metadata authoring and previewing device has similar front panel menus and a (much easier to use) remote control program as well.

Main Setup Menu -> Audio Service -> Channel Mode -> Choose one of *1/0, 2/0, 3/2*, etc.

Main Setup Menu -> Audio Service -> Dialog Level -> Set to *-1 dB to -31 dB*

Main Setup Menu -> BSI Parameters -> Center Downmix level -> Choose one of *-3.0 dB, -4.5 dB, -6.0 dB*

Main Setup Menu -> BSI Parameters -> Surround Downmix level -> Choose one of *-3.0 dB, -4.5 dB, -6.0 dB*

Main Setup Menu -> Dynamic Range -> Line Mode -> Choose one of: *Speech, Music Light, Music Standard, Film Light, Film Standard, Disabled*

Main Setup Menu -> Dynamic Range -> RF Mode -> Choose one of: Speech, Music Light, Music Standard, Film Light, Film Standard, Disabled

F.2.4 DRC Control Available at the Decoder

- Very little commonality between the names used by consumer devices for Line, RF or "Off" DRC modes.
- Typical decoder calls Line Mode "Normal", RF Mode "Late Night" and no dynamic range compression "Theatre".
- The individual consumer can typically choose one of these three modes.
- In some (higher end) decoders, the listener can adjust the fraction of the Line or RF mode compression applied.

Annex G: AC-3 Metadata Parameters

Default values set in some AC-3 encoders may have no correlation to specific Operator Content, and should not be relied upon. The AC-3 encoder manufacturer should be consulted for guidance.

Table G.1 Critical Control Metadata.

Function	Bitstream Variable	Description
Dialog Level	dialnorm	Controls dialogue normalization.
Channel Mode	acmod	Designates the number and type of active channels.

Table G.2 Basic Control Metadata – Encoder.

Function	Description
Bitrate	Encoded bitrate.
RF Overmodulation Protection	When enabled the AC-3 encoder includes pre-emphasis in its calculations for RF Mode compression.
DC Filter	Applies a DC-blocking 3 Hz high-pass filter before AC-3 encoding.
Lowpass Filter	Applies a lowpass filter to the main input channels before AC-3 encoding.
LFE Lowpass Filter	Applies a 120 Hz lowpass filter to LFE channel before AC-3 encoding.
Surround 3 dB Attenuation	Attenuates surround channels 3 dB before encoding.
Surround Phase Shift	Before encoding, creates the phase-shifted surround channels necessary to create an Lt/Rt output in a decoder that can be decoded using Dolby Pro Logic to L, C, R, S.

Table G.3 Basic Control Metadata – Decoder.

Function	Description
LFE Channel	Indicates whether LFE channel is present. Available only in channel modes 3/2, 3/1, 3/0, 2/2, and 2/1.
DRC Line Mode Profile	Designates preset Dynamic Range Control (DRC) compression configuration for line-mode decoding.
DRC RF Mode Profile	Designates preset Dynamic Range Control (DRC) compression configuration for RF-mode decoding.
Dolby Surround Mode	Indicates whether a two-channel encoded bitstream contains a Dolby Surround (Lt/Rt) program and requires Dolby Pro Logic decoding.
Dolby Surround EX Mode*	Identifies if audio is encoded as a Dolby Surround EX stream and requires Surround EX decoding.
Preferred Stereo Downmix*	Designates preference for Lt/Rt (Pro Logic encoded) or Lo/Ro (stereo only) downmix.
Center Downmix Level	Designates downmix level for the C channel when end user has no center speaker.
Lt/Rt C Downmix Level*	When the stereo downmix is Lt/Rt, designates downmix level for the C channel when end user has no center speaker.
Lo/Ro C Downmix Level*	When the stereo downmix is Lo/Ro, designates downmix level for the channel when end user has no center speaker.
Surround Downmix Level	Designates downmix level for surround channels when end user has no surround speakers. Note that the 0 (-999 dB) setting discards the surround channels.
Lt/Rt S Downmix Level*	When the stereo downmix is Lt/Rt, designates downmix level for surround channels when end user has no surround speakers.
Lo/Ro S Downmix Level*	When the stereo downmix is Lo/Ro, designates downmix level for surround channels when end user has no surround speakers.
*Extended bit stream parameters. Not supported by all decoders.	

