Loudness and Dynamic Range in broadcast audio

— the Dolby solution

Tony Spath
Dolby Laboratories, Inc.

Digital delivery media offer a wider dynamic range for audio than their analogue predecessors. This entails adopting a larger difference between the average levels (and thus the implied loudness) and the signal peaks. Although it is possible to implement this larger difference in a TV station or media studio, problems will occur in the home – due to inconsistent loudness and electrical levels within the consumer receivers and audio equipment.

The audio delivery system, Dolby Digital ™, includes specific tools to overcome these problems, while allowing the full dynamic range of digital audio to be delivered. These tools – Dialogue Normalization and Dynamic Range Control – are described here with particular reference to digital TV.

Reference levels and loudness

Audio levels (such as programme loudness, peak programme levels and reference levels) are typically measured as a decibel ratio. Broadcast audio engineers are nearly always dealing with instantaneous or short-duration signal levels in programmes; historically, the motive will have been to be aware of how much headroom the system has over the audio peaks, and how far from the system noise floor are the details in the audio signal.

This has led to a well-known set of reference levels – e.g. 0 VU, 0 dBu, 20 kHz RF deviation – which are used to line up a signal path for unity gain, and a set of metering practices and standards against which the audio signal in question can be measured. For example:

- 0 dBm (where m is 1 mW into 600 Ω);
- 0 dBu (where u is 0.775 V_{RMS});
- PPM 4 on a BBC peak programme meter (the peak level produced by a 0.775 V_{RMS} sine wave);
- 100% on a European peak meter (corresponds to +6 dBu).

Typically, we want to know where the peak levels in programmes are, so we don't cause overload. For analogue TV in the UK, for example, the broadcast maximum level is PPM 6 (= +8 dBu). This level has to relate to the relevant limits through the whole of the broadcast chain, so we find:

- +8 dBu = 510 nWb/m on (analogue) tape;
- +8 dBu = –10 dBFS peak on digital tape (0 dBu reference level is –18 dBFS peak in the UK);
- +8 dBu = ±50 kHz deviation in PAL;
- +8 dBu = 2 V_{RMS} (the maximum level in consumer electronics equipment).

