

INTER-PROGRAM LEVEL JUMPS IN BROADCAST

Thomas Lund

TC Electronic, Denmark

ABSTRACT

Around the world, DTV has experienced a growing problem with level jumps; especially when wide dynamic range programming is followed by narrow dynamic range material; for instance, commercials, promos or sports. Important contributors to such inter-program level jumps have been investigated and are reported.

The paper discusses how to describe audio programs with different dynamic range signatures adequately using loudness visualization and a few long-term descriptors - at the same time avoiding distortion and unacceptable level fluctuations. It is shown how descriptors derived from the open ITU-R BS.1770 standard can be employed to streamline the flow of audio in music, film and broadcast.

The Universal Descriptors presented are based on more than four years of research, and they are valid for any type of program - talk, music, film, commercials etc. Universal Descriptors may be used for delivery specifications and internally at the station for low labor, tight control of loudness and dynamic range across broadcast platforms such as HDTV, SDTV, Mobile TV and IPTV. The Universal Descriptors improve on DTV audio predictability and listener experience, regardless of the type of data reduction codec used for a certain platform (e.g. MP3, AAC or AC3).

INTRODUCTION

Current methods for measuring level in broadcast are based on peak level detection [1], which makes low dynamic range material appear loud, see *Fig 1 center*. Therefore, blind clipping and limiting is employed excessively in CD and commercial production today. In previous work [2, 3, 4, 5, 6], distortion in consumer equipment, sample rate converters, and data reduction codecs has been documented as side effects of such practice, and a shift from measuring level to measuring loudness is therefore recommended.

Unlike level, *loudness* is subjective, and listeners weigh the most important factors differently: Sound pressure level, Frequency contents and Duration. In search of an “objective” loudness measure, a certain Between Listener Variability (BLV) and Within Listener Variability (WLV) must be accepted, meaning that even loudness assessments by the same person are only consistent to some extent, and depends on the time of day, her mood, the degree of attention etc. BLV adds further to the blur, when sex, culture, age etc. are introduced as variables. In the real world, unknown reproduction systems, of course, add even more blur.

Because of the variations, a generic loudness measure is only meaningful when it is based on large subjective reference tests and solid statistics [7]. Together with McGill University in Montreal, TC Electronic has undertaken loudness model investigations and evaluations. The results denounce a couple of Leq measures, namely A and M weighted, as trustworthy, generic measures of loudness. In fact, a quasi-peak (“PPM”) meter showed better judgment of loudness than Leq(A) or Leq(M). Even used just for speech, Leq(A) is a poor pick [7, 8, 9], and it performs worse on music and effects. An appropriate choice for a low complexity, generic measurement algorithm has been known as Leq(RLB), and is now part of the ITU-R BS.1770 standard, which defines the measurement of loudness and of true-peak level.

DYNAMIC RANGE TOLERANCE

DTV has the potential to carry wide dynamic range audio, but this aspect is not important to the general consumer [5]. What matters most is consistency of loudness within programs and between programs, and speech intelligibility. Consumers have a defined Dynamic Range Tolerance, DRT, specific to their listening situation, see *Fig 1 left*. The DRT is defined as a Preferred Average window with a certain peak level Headroom above it. The average level has to be kept within certain boundaries in order to maintain speech intelligibility, and to avoid music or sound effects from getting annoyingly loud or soft.