Table G.4 Informational Metadata

Function	Description
Audio Production Info	Indicates whether the Mixing Level and Room Type parameter settings are carried in the bitstream.
Bitstream Mode	Describes the audio service carried in the bitstream.
Copyright	Indicates whether the encoded bitstream is copyright protected.
Mixing Level	Indicates absolute acoustic SPL of main dialogue channel during final mixing session.
Room Type	Indicates size and calibration of the mixing room used for the final mixing session.
Original Bitstream	Indicates whether the encoded AC-3 bitstream is the master version.
A/D Converter Type*	Identifies the setting for type of A/D converter.

Annex H: Quick Reference Guide for TV Station and MVPD Engineers on Loudness Management

H.1 INTRODUCTION

This Annex summarizes the recommendations in ATSC A/85 and provides guidance to broadcasters and other video program distributors on controlling and maintaining consistent audio Loudness of their TV stations and channels.

H.2 SCOPE

This Quick Reference Guide is not intended to replace the complete RP. Its scope is limited to a 'how to' guide for DTV station Operators and MVPDs. Readers of this document are encouraged to review the complete RP for more detailed information and the background to this Guide. In the event of a conflict between the Guide and the RP, the RP takes priority over this Quick Reference Guide. This Quick Reference Guide is based on the use of a fixed metadata system[13] (see Section 7.3).

H.3 DEFINITIONS

Anchor Element – The perceptual Loudness reference point or element around which other elements are balanced in producing the final mix of the Content, or that a reasonable viewer would focus on when setting the volume control.

BS.1770 – Recommendation ITU-R BS.1770 [3]. This specifies algorithms that provide numerical values indicative of the perceived Loudness and True-Peak levels of the audio Content that is measured. Loudness meters and measurement tools that have implemented the BS.1770 algorithm will report Loudness in units of LKFS, and True-Peak levels in units of dBTP.

Dialogue Level – The Loudness, in LKFS units, of the Anchor Element[14]. When referring to an AC-3 encoder parameter, the term **Dialog Level** is used.

dialnorm – An AC-3 metadata parameter, defined in A/52 [4], that is carried in the AC-3 bit stream. This is used to indicate how far the average Dialogue Level is below 0 LKFS. Valid values are 1 to 31. Loudness values outside this range cannot be expressed by dialnorm. The value of dialnorm is numerically equal to the absolute value of the Dialogue Level. The value of 0 is reserved.

LKFS – Loudness, K-weighted, relative to full scale, measured with equipment that implements the algorithm specified by BS.1770 [3]. A unit of LKFS is equivalent to a decibel.

Target Loudness – A specified value for the Anchor Element (i.e., Dialogue Level), established to facilitate Content exchange from a Content supplier to an Operator.

H.4 LOUDNESS MANAGEMENT

Key Idea: Goal is to present to the viewer, consistent audio Loudness across commercials, programs and channel changes.

[13] Agile metadata is an alternative to using the fixed metadata system approach. See Section 7.5 of the RP for details concerning the agile metadata approach.

[14] The term "Dialogue Level" is based on dialogue's widespread use as the anchor for mixing of Content.

H.5 FCC REQUIREMENT

ATSC document A/53, Part 5 [1] mandates the carriage of dialnorm and correctly set dialnorm values.

> Key Idea: Set the AC-3 encoder's dialnorm to match the Loudness of average Dialogue Level of the Content.

H.6 MEASUREMENT OF CONTENT AS DELIVERED

See Section 5.

H.6.1 Long-Form Content

The average dialogue Loudness must be measured and reported as the Loudness for Long-form Content. This should be performed using BS.1770 (without the relative-level gate[15]) along with a speech gating algorithm and measured over the entire duration (not act by act) of the Content's composite mix. (Versions of BS.1770 without relative-level gating can be referred to as BS.1770-1). This practice yields a speech-gated Loudness measurement that represents the Loudness of Long-form Content during postproduction and downstream. It is vital that this average dialogue Loudness value is reported as the Loudness of the Long-form Content.

If measuring the average dialogue of the entire program as described above is not possible, a section of the Content that is representative of the Anchor Element (typically dialogue) should be isolated and measured and reported as Loudness of the Long-form Content**Error! Bookmark not defined.**. See Section 5.2.3.

