

## REPORT ITU-R BS.2054-1

**Audio levels and loudness**

(2005-2008)

## CONTENTS

	<i>Page</i>
1 Introduction .....	2
2 Simulcasting of analogue and MPEG-1 layer II digital audio.....	2
3 Loudness considerations.....	2
4 Operating in the analogue domain.....	2
5 Operating in the digital domain.....	4
6 Use of audio meters.....	4
6.1 VU meters.....	4
6.2 Peak meters.....	5
6.3 Loudness meters.....	5
7 Harmonization of audio alignment levels for digital programme exchange – Adoption of SMPTE RP155 .....	5
8 Peak audio level.....	6
9 The studio environment.....	6
10 Application of volume compression in post-production following the final mix of a television commercial soundtrack .....	7
11 Ingest of soundtracks into the television broadcasting chain .....	8
12 Tests performed at CBS.....	9
13 Summary.....	10

## 1 Introduction

This Report describes advice for the broadcasting of television programmes which include pre-recorded television advertisements (commercials) with particular reference to audio levels and loudness. It considers the several processes of studio production, recording onto storage media, transport of the media, and broadcast via a television presentation and transmission system. This descriptive material is provided as guidance. It describes one administration's approach to dealing with the ingest and transmission of television soundtracks, in particular the factors which contribute to loudness.

Question ITU-R 2/6 *decides* the following Questions should be studied:

- a) What audio metering characteristics should be used to provide an accurate indication of signal level in order to assist the operator to avoid overload of digital media?
- b) What audio metering characteristics should be used to provide an accurate indication of subjective programme loudness?

While the work of the Rapporteur on Level Metering within Working Party 6P is progressing, some administrations are providing interim measures to address audio levels and loudness. For instance, Australian television broadcasters have established a common alignment level,  $-20$  dBFS, in accordance with SMPTE RP155. Guidelines have been introduced specifying that volume compression where used **after the final mix** of a television commercial soundtrack, be restricted to a slope of 2:1 with an onset point of  $-12$  dBFS.

## 2 Simulcasting of analogue and MPEG-1 layer II digital audio

Many administrations are currently progressing through or planning a transition to digital broadcasting. During the transition a requirement exists for simultaneous broadcasting in analogue and digital form.

Analogue and digital broadcasting systems have different parameters which in turn influence the operational characteristics of audio processors and, consequently, affect the perceived loudness of the analogue and digitally transmitted sound.

## 3 Loudness considerations

Within most broadcasting systems, television broadcasters transmit material of varying programme genres contiguously and interspersed with inserted material, including advertisements. There is a potential for variations in the perceived loudness of adjacent audio segments.

The factors contributing to perceived loudness are complex but the correct alignment of audio levels through the various stages of production and transmission and the careful management of dynamic range and spectral content all contribute to preventing extreme variations in perceived loudness.

