Control of Loudness in Digital TV

THOMAS LUND

TC Electronic A/S
Risskov, Denmark
thomasl@tcelectronic.com

ABSTRACT

To facilitate better consistency between programs and stations, ITU, EBU and ARIB have investigated the standardization of broadcast loudness. This paper examines some consequences of a global loudness standard with regard to metering and control at the Ingest, Production and Transmission stages.

Findings are reported from the latest research into mono, stereo and multichannel loudness measurement of real-world broadcast sounds. The improvements achieved by the new loudness models are quantified against previous level descriptors, such as, for example, PPM and Leq(A).

Besides from reducing consumer annoyance with jumping levels, less engineering time needs being spent per audio stream. This too, is important because digital broadcast means a significant proliferation of the number of channels and the number of platforms. Each platform, such as TV, radio, internet, podcast, and other personal entertainment systems, has its own requirements for dynamic range, frequency range and speech intelligibility, so more automated handling is simply a necessity.

The paper also introduces the term Dynamic Range Tolerance, DRT, which specifies the most desirable audio treatment for various broadcast platforms, and therefore is a practical tool for optimizing listener pleasure in digital broadcast. Examples of digital broadcast installations relying on static and dynamic metadata for loudness control are given.

This paper is targeted radio and TV production, installation and management professionals. It provides user and technical info, and does not endorse or promote commercially available equipment.

DYNAMIC RANGE TOLERANCE AT THE CONSUMER

DTV has the potential to carry more ambitious audio than ever before. Ambitious, with regards to formats, dynamic range, and frequency response.

For example, feature films may be presented more like they were mixed and edited, with fewer compromises on the picture as well as on the audio side. However, audio still needs optimization for a presentation environment different than a cinema, like the picture still needs color space, rate and resolution corrections.

While DTV in itself must be able to cover several consumer situations, other emerging digital broadcast platforms widen the dynamic range target even further.

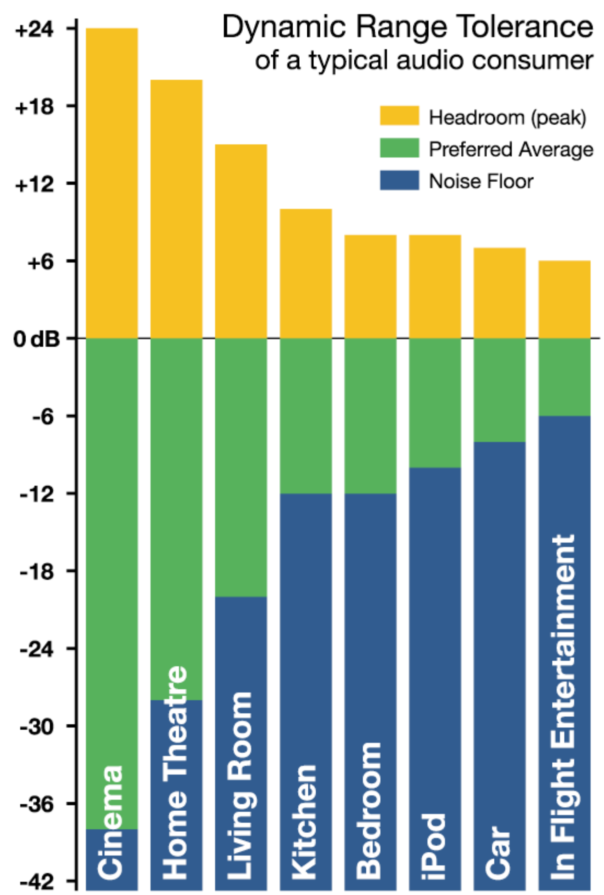

Fig 1. Dynamic Range Tolerance for consumers under different listening situations.

According to recent studies by the author, consumers have a well defined Dynamic Range Tolerance, DRT. When the average level is within certain boundaries, sentences or phonemes (speech sounds making up words) are correctly identified by a listener, the main instruments in a piece of music are heard; and sudden disturbances, such as loud effects, distortion, or other unacceptable sounds, do not appear. If the level fluctuates outside the tolerance area too often, the listener gets annoyed.

The DRT is defined as a Preferred Average window plus a peak level Headroom. The DRT depends on a consumer's listening environment as detailed in *Fig 1*.

In situations with significant background noise, such as inside various transports or urban environments, see *Table 1*, it's a challenge to get a wide dynamic range message across - be it music or spoken - without reproduction distortion being added, or damaging the listener's ears. The latter is becoming important as recent studies suggest that headphone levels may be set 5-10 dB above the same person's preference when listening through speakers, see *Fig 2*. If the same holds true in noisy environments, where iPods are often used, mobile platforms could pose a threat to hearing.

	SPL A weighted	SPL C weighted
Living Room, Suburban	45 dB	
Living Room, Urban	55 dB	70 dB
Inside Car	65 dB	85 dB
Inside Jet	75 dB	90 dB
Walk in Traffic	80 dB	92 dB
Subway	90 dB	100 dB

Table 1. Typical noise levels measured by the author. All environments are realistic for broadcast consumption today.