> Key Idea: Measure the Long-form Content Loudness when dialogue is present. This value is the Dialogue Level of the Content. The Dialogue Level (in units of LKFS) should match the dialnorm value of the AC-3 encoder.

H.6.2 Short-Form Content

See Section 5.2.4.

> Key Idea: Measure the Loudness of all audio channels[16] and all elements of the soundtrack integrated over the duration of the Short-form Content. The value of the Loudness measurement (in units of LKFS) should match the dialnorm value of the AC-3 encoder.

H.6.3 Newscasts or Other Live Programming

The principle of measuring the Loudness of the dialogue of the Content applies to live productions done in real time as the production progresses.

The intent of Loudness measurements made during a live event is to guide the mixer to produce the Content at a Loudness that matches the dialnorm setting of the AC-3 encoder.

A BS.1770 Loudness meter may be helpful when mixing in noisy environments, or when a consistent monitor level cannot be maintained. See Section 5.2.2.

[15] A relative-level gate was added to BS.1770-2 and subsequent versions.

[16] The LFE is not included.

Key Idea: Use a BS.1770 meter to help ensure that real-time Content Loudness matches the dialnorm setting of the AC-3 encoder.

H.6.4 File-Based Content

File based storage makes it practical to automate the Loudness measurement and to adjust the Content Loudness and/or the dialnorm value (if any) that may have been assigned to the Content. See Section 5.2.6.

Key Idea: Ensure file-based Content's Loudness matches the dialnorm setting in the AC-3 encoder.

H.7 TARGET LOUDNESS TO FACILITATE PROGRAM EXCHANGE

See Section 6.

Target Loudness is a specified value for the Dialogue Level established to facilitate Content exchange from a supplier to an Operator.

For delivery or exchange of Content without metadata (and where there is no prior arrangement by the parties regarding Loudness), the ATSC specifies a Target Loudness value of -24 LKFS, which serves to establish a common operating level for use with that fixed value of dialnorm. Measurement tolerance of up to approximately ±2 dB around this value is anticipated, due to measurement variations. Content Loudness should **not** be targeted to the high or low side of this tolerance.

Key Idea: For Content without metadata, use the Target Loudness value of -24 LKFS.

H.8 METHODS TO EFFECTIVELY CONTROL SHORT-FORM TO LONG-FORM LOUDNESS

Large Loudness variation during transitions can be effectively managed by adhering to the following practices:

For Operators using a fixed dialnorm system, see Section 7.3:

a) Ensure that all Content meets the Target Loudness and that the dialnorm value matches this value.
b) Employ a file-based scaling device to match Dialogue Level of non-conformant Content to the Target Loudness.
c) Employ a real-time Loudness-processing device to match the Dialogue Level of non-conformant Content to the Target Loudness.

Key Idea: Ensure that all program and commercial audio Content matches the dialnorm value of the AC-3 encoder. Use a BS.1770 meter to verify the Dialogue Level of the Content.

H.9 AFFILIATE DIALNORM SETTING

See Section 7.3.5.

An Operator (affiliate, TV station, MVPD, etc.) receiving Content that is delivered at a fixed Loudness, where there is no gain adjustment or processing after the receiver, should set the value of dialnorm in the Operator's AC-3 encoder to match the network originator's specified Dialogue

Level. If a fixed gain or loss is applied in the signal chain, the AC-3 encoder dialnorm value needs to be offset accordingly from the originator's Dialogue Level.

If Loudness processing is applied to the originator's audio, the processor's Target Loudness value should match the Operator's AC-3 encoder 's dialnorm value. See Section 9.3 for additional background on audio processing.

> Key Idea: Set the AC-3 encoder's dialnorm value to the originator's Dialogue Level (as adjusted).

H.10 TV STATION OR MVPD CONTENT INSERTION

In the case of TV station or MVPD insertion of local commercials or segments, the Operator should ensure that the Dialogue Level of the local insertion matches the dialnorm setting of the inserted audio stream.

> Key Idea: Ensure that the Dialogue Level of inserted Content matches the dialnorm setting of the inserted audio stream.

If the network originator's feed is decoded to baseband, the Loudness of the decoded audio is to be measured and the value of the re-encoder's AC-3 dialnorm value is set to match the measured Loudness for the next stage of encoding. In this case either the Operator needs to modify the network originator's Loudness to match the target value of the Operator's system, or the originator's Loudness value (as measured) will be used as the dialnorm value in the next stage of AC-3 encoding. At this re-encoding stage the Operator needs to also ensure that the other audio metadata parameters are set appropriately.