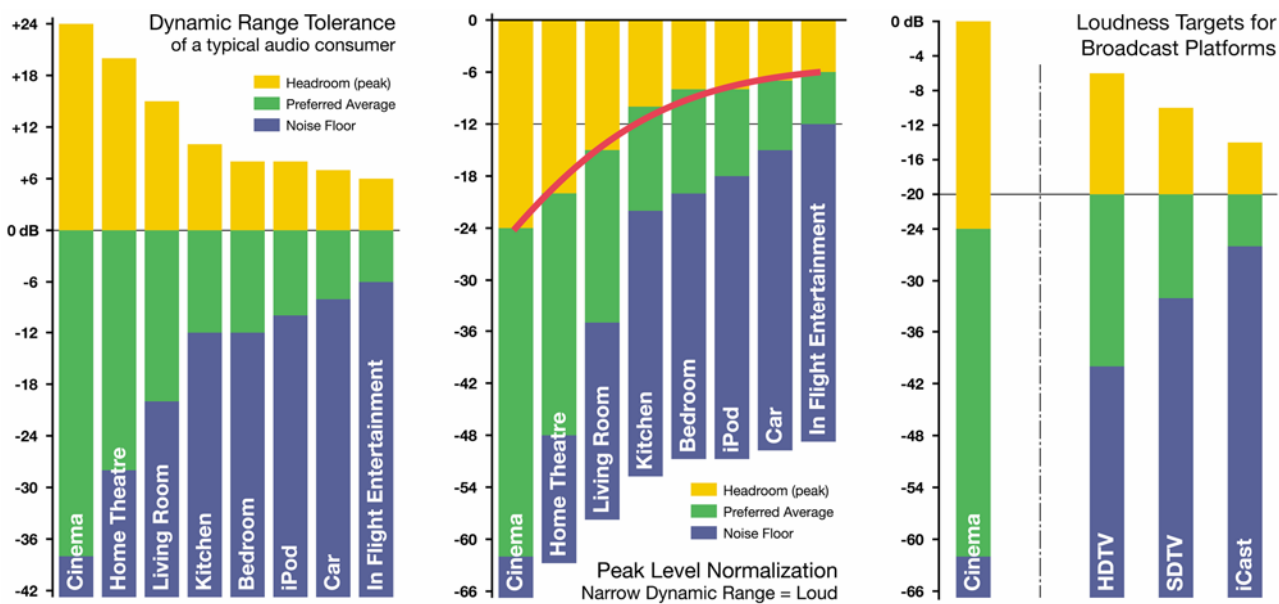


Fig 1, Consumers prefer low loudness variation under some listening conditions. Left: Dynamic Range Tolerance of the average consumer under different listening conditions. Center: When peak normalization is used, low dynamic range material end up the loudest (red line). Right: Suitable loudness variation for various broadcast platforms: HD, SD and Mobile TV.

It should be noted that TV listeners more often object against audio when the dynamic range is too wide, than when it is too restricted. The only reproduction situation where a wide dynamic range is a recognized benefit to the general public is in a cinema. Consequently, it is a main concern for the broadcaster to get speech intelligibility and consistency of loudness catered for not only on HDTV, but across all platforms.

Loudness history and end-listener Dynamic Range Tolerance in combination should therefore be visible from production onwards. The engineer, who may not be an audio expert, should be able to identify and consciously work with loudness developments within the limits of a target distribution

platform, and with predictable results when the program is transcoded to another platform, see *Fig 1 right*. A loudness meter might consequently use a color-coded display so it's easy to identify target level (green), below the noise-floor level (blue), or loud events (yellow).

The aim is to center dynamic range restriction around average loudness, in the case of *Fig 1 right* – 20 LFS, thereby automatically avoiding to wash out differences between foreground and background elements of a mix. When production engineers realize the boundaries they should generally stay within, less dynamics processing is automatically needed during distribution, and the requirement for maintaining time-consuming metadata at a broadcast station is minimized. Note how different the broadcast requirements are from those of Cinema.

THE ITU-R BS.1770 STANDARD

ITU-R BS.1770 has been specified for measuring mono, stereo and 5.1 material. The loudness part is not really measuring loudness, but rather an estimate of the *gain offset* required to match the loudness of one sound clip to that of a reference sound.

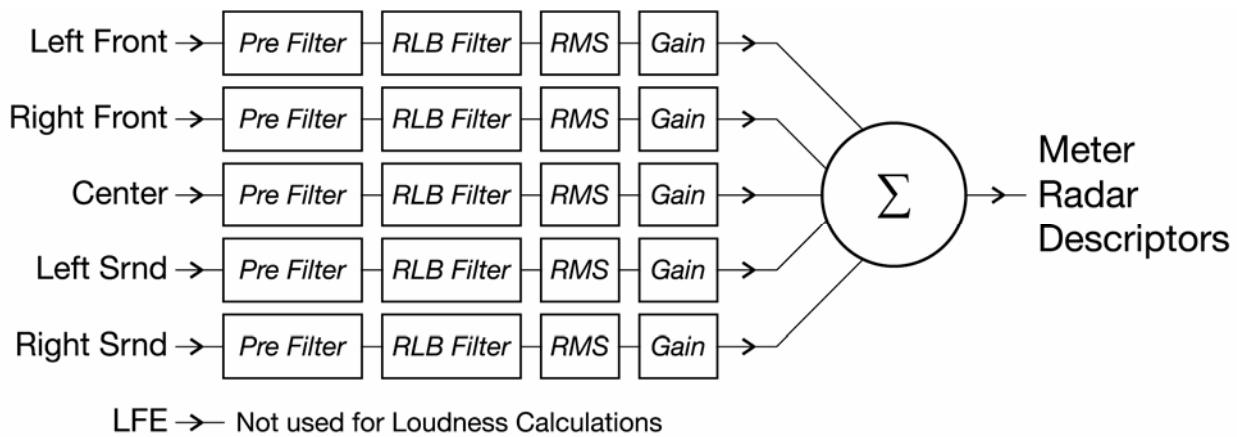


Fig 2, The ITU-R BS.1770 measurement used for short-term, mid-term and Universal Descriptors. Note how the LFE channel is not part of the loudness calculation.