According to health studies [1], modern life exposure to noise typical of urban (non-aircraft and non-highway) environments produces widespread annoyance, speech interference, and sleep disturbance; and there is some evidence that human response to noise exposure at Ldn values in excess of 70 dB is more acute than at lower levels.

It should be noted that TV listeners typically object against too wide dynamic range, rather than when it is too restricted. Lack of speech intelligibility is the second worst offender, and often the cause for requesting more dynamic range limitation.

Against the hopes of most audio aficionados, as more people are also listening through headphones (iPods and other personal entertainment systems), the DRT trend is therefore currently moving towards more dynamic range restriction in broadcast.

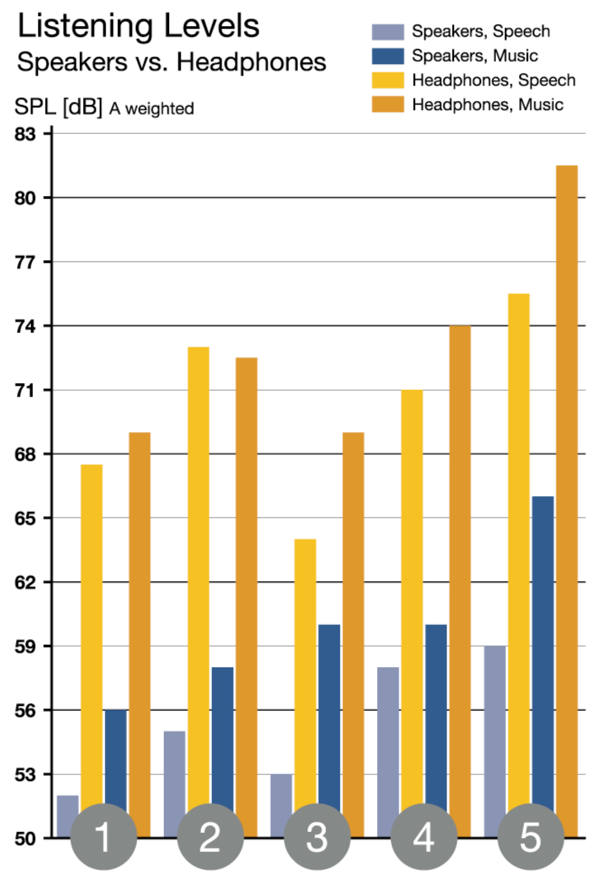

Fig 2. Preferred listening levels for different groups of employees at the Danish Radio & TV [2].
1. Administration (non-engineer)
2. Journalists (non-engineer)
3. Classical music engineers
4. Pop/Rock music engineers
5. Noise engineer

Fig 3 shows spectral noise conditions inside a car [3]. Low frequency noise from the road-tire contact is the main source, as long as the windows are kept closed.

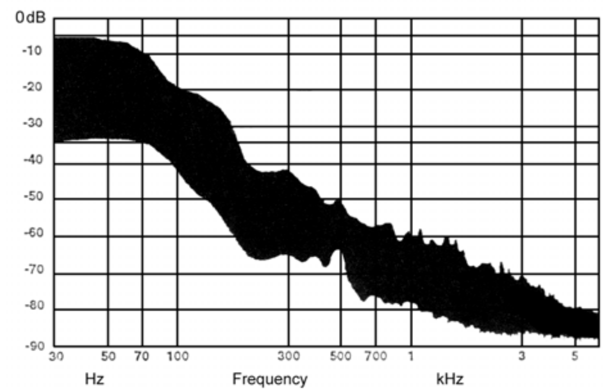

Fig 3. Typical noise spectrum in a moving car with the windows closed (upper trace), and when idling (lower trace).

DYNAMIC RANGE OF BROADCAST MATERIAL

Program material for TV broadcast is generally aimed at a listener in the Living Room or Kitchen region, see *Fig 1*. This kind of material should be thought of as having a *normal* broadcast dynamic range signature.

Commercials, promos and consumer CDs typically have a more restricted dynamic range, and therefore appear loud on TV, where normalization is based only on peak content. This kind of material should be thought of as having a *hot* dynamic range signature.

On the opposite side we have film production, aimed at a completely different listening scenario, where much softer and much louder level than the average can be reproduced and heard. Production for wide dynamic range listening can also include classical or acoustic music. All material of such nature should be thought of as having a *soft* dynamic range signature.

Music and entertainment radio is typically aimed at Car listening, so the dynamic range signature is generally hot. The only type of radio with a wider dynamic range typically carries classical music, drama and low key, talk based programming.

It should be noted that consciousness about what dynamic range a program is suited for may be built into a broadcast installation by using calibrated meter and loudspeaker environments [4, 5].

To summarize, broadcast material is produced in a way that fits the listening conditions of a wide majority of consumers in the best possible way. The most dramatic difference between program material and consumer requirements concerns feature film. To have a feature film align with domestic listening conditions without loosing too much detail, or distorting the loud parts, low level may need to be brought up by 12-20 dB, and the headroom restricted by 12-16 dB, see *Fig 1*.

