

LOUDNESS CONTROL IN DIGITAL BROADCAST

T. Lund

TC Electronic A/S, Denmark

ABSTRACT

Worldwide, broadcasters are looking for practical approaches to digital audio production and delivery. One of the main dilemmas is to assess and control loudness with precision but without latency. Concerns include the added production time predicted with some suggestions of pre-transmission loudness measurements and format control. More TV channels with the same amount of viewers is not an indication to employ extra time in production.

In this paper, new realtime approaches to securing loudness consistency, speech intelligibility, audio quality and format control are described. Experience from broadcast stations will be reported, including strategies for optimizing perceptually data-reduced audio, and targeting it for specific groups of listeners.

1 INTRODUCTION

It still remains to be seen what aspects of DTV users will find the most attractive. Number of overall TV channels, consistency and quality of audio and picture, new services or a combination of the above. The optimum balance will be different from country to country, and between different groups of audiences.

This paper addresses current audio issues and projected experiments, discuss current trends, and examine how analog broadcast can co-exist with digital in these times of transition.

Section 2 is about parameters to ensure audio quality, while sections 3-5 deal with current tendencies and future needs in the broadcast chain.

2 AUDIO QUALITY

The entire digital broadcast signalpath prior to transmission should aim for a resolution of at least 20 bit, 48kHz linear audio. The latest data compression codecs for transmission may take advantage of more than 20 bits of linear audio, even though a data-reduced transfer rate of 256kbps, 192kbps, 160kbps or less is used.

With this in mind, we will examine aspects of the current and future audio signalpath at the broadcaster.

2.1 AD and DA conversion

The broadcast signalpath is still typically a mix of analog and digital equipment.

From a resolution point of view, use the bits available in any A to D and D to A converter. When all the bits are activated, the AD converter produces a Digital Full Scale signal, or 0 dBFS. Audio processors should let the user calibrate the scaling of the AD and the DA and keep that scaling stored with presets. Gain must be applied before the AD converter and after the DA, so the full

dynamic range of the converters can be utilized regardless of the absolute analog level handling required. An example of flexible AD and DA analog domain scaling is shown in **Fig 1**.

The analog level required for 0 dBFS can vary from country to country. EBU R68 is used in most European countries. This standard specifies +18 dBu at 0 dBFS, while US installations use +24 dBu for 0 dBFS. In Japan, France and some other countries, converters may be calibrated for +22 dBu at 0 dBFS. Note how different scaling may affect the amount of headroom available above 0 dBFS discussed later in this paper.

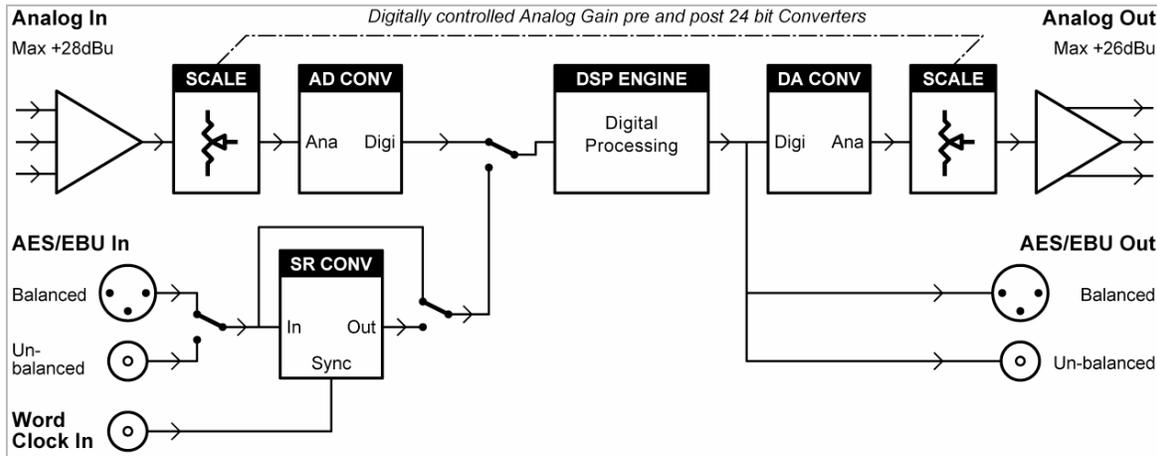


Figure 1: Analog domain scaling of AD and DA converters.

