SMPTE Meeting Presentation

Loudness vs. Speech Normalization in Film and Drama for Broadcast

Thomas Lund

TC Electronic A/S, Risskov, Denmark. tl@tcgroup.tc

Esben Skovenborg

TC Electronic A/S, Risskov, Denmark. esk@tcgroup.tc

Written for presentation at the

SMPTE 2014 Annual Technical Conference & Exhibition

Abstract. *This paper extends a previously published study of the differences between level normalization of programs using the two dominant methods: Loudness normalization and speech ("dialog") normalization. Instead of adding to the continuing debate of the subjective merits of one method over the other, important technical aspects are examined empirically.*

The difference in normalization level between Loudness and speech measures was up to 14 dB. For all films, the Loudness method provided the greatest headroom. Half the films could be broadcast at a fixed target level of −24 LKFS (loudness, K-weighted, relative to full scale) without dynamics processing.

When it was speech normalized, not a single film could be broadcast at the same target level without applying dynamics processing. The study furthermore found a systematic difference between manual speech measurement and automatic speech measurement.

The measured movies were also compared to the 2013 season of a high profile TV drama production. The loudness properties of the drama were found to be comparable to the movies. In addition, different broadcast/playback paths were found to have markedly different effects on the Loudness Range of the drama series.

Uncertainties in classification, definition, and measurement are summarized and compared to the requirements for precision in Advanced Television Systems Committee (ATSC) and the European Broadcasting Union (EBU) loudness-based standards. Finally, consequences of these findings are discussed relative to ITU-R BS.1864, the International Telecommunication Union's standard on broadcast program exchange.

Keywords. Loudness normalization, speech normalization, dialog intelligence, metadata, PLR, headroom, measurement uncertainty, Loudness Range, ATSC A/85, EBU R128, ITU-R BS.1770, ITU-R BS.1864.

Introduction

For decades, the music and broadcast industry has primarily been using peak-level meters for measurement and normalization. Consequently, commercials have been considerably louder than ordinary programs, and hyper-compressed music tracks have been systematically louder than tracks with a more natural distribution of level. [1-3]

To prevent excessive inter-program level jumps from damaging digital television (DTV), the International Telecommunications Union Recommendations Section (ITU-R) specified an open method for measuring Loudness rather than level. [4] Similar efforts had been tried before, but they had never caught on widely in analog broadcast. [5]

Loudness is a perceptual property of an audio signal when it is reproduced acoustically and listened to. It is a complex, nonlinear function of amplitude, frequency, and bandwidth. ITU's loudness measure is basically an Leq calculation with a high-pass frequency weighting. The original monophonic Leq(RLB) measure has subsequently evolved into Leq(K), which can be used with mono, stereo, and 5.1 formats. In 2010, ITU-R decided to adopt a cross-genre friendly adaptive level-gate into the standard, then labeled BS.1770-2. Broadcast organizations around the world readily incorporated this significant improvement. In 2013, ATSC followed suit, thereby making the same definition the foundation of loudness regulation everywhere.

At the time of writing, the BS.1770-3 revision is in effect. This latest update from 2012 is identical to revision 2 concerning Loudness, but it contains a less ambiguous definition of the "True-peak" measurement than the previous version.

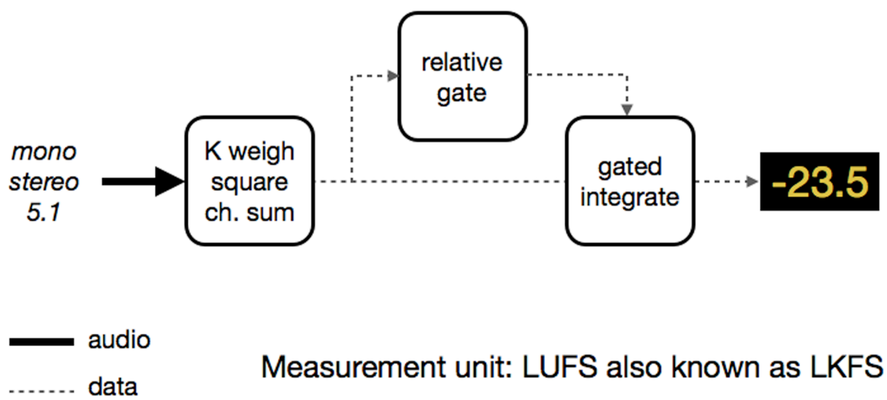

Figure 1. ITU-R BS.1770 (Program) Loudness including adaptive level-gate.

Terminology

Figure 1 shows the BS.1770 measurement with the first block consisting of the "K" frequency weighting of the input samples, squaring and summing across channel. The preprocessed samples are integrated into the overall loudness level, controlled by a gating based on a threshold level relative to the overall average.

The integrated gated loudness level, measured over a program, is denoted "Program Loudness", or in this paper just "Loudness".

The peak to Loudness ratio (PLR) measures the ratio (in dB) between maximum True-peak level and Loudness of a program, film, or music track. Both metrics are defined in ITU-R BS.1770.

Headroom denotes the ratio between the maximum peak level a signal path or a system can handle and its target loudness level. For example, in ATSC A/85, headroom, the ratio between the target at −24 LKFS and the maximum True-peak level at −2 dBTP, is 22 dB. Note: The unit [LKFS] (loudness, K-weighted, relative to full scale) is the same as [LUFS] loudness units related to full scale), the latter being International Organization for Standardization (ISO) compliant.

Sometimes “dynamic range” is wrongly used about one of the above. However, dynamic range should be reserved to describe the difference between the lowest and highest levels handled by a signal path, which is its traditional meaning in pro audio.