DEFINING LOUDNESS

Unlike electrical level, Loudness is subjective. Listeners weigh the most important factors differently:

- Sound pressure level
- Frequency contents
- Duration

Therefore, defining the loudness of a sound shows a certain Between Listener Variability (BLV), even within homogenous groups [6], while differences in age, sex, culture etc. can add further to the variation. Also, individual loudness assessments by the same person are only consistent to some extent, and depends on the time of day, mood, attention etc. This type of variation is called Within Listener Variability (WLV).

Because of the variations, a generic loudness measure is only meaningful if it is based on large subjective reference tests and solid statistics.

For the past five years, ITU has investigated "Audio Metering Characteristics suitable for use in Digital Sound Production". Questions studied [7] by a special rapporteur group, SRG-3, include:

1. Meter characteristics to *avoid overload*,
2. Meter characteristics to indicate *subjective loudness*,
3. *Display* characteristics for efficient use,
4. Methods to *evaluate* meters and displays.

Subjective reference tests were conducted to find the loudness model which could deliver the best level dose description of various audio segments. This would answer part of question 2, and enable the design of a simple level indicator, generating just one average dose number from any length of audio segment.

The tests suggested that a relatively simple Leq measure close to a C weighing, labeled "Leq(RLB)", under certain conditions was a good predictor of perceived loudness.

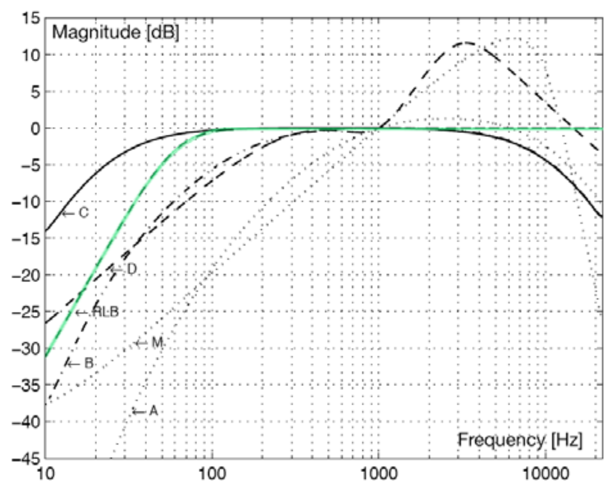

Fig 4. Weighting filters used in combination with Leq measures. A, B, C, D, M and RLB (green).

After taking part in this dose exercise, together with McGill University, Montreal, TC conducted additional listening tests in order to design a more precise loudness model suited for both short-term and long-term measurements.

We were concerned that describing the level variations of an entire program using just one number was an over-simplification, and counterproductive to the ITU objective of designing a realtime *meter*, because programs or segments with very different dynamic range properties could be assigned the same number, se *Fig 5*. Other aspects of the first round of listening tests also needed further investigation [8].

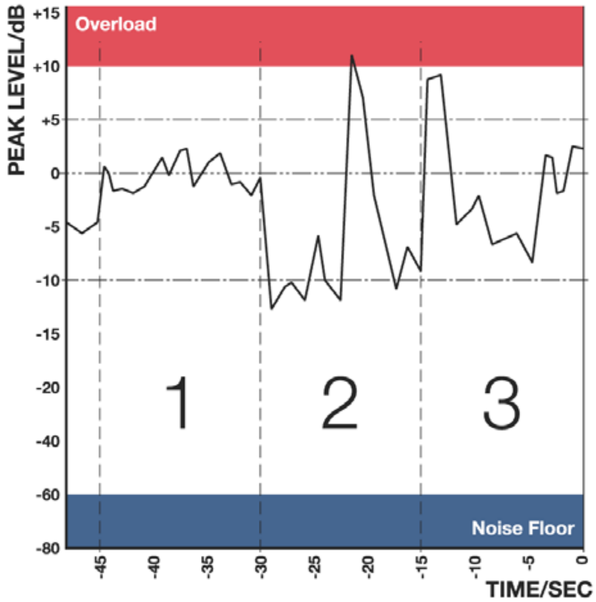

Fig 5. A dose approach to Loudness. Audio segment 1, 2 and 3 may produce the same dose measure, even though their level profiles over time are quite different.

The new large scale tests at McGill, in combination with extra trials at TC, plus the lessons learned from the first round of ITU tests, provided a solid foundation for evaluating different loudness models' performance on time varying broadcast type signals. Using the combined data, it is possible to assess different models' precision on speech, music and effects, thereby extending the somewhat narrow aim (predominantly well-controlled speech) of the first ITU tests. A summary of the results are shown in *Fig 7*.

The evaluation measurements suggest a grouping of the loudness models into classes. Somewhat surprisingly, three rather widespread loudness measures, the Zwicker model, Leq(A), and Leq(M), cannot be recommended for standardization, and would fall outside any classification. Actually, the worst performance was observed from the Leq(A) and Leq(M) measures, which are implemented in some sound level meters and signal analyzers, and applied for loudness measurements in broadcast and cinema. It should be noted that all of the above performed worse than the quasi-peak measure used in PPM type meters (IEC 268-10).

To the ITU and EBU, it has therefore been suggested to define a simple Leq(RLB) as a baseline measure, which could be the foundation of standardized loudness model.