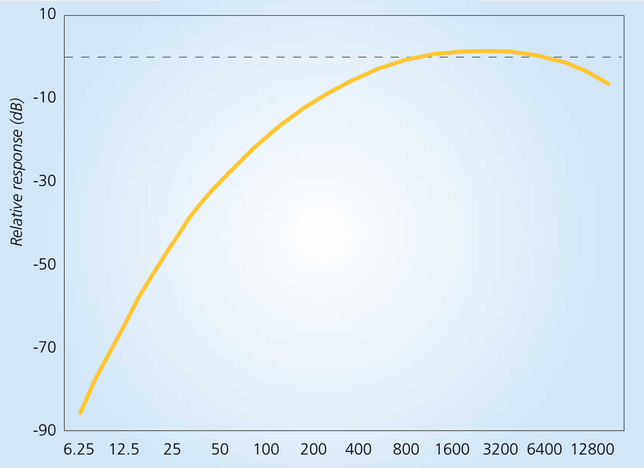

Figure 1
Leq(A) measurement of loudness

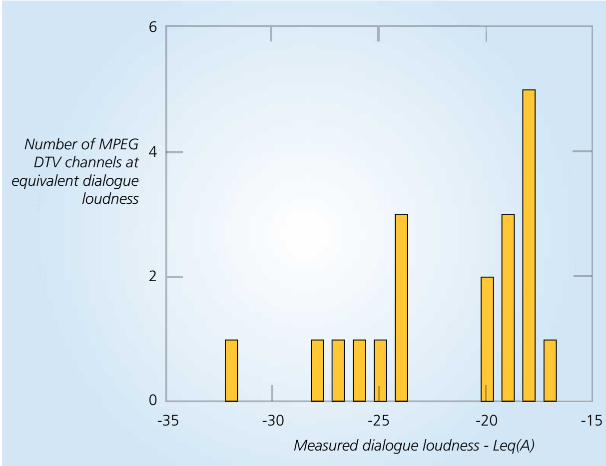

Figure 2
Loudness variations from European DTV statio

A level that should be measured by programme makers, but which is often overlooked, is *programme loudness*. This is used by the consumer, who has no knowledge of signal peaks or noise floors, to set the volume level. It may conveniently be measured as dBFS Leq(A). Leq(A) is a long-term average measurement of the loudness of sound, referenced to 0 dBFS, using an A-weighting to match the frequency response of the ear *(see Fig. 1.)*.

Reference level and loudness are not the same thing. Programmes with the same reference level can have different loudness values. *Fig. 2*, taken from Gundry [1], shows a sample of DTV channels available on a European satellite, measured for speech loudness using Leq(A) during the same time period. Note that, amidst the differences, there are broadly two groupings of level, reflecting the use of two different digital reference level standards.

Relating the reference levels and loudness

In the cinema, for some while, loudness has been linked to the electrical programme level (reference level) in the way the mixing room and the replay environment (the cinema) have been calibrated. As a result, speech levels and thus the linked elements of music and atmospheres show some consistency from film to film. Despite the lack of explicit calibration of loudness level to a reference signal level, loudness on media designed for the consumer has hitherto ended up with some degree of consistency. Consistent loudness has been achieved, by and large, by one of two methods: either by conventional taste (e.g. a lot of pop music sounds "wrong" if it has a large peak-to-average value, so it ends up with just around 6 dB between the average and peak levels: not in itself a bad thing); or because the medium itself imposes performance limitations on the programme that dictate an optimum use of the medium. As a result, this creates a de facto consistency on audio programme loudness levels. Examples exist, of course, of small level differences being apparent: adverts, for example, frequently sound "louder" than the programmes during which they occur.

As an example of the limitations of the (analogue) medium, when dictating a de facto programme loudness level, consider an analogue TV or radio transmission. Peak level is +8 dBu. Reference level is 0 dBu. And the peak-to-reference range is 8 dB – a very small range but one that has come about due to the need to optimize signals to the restricted medium of analogue broadcasting. Consumer electronics equipments, including

Abbreviations

DRC	Dynamic Range Control
DTV	Digital Television
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
MPEG	(ISO/IEC) Moving Picture Experts Group
PCM	Pulse Code Modulation
PPM	Peak Programme Meter
RF	Radio-Frequency
VU	(Audio) Volume Units

TV receivers, are designed with this figure in mind and, as stated earlier, +8 dBu (= 2 V_{RMS}) usually represents the maximum audio level a consumer electronics device can take without distortion.

In a limited transmission medium with issues such as noise and limited headroom (i.e. limited usable audio dynamic range), there is a good approximation between the reference level and loudness because, in setting the peak level, you are also approximating the average level. So there is little need to measure the loudness – an average reference level is a close enough approximation. This is probably why measuring the loudness seems such a new concept.

The consumer's audio system

Consumers use programme loudness – typically dictated by speech levels – to set their volume level. But clearly there will be a relationship between the acceptable range of average speech loudness and the reference levels / peak levels in consumer audio systems. There are some overriding factors that a consumer expects from the audio system in the home. Broadly speaking, the basics are:

1) outputs do not distort;
2) the volume levels match between different audio sources, e.g. between TV channels and between analogue TV and anything else that plays through the same system.

Fig. 3 shows the basic requirements for a consumer TV audio system.