2.2 Digital interfacing

If audio equipment is interfaced digitally, there is no need to worry about level calibration. A digital full scale on a DAT recorder is the same as a full scale on an expensive digital console. Levels are easier maintained with digital interfacing than in the analog domain. Also, differences between level standards are less troublesome in the digital domain, because the uncertainty of what analog level represents 0dBFS in a particular piece of equipment is avoided.

The main problem with digital interfacing is format incompatibility if an industry standard like e.g. AES3 isn't used. Also, jitter build-up can lead to loss of signal, and level close to full scale can create overloads in the receiving device. More info about the last phenomena can be found in the 0 dBFS+ paragraph and in the references [11, 12, 13, 15].

It should also be noted that the use of sample rate converters ought to be kept at a minimum. Synchronous routing and interfacing should be used to avoid signal degradation from jitter or 0 dBFS+ signals [13, 15].

2.3 Level Meters

Analog peak meters, often called PPM's, are normally found in European broadcast and film. They use an integration time of 5-10 ms and therefore don't show the shortest program peaks. This family of meters is defined in IEC 268 [4] and DIN 45406.

VU meters, invented by Bell Labs in the late 1930's, show very little transient information at all, because 99% deflection isn't reached before a signal has lasted for 300 ms. In the digital domain, hundreds of samples pass by without the highest value showing up on a PPM or VU meter, but an overload can still produce alias distortion and other unpleasant artifacts in the digital domain.

Consequently, sample based digital meters show a peak read-out 3-5 dB higher than analog PPM's and 16-20 dB higher than VU's, even though a sine tone is shown at a reference level on all instruments.

2.4 Reference level

In film, post and broadcast production, a certain reference or line-up level is often defined to specify what is considered a normal operating level. The reference level is typically the level around which dialog is centered. It is also the standard level for alignment tones.

In European broadcast, the reference level is -18 dBFS (EBU R68), in the US -20 dBFS has been chosen (SMPTE RP155).

2.5 Level above 0 dBFS

In the digital domain, the peak level may deviate from the peak level in the analog domain. If a digital signal represents level which exceeds a normal sine wave peaking at 0 dBFS when upsampled or converted to analog, we will define such level to be 0 dBFS+. It has been recently shown that 0 dBFS+ peaks are readily derived from commercial CDs, that they happen more and more frequently, and that such peaks may create obvious distortion in consumer equipment [13] as well as professional equipment for domain and sample rate conversion [14], dynamics processors and data reduction encoders and decoders [15].

0 dBFS+ peaks on recent releases can happen hundreds of times on just a single song, leaving reproduction equipment or data-reduction codecs in more or less a permanent state of distortion. Distortion figures can amount to more than 10% plus a build-up of listening fatigue. An example of the type of signals a digital signalpath in 2003 has to handle is shown in **Table 1**. More extensive information can be found in [15].

Track	Artist	Year	Max Dig. dBFS	Hot 1 no/10 s	Hot 2 no/10 s	Status
Lose Yourself	Eminem	2002	0.0	>25 sus	>25	4
La Fiesta De Amadito	Amadito Valdez	2002	0.0	2	0	1
Don't Stop	Anastacia	2001	0.0	>25 sus	15	4
Played Alive	Safri Duo	2001	0.0	>25	16	3
The Call	Backstreet Boys	2000	0.0	>25 sus	18	4
I Got a Girl	Lou Bega	1999	0.0	>25 sus	3	4
Let's Get Loud	Jennifer Lopez	1999	0.0	>25 sus	10	4
Smooth	Santana	1999	0.0	20 sus	15	4
Believe	Cher	1998	0.0	10	4	3
Vissa Har Det	Bo Kaspers Ork.	1998	0.0	1	0	1
That Don't Impress Me Much	Shania Twain	1998	0.0	3	0	2
Miami	Will Smith	1997	0.0	17	9	3
El Cuarte de Tula	Buena Vista SC	1997	- 0.2	0	0	1
Block Rockin' Beats	Chemical Bros.	1997	0.0	8	5	3
Wannabe	Spice Girls	1996	0.0	5	0	2
We'll be Together	Sting	1994	- 0.2	1	0	1
Off the Ground	Paul McCartney	1993	0.0	1	0	1
I've Been to Memphis	Lyle Lovett	1992	- 0.9	0	0	1
Mysterious Ways	U2	1991	- 0.1	0	0	1
Black or White	Michael Jackson	1991	- 0.2	0	0	1
The End of the Innocence	Don Henley	1989	- 2.2	0	0	1
Nick of Time	Bonnie Raitt	1989	- 2.1	0	0	1
Dirty Blvd	Lou Reed	1988	- 0.2	0	0	1
Graceland	Paul Simon	1986	- 3.4	0	0	0
Two Tribes	Frankie Goes...	1984	- 0.7	1	0	1
She Took Off My Romeos	David Lindley	1981	- 1.9	0	0	1
Little Sister	Ry Cooder	1979	- 8.7	0	0	0