TC and others have requested to also standardize a classification, whereby models with more precision than Leq(RLB) could be taken advantage of for critical applications. Technical standards all around us, from the old Meter in Paris to the duration of a second, are measures we have been able to use with more precision as we got cleverer. A few years back, Leq(M) was believed to be perfect as a theatrical dose measure. Until recently, Leq(A) was considered a valid measure for assessing the loudness of speech. Consecutive sample counting was chosen as a level measure in CD production. In all cases, we now know better. A couple of years from now, we will have learned more again.

Consequently, Leq(RLB) will probably be chosen by ITU as *the* Class 0 Loudness measure. The standardization will hopefully also include descriptions of how future models could earn a Class 1, 2 etc. grading, while still maintaining a valid reference to Class 0.

At the time of writing, the ITU standardization procedure was not yet completed. Detailed information about the additional McGill and TC subjective tests can be found in [6, 8].

Finally, it should be noted that the idea of a perceptually based level calculation is not new. An aging, but respectable measure such as "CBS Loudness", is still being used with success for automated level control [9]. This model has served as a de facto reference for objective loudness measurement, in the broadcast community.

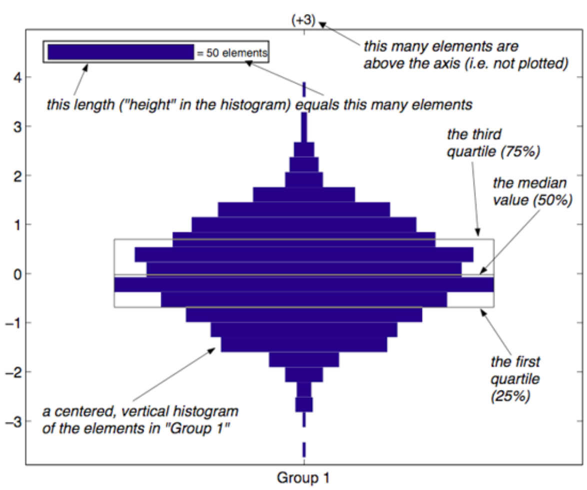

Fig 6. A guide to reading the Loudness Model Evaluation Diagram of *Fig 7*.

A MULTICHANNEL LOUDNESS MEASURE

The subjective tests detailed previously were done using mostly mono material. At the time of writing, only a small, hasty experiment has been carried out to shed light on the loudness of real-world, multichannel signals [10]. Unfortunately, this work is not consistent with earlier tests. It introduces new Leq weighting curves, which could be referred to as Leq(R2LB). (The "R" in the original "RLB" stands for revised).