Headroom, PLR, and Audio Quality

PLR is a measure of how demanding on headroom a program or music track will be for the downstream signal path. As long as PLR is lower than the available headroom, a static gain correction ("normalization") without peak limiting or clipping can be performed.

Modern pop/rock music and broadcast commercials generally have the lowest PLR values, which is a sign of extensive use of compression and limiting in the production process. A recent study compares PLR over time of the most popular music tracks in the U.S., U.K., and Germany (Fig. 2). It reveals a maximum around the introduction of compact disc (CD) in the mid-1980s, and a significant decline ever since. Some tracks today have a PLR of less than 8 dB. [6] At the opposite end of the scale, feature films and classical music have the highest PLR values, sometimes over 30 dB. [7]

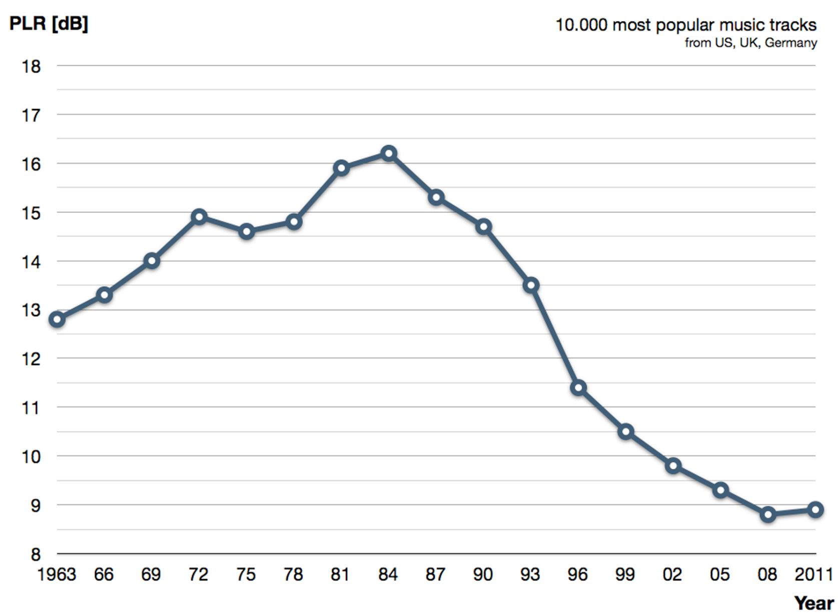

Figure 2. PLR trend in pop music since 1963, adopted from Ortner. [6]

Distribution platforms also have different amounts of headroom. Whereas analog broadcast on a good day has a small headroom of 12 dB, DTV has a respectable 22 dB as default. Mobile TV and pod devices are under certain restraints and typically provide less. [8] Sony offers 18 dB headroom for its portable players. [9] Apple iPods and iTunes Radio are also on the decent side with 16.2 dB, while many streaming services and most Internet radio stations are bad for audio quality because of the measly headroom they offer.

If headroom is too low, users either have to live with level jumps between tracks or programs, or with a disgusting sound quality when fine loudspeakers or headphones are used. Today, the audio quality bottleneck for a home listener with a decent reproduction system is typically content with a low PLR, a distribution platform with a low headroom, and/or audio that has undergone lossy data reduction.

Dialog, Anchor, and ATSC A/85

While analog broadcast generally relied on peak-level metering, audio for cinema movies has traditionally been centered on regular speech or “dialog.” The optical sound track provided only a limited dynamic range, so a softly mixed dialog would drown in noise, and there wasn’t headroom for loud music or effects. In today’s digital television and in digital cinema, the potential differences in loudness between elements of a mix are higher because much more headroom is now available.

Original ATSC audio guidelines for DTV, A/53 from 1995, adopted the cinema method and called for “average spoken dialogue” as the normalization technique to be used in broadcast. In 2009, a new recommended practice, the A/85, was released. ITU’s BS.1770 replaced the previous Leq(A)-based measure, and the term anchor was used to describe the normalization reference for a program rather than just its dialog. Because of the CALM Act, revisions of A/85 from 2011 furthermore changed the requirement when measuring commercials to take all sources into account, and not just “dialog” or “anchor.” [10]

This paper investigates transparency and headroom consequences of using Loudness versus dialog measurements when normalizing feature films and broadcast programs.

Loudness Range

The Loudness Range (LRA) is a descriptor which can quantify the variation in a time-varying loudness measurement. The Loudness Range measures the variation of the loudness on a macroscopic timescale, in units of LU (i.e. on a dB scale) [17].

The LRA value remains unchanged if programs are gain-normalized, or if they are encoded with lossy data reduction. LRA also typically remains unchanged if a program undergoes peak level limiting, but it decreases notably if dynamic compression is applied somewhere along a given signal-path [16]. An LRA measurement can also be applied to a single component or stem of a program, e.g. speech or music.

In this study, LRA measurements were made to determine whether consumers got the same audio when listening to an episode of TV drama via a DVD or via HBO Nordic. Furthermore, the LRA was measured for a modern drama production, and the LRA was compared with contemporary pop music and movie genres.

Test Method

Loudness, PLR, and speech level were measured for
- the 35 feature films on the Internet Movie Database (IMDb) Top100 list since 1995 [7]

Loudness and LRA were measured for
- all 10 episodes of ubiquitous, high-profile television drama series Game of Thrones, from 2013.

Loudness and PLR were measured in

- 127 British, German, Brazilian, French, American, Japanese, Chinese, Norwegian, and Danish broadcast programs produced from 2010 to 2013, all of which were regular programs with a duration of 20 min or more.