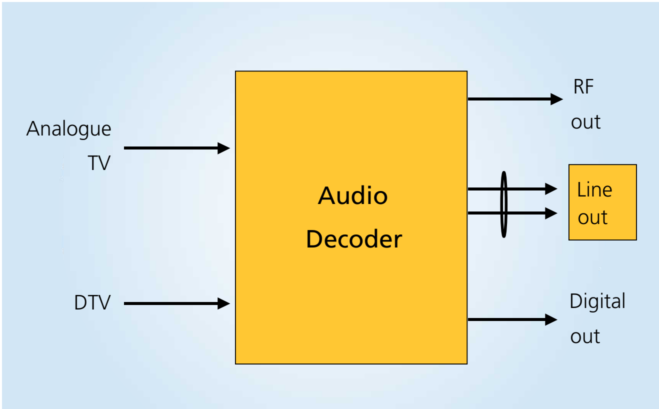

Figure 3
Audio basics of a consumer set-top box

The three outputs shown are designed to feed three different applications. The Line Output drives a hi-fi system; the RF Output is tailored specifically to match the audio levels that result from typical RF-demodulated TV audio. The Digital Out socket supplies the undecoded bitstream to a home cinema system.

To satisfy condition (1) above, the peak audio output from the broadcaster should always be ≤ +8 dBu. To satisfy (2), the broadcast signal must always have the same properties as analogue FM audio to match the RF output requirements (because the reference level does not measure loudness, and signals cannot exceed +8 dBu on a consumer receiver). This leads us to the unpleasant conclusion that, unless the broadcaster and receiver have some new way of dealing with the relationship between audio programme levels and programme loudness, the broadcast signal must have the same properties as analogue FM. Most disturbing is that this applies also to digital media – even when using MPEG stereo audio on digital TV, and MPEG or PCM audio on digital packaged media. This rather cuts across the broadly-held belief that digital media offer the promise of a new level of audio quality.

A new approach

Once we leave analogue and employ digital media such as DVD or DTV [2][3] with wider dynamic range, the programme maker has a choice – either to continue in "the analogue way" of setting an unnaturally small level difference between the peak and average level (and thus the implied norm for matched loudness), or work with the wide dynamic range of digital, which means permitting a bigger peak-to-average ratio. In the latter case, changing the peak-to average ratio may be acceptable in the broadcast station or media preparation studio, but a problem occurs when you send that audio into the home, as the consumer receivers and audio equipment currently in use still receive analogue sources with narrow dynamics. The choice then for digital content is between matching the analogue audio levels so there is consistency in the home, or taking advantage of digital audio's dynamic range and thus losing consistency of levels between programmes and sources in the home.

One audio delivery system has been designed to overcome this dilemma by providing a way to make new programmes fit old replay systems, where this is important, and to unlock the potential of digital audio for those who want that benefit. The Dolby [1] Digital audio system [4][5] includes specific metadata values, carried within the encoded audio stream, to allow loudness levels and dynamic range to be optimized, in each home decoder, to the specific audio system and requirements of each consumer. Importantly, these parameters are set as part of the production process, so this optimization is under the control of the programme maker.

Dialogue Normalisation (Dialnorm) is an attenuation signal designed to normalise the loudness levels between programmes using speech as the common reference. Speech is appropriate for setting the loudness levels due to its commonality and consistency of use in programming, and the widespread and instinctive use of speech by consumers as a touchstone for programme volume in the home. Dialnorm exists as a metadata parameter within a Dolby Digital audio stream and is used by the audio decoder to set the level at which the audio replays in the home. Dialnorm can be effectively measured and set with an Leq(A) meter, and specific tools relating Dialnorm to loudness are now becoming available.

The LM100 Broadcast Loudness Meter from Dolby, for example, allows broadcasters using Dolby Digital to set Dialnorm automatically. In addition, broadcasters using MPEG audio can use the LM100 to measure the speech level of a programme and to control suitable level-setting equipment. Note that when the LM100 is used to line up speech levels for analogue or MPEG digital broadcasts, the dynamic range constrictions of those formats will still be present. Though based on speech at present, future methods of setting Dialnorm are envisaged, that will be capable of accounting for the perceived loudness differences that result from spectral and other characteristic differences between speech and non-speech signals.