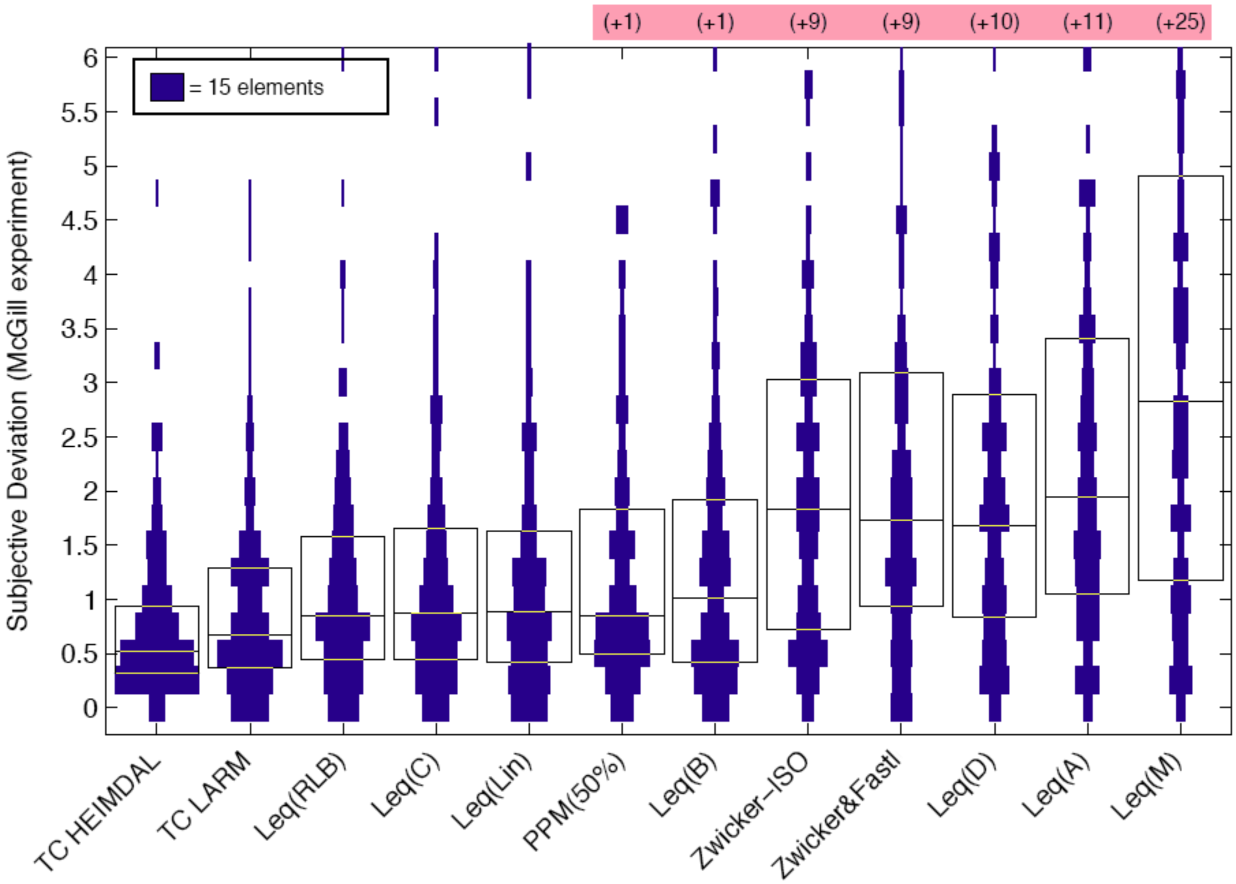

Fig. 7. Evaluation of different Loudness Models (names at the bottom) using a wide range of broadcast audio material [8].
Loudness models to the left are in better agreement with human listeners than models to the right of the chart.
Red indication at the top signifies outlier audio segments, misjudged by more than 6 dB of a particular loudness model.

Evaluating the Leq(R2LB) measure against our combined databases, not surprisingly, reveals different results than using Leq(RLB).

It would be unfortunate if the weighting needs to be adjusted when the program material or number of channels change. Therefore, more substantial investigations into the loudness of multichannel signals are indicated. Some of the questions needing answering include 1) define basic measure and associate with previous tests, 2) directivity and listener orientation issues, 3) channel weighing, and 4) short-term vs. long-term performance and correlation analysis.

SHORT-TERM AND LONG-TERM LOUDNESS

The most important omission so far of the loudness standardization process has been the lack of short-term versus long-term measure investigations.

By focusing exclusively on long-term, retrospective use, the measure may be suited for program level off-sets, logging, or for the gathering of metadata, but it neglects the even more relevant realtime metering and control applications.

The author therefore has been involved with the development of two new models of loudness, “LARM” and “HEIMDAL”. The main objective of the new models has been to produce an accurate and robust estimate of the perceived loudness of sound segments consisting of speech and/or music. Both models be can be used to compute a short-term loudness, or for calculating long-term average by adjusting their analysis-windows.

Fig 8 shows a screen shot from a proof of concept study, offering a realtime, short-term plus long-term loudness indication in one compact view.

The angular view on the outer ring makes it possible to scale down the display, because it’s not necessary to be able to read the LU numbers. This is a useful feature for picture overlay purposes. The speed of the inner radar view (long-term loudness display) can be adjusted for a revolution per, for instance, 1, 2, 4, 12 and 60 minutes. The display handles any number of audio channels, and provides compact loudness information, but for inter-channel balancing a traditional bar-graph display should also be available. Note that the meter is not a commercially available product.

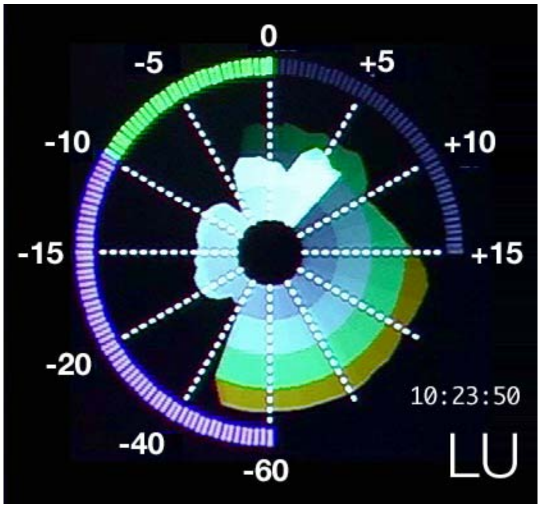

Fig 8. Example of Loudness Meter combining a realtime measure in the outer ring with a history in the “radar view”.

STATION STRATEGIES FOR CONTROLLING LOUDNESS AND AUDIO FORMATS

Analog TV already has its problems with jumping levels between programs and stations. These problems will get bigger if HDTV stations start transmitting feature films with a less suitable dynamic range than today. Film fall way outside the Dynamic Range Tolerance of the average consumer (see previous sections) under her domestic listening conditions.

Consequently, dynamic range restriction must take place either at the station, or inside the consumer’s receiving device.

As seen on *Fig 1*, the dynamic range translation should deal with both overly soft and overly loud parts. Ideally, the perfect re-mapping should happen at the receiving end to accommodate a wide range of listening conditions.

Metadata in conjunction with, for instance, Dolby AC3, provides some of these capabilities. However, even if the consumer knows how to adjust the dynamic range of a film to her current listening conditions, the optimum dynamics treatment unfortunately exceeds the capabilities of an AC3 decoder. The dynamic range control in the codec is acceptable for cut and boost ranges of less than 6 dB, but preparing a feature film for broadcast needs considerably more than this (*Fig 1*).