- 650 pop/rock tracks from the U.S., U.K., Brazil, Sweden, and Denmark released in 2008 to 2012; all the tracks were from a personal collection.

The movie and television drama analysis was based on the AC-3 streams found on the digital video disc (DVD) versions of the films (i.e., Dolby Digital). The stream containing the 5.1 channel audio with the original language of each film was used. The AC-3 streams were decoded and analyzed without applying any dynamic compression (DRC). AC-3 decoders in consumer equipment, such as DVD players, would generally apply DRC based on AC-3 metadata, according to different profiles and depending on options in the device (such as “night mode”). To avoid such ambiguity and compression, DRC processing was bypassed in the present analysis.

Regular broadcast analysis was based on linear PCM (48 kHz, synchronous) with programs being either stereo or 5.1. Music was measured bit cloned from CD as stereo linear pulse code modulation (44.1 kHz, synchronous).

Measuring Loudness and Speech Level

Loudness and True-peak level were in these studies measured according to ITU-R BS.1770-3. Speech level may be measured manually, by an operator isolating some intervals that he considers as representative speech for the program in question, and then measuring the loudness of those intervals with a Loudness meter.

Alternatively, it is possible to measure the loudness of speech automatically. In 1969, Belger outlined “a device for automatic discrimination between speech and music” and described how that device could be coupled with a loudness meter. [11]

Nevertheless, the integration of any (automatic) speech classifier into the loudness measure opens up several uncertainty issues, described in detail in [7]:

- Definition of multichannel *topology*: Should there be individual speech classifiers in all channels?

- *Classification* of speech: Should e.g. shouting and speech over action sounds be excluded from the measurement?

- *Measurement* of speech: Should the level-gate in BS.1770-3 be applied in addition to the speech ‘gate’?

Results

The biggest difference between measuring Loudness and speech level was seen in feature films. The consequences for headroom and transparency are described below. Further results were provided in Skovenborg and Lund. [7]

Loudness vs. Speech: Headroom

For all 35 movies, normalization based on Loudness was found to require less headroom during distribution than that had the same movie been normalized using speech level. Results are shown in Fig. 4. Bars above the 0 line indicate that limiting is needed, or clipping will occur when a target level of −24 LUFS/LKFS is observed. The automatic speech measurement, used here, comprises the speech classifier of the "Dialogue Intelligence" (DI) algorithm, using an Leq(K) measurement as loudness integration, gated by the output of the speech classifier [7]. Speech Normalization refers to setting the normalization level for the program to the DI-gated loudness level.

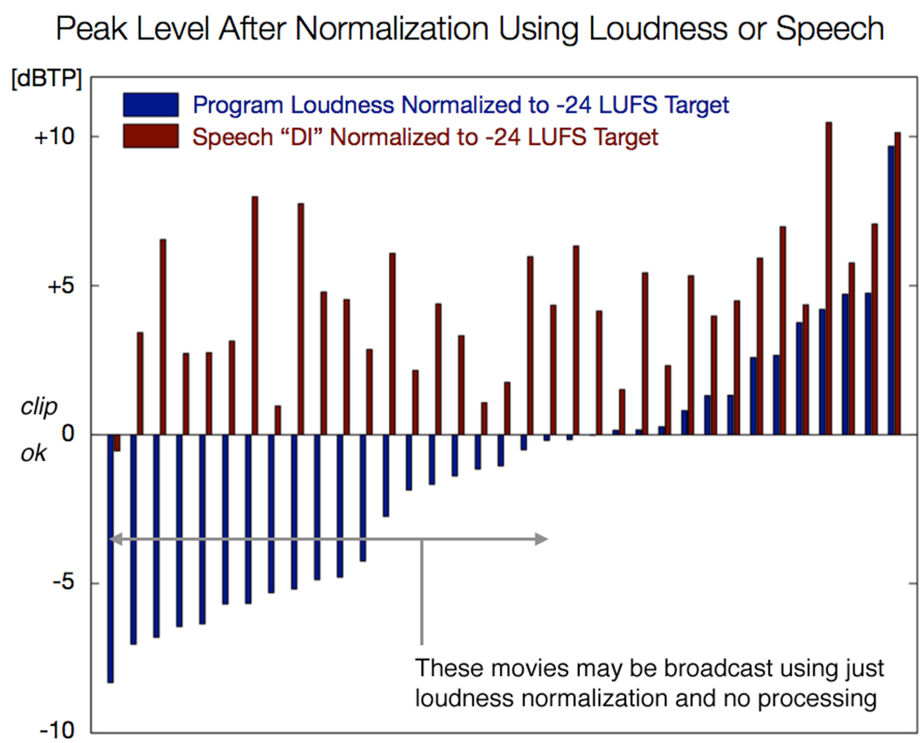

Figure 3. Max True-peak level in movies after normalization by Loudness (left bar/blue) or normalization by speech (right bar/red). Adopted from [7]

The median peak level after normalization on Loudness is −1.0 dBFS, whereas for dialog normalization, it is +4.0 dBFS. The figure shows that more than half the movies could be distributed without any processing when Loudness is normalized to −24 LUFS. For all the movies, processing would be needed to do the same in the case where normalization is based on auto-speech "DI."

The target level for Loudness normalization is −24 LUFS in ITU BS.1864, ATSC A/85, and TR-B32, while it Is −23 LUFS in EBU R128.12 These levels represent a compromise: The target should be low enough to provide sufficient headroom for programs of wide Loudness Range or PLR, yet not too low, because the gap to previous operating levels of broadcast platforms, infrastructure, and procedures would then be unacceptably wide.

Even though DI normalization seems to demand a lower target level than Loudness normalization, implementing this would need some critical metadata to specify the method/level that had been used, which again could cause problems with compatibility and robustness. Hence, normalizing by aligning the actual audio of programs to the target level is often preferred.