Table 1: Commercial CD Level going up.

Table 1 abbreviations:

Max Dig.: Max repeated digital level encoded on CD.

Hot 1: Typical number of occurrences of hits between 0 dBFS and +1 dBFS per 10 seconds.

Hot 2: Typical number of occurrences of hits above +1 dBFS per 10 seconds.

Status: 0: Low level, 1: Good level, 2: High level, 3: Probable overload of a broadcast signalpath, 4: Certain overload of normal broadcast signalpath.

In the IEC standard for digital peak level meters [4, 5] the problem is recognised, but it is accepted that a meter can be inaccurate at certain frequencies related to the sample rate. Newer digital peak meters overcome this problem by estimating the signal between the samples.

Not only when converting from digital to analog the signal is reconstructed with a potentially higher peak value as immediately seen, but also when passing the signal through a sample rate converter or a data reduction codec there is a risk of overload [15].

2.6 Alias distortion

In a digital system, when higher frequencies than $F_s/2$ are passed from the input or created in the signal processing itself, alias signals are mirrored at $F_s/2$. Like in the analog domain, digital processing does produce distortion due to lack of headroom, rounding errors and other limitations of precision. But the higher the sampling rate, the less the risk of significant signals hitting $F_s/2$ and mirroring back as non-harmonic alias distortion into the audible frequency band, as long as signals are not hard clipped in the digital domain.

Alias distortion is difficult to handle for a data compression codec, and should be minimized whenever possible. Even if 32kHz sampling is used for digital transmission, higher sampling rates should be used in the upstream signalpath to ensure a clean signal from which to extract information. An example of alias distortion is shown in **Fig 2**.

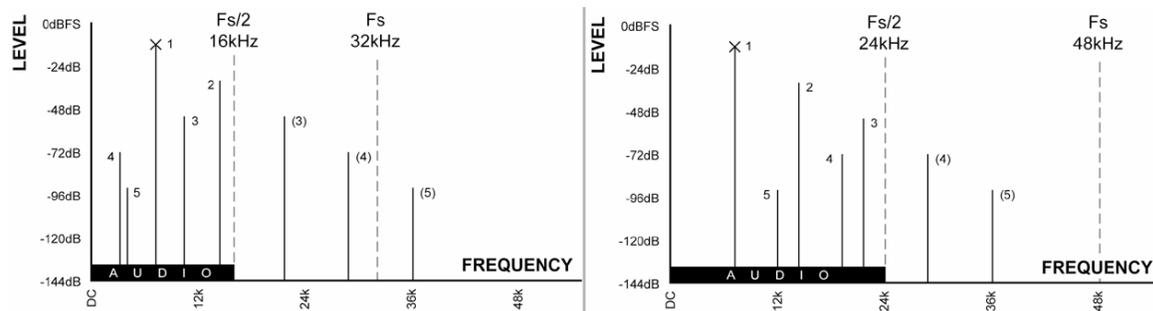


Figure 2: Alias Distortion

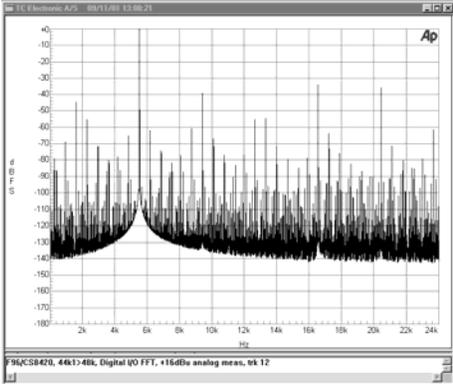
Alias distortion on Sine tone (marked "x") with processing sample rates of 32 and 48 kHz.