If such a large correction is left only to the AC3 decoder, the wide-band gain changes can be quite audible, especially when they happen before they were supposed to (because of the discrete intervals in the controlling). Also, film and music corrections require a multiband structure so listeners don’t sacrifice speech intelligibility, or get excessive spectral intermodulation added to their music.

Fig 10 summarizes three different DTV strategies for controlling loudness, dynamic range and audio formats during Ingest, Production and Transmission.

In *Fig 10, drawing no. 1*, the Ingest Gate (i1) is used to bring imported programming into the DRT (Dynamic Range Tolerance) of the station’s DTV transmissions. Audio material supplied data reduced (e.g. Dolby E, dts or other codecs) is converted to linear audio, and dynamic range translated. Downstream of Ingest, metadata need not be dealt with, and can be discarded.

Production is business as usual, using mono or stereo equipment, and not relying on the generation of metadata. 5.1 programming may be discretely mixed, or 5.1 can be produced using up-conversion of a stereo mix in combination with, for instance, extra sports stadium atmosphere. If datareduction is needed for OB or other live feeds, this is dealt with as an isolated encode/decode situation.

The Transmission Gate (T1) acts as realtime loudness corrector, see *Fig 9*, with special attention to junctions between programs, and carries out further dynamic range and format conversions for ATV, Pod and Web services. DTV transmission is datareduced according to regional standards, and passed with metadata that only changes if the audio format does (e.g. from stereo to 5.1). DTV may be up-converted to 5.1 where indicated (sports, game shows etc.)

Routing internally at the station is based on linear digital audio, typically using AES/EBU and/or SDI transports.

Fig 9. Example of Dolby LM100 meter measurement before and after automatic loudness correction during transmission. Challenging 20 sec broadcast segments butt edited over 5:30 minutes.

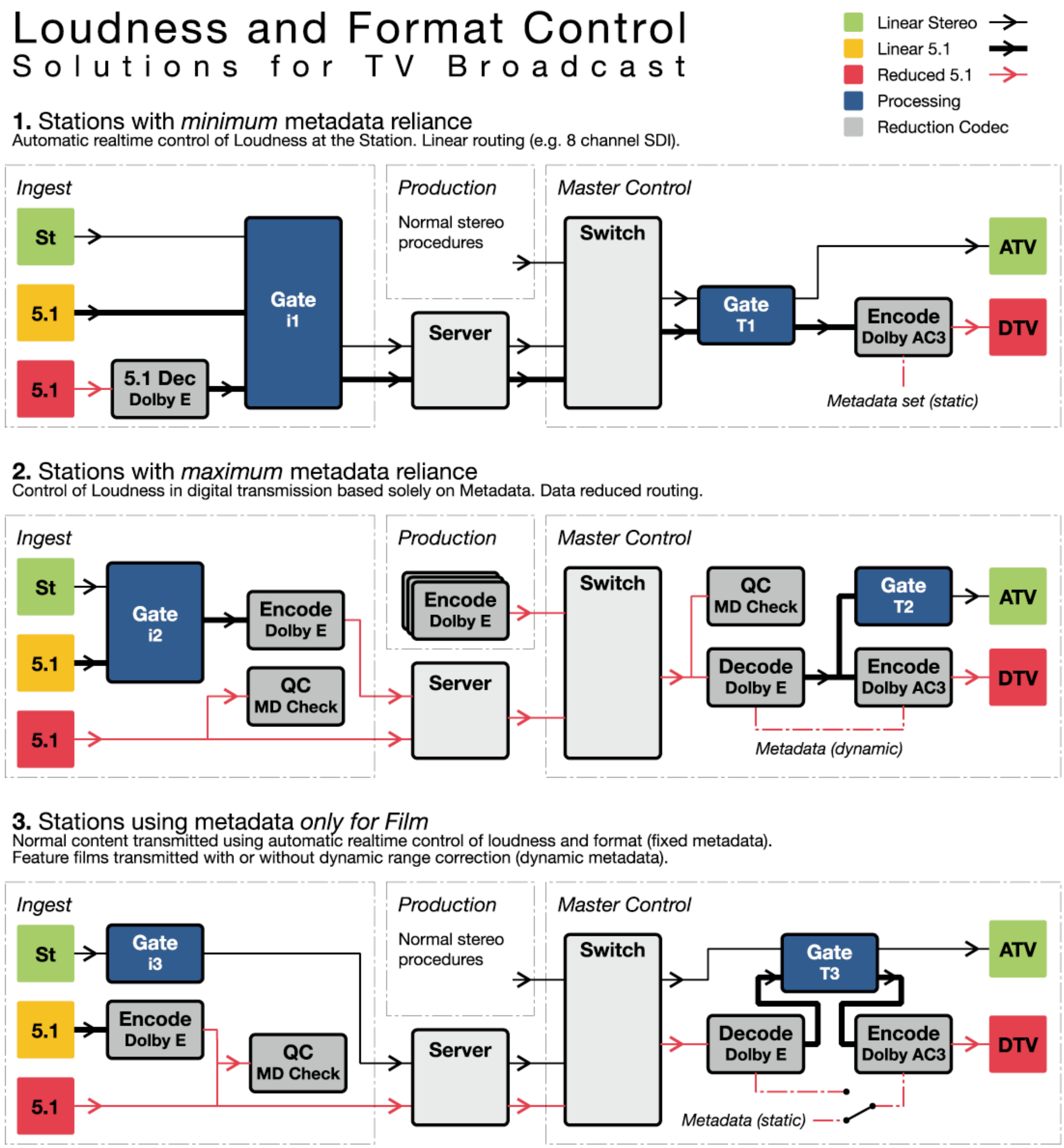

Fig. 10. Three different ways of handling Loudness Control, Multichannel audio and Data Reduction in digital broadcast.