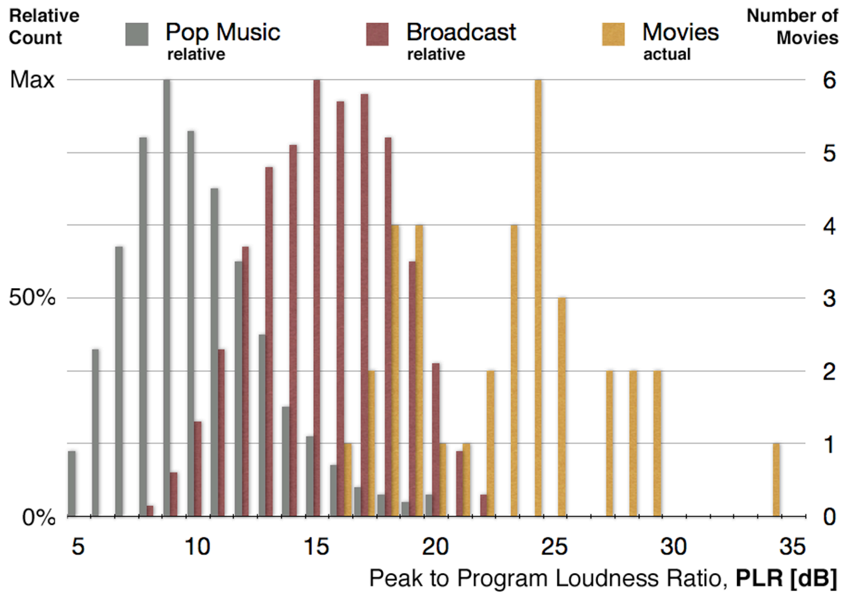

Fig 4. Measured PLR in new Pop Tracks, Programs and Movies

Figure 4 shows the PLR of all 35 movies and summarizes the results from the broadcast program and pop/rock track investigations. PLR numbers reflect the amount of headroom needed to convey each program when normalized based on Loudness.

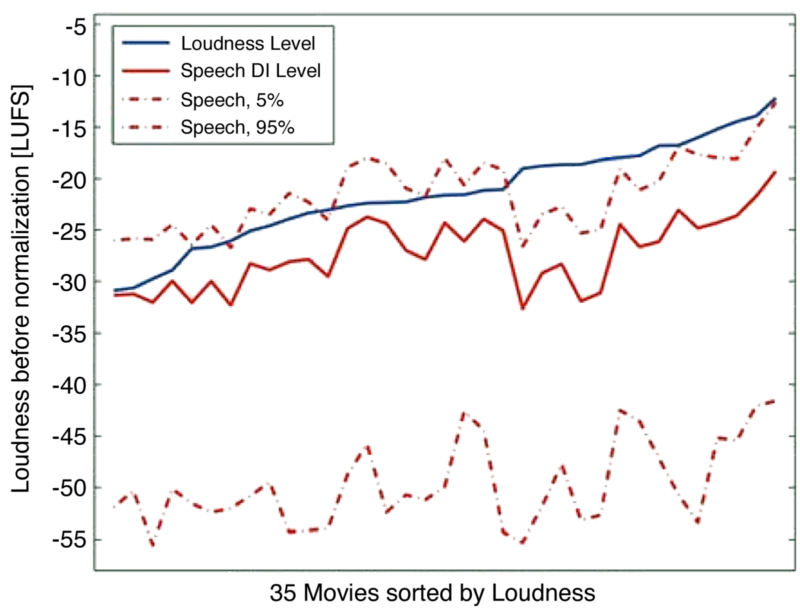

Figure 5. Loudness variation of blocks classified as speech, for each movie. Adopted from [7]

Loudness variation in blocks classified as speech

The loudness variation in blocks classified by DI as speech, after discarding the top 5% and bottom 5%, is between 23 and 35 LU, Figure 5.

The time-varying loudness is measured in overlapping blocks of 0.4 s length, i.e. the Momentary loudness [12]. These loudness blocks are what the integrated, speech-gated loudness level is calculated from.

The integrated speech level (solid red line, Fig 5) is much closer to the top 5% than to the bottom 5% of the blocks, which is due to the power-domain averaging taking place in the Leq calculation.

Manual vs. Automatic Speech Measurement

Whereas the Loudness measurement and metering of programs are specified in international standards, the measurement of speech level is more loosely defined - and proprietary technology may be involved.

In order to further assess the automatic dialog measurement, the so-called speech anchor of 10 movies was separated by hand and used as a basis for an independent measure of the films' dialog level. [7]