> Key Idea: If the network originator's feed is decoded to baseband, ensure that the measured Dialogue Level of the Content matches the dialnorm setting of the next stage of AC-3 encoding.

H.11 AC-3 DYNAMIC RANGE CONTROL (DRC)

The AC-3 system includes DRC profiles for "Line mode" and "RF mode." While choosing these parameters may be useful to the Operator and viewer for limiting the overall Loudness range, DRC should not be relied upon to correct Loudness variations between programs, programs and commercials or between TV stations or cable channels and during channel changes. (See Section 8.3 and Annex F.)

> Key Idea: AC-3 Dynamic Range Control should not be relied upon to mitigate Short-form to Long-form Content or TV channel-to-channel Loudness variations.

Annex I: Quick Reference Guide for Audio Mixers and Editors Creating Content

I.1 INTRODUCTION

This Quick Reference Guide summarizes the recommendations in this RP to provide guidance to audio mixers and editors creating audio for DTV. It promotes a goal of managing audio Loudness of Content consistent with artistic intent.

I.2 SCOPE

This Quick Reference is not intended to replace the RP. The scope is limited to a "how to" guide for audio mixers and editors involved with the creation of audio for DTV. Readers of this Annex are encouraged to read the full RP to educate themselves on the details and background to this guideline. In the event of a conflict between the Quick Reference Guide and the RP, the RP takes priority[17].

I.3 DEFINITIONS

BS.1770 – Recommendation ITU-R BS.1770 [3]. This specifies algorithms that provide numerical values indicative of the perceived Loudness and True Peak levels of the audio Content that is measured. Loudness meters and measurement tools that have implemented the BS.1770 algorithm will report Loudness in units of LKFS, and True Peak levels in units of dBTP.

dBTP – Decibels, True-Peak relative to full-scale

Dialogue Level – The loudness, in LKFS units, of the Anchor Element[18]. When referring to an AC-3 encoder parameter, the term **Dialog Level** is used.

dialnorm – An AC-3 metadata parameter, defined in ATSC A/52 [4], which is carried in the AC-3 bit stream. It is used to indicate how far the average Dialogue Level of the Content is below 0 LKFS. Valid values are unsigned integers within the range of 1 to 31. Loudness values outside this range cannot be expressed by dialnorm. The value of dialnorm is numerically equal to the absolute value of the Dialogue Level. The value of 0 is reserved.

LKFS – Loudness, K-weighted, relative to full-scale, measured with equipment that implements the algorithm specified by BS.1770 [3]. A unit of LKFS is equivalent to a decibel.

Target Loudness – A specified value for the Anchor Element (i.e., Dialogue Level), established to facilitate Content exchange from a Content supplier to an Operator.

I.4 MONITORING ENVIRONMENT

A correct monitoring environment is critical to satisfactory mix results. Please read the background discussion in Section 10.1 as well as Section 10.2 for a full understanding of this topic.

It is essential that the correct monitor level be used. This level varies with room volume. See Section 10.4 for additional information and Table I.1 below:

[17] This Quick Reference Guide is based on the use of a fixed metadata system (see Section 7.3 in the RP). Agile metadata is an alternative to using the fixed metadata system approach. See Section 7.5 of the RP for details concerning the agile metadata approach.

[18] The term "Dialogue Level" is based on dialogue's widespread use as the anchor for mixing of Content and historically, it was felt that for most programs, dialogue would be the Anchor Element.

Table I.1 Reference Sound Pressure Level

Categories	Room Volume in Cubic Feet	SPL in dB re 20 μN/m2
I, II	> 20,000	85*
	10,000 < 19,999	82
	5,000 < 9,999	80
	1,500 < 4,999	78
	< 1,499	76
III	Depends on room usage. For editing purposes, may be controlled by the editor for use with the material at hand. For final program mixing, follow the recommendations for categories I, II above.	
IV	< 1,500	76
V		Use 2 cc coupler and set 440 Hz level to 74 dB.
* Per SMPTE RP 200 [6]		

Note that Table I.1 reproduces Table 10.2 for the reader's convenience. Consult Annex D for a synopsis of mixing room setup and calibration.