Fig 3 shows an example of alias distortion from a sample rate converter subjected to a sine wave of 5.512,5 kHz ($f_s/8$) with a starting phase of 67.5 degrees. The theoretical peak should have been +0.69 dBFS, but instead wideband distortion is the result. At the time of writing, the author has not seen a

commercial sample rate converter chip not producing this type of result.

Figure 3: Alias Distortion from 0 dBFS+ signal

The CS8420 chip fed with an $f_s/8$ sine wave.



3 BROADCAST PROCESSING STRATEGIES

If ordinary consumers of TV are asked, the most often heard requests for better audio are:

- being able to always hear the words,
- not having to constantly change level with the remote.

In a digital system there are no technical reasons why a wider dynamic range could not be allowed, but the average end listener doesn't want wide dynamic range if she hasn't got an easy to use way of restricting it.

The broadcaster has to choose the processing strategy that fits most of the viewers best. The strategy may be adjusted, if better domestic processing becomes the rule years from now, but the end-listener would need multiband processing then, to avoid a change for the worse.

3.1 Speed and Quality

Broadcast stations are specialized in getting the dynamic range right for domestic listening. Digital multiband processors are now not only being used for transmission, but also in production and OB environments because they can ensure speech intelligibility and consistent audio quality despite a high working speed.

They can optimize loudness but may also be used as dynamic spectral balancers and brickwall limiters, assisting the mix engineer and allowing him to work faster. Multiband processing in production works well for analog and digital transmission and is a valid way of integrating the requests of the end-listener into the sound design.

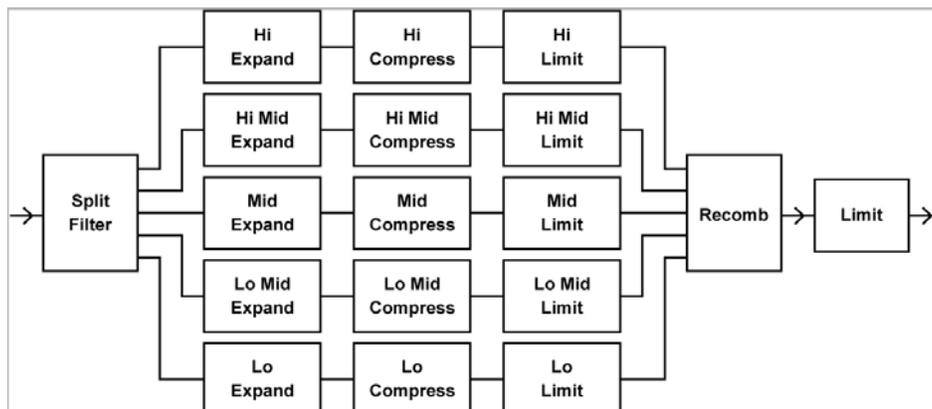
3.2 Multiband dynamics

In a multiband processor the audio band is split into a number of frequency ranges, each compressed, limited and expanded individually. An example is shown in **Fig 4**.

Digital processors enable look-ahead detection and perfect re-combination of the bands, thereby overcoming the major problem with analog multiband systems.

Advantages of multiband processing:

- maximum speech intelligibility.
- modulation and loss of high frequency signal is avoided.
- higher loudness and less obtrusive limiting may be obtained.
- spectral balancing may be used to help obtain a specific frequency response at a specific level.



*Figure 4: Multiband dynamics processor.
Five frequency bands and re-combination limiter at the output.*

3.3 Spectral balancing

Some of the user-friendliness built into a signalpath relying on analog tape has been lost in digital production. Most noticeably, there is now no longer a restriction of high levels at high frequencies, so human voice can get overly bright because of equalization applied to obtain clarity at low levels.

In multiband processors, dynamic spectral balancing may be used to compensate for our ear's increasing non-linearity at low levels, and to avoid too much brightness at high levels. Using this technique, it is possible to boost e.g. low and high frequencies at low levels but have the frequency response flat or even cut at high levels.

Dynamic spectral balancing is achieved by using different compressor thresholds, ratios and levels in the different bands of a multiband processor. Independent controls for all the bands should be available for maximum control over spectral balance, and the number of bands should be selectable.