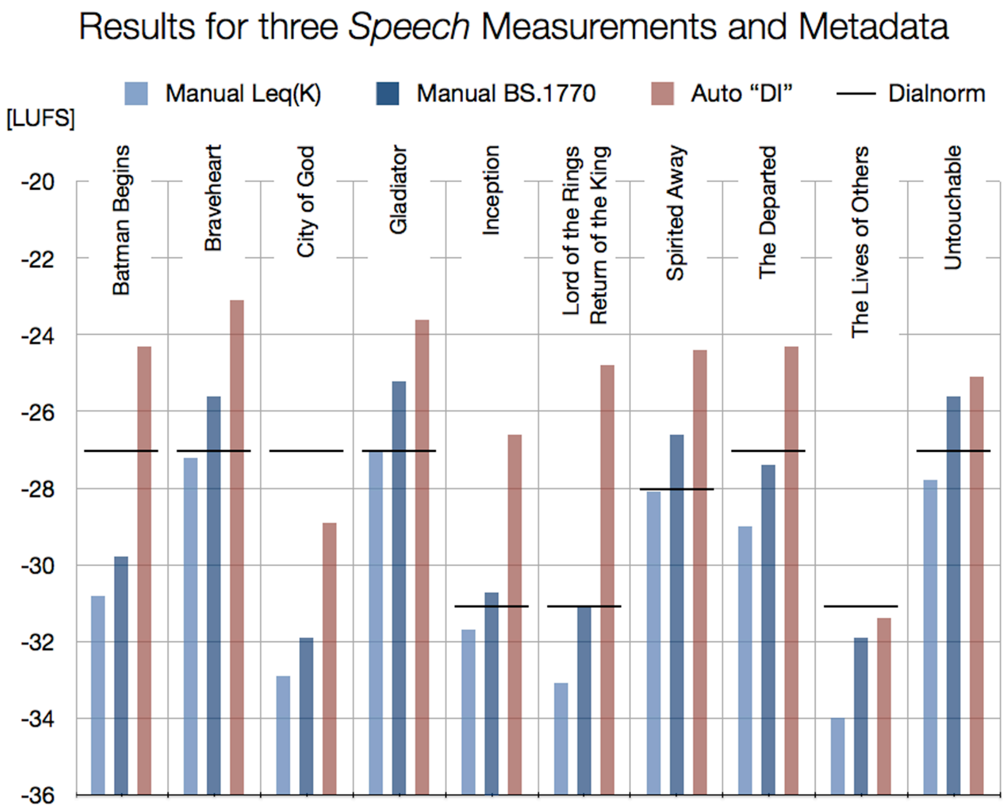

Figure 6. Measuring speech level in movies manually or automatically; and Dialnorm value on DVD.

All editing and listening were performed blindly; the experienced subject/editor did not know anything about the results of the automatic measurements when selecting and measuring suitable dialog segments from each movie. The editing criterion was normal speech, that is, not whispering, not shouting, not singing, and with no competing noise or music. Atmospheric sounds such as wind, birds, traffic, air-con, and background music were allowed in the assemblies.

Figure 6 summarizes the differences between measuring 5.1 speech assemblies with BS.1770-1 or BS.1770-3 and running the movie through an automatic "DI" measurement [7]. It also indicates the Dialnorm value encoded on each DVD.

After completing the manual speech segmentation, the performance of the automatic speech detector was evaluated subjectively on some of the same movies. The algorithm performing the speech-gated loudness measurements was modified to also process the audio such that the output would be attenuated by 20 dB when the audio was classified as non-speech and simply bypassed when classified as speech. Thus, the "speech" segments would appear noticeably louder than the surrounding “non-speech,” and the performance of the speech-classifier could thereby be evaluated by listening to the actual film mix.

By applying human intelligence in evaluating the automatic speech detection, it was simple to spot the system’s susceptibility to false positive and false negative classifications. False positives were found to be the most common source of disagreement between the automatic and subjective detection: The auto-detection was too tolerant of segments that clearly didn’t fulfill the criteria for “normal speech,” thereby including some shouting, fighting noise, loud effects, and music. As for the false negatives, some scattered or very softly spoken dialog remained undetected.

Loudness vs. Speech: Transparency

While Loudness is precisely defined in ITU-R BS.1770, there was a difference in the resulting speech loudness level from using different multichannel topologies. The difference between the two was on average 1 dB (up to 4 dB) per film [7].

The difference between the dialog level based on the manually edited speech assemblies and on the automatically speech-gated measurements is 4.5 LU (average), up to 8.3 LU, with the automatic measurement higher in 10 of 10 cases. Comparison of Leq(K) with BS.1770-3 measurements, both based on the speech assemblies, can answer the question: Does the -10 LU relative-level gating in BS.1770-3 matter in the measurement of speech? The answer is yes, the gated measurements are 1.6 LU higher, on average.

Because these three sources of uncertainties are independent, they may add up, as shown in Fig. 7. Widening the scope to other types of speech programs, the uncertainties of measuring this property would generally be reduced. Table 1 gives a “guesstimate” of the slack to be expected as genre and Loudness Range (LRA) vary.

Type of program	Typical LRA [LU]	Definition [LU]	Classification [LU]	Measurement [LU]
Commercial	4-8*	0.2	0	0
Regular Broadcast	6-12	1.0	3.0	1.0
Broadcast Drama	10-22	1.4	5.0	2.5
Feature Films	16-30	1.6	8.3	4.0

Table 1. Estimated uncertainty when normalization is based on dialog.
* For programs <1 min, LRA is secondary to e.g. Max S

Table 1 is based on ATSC A/85 criteria, where promos and commercials are Loudness normalized, and other programs are normalized to dialog level, not counting further uncertainty added as a result of the choice of the anchor on which to rely. In a multi-format environment, extra uncertainty is added to both speech and loudness measurements. [13]

Loudness and Loudness Range of TV Drama

To extend our previous studies’ investigation into DVD movies and music tracks, season 3 of the drama series "Game of Thrones" (GoT) was measured, using the same methodology. The Loudness and the Loudness Range were measured of the 10 episodes, using the 5.1 audio from DVD, English language.

Figure 7 shows the Loudness, as per BS.1770-3, for the 10 episodes, together with 3 examples each of pop music and movies, selected as typical for their genre. The dialnorm for all 10 episodes was set to a value of 27, which we found to be a common choice (or default?) for DVD programs [7].