In *Fig 10, drawing no. 2,* the Ingest Gate (i2) is used to datareduce import programming, and to inspect metadata associated with it. Downstream of Ingest, metadata must always be available and preserved, meaning no analog transfers or sample rate converters.

In production studios, metadata has to be attached to all programs. Production can be native mono, stereo or 5.1 as required. OB and Live production can be incorporated using fixed metadata with appropriate upstream dynamics processing.

The Transmission Gate (T2) is used as a dynamic range and format converter for ATV, Pod and Web services. DTV transmission relies solely on metadata when it comes to loudness control and speech intelligibility.

Routing internally at the station is based exclusively on datareduced, synchronous digital audio. Data encoders and decoders are used for breakouts and monitoring. Audio/video synchronization needs special attention in designs where an arbitrary number of monitoring posts are needed.

In *Fig 10, drawing no. 3,* the Ingest Gate (i3) is used to bring import mono and stereo programming into the DRT of the station's DTV transmissions. 5.1 material is datareduced before being transferred to the server.

Ingested 5.1 material is assigned metadata if it doesn't already have it. Downstream of Ingest, metadata is only used for selected 5.1 transmissions.

Production is business as usual, using mono or stereo equipment, and not relying on the generation of metadata. 5.1 programming may be discretely mixed, or 5.1 can be produced using up-conversion of a stereo mix in combination with, for instance, extra sports stadium atmosphere. If datareduction is needed for OB or other live feeds, this is dealt with as an isolated encode/decode situation.

The Transmission Gate (T3) acts as realtime loudness corrector (*Fig 9*) with special attention to junctions between programs, and carries out further dynamic range and format conversions for ATV, Pod and Web services. In DTV, metadata normally only changes if the audio format does. It may be up-converted to 5.1 where indicated (sports, game shows etc.).

Selected 5.1 programs, e.g. feature films, may be transmitted using the original dynamic metadata, while the Transmission Gate simultaneously provides suitable format and dynamic range conversions for ATV, Pod and Web services.

Mono and stereo routing internally at the station is based on linear digital audio. Selected 5.1 programs are routed datareduced for preservation of metadata.

DYNAMIC RANGE TRANSLATION

The main part of dynamic range translation and loudness control should be done at the station, leaving only smaller corrections to be performed at the consumer.

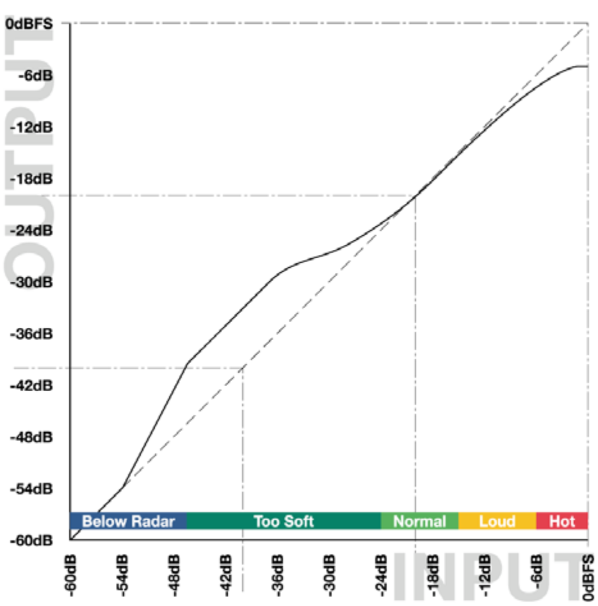

Fig 11. Example of dynamic range re-mapping: From Home Theatre/DVD to Living Room listening conditions (*Fig 1*).

This ensures competitive audio with regards to quality and speech intelligibility, and prevents asking more from the AC3 decoder than it can deliver in a civilized manner.

Fig 11 and *fig 12* show rational transfer characteristics complying with the DRT of the consumer, without affecting level already on target.

Fig 12. Example of dynamic range re-mapping: From Home Theatre/DVD to Living Room listening conditions (*Fig 1*).

Features films in 5.1 may have their range optimized for broadcast using the example in *fig 13*. This particular transfer curve has been used successfully at stations paying special attention to speech intelligibility.

Compare against the DRT chart, *fig 1*, and note how the Center channel is given an extra low level advantage compared to the four lateral channels. This ensures that dialog can still be heard when the words could otherwise be lost to listening room noise. The lateral channels are linked two and two, or all in one group.