4 CONTROLLING LOUDNESS

If program material is limited to human voice only, simple loudness assessment based upon IEC 268-10 (PPM level meter) or a weighted Leq measurement may be all that is needed to ensure level consistency and balance within and across broadcast programs. Experienced engineers can, for example, achieve good results with PPM meters when balancing dialog, because an "almost peak" approach is quite consistent with the apparent loudness of human voices.

With recent years advances in realtime processing capabilities, it is now possible to measure and control the loudness of signals more complex than speech.

When there is a desire to control loudness automatically in realtime, it is important that different equipment manufacturers don't try to invent their own loudness standard, but rely on the extensive basic research done by pioneers such as Zwicker and Fastl [8] more than two decades ago. It seems that one of the dose approaches, called Leq(m), has now been retracted, but new "truths" with dubious technical foundation may emerge.

Loudness measurement of complex signals, instead should be based on an international standard, such as ISO 532, taking the spectrum and masking factors of the sound into account. A set of contours of equal-loudness is defined in ISO 226 (Fletcher and Munson).

An example of a realtime loudness processor with multiband capabilities is shown in **Fig. 5**. The algorithm uses wide dynamic range with internal headroom well above 24 bits to avoid short-term overload and clipping without sacrificing resolution. Latency of around 1 ms is made possible by using a combination of loudness estimates and long term adjustments.

Quick estimates of the loudness are based upon weighted measurements, and level is corrected accordingly using adaptive time constants. Short-term effects are moderated by a multiband processor and instant output limiter, while long-term loudness can be subject to further adjustments by including the perceptually based feedback loop as shown. An algorithm with a topology for the most part following the example has shown remarkably consistent loudness and spectral results.