> Key Idea: Goal is to correctly setup your listening environment once and make sure you are always listening at this level when creating Content. This is true even if you must use headphones to monitor.

I.5 BS.1770 LEVEL MONITORING

BS.1770 provides a new measurement technique for monitoring audio levels. The use of measurement tools which support the BS.1770 measurement methods by all involved in audio production will assist the industry to manage audio Loudness of Content consistent with artistic intent. See Section 5.2 for a short discussion of how best to use the technique, and Annex A for a detailed discussion of BS.1770 and how it works.

> Key Idea: Use BS.1770-compliant measurement tools!

I.6 CONTENT LOUDNESS DURING MIXING

> Key Idea: With the monitor level set correctly, always mix relying on your hearing. Use a BS.1770 Loudness monitoring tool to confirm what you hear.

I.7 TARGET LOUDNESS FOR CONTENT WITHOUT METADATA

For delivery or exchange of Content without metadata[19], the Target Loudness value should be –24 LKFS. Measurement tolerance of up to approximately ±2 dB around this value is anticipated, due to measurement variations. Content Loudness should **not** be targeted to the high or low side of this tolerance. The True Peak level should be kept below -2 dBTP (± 0.5 dBTP) in order to provide headroom to avoid potential clipping due to downstream processing (such as audio coding used in delivery).

[19] See Section 6.

Key Idea: When generating Content and the program delivery level requirement is unknown or has not been specified, mix Dialogue Level to -24 LKFS with True Peaks below -2 dBTP.

I.8 FCC REQUIREMENT

ATSC document A/53, Part 5 [1] mandates the carriage of dialnorm and correctly set dialnorm values.

Key Idea: The AC-3 encoder's dialnorm will be set to match the Loudness of average Dialogue Level of the Content.

I.9 MEASUREMENT OF POST-PRODUCED CONTENT

See Section 5.2.1.

I.9.1 Long-Form Content

The average dialogue Loudness must be measured and reported as the Loudness for Long-form Content. This should be performed using BS.1770 (without the relative-level gate[20]) along with a speech gating algorithm and measured over the entire duration (not act by act) of the Content's composite mix. (Versions of BS.1770 without relative-level gating can be referred to as BS.1770-1). This practice yields a speech-gated Loudness measurement that represents the Loudness of Long-form Content during postproduction and downstream. It is vital that this average dialogue Loudness value is reported as the Loudness of the Long-form Content. See Section 5.2.3.

Key Idea: Measure the Long-form Content audio when typical dialogue is present and record this value as the Loudness of the Content.

I.9.2 Short-Form Content

See Section 5.2.4.

Key Idea: Measure the Loudness of all audio channels[21] and all elements of the soundtrack integrated over the duration of the Short-form Content.

[20] A relative-level gate was added to BS.1770-2 and subsequent versions.

[21] The LFE is not included.

Annex J: Requirements for Establishing and Maintaining Audio Loudness of Commercial Advertising in Digital Television

J.1 INTRODUCTION AND SCOPE

The recommendations in this Annex are based on other sections of this Recommended Practice. This Annex contains all the courses of action necessary to perform effective Loudness control of DTV commercial advertising.

J.2 LOUDNESS MANAGEMENT

The Operator's goal is to present to the audience consistent audio Loudness across Long-form and Short-form Content. The Operator should, whenever possible, present the Content to the audience with the most accurate and highest quality sound, free of any type of audio artifacts not contained within the original material as delivered by the Content supplier.

J.3 DIALNORM FOR AC-3-BASED SYSTEMS

ATSC document A/53 [1], Section 5.5, mandates the carriage of dialnorm and correctly[22] set dialnorm values.

J.4 COMMERCIAL ADVERTISING LOUDNESS

It is vital that, when Loudness of Short-form Content (e.g., commercial advertising) is measured, it be measured in units of LKFS including all audio channels[23] and all elements of the soundtrack over the duration of the Content.

J.5 COMMERCIAL ADVERTISING AT THE POINT OF INSERTION

In the case of insertion of Short-form Content[24] (e.g., commercial advertising), it is vital that the Loudness, measured as per Section J.4, of the inserted Short-form Content match[22] the dialnorm setting of this inserted AC-3 audio stream, per Section J.3.