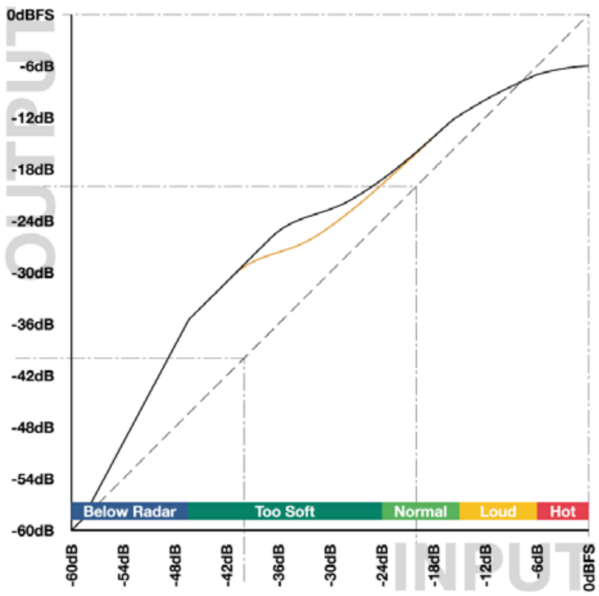

Fig 13. Example of multiband dynamic range re-mapping of a 5.1 feature film to domestic listening conditions (*Fig 1*). Black curve: Center channel. Orange curve: L, R, Ls, Rs.

Note that film mixing facilities typically make use of the same dynamic range processing techniques prior to data-reduction to optimize codec performance with regard to conveying space and suppressing artifacts, for instance hot level generated distortion and listener fatigue at the consumer [11, 12, 13, 14].

CONCLUSION

Better and automated control of broadcast loudness is made possible using new perceptually based measures. Thanks to efforts of the ITU and regional broadcast organizations, one global measure may eventually become standard.

The guiding principle for digital TV and Radio station level control is to adjust the dynamic range of imported material during ingest to meet the end listener's DRT, and have an automated loudness control balance the final transmission, and transitions between programs.

The guidelines given eliminate listener annoyance with jumping levels, loss of speech intelligibility, and build-up of listener fatiguing distortion. Metadata based dynamic range control alone does not comply with the DRT of the listener. Nevertheless, good audio results, without station workload penalties, can be realized with AC3 using complementary processing at the station.

The bad news is that all improvements should be realized without engaging more broadcast personnel. New channels and new personal entertainment services have to rely on automated and low time-consumption procedures.

The good news is that with special attention to the Gates at Ingest and Transmission, mono and stereo production can be maintained, while taking advantage of loudness metering and control. Sensational 5.1 transmissions can effortlessly be derived from stereo production, when a multi-channel delivery is requested, but a discrete 5.1 mix not available. AC3 decoders respond predictably to audio format commands embedded in the metadata regardless of how the 5.1 content has originated.

From the failing level control in CD and Movie production our industry has learned a lesson of how not to measure level in digital media. We can do better than using unconscious sample counting or casual Leq measures. A well founded, global loudness standard valid for all types of sound will provide digital broadcast with a better chance of doing it right.

REFERENCES

[1] The Urban Noise Survey, Aug 1977, U.S. Environmental Protection Agency, Washington D.C.

[2] Brixen, E.B.: Report on Listening Level in Headphones. Document KKDK-068-01-ebb-1 for the Danish Radio, Copenhagen, 2001.

[3] Kässer, J. & Blum, P.: Audio Reproduction in Cars. Proceedings of Tonmeistertagung 18, Karlsruhe 1994.

[4] Katz, B.: An Integrated Approach to Metering, Monitoring and Level Practices. JAES, no. 9, 2000.

[5] Lund, T.: Monitoring Audio for Digital Broadcast. Proceedings of NAB BEC, Las Vegas, April 2005.

[6] Skovenborg, Quesnel & Nielsen: Loudness Assessment of Music and Speech. Proceedings of the AES 116 convention, Berlin 2004. Preprint 6143.

[7] ITU-R, WP6P: Audio Metering Characteristics Suitable for the Use in Digital Sound Production, Geneva, 2000.

[8] Skovenborg, Quesnel & Nielsen: Evaluation of Different Loudness Models with Music and Speech. Proceedings of the AES 117 convention, San Francisco 2004. Preprint 6234.

[9] Jones, B.L. & Torick, E.L.: A New Loudness Indicator for Use in Broadcasting. Proceedings of the AES 71 Convention, Montreux, 1982.

[10] Soloudre G. & Lavoie M.: Stereo and Multi-channel Loudness Perception and Metering. Proceedings of the AES 119 Convention, NYC, 2005. Preprint 6618.

[11] Nielsen, S. & Lund, T.: Level Control in Digital Mastering. Proceedings of the AES 107 convention, New York 1999. Preprint 5019.

[12] Nielsen, S. & Lund, T.: Overload in Signal Conversion. Proceedings of the AES 23 conference, Copenhagen, 2003.

[13] Lund, T.: Distortion to The People. Proceedings of the Tonmeistertagung 23, Leipzig, November 2004. Paper A05.

[14] Submission to the Australian Broadcasting Authority: Loud Advertisements on Television. Audio, Video & Post Production Industries of Australia, 2002.