Multiband Loudness

Block Diagram

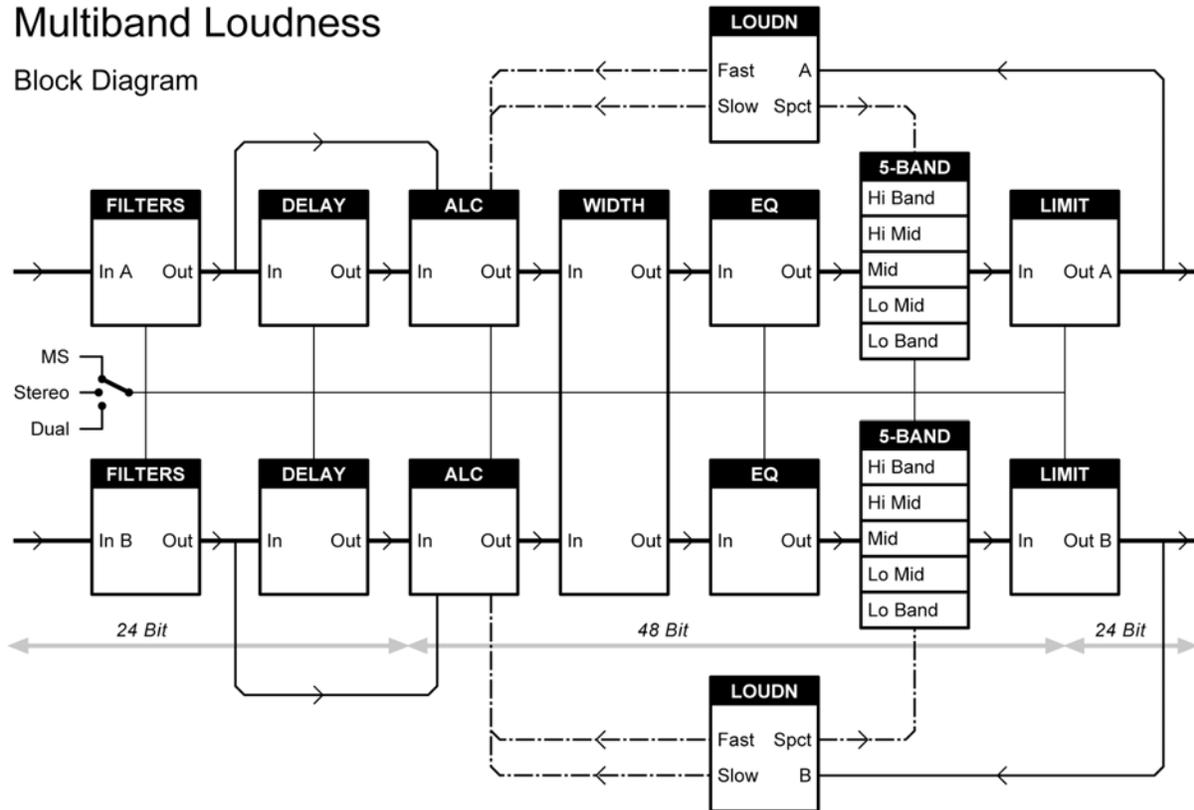


Figure 5: Realtime Loudness and Multiband processor

4.1 Multiple formats

In digital broadcast, audio will change between multiple mono, stereo and multichannel, and even a mix of the formats. Consequently, broadcasters around the world are installing processing and interfacing to cope with such demands in key areas of their infrastructure.

For multichannel dynamics processing, different linking topologies should be available. The 5 main channels need different linking for film as compared to music, while the LFE channel normally should not be linked to the others. Because 5.1 programs will mostly be film or high profile TV programs, multiple slopes must be allowed in order to e.g. bring up low level dialog, leave reference level material unaffected and limit headroom.

4.2 Data Reduction

Linear digital audio contain plenty of redundant information, so in order to conserve bandwidth most transmission formats take advantage of data reduction for audio and video.

Among informed audio consumers it is generally agreed that high quality audio may be picked up from a 1:6 or 1:8 data reduction transmission, if audio upstream from transmission is kept at a high resolution and free of alias distortion. With the AAC codec, the transmission bandwidth required to maintain “transparent audio” has now probably dropped further to around 64 kbps per channel.

Recent tests at TC Electronic with some of the most relevant perceptual codecs are thoroughly dealt with in [15] with an excerpt re. peak handling shown in **Table 2**.

Algorithm	Mode	Datarate [kbit/s]	Avg. per ch. [kbit/s]	Max peak re. 0.5
MPEG-1 L II	stereo	384	192	+1.3 dB
MPEG-1 L II	stereo	224	112	+1.3 dB
MPEG-1 L III	stereo HQ	320	160	+1.7 dB
MPEG-1 L III	stereo HQ	160	180	+2.3 dB
MPEG-1 L III	int-st HQ	128	64	+5.3 dB
MPEG-1 L III	int-st fast	128	64	+3.0 dB
MPEG-1 L III	int-st HQ	96	48	+4.7 dB
MPEG-2 L III	22.05 kHz, i-st HQ	80	40	+1.7 dB
DTS	6 ch.	1234	206	+0.6 dB

Table 2: Maximum peak values observed in 12 hot CD excerpts (length 14-33 s) perceptually coded with various algorithms, data rates and modes.

4.3 Metadata

In digital transmission, audio signals can be accompanied by data about the audio, so-called metadata. The audio itself can be linear or data-reduced as required, but the reduction format can often be regarded separately from the metadata.

Carriers like MPEG-2 or MPEG-4 can be combined with audio data reduction such as AAC to form a complete industry standard video and audio broadcast system. Metadata can be of different complexity. From basic, like in older developments such as the AES3 interface, Eureka 147 DAB [3] or Dolby AC3, to extensive content descriptors, like in MPEG-4 and MPEG-7. Basic audio metadata include information about how the channels are used, but new designs also enable the user to decompose or remix the audio, spatialize it, synthesize translations etc.

It remains to be seen how many of these possibilities will be used. It is worth noting that several broadcasters currently shy away from using metadata to do more than change between e.g. multiple mono, stereo and (in some cases) 5.1 reproduction.

5 DOMESTIC AND BROADCAST STATION AUDIO DEVELOPMENTS

Consumers are getting more spoiled by the minute. They demand content when they need it, with audio and video perfectly scaled for their present situation. If the family is watching an action film, you might specify theatre-like surround sound with wide dynamic range, but if you're watching alone at night with the children asleep, very restricted dynamic range and mono is probably chosen. Therefore, extensive processing capabilities at the listener will soon be taken for granted. Listeners will request better speech intelligibility than now, so outdated wideband dynamics processing at the home will not be acceptable. Instead, multiband dynamics processors will find way into the homes alongside processing to maximize envelopment and spaciousness for a highly scalable speaker set-up.