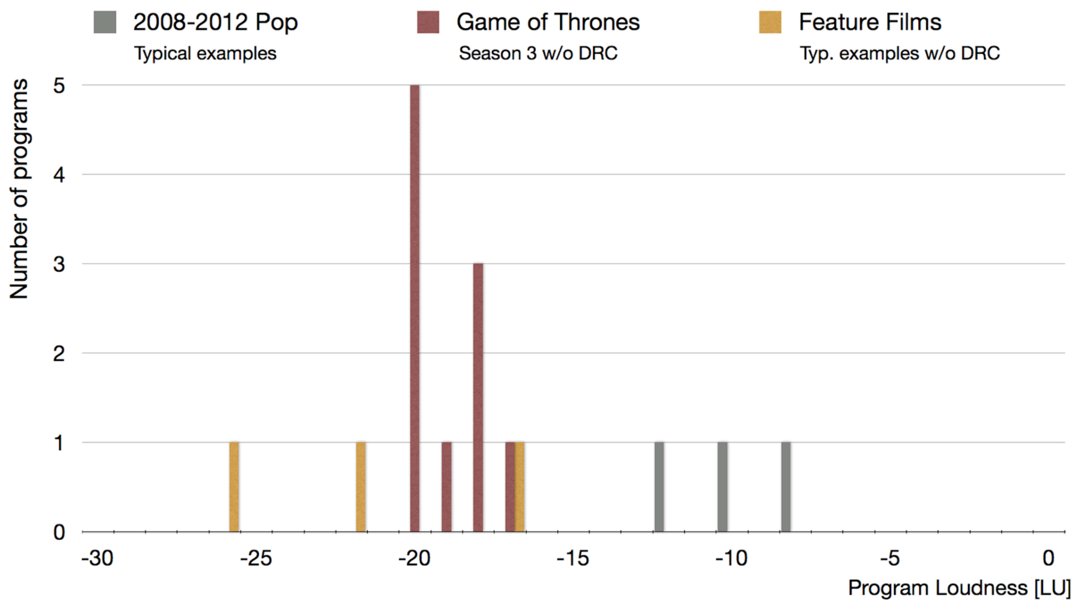

Figure 7. Loudness for the 10 episodes of Game of Thrones
compared with typical examples of Loudness in Film and in Pop Music.

The Loudness of each episode is within -20.4 and -17.3 LUFS. This interval is comparable to DVD movies, but it also reveals that the episodes were not quite Loudness-normalized within the season. All episodes, however, could be transmitted Loudness normalized without further processing under the 22 dB of headroom offered by ATSC A/85 and EBU R128.

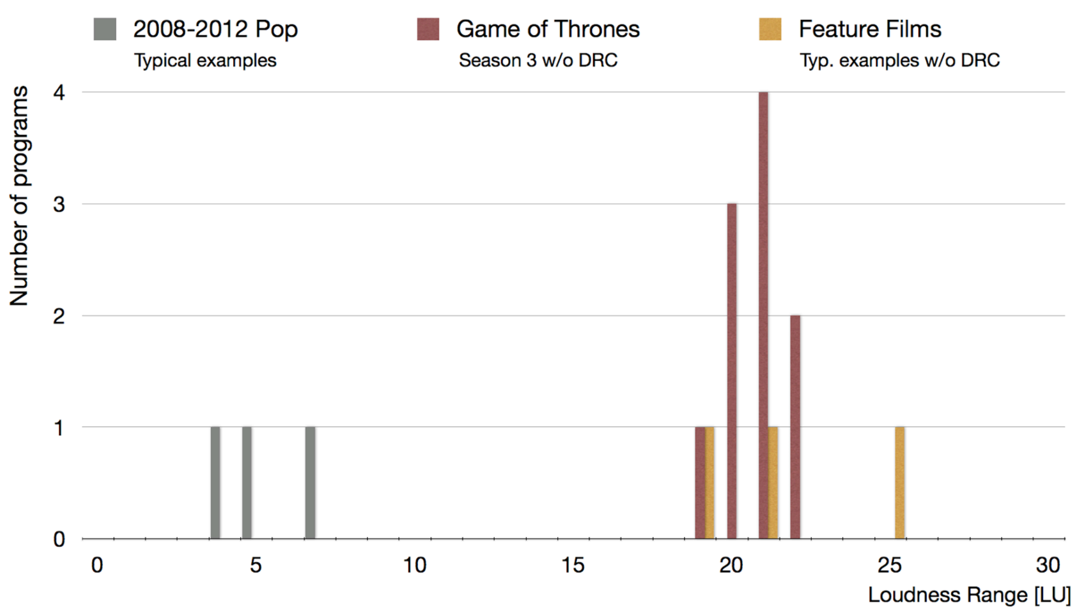

Figure 8. Loudness Range for the 10 episodes of Game of Thrones
compared with typical examples of Loudness Range in Film and in Pop Music.

In Figure 8, the Loudness Range (LRA) of the same Game of Thrones episodes are between 18.7 LU and 22.1 LU, so episodes also fall within a relatively narrow interval on that parameter. The LRA was furthermore comparable with the LRA of DVD movies, although some movies, such as The Matrix and Lord of the Rings have LRA as high as 25 LU.

Because the LRA basically measures the distance from the soft to the loud scenes, an increase of even a few LU may have a noticeable impact, if played on typical home entertainment systems. Such setups might apply varying amounts of dynamic compression (e.g. "DRC"), with or without the listener knowing, that would inevitably reduce the resulting LRA.