BS.1770 has been designed for use across audio formats from mono to 5.1, with more headroom per channel the higher the format, when the Loudness is kept constant. See block diagram in *Fig 2*. It has been debated if the loudness part of the new standard is robust enough. As a global loudness reference, it will be exploited where possible. However, with homogenous mono material, Leq(RLB) has been verified in independent studies to be a relatively accurate measure, taking its simplicity into account [7, 8]. It should be noted that BS.1770 ended up not incorporating Leq(RLB), but Leq(R2LB), now formally known as Leq(K).

The other aspect of BS.1770, the algorithm to measure *true-peak*, is built on solid ground. Inconsistent peak meter readings, unexpected overloads, distortion in data reduced delivery and conversion etc. has been extensively described [2, 3, 4, 5, 6]. In liaison with AES SC-02-01, an over-sampled true-peak level measure was specified.

Other combined loudness and peak level meters exist already, for instance the ones from Dorrroughs, but BS.1770 offers a standardized way of measuring these parameters. A loudness meter may use either the measurement unit of LU (Loudness Units) or LFS (Loudness Full Scale). LU and LFS are measurements in dB, reflecting the estimated gain offset to arrive at a certain Reference Loudness (LU) or Maximum Loudness (LFS) as defined in BS.1770. A Reference point for LU has not been fixed at the time of writing. Until this value has been set in stone, LFS (or

“LKFS”, pointing to the Leq(R2LB) weighting of BS.1770), should be chosen to avoid ambiguous use of the term LU.



Fig 3, Loudness meter adhering to ITU-R BS.1770.
 Left: Feature film (Pirates of the Caribbean): Low Loudness Consistency typical of film.
 Center: Broadcast (German national broadcaster): Consistency suitable for broadcast.
 Right: Pop music (Madonna): Overly high Consistency typical of pop. Generates listener fatigue.

To control loudness developments consistently over time, a compelling solution is to have the loudness history visualized from production onwards. This way, a mixing engineer or a journalist is able to arrange and identify long-term as well as short-term loudness developments inside a program. A meter example with such virtues is shown in *Fig 3*.

UNIVERSAL DESCRIPTORS

Visualization of loudness developments can be combined with program duration descriptors also rooted in ITU-R BS.1770. Such Universal Descriptors may be used during ingest or inside broadcast servers to define a level offset for each program, and for delivery specifications.

It has been suggested to reference broadcast programming to the level of its dialog, which to some extent works for film. However, this has bad consequences in other types of production, where mixing esthetics between programs may vary significantly, where dialog not always take center stage, where any type of sound may be disturbing, and where the consumer Dynamic Range Tolerance is lower than in a cinema. The sound of a phone ringing in a commercial, John Frusciante’s guitar, or a fighting scene in Pirates of the Caribbean can all make some people grab the remote, and should naturally have an influence on the loudness of a program. Even if a TV station is news only, documentaries or drama, it will still have accompanying sounds that can be annoying.