[22] See Section 7.1 of this document.

[23] Per BS.1770, the LFE is not included.

[24] See Section 8.4 of this document.

Annex K: Requirements for Establishing and Maintaining Audio Loudness of Commercial Advertising in Digital Television When Using Non-AC-3 Audio Codecs

K.1 INTRODUCTION AND SCOPE

The recommendations in this Annex are based on other sections of this Recommended Practice. This Annex contains the courses of action necessary to perform effective Loudness control for DTV commercial advertising when using non-AC-3 audio codecs.

K.2 LOUDNESS MANAGEMENT

The Operator's goal is to present to the audience consistent audio Loudness across programs, commercials, promotional material, and public service announcements.

K.3 LOUDNESS FOR NON-AC-3 CHANNELS

It is vital that the delivery channel operate at an Operator-selected Loudness target value (as measured in units of LKFS) for Content on the channel.

K.4 COMMERCIAL ADVERTISING LOUDNESS MEASUREMENT

It is vital that, when Loudness of Short-form Content (e.g., commercial advertising) is measured, it be measured in units of LKFS including all audio channels[25] and all elements of the soundtrack over the duration of the Short-form Content.

K.5 COMMERCIAL ADVERTISING AT THE POINT OF INSERTION

In the case of insertion of Short-form Content (e.g., commercial advertising), it is vital that the Loudness, measured as per section K.4, of the inserted Short-form Content match the delivery channel's Loudness target value within +/-2 dB.

[25] Per BS.1770, the LFE is not included.

Annex L: Guidelines for Establishing and Maintaining Audio Loudness of Internet Streaming Services When Using Metadata-based and Non-metadata-based Codecs

L.1 INTRODUCTION AND SCOPE

The recommendations in this Annex are based on other sections of this Recommended Practice. This Annex contains the courses of action necessary to perform effective Loudness control for Internet Streaming services when using Loudness-metadata-based audio codecs (e.g., AC-3, AC-4, DTS-UHD, E-AC-3, MPEG-H Audio, xHE-AAC) and non-Loudness-metadata-based audio codecs (e.g., AAC, MP2, MP3).

L.2 LOUDNESS MANAGEMENT

The Operator's goal is to present to the audience consistent audio Loudness across Long-form and Short-form Content. The Operator should, whenever possible, present the Content to the audience with the most accurate and highest quality sound, free of any type of audio artifacts not contained within the original material as delivered by the Content supplier.

L.3 LONG-FORM AND SHORT-FORM CONTENT LOUDNESS MEASUREMENT

For metadata-based and non-metadata-based codecs, it is vital that Long-form Content Loudness be measured using dialogue as the Anchor Element, and Short-form Content Loudness be measured using the Full-program Mix as the Loudness value. See Section 5.

L.4 LOUDNESS FOR METADATA-BASED CODECS

For metadata-based systems, Loudness metadata must match the Loudness of the Content. As an example, ATSC document A/53, Part 5 [1], Section 5.5 mandates the carriage of dialnorm and correctly[26] set dialnorm values.

L.5 LOUDNESS FOR NON-METADATA-BASED CODECS

It is strongly recommended that all Content (Long-form and Short-form) on a Streaming delivery service be presented at **only one** specific and consistent Target Loudness[27]. Selecting a Loudness value between -23 and -27 LKFS[28] is recommended, unless there is a prior arrangement otherwise. See AES71, AES Recommended Practice "Loudness Guidelines for Over-the-Top Television and Online Video Distribution," [37] Annex D, Tables D1 and D2, for additional guidance.

[26] See Section 7.1 of this document for information on the use of dialnorm.

[27] It may not always be possible for a service to be operated at a single Target Loudness. What is vital, however, is that the Loudness of Short-form Content not be greater than that of the Long-form Content it accompanies.

[28] -23 to -27 LKFS represents a practical range that enables Loudness consistency while facilitating creative intent. For example, some Streaming Operators use -23 LKFS (conformant to EBU R128 s2 [36]), while others use -27 LKFS (provides greater headroom), and still others use -24 LKFS (the ATSC A/85 recommendation for TV broadcasting, or as a default in the absence of any other agreed-upon value).