Preservation of Loudness Range in Broadcast

To examine if a consumer should expect processed audio additional to lossy data reduction, Loudness and Loudness Range were measured for one particular episode of Game of Thrones in 3 different versions of the program:

1) from the AC-3 bit-stream on the DVD (i.e. without DRC)
2) off the DVD, via MacBook digital optical output
3) from HBO Nordic's digital stream, via MacBook digital optical output

	Loudness	Loudness Range
DVD, AC-3 file rip w/o DRC	-19.5 LUFS	21.6 LU
HBO Nordic, Mac optical out	-23.7 LUFS	21.8 LU
DVD, Mac optical out	-18.9 LUFS	16.5 LU

Table 2. One episode (GoT season 3, #8,: “Second Sons”), different results.

Assuming HBO has used the same source material as the DVD, it the HBO stream was found to be free of processing but merely normalized. When playing the episode from DVD, however, dynamics processing had clearly been added. LRA being lower indicated this, and it was also confirmed by listening. Six other episodes of GoT from season 2 and 3 showed the same LRA difference of around 5 LU between DVD and HBO, with the higher range from HBO in every case.

Discussion

In its A/85 recommended practices for audio in DTV, ATSC has been somewhat film-oriented by carrying on that genre’s tradition for dialog centrism - a principle that undeniably has served the film industry well for decades, and allowed for calibrated reproduction systems to be deployed all over the world.

However, analog audio in cinema, like in broadcast, had a reduced-size playing field, which put a limit to how pronounced balancing differences would manifest at the end-listener. In digital cinema as well as in digital broadcast, it is possible now to create extreme difference between a regular and a loud sound, such as between mezzoforte (or speech level) and forte fortissimo (fff, or an action scene).

Obviously, producers of short commercials should not have access anywhere near fff, and as audio professionals, we have an obligation to prevent a new quality-depressing loudness war from building in broadcast or in film like it did in music.

The question is, however, has time run out for dialog centrism? In cinema, complaints from the audience can be taken for granted with any new action movie, unless the replay gain is reduced by 8 to 12 dB in the movie theatre. Sadly, a reduction in playback gain also represents a loss of headroom. Considering audio quality, an effective level normalization method would be preferable.

In broadcast, with the extra challenge of different genres having to coexist, it seems indicated to go back to physiology and get the priorities right: Our number one audio concern is general loudness. Hearing loss can develop from any type of sound, and auditory reflexes do not discriminate either. [14] Speech level is not even second on the human priority list. Conscious bandwidth assignment is based on our wish to receive a verbal message, our understanding of the language/dialect, and speech intelligibility. [15] To some extent, speech level may correlate with speech intelligibility. In this context, however, speech level is not the primary variable but a confounder.

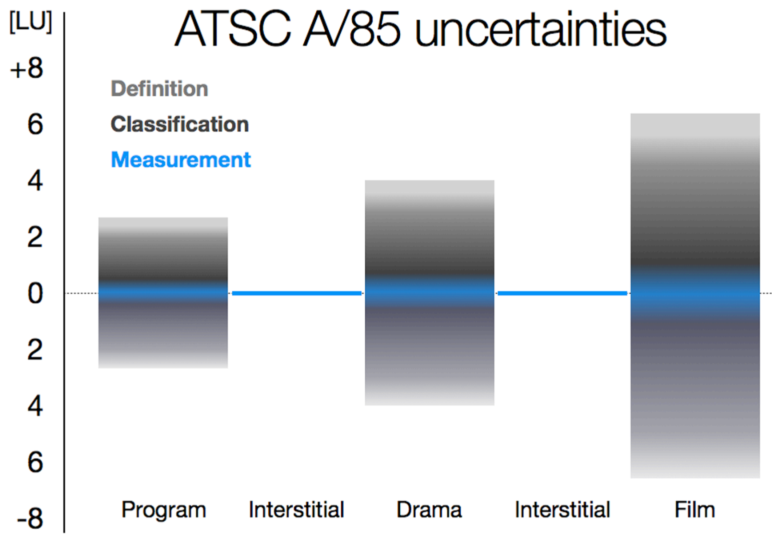

Figure 9. Measurement slack per program.

Taking these factors into account, it would not appear wise to have broadcast revolve around a secondary audio criterion, based on proprietary IP, with a high intra-program variation (Fig. 5), that introduces a significant amount of slack, wastes headroom, and does little to help end the loudness wars (Fig. 2).

With its R128 recommendation, instead the EBU has defined a simple and more transparent approach to broadcast leveling where regular programs as well as commercials are normalized using the genre-neutral BS.1770 method. ASWG under Sony games has released production guidelines based entirely on the same principles [9], and according to experiments by the authors, iTunes Sound Check and iTunes Radio also employ all-source energy-based normalization. The simple solution is possible, and it gives a credible result for a vast majority of broadcast programs.

If programs have wide Loudness Range, no normalization technique can prevent annoying inter-program level jumps from sometimes occurring [16]. Still, compared to speech normalization, BS.1770 minimizes the headroom required to convey each track or program, and it minimizes the uncertainty associated with the measurement. It would therefore also be an obvious choice to adopt the ITU-R BS.1770 measurement as default in ITU-R BS.1864 on international program exchange. The typical exchange program is news, sports or music where BS.1770 delivers a precise answer.

Figure 9 shows a "guesstimate" of the per-program measurement uncertainty under a mixed leveling regime like ATSC A/85 where interstitials are measured differently from regular programming. Under those circumstances, commercials and promos are assessed with a tight tolerance (±0.1 LU), which is well below the aim of various standards calling for between ±0.5 and ±2 LU precision. However, regular programming can have a measurement slack of up to 14 LU. In reality, only interstitial leveling can therefore be enforced under a mixed regime.

With the large playing-field of digital film, drama and broadcast, more focus on mixing balance could be a place to start. For instance, not generating too much difference between Loudness and speech level, and not exaggerating Loudness Range just because it's possible. At the same time, distribution needs to become less ambiguous than what it is today with e.g. AC3.