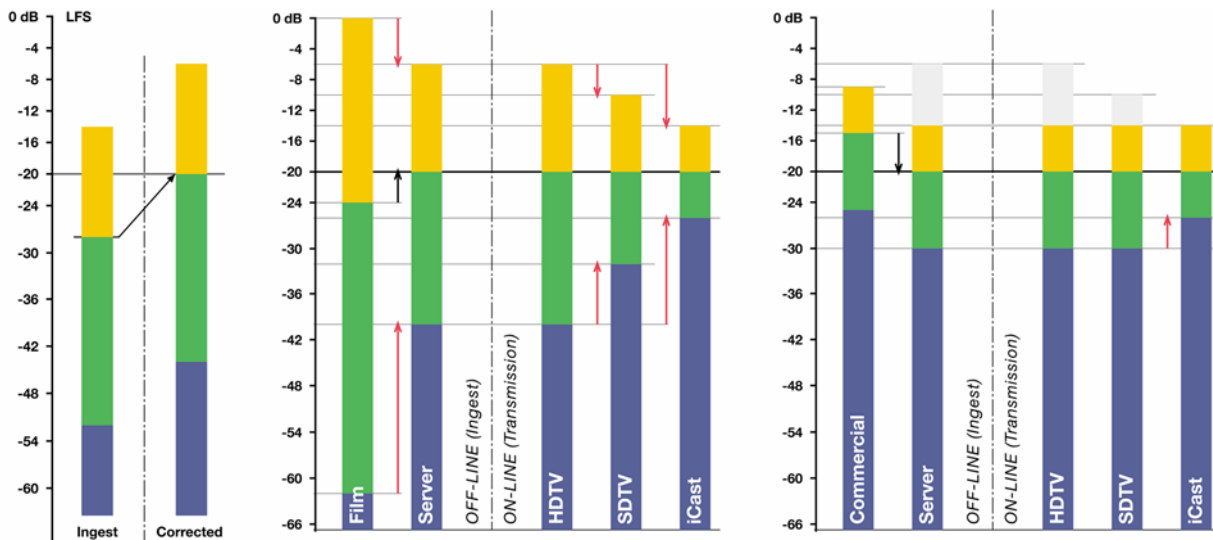


Fig 4, Center of Gravity specifies an optimum gain offset per program (black arrows).. The illustration also shows additional trickle-down processing from HD to SD to mobile platforms. Left: Center of Gravity is 8 dB too low, which is corrected during ingest or on the broadcast server. Center: Feature film requires Center of Gravity correction plus dynamic loudness processing (red arrows). Right: A commercial typically needs little more than Center of Gravity correction.

For audio meant to be distributed over a number of platforms, it is fundamental to define its Center of Gravity. With this center point well defined, it is simple to transcode a given program to any platform, see Fig 4. However, studies of dialog from broadcast and film, music, commercials and effect sounds have led to the conclusion that at least one more telling parameter than Center of Gravity should be used for program delivery specification, namely Consistency, which is a long-term statistical measure also rooted in the loudness part of BS.1770. Consistency has been designed to indicate intrinsic loudness variations inside a program. A combination of Center of Gravity and Consistency is valid across a wide range of program material.

With the variety of audio platforms we use on a daily basis, chances are that a certain film, piece of music or broadcast program hasn't been produced with our immediate Dynamic Range Tolerance in mind. Audio professionals should therefore try to anticipate the downstream processing that might happen to their production, no matter if the changes take place at a broadcast station or at a consumer.

When dynamic range transcoding occurs only below and above the Center of Gravity area of a music track or a broadcast program, as shown in Fig 4, its core is automatically preserved; thereby avoiding foreground and background elements of a production to get messed up. Note how the system works coherently from production onwards without the need for setting or controlling metadata. When it comes to a broadcast station, long-term loudness normalization (level offsets) should ideally be taken care of during ingest or inside a file server.

INVESTIGATION OF INTER-PROGRAM LEVEL JUMPS

In order to determine the best method to avoid inter-program level jumps without employing dynamics processing, more than 1500 program transitions between wide dynamic range material and narrow dynamic range material all containing dialog have been investigated. Audio test programs consisted of: 1) Wide Dynamic Range material, *WDR*. 36 different full length theatrical movies. 2) Medium Dynamic Range material, *MDR*. 70 different full length programs targeted for DTV broadcast. Documentaries, drama, sports. 3) Narrow Dynamic Range material, *NDR*. 115 different, broadcast promos and commercials.

All WDR material was 5.1. Duration per item between around 90 and 150 minutes. Film came from US/Can (25), Europe/Australia (7) and Japan (4). 7 were based on theatrical releases, 30 on unprocessed DVD's. When linear PCM was available, this was used. Second choice was DTS data reduction, third choice Dolby AC3. MDR material consisted of a mix of stereo and 5.1. Duration per item between 15 and 60 minutes. 62 items were linear PCM, 8 were Dolby E. NDR material consisted of a mix of stereo and 5.1. Duration per item between 20 secs and two minutes. 109 items were linear PCM, 6 were Dolby E.

Each WDR item was broken randomly up into 10 parts, each part at least 4 minutes long, making room for 9x2 transitions to/from NDR interruptions. Tests were performed with all audio items normalized using 1) dialog and Leq(A), 2) dialog and Leq(K), and 3) non-discriminating Leq(K). The test was repeated using MDR items instead of film. MDR items were broken up into 8 parts, each part at least 2 minutes long. The level change across each transition was measured using the Leq dose measure specified in ITU-R BS.1770 with a pre-roll window of 15 secs, and a post-roll window of 7 secs. Other measures were also employed, but they are not reported here.