Dynamic Range Control (DRC) is a gain control signal, designed to allow different amounts of overall dynamic range when decoding. It exists as a metadata parameter within a Dolby Digital audio stream and is used by the decoder to alter the programme dynamic range, either because the capabilities of the replay equipment are in some way inadequate, or because the listener doesn't want the full dynamic range for some reason. DRC provides simple, yet flexible and comprehensive, control of peaks and low-level audio elements *(see Fig. 4)*.

Dolby Digital allows consumer audio systems to provide two different outputs to be taken with different dynamic range profiles: the Line Output drives hi-fi systems, the RF Output is tailored specifically to match the audio levels that result from typical RF-demodulated TV audio. Accordingly, the metadata in Dolby Digital conveys two DRC words, one principally for the line outputs and the other principally for the RF modulator. Different compressions may therefore be applied, optimized for the different functions. Note that the unity-gain segment in the middle of the compression profile *(in Fig. 4)* allows normalized speech to pass without level modification; also, note that the compression shifts abnormally loud or soft passages towards that speech level. An additional benefit of the correct use of Dialnorm and DRC is that multichannel audio mixes can be downmixed more faithfully onto two-channel systems.

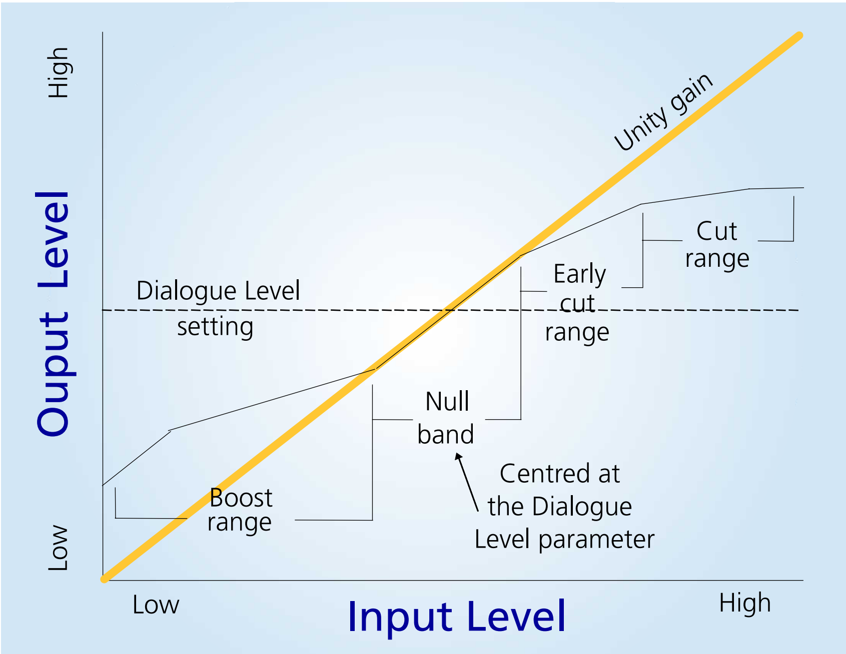

Figure 4
Characteristics of Dynamic Range Control

More details of Dialnorm, DRC and the ability to downmix have been well described by Vernon [5].

1. Dolby is a registered trademark of Dolby Laboratories.

Dialnorm and DRC are implemented in some 250 million decoders as part of the DVD and DTV standards, making these tools available to programme makers today.

How Dialnorm and DRC improve loudness and level matching

Continuing the examination of digital TV audio which, with MPEG audio, limits us to broadcasting a narrow dynamic range, how does digital broadcasting fare when using Dialnorm and DRC in Dolby Digital?

The RF output of a consumer set-top box has a reference-to-peak range of close to 8 dB. Dolby Digital can use DRC to give the signal an "analogue-like" 8 dB reference-to-peak range. Note that broadcast audio can now possess a greater dynamic range – as the RF output is created in the receiver. As a result, we can broadcast wide-dynamic-range audio that can be listened to with that full dynamic range, tailored to suit a good-quality hi-fi system or matched to the typical dynamic range of analogue (and today's digital) TV equipment. Also, through the use of Dialnorm, there are no major loudness differences when switching channels.