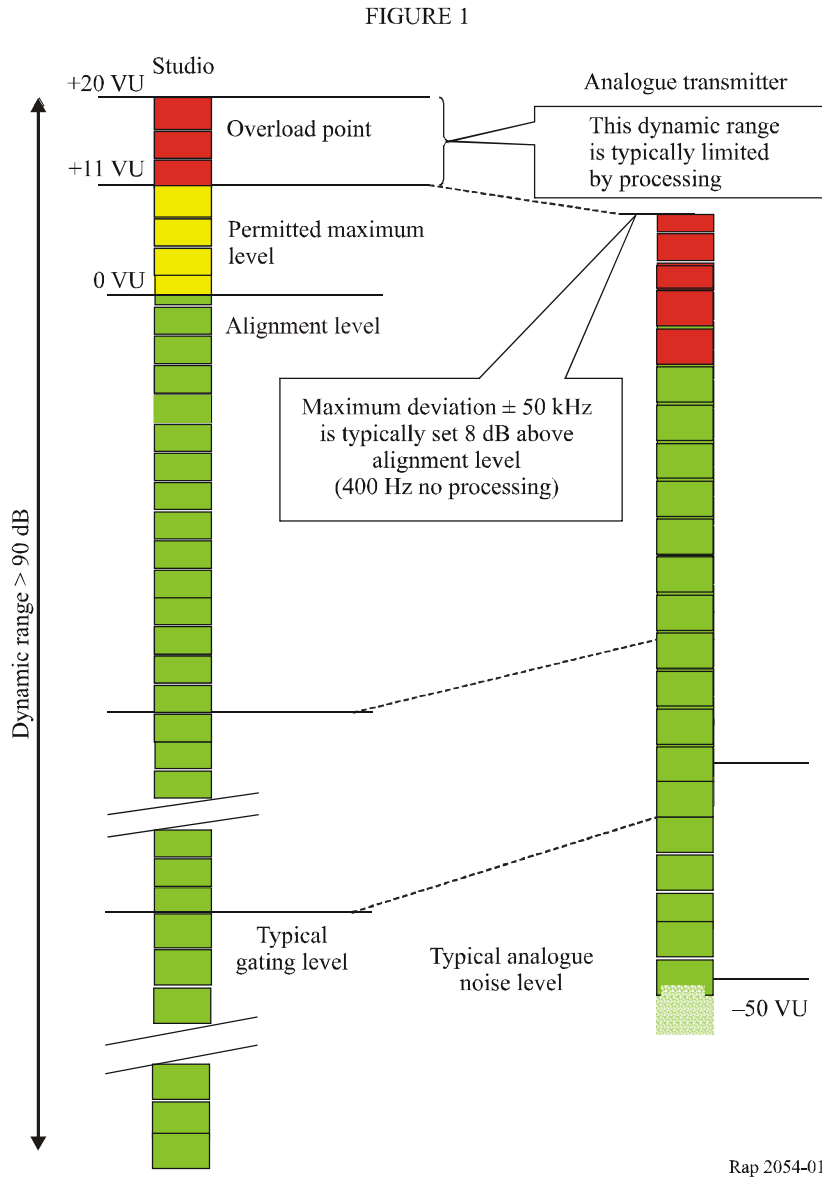
## 4 Operating in the analogue domain

In the analogue domain, the amount of headroom available in recording devices is limited by the amount of distortion that may occur at high levels of audio.

Magnetic recordings will be limited by the integrity of the magnetic image at high levels of coercion.

The analogue television transmission system is limited by the allowable deviation of the FM sound carrier.

Figure 1 describes the parameters and limits of the analogue television transmission audio system as adopted in Australia.



The lower limit of the analogue studio audio system is determined by the level of the system noise. Typically, this is maintained at least 50 dB below the alignment level providing some 70 dB of dynamic range. Although a studio analogue system and recording devices may provide adequate headroom for dynamic production, the limiting factor in an analogue broadcasting chain is the allowable deviation of the FM transmitter and broadcasters have to apply some compression and limiting of the audio level at the transmitter input to avoid over deviation, particularly at high frequencies where pre-emphasis can reduce the clipping level significantly compared to that at 400 Hz.

It is necessary to define the nature of peak level. Peak level in this context and as used throughout this document, really means what is commonly called “quasi peak” levels – the levels as measured by a PPM type meter having (typically) an integration time of 10 ms.

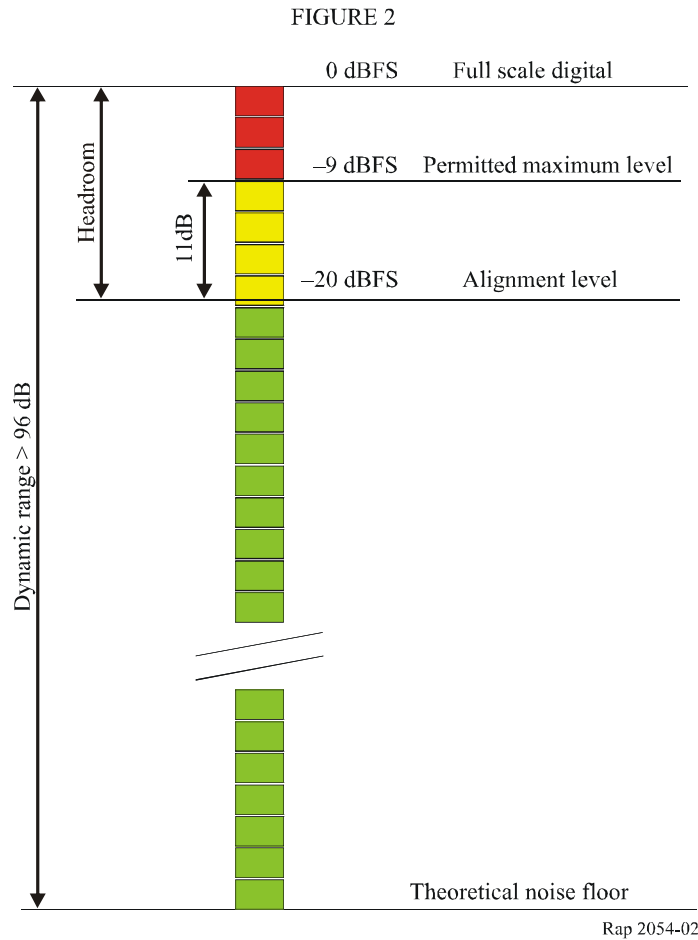
True peak level of programme content is usually not measured in the context of analogue audio systems.