L.6 LOUDNESS FOR MIXED METADATA-BASED AND NON-METADATA-BASED CODECS

Streaming services that use both metadata- and non-metadata-based codecs will provide consistent Loudness if Content is delivered as described in both L.4 and L.5 above. When further amplification is required for comfortable listening on devices, refer to ANSI/CTA-2075 "Loudness Standard for Over-the-Top Television and Online Video Distribution for Mobile and Fixed Devices" [38] for guidance.

L.7 LOUDNESS WHEN SWITCHING BETWEEN METADATA-BASED AND NON-METADATA-BASED SERVICES

When Streaming services switch between metadata-based and non-metadata-based coding, use of the methods in L.4 and L.5 above for the delivery of Content will help provide consistent Content Loudness.

L.8 SHORT-FORM CONTENT LOUDNESS MEASUREMENT

It is vital that, when Loudness of Short-form Content is measured, it is measured in units of LKFS using Full-program Mix, i.e., including all audio channels[29] and all elements of the soundtrack over the complete duration of the Short-form Content.

L.9 SHORT-FORM CONTENT LOUDNESS AT THE POINT OF INSERTION

In cases of insertion of Short-form Content, including Programmatic Advertising insertion, it is vital that Loudness of the inserted Short-form Content, as described and measured in Sections L.4 and L.5 above, matches the delivery channel's Loudness target value within +/-2 dB.

[29] Per BS.1770, the LFE is not included.

Annex M: Loudness and True Peak Quick Reference

Table M.1 ATSC A/85 Loudness and True Peak Quick Reference

<table>
<tr><th rowspan="2">Content type</th><th colspan="4">Loudness Measurement Method
(Anchor Element)</th><th rowspan="2">Maximum True Peak[30] (dBTP)</th></tr>
<tr><th colspan="2">Integrated Dialogue Loudness[31] (LKFS)</th><th colspan="2">Full-Program Mix Loudness (LKFS)</th></tr>
<tr><td>Long Form</td><td>Without metadata: -24*</td><td>With metadata: dialnorm** = Loudness</td><td colspan="2">Content Dependent[32]</td><td>-2</td></tr>
<tr><td>Short Form</td><td colspan="2">N/A</td><td>Without metadata: -24*[33]</td><td>With metadata: dialnorm** = Loudness</td><td>-2</td></tr>
</table>

It is strongly recommended that **all** Content be presented to the audience at **only one** specific and consistent Target Loudness[34]. Long-form and Short-form Loudness measurements should be performed per Table M.1, regardless of the source of the Content (e.g., including downstream-inserted Short-form content).

For delivery or exchange of Content without metadata (and where there is no prior arrangement by the parties regarding Target Loudness), the Target Loudness should be -24 LKFS*.

For **Streaming delivery services only**, a single Target Loudness between -23 to -27 LKFS is recommended[35], unless there are other prior arrangements between Content-exchange or delivery parties.

[30] Minor measurement tolerance for True Peak values of up to approximately ±0.5 dB is anticipated, due to meter variations, and is acceptable.

[31] It is vital that the average **Dialogue Level** be measured and reported as the Loudness for Long-form Content. This should be performed using BS.1770-1 (i.e., without the relative-level gate added in subsequent versions), along with a Dialogue-Gating algorithm, and measured over the **entire duration** (not act by act) of the Content's composite mix. (See Section 5.1.1.)

[32] Full-program Mix should **not** be used for Loudness measurement of Long-form Content except in the rare case that the Long-form Content does not contain any dialogue whatsoever. See Section 5.2.4.

[33] Applies to Short-form Content using the ITU-R BS.1770 (-3 or later) measurement method. See Section 5.1.1.

[34] It may not always be possible for a service to be operated at a single Target Loudness. What is vital, however, is that the Loudness of Short-form Content not be greater than that of the Long-form Content it accompanies.

[35] This represents a practical range enabling Loudness consistency while facilitating creative intent. For example, some Streaming Operators use -23 LKFS (conformant to EBU R128 s2 [36]), while others use -27 LKFS (provides greater headroom), and still others use -24 LKFS (the ATSC A/85 recommendation for TV broadcasting, or as a default in the absence of any other agreed-upon value).

* Measurement tolerance of up to approximately ±2 dB around this value is anticipated, due to measurement variations. Content Loudness should **not** be targeted to the high or low side of this tolerance. (See Section 6 for details.)

** Or equivalent Loudness metadata.

<End of document>