Fig. 5 illustrates the comparative dynamic range that can be received in different applications through this approach. Note that the MPEG broadcast has to be compressed at source to allow for the lowest common denominator – that of the basic analogue mono compatibility. The home cinema replay uses the digital bitstream output to decode the audio in an A/V amp, allowing the full dynamic range to be experienced by those who want this.

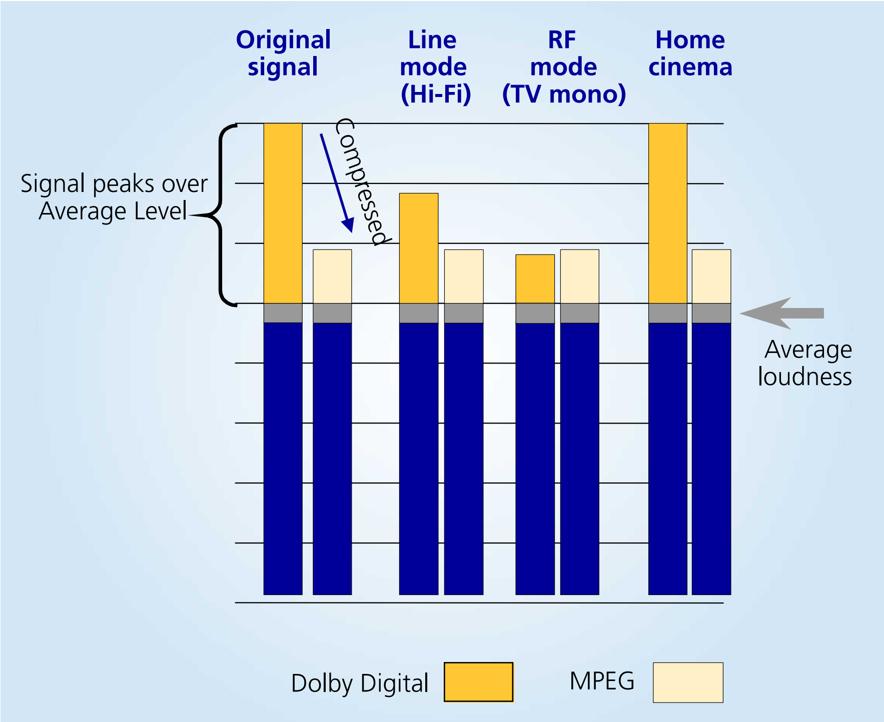

Figure 5
Comparative Dynamic Range in different types of audio system

Programme loudness levels are dictated by the need to match average loudness to a consistent electrical operating level in different replay environments: this forces simple audio systems such as MPEG or PCM to reach a compromise. The compromise can give slightly improved dynamic range around a lower average loudness level in hi-fi and home theatre environments, at the expense of "quiet dialogue" in applications where the programmes will be compared directly with today's analogue sources in TVs, for example. However, this compromise results in the level variations typically found between MPEG and PCM programmes. This is because programme makers have to make their own decisions as to "how loud the programme plays", and clearly different programme makers have different ideas, aims and benchmarks for this. Use of Dialnorm and DRC allows ideal placement of levels to suit the replay environments.

Fig. 6 shows how loudness levels within home equipment relate to the broadcast, using simple (MPEG or PCM) audio systems and systems using Dialnorm and DRC to tailor loudness and dynamic range in the receiver. Use of Dialnorm and DRC, along with separate outputs for Line, RF and Home Cinema, allows the operating-level-to-peak ratio to be matched to the replay equipment's electrical levels. Note that in practical TV (RF) replay, the Dolby Digital RF output is very close (typically only 1–2 dB different) to the MPEG/PCM and analogue operating level.