## 5 Operating in the digital domain

Where production, post-production, switching and mixing of television programmes and advertisements is carried out in the digital domain, the digital audio system is aligned using a 1 kHz sine wave reference signal 20 dB below full-scale digital “0 dBFS” (i.e. 20 dB below the level at which digital clipping of the sine wave signal commences).

The alignment level –20 dBFS is usually equated to zero on the station’s VU meters or “4” or “TEST” on PPM metres.

Figure 2 describes the parameters and limits of the digital audio system as adopted by Australia.



In a 16-bit digital audio system where there is a theoretical dynamic-range of some 96 dB, if the average audio level is zero VU and the quasi peak level is some 11 dB above alignment level then there will be some 9 dB range before digital clipping occurs.

## 6 Use of audio meters

### 6.1 VU meters

Standard VU (Volume Unit) meters are commonly used in television facilities in many administrations to adjust audio levels during the recording, playback and transmission of audio material. A VU meter will indicate an “average” audio level, but cannot indicate the instantaneous level of audio peaks. The standard VU meter has an integration time of 300 ms and thus its “average” reading relates to that particular integration time.

A VU meter scale is not capable of directly representing the full dynamic range of audio signals in analogue or digital systems but VU meters provide a convenient means of ascertaining that the audio level is within normal parameters relative to the alignment level and they provide some limited indication of loudness.

It should be noted that in production and broadcasting establishments there is a wide range of audio level indicators including LED and on-screen devices commonly referred to as “VU meters”. Although most audio level indicators will return consistent measurements under steady state conditions such as that produced by a sinusoidal alignment tone, there may be inconsistencies between instruments where they are used to measure programme levels in dynamic material.

## **6.2 Peak meters**

Where meters such as Peak Programme Meters (PPMs) are used, they indicate quasi peak levels more accurately than the VU meter because their internal time constants are optimized for such measurement.

Since Peak Programme Meters are somewhat better indicators of quasi peak audio level they are more useful in the management of system headroom but are of limited use in assessing loudness.

True peak meters indicate the absolute peak level of a signal to a limit of single digital sample values. As such they have no value or use in assessing loudness. However, their value is in controlling digital overload issues. A suitable true peak meter algorithm is described in Recommendation ITU-R BS.1770.

## **6.3 Loudness meters**

The development of new meters capable of measuring the transmitted electrical signal in a way that will correlate to the human perception of loudness when that electrical signal is reproduced on loudspeakers in a typical “domestic” listening environment has led to the development of Recommendation ITU-R BS.1770.

It is envisaged that such meters will eventually provide producers and broadcasters with an objective means of comparing the perceived loudness of adjacent programme segments or commercial/programme junctions or differences between services solely by measuring the electrical signal level.

## **7 Harmonization of audio alignment levels for digital programme exchange – Adoption of SMPTE RP155**

Australian television broadcasters have adopted SMPTE RP 155 audio levels for the digital audio interface.

For television recordings a sinusoidal steady state tone at 1 kHz representing the alignment level of –20 dBFS should precede programme material presented for broadcast. This level is usually equated to zero (zero VU) on the station’s VU type audio level meters and is used to align the broadcaster’s recording and transmission equipment to the same reference level as the originating equipment.

When measured with a VU type meter, the normal audio level of the programme material that follows the alignment signal should be approximately zero VU.

When measured with a PPM or digital equivalent type meter, the normal quasi peak audio level of programme material will, depending on the level of processing used, vary typically in the range of +2 to +9 dB above alignment level (heavily compressed commercials and pop music may not peak above 4 or 4.5 on a PPM. In contrast wide range classical music might read ppm 6 to 6.5).