Conclusion

In this paper we have extended our empirical study of the two dominating methods of audio level alignment of feature films in broadcast: Loudness normalization versus dialog normalization. The integrated loudness method was ITU BS.1770-3, whereas the dialog measure employed the speech classifier known from "Dialog Intelligence" (DI) products.

Uncertainty when assessing Loudness and speech level was investigated. For Loudness, a single source of uncertainty exists, namely, the measurement itself, amounting to 0.2 LU. For speech level, three independent sources were identified: measurement, definition, and classification at 1.6 LU, 4 LU, and 8 LU, respectively.

The automatic DI measurement of speech level was compared to manual separation and measurement of the dialog anchor of 10 films. We found a typical difference of 4.5 dB, with the automatic measurement producing the higher level in all cases. We then introduced an algorithm to audition the speech classifier directly on the film mixes, and an informal listening test indicated that the DI measure was positively biased, mainly due to its inability to distinguish between "normal speech" and speech on top of, or surrounded by, "action" sounds.

In addition to feature films, Game of Thrones (GoT) television drama, interstitials, broadcast programs, and music tracks were analyzed. With a target Loudness level of −24 LUFS (=LKFS), and when normalization was based on Loudness, half the films and all the other programs and tracks could be passed without the need for dynamics processing at the station or at the consumer. For the same target loudness, when normalization was based on DI, dynamics processing was required for every film.

Season three of GoT was furthermore found to have Program Loudness and Loudness Range comparable to cinema movies, and those parameters showed limited variation between the episodes. Improved headroom in broadcast, instituted by ATSC, EBU and other organizations, is clearly required for conveying content such as GoT faithfully. End-listeners, however, can expect more variation in the resulting LRA than between episodes just depending on which platform they listen to. In this case, HBO streaming had a higher Loudness Range than a DVD. The unpredictable end-listener experience seemed to be caused by the AC3 codec or how it was implemented.

For all types of programs, normalization based on Loudness was found to demand less headroom in distribution, linking, and transmission than normalization based on speech level. The difference in headroom requirements was up to 14 dB. Less hunger for headroom and less uncertainty in the measurement makes Loudness normalization an obvious first choice in ITU-R BS.1864.

Finally, the measurement uncertainties under the ATSC A/85 mixed regime means that only interstitial leveling can reliably be enforced. Whether freedom in the level normalization of regular programs is good or bad depends on the point of view. The U.K. has enjoyed some success with comparable BCAP practice, where regulation of interstitials-only was instituted in 2008, though the U.K. has now adopted EBU R128. For the time being, the liberal A/85 approach may however be enough as it addresses the main concern of the CALM Act: Overly loud commercials.

Still, such considerations do not change the conclusion of this paper: DI should be regarded as a headroom-hungry, rubber-band measure. Time has run out for dialog centrism in broadcast, especially when based on metadata that cannot be trusted.

Acknowledgements

We would like to thank Leon and Lukas “film” Mortensen for their tireless assistance in analyzing the DVDs for this study.

References

1. S. H. Nielsen and T. Lund, “Level Control in Digital Mastering,” Proc. 107th AES Convention, New York, 1999.

2. S. H. Nielsen and T. Lund, “Overload in Signal Conversion,” Proc. 23rd Int. AES Conference, Copenhagen, Paper #11, 2003.

3. T. Lund, “Specifying Audio for HD,” Proc. NAB BE Conference, Las Vegas, 2007.

4. ITU-R, “Recommendation ITU-R BS.1770-3: Algorithms to measure Audio Programme Loudness and True-Peak Audio Level,” International Telecommunication Union, Geneva, 2012.

5. B. L. Jones and E. L. Torick, “A New Loudness Indicator for Use in Broadcasting,” Proc. 71st AES Convention, Montreux, 1982.

6. R. Ortner, “Je lauter desto bumm!—The Evolution of Loud,” M.A. Thesis, Donau Universität, Krems, 2012.

7. E. Skovenborg and T. Lund, “Level-Normalization of Feature Films Using Loudness vs Speech,” Proc. 135rd AES Convention, New York, 2013.

8. T. Lund, “Audio for Mobile TV, iPad and iPod,” Proc. NAB BE Conference, Las Vegas, 2013.

9. Sony Computer, “Average Loudness and Peak Levels of Audio Content on Sony Computer Entertainment Platforms,” Recommendation ASWG R001, London, 2013.

10. ATSC, “Techniques for Establishing and Maintaining Audio Loudness for Digital Television,” Doc A/85, Advanced Television Systems Committee, Washington, DC, 2013.

11. E. Belger, “The Loudness Balance of Audio Broadcast Programs,” Journal of the Audio Engineering Society, 17(3): 282–285, 1969.

12. EBU, “EBU Technical Recommendation R128. Loudness Normalisation and Permitted Maximum Level of Audio Signals,” European Broadcasting Union, Geneva, 2010.

13. T. Lund, “DTV, Mobile TV and the Downmix Dilemma,” Proc. of the 27th Tonmeistertagung, Cologne, 2012.

14. W. Boron and E. Boulpaep, Medical Physiology (2nd ed.), Elsevier: Amsterdam, 2011.

15. K. Küpfmüller, “Nachrichtenverarbeitung im Menschen,” in Taschenbuch der Informatik, Springer: Berlin, 1974.

16. E. Skovenborg and T. Lund, “Loudness Descriptors to Characterize Wide Loudness-Range Material”, Proc of the 127th AES Convention, New York, 2009.

17. E. Skovenborg, “Loudness Range (LRA) – Design and Evaluation”, Proc of the 132nd AES Convention, Budapest, 2012.