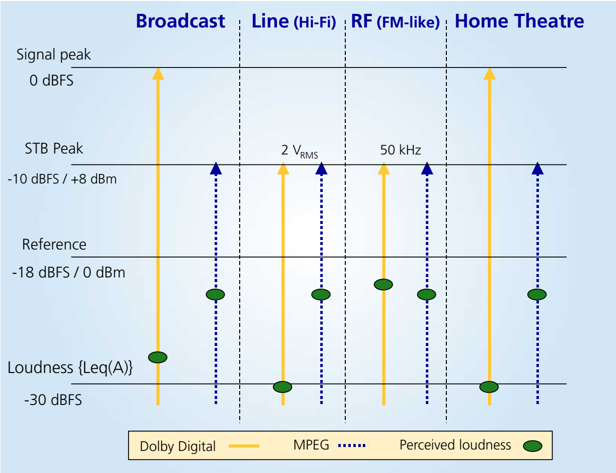

Figure 6
How loudness levels in the home match the broadcast - I

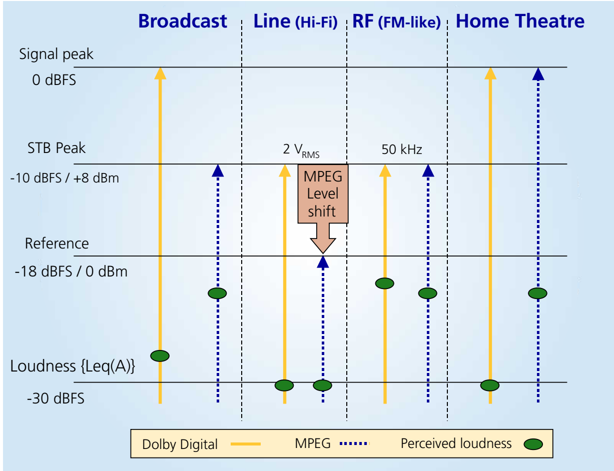

Figure 7
How loudness levels in the home match the broadcast - II

Note in particular that the potential dynamic range of the home cinema / top-end home audio system is at least as great as that of the broadcast, allowing where necessary 31 dB between average and peak levels.

The remaining inconsistency in *Fig. 6* is in the hi-fi replay levels. In order to match the levels in hi-fi, a practical solution is to reduce the level of the decoded MPEG audio in decoders *(see Fig. 7)*.

A practical illustration of how this can be achieved in TV audio decoders is shown in *Fig. 8* [1].

Practical Implementation

The conclusion for digital TV – and other audio delivery formats – is that *reference level* as an approximation for *loudness* will no longer work if we have a *wide dynamic range* signal. Therefore, we need to measure