## 8 Peak audio level

In the digital domain quasi audio peaks should not exceed  $-9$  dBFS, i.e. quasi peak excursions should not be more than 11 dB above the alignment level. It must be understood that +11 dB in this context is not a deliberate aim point for production levels, but is a technical limit to be observed. This limit will help ensure that short-duration true peaks do not reach 0 dBFS (full scale).

In an analogue FM transmission system, quasi audio peaks should not exceed the alignment level by more than 8 dB.

These levels are recommended for optimum use of the available headroom in the analogue and digital systems.

## 9 The studio environment

Studio analogue sound systems are capable of mixing, recording and reproducing material with dynamic ranges extending from the level of the audible system noise to the level at which distortion is unacceptable. For practical purposes, this represents a dynamic range of some 70 dB.

Studio digital sound systems typically operate with dynamic ranges of more than 90 dB. The lower limit in a digital audio system is determined by the theoretical digital noise floor where there is no meaningful data. This lower limit is principally determined by the audio word length (16, 18, 20, 24 bits).

The upper limit in a digital audio system is defined as the full-scale digital level, 0 dBFS. At that point, digital clipping occurs because the audio signal cannot be adequately represented by the finite number of data bits available.

Using a VU metering system, programme audio material should be recorded such that the normal programme level is around zero VU with occasional louder passages allowed to exceed this level by 2 or 3 dB (+3 dB being the limit on most VU meters). In a normal broadcast audio mix of speech and music and/or sound effects (not a proprietary multichannel surround mix), the dialogue level will typically fall at around  $-2$  to  $-3$  dB below the alignment level. For significantly processed material such as commercials or pop music the VU meter reading should not be permitted to exceed zero VU.

In both the production and transmission phases of audio, it is common to employ various forms of audio processing. In the production phase processing is a normal part of the creative process. However, both the production and emission processing will have a common aim, i.e. to provide material to the viewer so that the loudest and softest passages of the material can be enjoyed without the need to adjust the receiver volume control. An extreme example of this is the emission processing of cinema style audio mixes which require compression for comfortable listening in the home environment.

The consequence of compression is reduction of the ratio between the peak and “average” level of the content. Increasing the “average” level will increase the apparent loudness. The human ear tends to be more sensitive to frequencies in the mid range and if these frequencies are artificially boosted, then again the apparent loudness will increase. The use of audio processing must be judicious so that the compression of the dynamic range of the soundtrack plus any other processing employed does not produce excessively loud or strident material.

Soundtrack production studios often employ gates to attenuate or eliminate the sounds below a lower threshold, and peak limiters to prevent audio exceeding the level that causes distortion, or digital clipping. These devices should not be used for the purpose of increasing the relative loudness of the material.