WDR-NDR test results

Differences between the three normalization methods were found. The level jumps measured using dialog Leq(K) normalization produced slightly improved results over dialog Leq(A). On the average a 0.8 dB improvement. The level jumps measured using non-discriminating Leq(K) gives a more pronounced improvement, typically amounting to 5.5 dB. No significant difference was found between WDR material using linear PCM, DTS or AC3 with regard to level normalization.

MDR-NDR test results

Differences between the three normalization methods were found. The level jumps measured using dialog Leq(K) normalization were lower than with dialog Leq(A), on the average a 1.2 dB improvement. The level jumps measured using non-discriminating Leq(K) again gave a more pronounced improvement, typically amounting to 4.2 dB.

CONCLUSION

Systematic differences between the ability of three level normalization methods to minimize inter-program level jumps were found. Neither method could guarantee inter-program transition-jumps of less than 6 dB, but program normalization based on Center of Gravity was found to be a significant improvement over normalization using either dialog Leq(A) or dialog Leq(K). Different long-term normalization methods for use in digital broadcast were ranked from best to worst based on their ability to reduce inter-program level-jumps:

1. Experienced engineer hands-on
2. Center of Gravity normalization based on Leq(K)
3. Non-discriminating PPM 50% normalization
4. Dialog based Leq(K) normalization
5. Dialog based Leq(A) normalization
6. Sample peak level normalization

Items 1-3 perform notably better than items 4-6. Item 3, the PPM 50% measure, has been suggested for audio descriptor use by IRT in München, Germany. Its performance was found to be somewhat close to that of Leq(RLB).

For mono, stereo and 5.1 broadcast programming, it is concluded that *inter-program* level jumps cannot be avoided if only static level offsets are employed. However, level jumps between typical

broadcast programs may be significantly reduced when normalization is based on the simple, open standard dose measure included with ITU-R BS.1770 instead of normalization based on peak level or dialog level. With efficient normalization procedures in place for any kind of audio material, less dynamics processing is therefore required at the station. Furthermore, for broadcast platforms where a lower dynamic range is indicated, see *Fig 1*, *within-program* level jumps may be reduced consistently and automatically without sacrificing audio quality. This principle is referred to as “trickle-down processing” in *Fig 4*.

Center of Gravity and Consistency, in combination with the short-term meter and loudness history display presented here, provides an easy, standardized and holistic solution which can be used for any type of audio production, and regardless of which equipment or data reduction system is employed. The concept co-exists with systems requiring metadata, though such data need not necessarily be collected or maintained when the guidelines given here are followed.

In conclusion, Center of Gravity alignment enables audio with different dynamic range signatures to be mixed without having low dynamic range content, such as an aggressive commercial or a pop CD, emerging louder than everything else. Used at a broadcast station - during ingest or on the station server - Center of Gravity alignment minimizes inter-program level jumps, and enables a processing-free window to be defined across broadcast platforms. Automatic trickle-down processing during transmission can ease the workload otherwise required to maintain a multitude of broadcast platforms.

REFERENCES

- [1] Katz, B. “Mastering Audio. The Art and The Science”. 2nd edition. Focal Press, 2007.
- [2] Nielsen, S.H. & Lund, T. “Level Control in Digital Mastering”. Proc. of the 107th AES convention, New York 1999. Preprint 5019.
- [3] Dunn, J. “Digital Filter Overshoot and Headroom”. DA Tech paper for Audio Precision, June, 2000.
- [4] Nielsen, S.H. & Lund, T. “Overload in Signal Conversion”. Proceedings of the 23rd AES conference, Copenhagen, 2003.
- [5] Lund, T.: “Distortion to The People”. Proc. of the 23rd Tonmeistertagung. Leipzig, 2004. Paper A05.
- [6] Lund, T. “Stop Counting Samples”. Proc. of the 121st AES convention, San Francisco, 2006. Preprint 6972.
- [7] Skovenborg, E. & Nielsen, S.H. “Evaluation of Different Loudness Models with Music and Speech Material”. Proc. of the 117th AES convention, San Francisco, 2004. Preprint 6234.
- [8] Lund, T. “Specifying Audio for HD”. Proc. Of NAB Broadcast Engineering conference, Las Vegas, 2007.
- [9] Moore, B.C.J., Glasberg, B.R. & Stone, M.A. "Why Are Commercials so Loud? - Perception and Modeling of the Loudness of Amplitude-Compressed Speech". JAES, vol.51:12, 2003.