Stations generally want to make sure new audio installations are build on industry standards rather than proprietary technology to efficiently distribute and control audio without latency, and to be able to maintain an installation without high costs. Audio installations today are typically designed around standard interface formats such as AES3 (balanced or on coax) or SMPTE 259 SDI.

Early moving countries within digital TV have experimented to find benefits of value to their viewers. For sports, multiple mono commentaries biased towards either team has generated interest,

and some countries are in the middle of experiments with a secondary stereo signal targeted at an elderly audience with the hearing difficulty known as presbycusis. Multichannel audio hasn't been the main use of digital broadcast up until now, besides from occasional transmission of feature film.

6 CONCLUSION

The broadcaster needs to maintain as much of the current working practices as possible through the period of analog to digital transition. Because analog transmission will be around for several years to come, and because the current broadcast workflow is efficient. Sacrificing *workflow efficiency* at the broadcast station would be dangerous with the number of TV channels multiplying within a near future.

Attempts of carrying *metadata* through from production to transmission should therefore be avoided. Extra work is implied in production, and metadata may get lost or entered incorrectly.

Instead, broadcasters should take advantage of recent year's improvements in *realtime loudness measurement and control* to improve in areas important to the end listener, while keeping production costs low. Many broadcasters have followed this principle, even if they are required to transmit data-reduced audio with metadata included. Metadata is frozen in transmission, and only changed when the format has to.

Several *experiments* will be needed before each broadcast-station finds its best use of the digital possibilities. Such experiments will have to be conducted while analog services are still active, and while parts of the digital signal chain operate normally. Therefore, *flexibility* in processing and format handling is another keyword of the transition period, and conclusions may not even be the same in all countries.

Stations that wish to take advantage of the improved audio qualities in digital transmission, should design the infrastructure and file management to maintain higher internal resolution than the data reduced transmission signal. The signalpath needs to maintain *headroom* to accommodate 0 dBFS+ signals which are becoming more and more frequent from production and commercial CDs. If such headroom is not implemented, the sound from the station will be at risk of generating listening fatigue, making viewers and listeners turn away. Linear audio at 48 kHz sample rate and 20 bit resolution with 6 dB of full scale headroom or higher is a safe standard to use inside the station.

To use proprietary data reduction systems inside the broadcast station is probably unwise, because of the danger of getting dependent on a single supplier. Better to rely on *international standard* solutions such as linear audio and SMPTE 259 or AES 3 interfaces. Current tendencies among broadcast stations worldwide widely follow the advice given in this conclusion.

REFERENCES

- [1] ITU (1995): Rec. ITU-R BS.412-7, Planning standards for FM sound broadcasting at VHF, section 2.3.
- [2] ATSC: Digital Audio Compression Standard A/52, 1995.
- [3] IEC (1990): IEC 268-17, Standard volume indicators, First edition.
- [4] IEC (1991): IEC 268-10, Peak program level meters, Second edition.
- [5] IEC (1995): IEC 268-18, Peak program level meters - Digital audio peak level meter, First edition.
- [6] A. Oppenheim & R. Schaffer (1975), Digital Signal Processing, Prentice-Hall, Englewood Cliffs.
- [7] ISO standards 226, 532.
- [8] E. Zwicker & H. Fastl (1990): Psychoacoustics - Facts and Models, Springer-Verlag, Berlin.
- [9] Carsten Hansen: Objective Reading of Loudness of a Sound Programme, AES technical paper, Copenhagen 1996.
- [10] Tomlinson Holman: Headroom of Various Media, Sound for Film and Television, 1997.
- [11] Soren H. Nielsen & Thomas Lund: Level Control in Digital Mastering. Proceedings of the 107th AES convention, New York 1999.
- [12] Soren H. Nielsen & Thomas Lund: 0dBFS+ Level in Digital Mastering. Proceedings of the 109th AES convention, Los Angeles 2000.
- [13] Soren H. Nielsen & Thomas Lund: 0dBFS+ Level in Audio Production. Proceedings of the 111th AES convention, New York 2001.
- [14] Thomas Lund: Dynamics Processing in Digital Audio Production. Proceedings of the NAB2000 convention, Las Vegas 2000.
- [15] Soren H. Nielsen & Thomas Lund: Overload in Signal Conversion. Proceedings of the AES 23 conference, Elsinore, 2003.