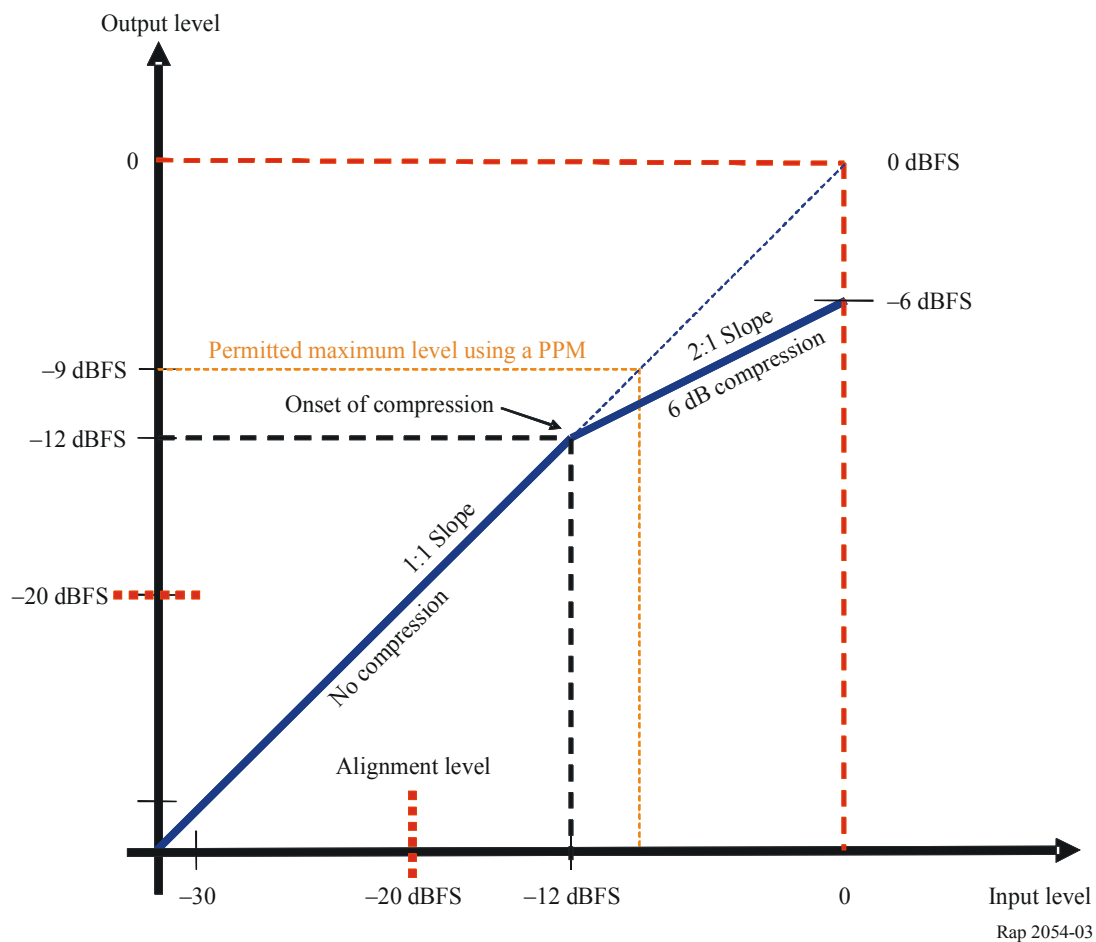
## 10 Application of volume compression in post-production following the final mix of a television commercial soundtrack

Australian free-to-air commercial television broadcasters have introduced guidelines specifying that volume compression should, where used **after the final mix**, be restricted to a slope of 2:1 with an onset point of  $-12$  dBFS.

Figure 3 provides a diagrammatic representation of this simple profile. In this profile, an onset of compression at  $-12$  dBFS allows for gentle compression of the upper 3 dB of the signal before reaching the maximum permissible peak level. If any further peak limiting were to be necessary, it would be provided automatically by the broadcaster's transmission processor.

The elements of a soundtrack, namely dialogue, music and effects are subject to various processes during production. Where these elements sit in the final soundtrack, with respect to audio levels and loudness, is the result of a final mix and effectively it is here that the loudness of the soundtrack will be principally influenced.

FIGURE 3



Material that has been compressed may sound louder, even though there is no increase in peak level. This is because compression of a soundtrack may raise the energy content of the sound by reducing the dynamic range (i.e. the difference between the loudest and softest levels of the sound) thereby making it more dense.

Many modern processors are not calibrated in dB, have constantly varying compression ratios and are likely to be multiband devices which apply different amounts of compression in different frequency bands. This makes it difficult for soundtrack producers to measure and quantify how much compression is applied to a soundtrack.

## 11 Ingest of soundtracks into the television broadcasting chain

As noted previously, audio material delivered for transmission should be preceded by a sinusoidal audio alignment signal of 1 kHz at a level of  $-20$  dBFS. The receiving station will align its systems to that signal so that it is equivalent to 0 VU, i.e. the level 20 dB below the point of digital clipping in the broadcasting station's audio system. Where PPMs are in use, this level is usually equivalent to "4" or the "TEST" level. A VU meter aligned to this reference level should read programme material in accordance with § 9 depending on the type of sounds and dynamic range of the material.

With file-based ingest of programme and commercial material, operational practices will need to be developed that reflect the nature of file-based operation.

As far as it is practicable, all stages of the broadcasting system should have unity gain and operate at the recommended levels for optimum headroom. It is intended that material provided to broadcasters should not require any level adjustment other than aligning the reference signal on the material to the broadcaster's zero reference.

In the analogue transmission chain, broadcasters must limit the extent of audio peaks to ensure that the FM sound carriers are not deviated beyond the allowable limit of  $\pm 50$  kHz.