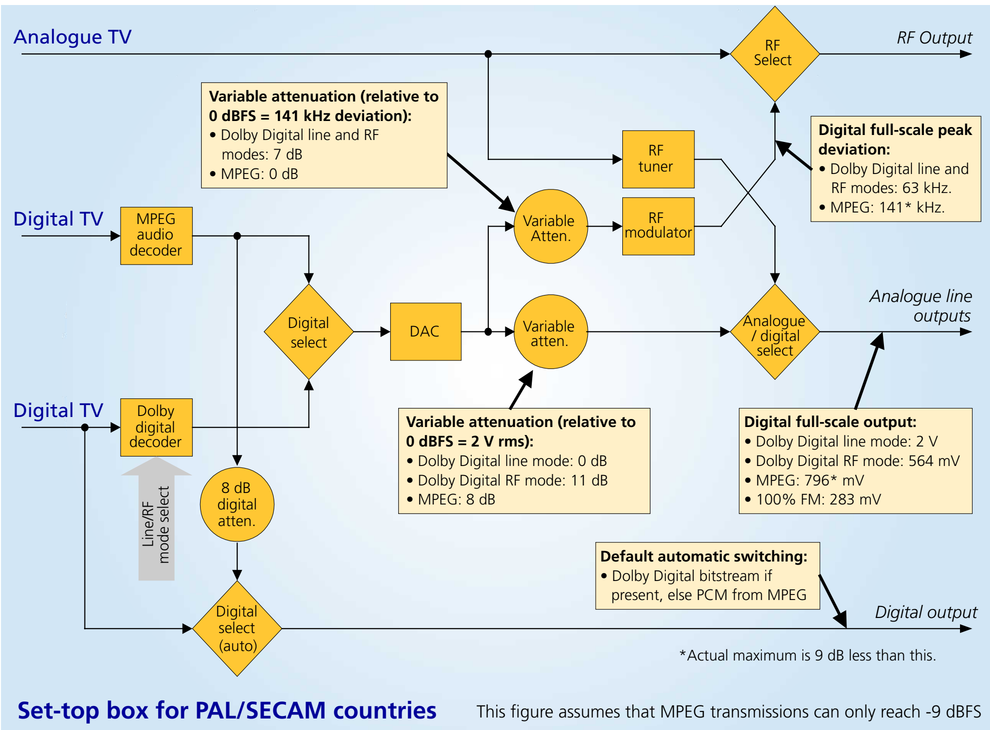

Figure 8
A practical DTV audio receiver / decoder

loudness and carry that measurement through to the home. Dialogue Normalization is a convenient (and widely implemented) way of doing this. Loudness can be measured with an Leq(A) meter; this value, relative to 0 dBFS, is entered as the Dialnorm value in a Dolby Digital bitstream, which then carries both encoded audio and loudness information to the home decoder. DRC parameters can then be carried in the same audio stream for further trimming of dynamic range, in particular to match audio dynamics in a narrow dynamic range RF stage that feeds a TV input.

Note that DTV and DVD are not the only applications where these concepts carry weight. Listening to classical music in a car, for example, shows quite effectively how "one size fits all" compression just doesn't work for everyone.

Conclusions

Reference Level ≠ Loudness Level. Dialogue Normalization provides an effective measure of programme loudness that can be carried along with the audio into the home. TV-specific RF outputs have limited dynamic range: MPEG and PCM audio dynamic range is limited by the need to operate through this RF output. Dynamic Range Control creates a range suitable for RF or line outputs without limiting the range of the content. Achieving this necessitates an appropriate Dialnorm setting to indicate loudness.

As a final note, audio professionals are usually interested in delivering better audio quality to their audience. A practical way to deliver audio that allows the better audio quality promised by digital to be appreciated by the consumer, has been described here. It allows the consumer to get the highest audio quality possible and is not limited by the need to suit the lowest-quality audio system. The logical way to achieve this is to carry out intelligent processing in the audio decoder. It requires a change in thinking by audio professionals but, for once, the consumer is ahead of the professional by being already equipped to handle this.

Tony Spath is responsible for marketing Dolby's technologies in Broadcast, Film and Consumer applications including DVD, Digital TV and Gaming platforms. He has worked with Dolby since 1985, during which time he has been closely involved with Dolby SR, AC-2 and AC-3 technologies and their practical application in broadcast production and post-production. He has also been involved in establishing and developing new markets around the world for Dolby cinema and film technologies.

Before joining Dolby, Mr Spath worked for 10 years as a music recording engineer and producer, and gained a Tonmeister degree from the University of Surrey in England.

Bibliography

[1] K.J. Gundry: **Dialogue Normalization in Dolby Digital and Co-existence With Other Broadcast Systems**
Proceedings of the International Broadcast Convention, September 2001.

[2] ETSI EN 101 154: **DVB, Digital Video Broadcasting: Implementation Guidelines for the use of MPEG-2 Systems, Video and Audio in Satellite, Cable and Terrestrial Broadcasting Applications**
http://pda.etsi.org/pda/home.asp?wki_id=6099

[3] **Standards Document A/52**, pp.76 - 81 and **Standards Document A/54**, pp. 49 - 52
United States Advanced Television Systems Committee.

[4] Mark F. Davis: **The AC-3 Multichannel Coder**
Presented at the 95th Convention of the Audio Engineering Society, October 1995.

[5] Steve Vernon: **Dolby Digital Audio Coding for Digital Television and Storage Applications**
Presented at the 17th International Conference of the Audio Engineering Society (High Quality Audio Coding), September 1999.