Broadcasters may also compress the dynamic range of the audio signal at the broadcast station output in both the analogue and the digital transmission chain, to ensure that the audio levels are consistent and that listeners can enjoy the softest and loudest passages of sound without having to adjust their volume controls beyond a comfortable setting. Other delivery platforms such as mobile devices and the Internet will require even more audio processing for satisfactory reproduction.

Where broadcasters use an audio-processing system, it is strongly recommended that it provide the following functions:

- automatic gain control (AGC);
- multiband compressor;
- capacity to adjust the attack and release time of the compressor;
- limiter (matched to the transmitter pre-emphasis in analogue transmission);
- adjustments to limit the range of AGC and compressor action to limit the gain applied to low-level passages; and
- ability to modify the action of the AGC and compressor to match future loudness measurement and control systems.

Figure 4 provides an Australian depiction of a typical television broadcasting station audio transmission chain.

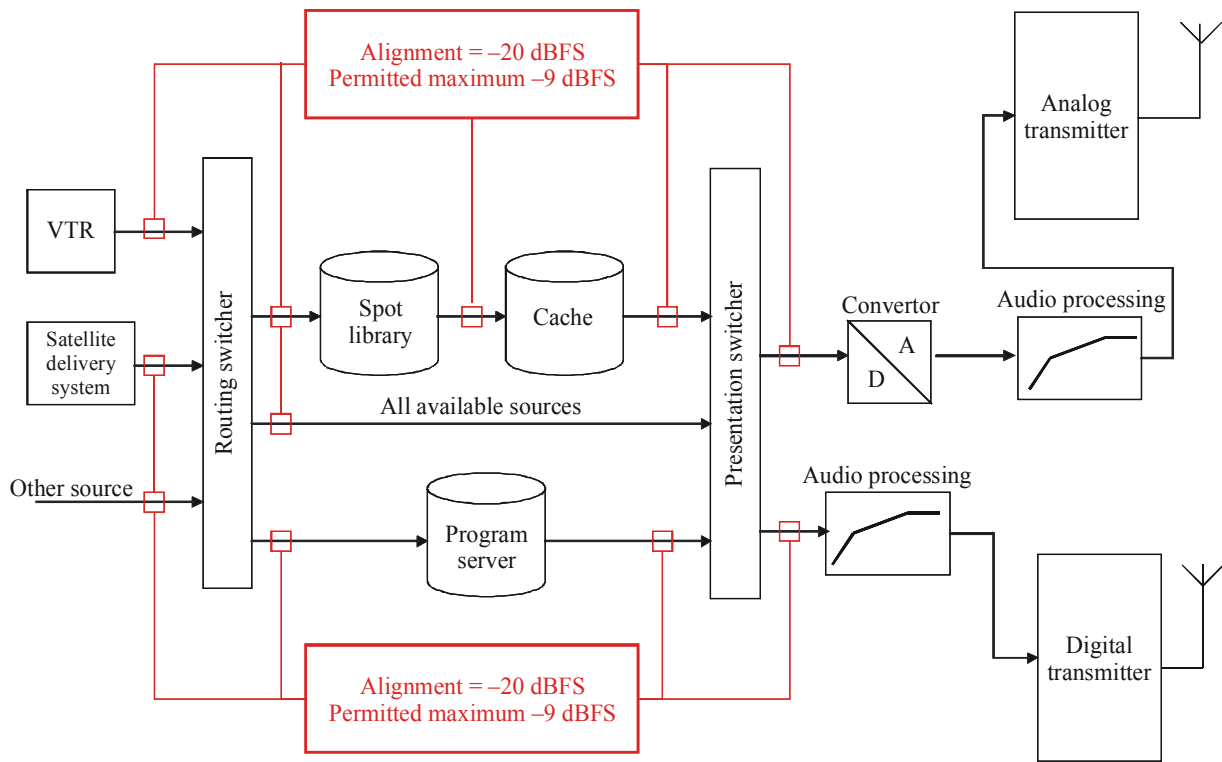
As illustrated in Fig. 4, where a broadcasting station's output is provided simultaneously in analogue and digital form, separate audio processing systems are employed to achieve an appropriate modulation characteristic at each transmitter. It is also recognized that the optimum dynamic range of the studio's audio output will depend upon the intended listening environment. Where the original sound may have been processed for the best mix in the studio environment, it is common practice to provide the appropriate contouring of the sound signal to suit each emission medium. The processing of the sound carried on the digital transmitter and destined for a high quality home entertainment system will be necessarily different to the analogue transmission that



will be received by (say) a small television set with integrated speakers. If the television station were also streaming audio out to an Internet service, for example, then the required processing of the streamed audio would probably be separately processed to match the reproduction characteristic of computers or, perhaps a mobile telephone service.

An alternative architecture to Fig. 4 may be used with centralized distribution to a very large number of distant transmitters. This involves distribution of only a digital signal which has been pre-processed. At each transmitter site, if it is necessary to also broadcast a parallel analogue television service on a separate channel, the audio modulation can be derived from the digital signal. Depending on the processing characteristic of the digital signal, it may be necessary to add some local compression and limiting before the analogue FM transmitter to optimize its modulation characteristic and prevent over deviation.

FIGURE 4



## 12 Tests performed at CBS

In an effort to better understand the problem of loudness in current television broadcasting and the best approach to avoid unwanted loudness jumps at programme transitions, CBS undertook a study of the loudness of programming on the CBS and CW Networks using available measurement techniques. The data set includes over 10 000 loudness samples and over 100 commercial breaks that were used to determine the range of loudness for a variety of programmes and commercials.

The effectiveness of the AC-3 parameter dialnorm<sup>1</sup> to harmonize loudness at the boundary between programmes and commercials was also investigated using the collected data. A “comfort zone”<sup>2</sup> excursion analysis was used to determine if the long-term loudness average, that dialnorm is based upon, accurately reflects the short term loudness differences experienced at programme-to-commercial boundaries.

CBS concluded that the test results indicate that while dialnorm may be helpful when the problem is to reasonably equalize loudness of programmes that have been produced at widely different average loudness levels, it is not very helpful when the problem is to equalize loudness at the transition between programmes or commercials. The reason is that dialnorm is based on a measurement of loudness averaged over the whole programme duration, while the perception of a loudness jump at a programme transition depends on the difference of loudness between the few seconds that precede the transition and the few seconds that follows it.

In addition, dialnorm, as currently implemented, does not take into account the contribution of the LFE channel to total programme loudness, and this can introduce important discrepancies between measured and perceived loudness on some types of programmes such as rock music.

The tests have led CBS to conclude that the operation of a loudness controller to be used on-line in order to avoid excessive jumps in loudness at programme transitions should be based on loudness measurements taken over a travelling window of a few seconds duration, with suitable attack and release times and gain slopes.

CBS further concluded that loudness measurements systematically taken over the entire duration of each programme, if available and reliable, may be useful to identify and correct major differences in loudness between different programmes but will not generally help to remove loudness jumps at transitions between programmes and commercials.

### 13 Summary

Some free-to-air commercial television broadcasters have adopted the following principles for programme and advertising material provided to a television broadcaster:

- a) Programme and advertising material shall be preceded by an audio alignment signal as specified below. The audio content as measured on a VU-type level meter shall in general be consistent with the alignment signal level. Ideally, it is intended that the television station equipment settings should remain fixed, so that there is a unity relationship between the alignment signal on the material, the ingest process and the transmission process.
- b) In digital systems the alignment level will be 20 dB below full-scale digital, i.e. –20 dBFS, in accordance with RP 155. The audio quasi peak level should nominally not exceed 11 dB above the alignment level.

---

<sup>1</sup> “Dialnorm” stands for “dialogue normalization”. The concept of dialogue normalization is that the measured loudness of a dialogue channel(s) (e.g. L/R, C, or all) is used to ensure equal psycho-acoustic perception of loudness, program to program. As originally intended, the measured value of dialnorm for each piece of content is to be transmitted as a metadata parameter and used to apply attenuation in a digital television receiver.

<sup>2</sup> The “comfort zone” of the viewer is the loudness interval above and below the nominal listening level that does not cause viewer discomfort. The “comfort zone” has been determined to be +2.5 dB above to –5.5 dB below nominal listening level. The “comfort zone” was established in:

RIEDMILLER, J.C., LYMAN, S. and ROBINSON, C. [2003] Intelligent Program Loudness Measurement and Control: What Satisfies Listeners? Audio Engineering Society Convention Paper, 2003 October 10–13, New York, New York, United States of America.

- c) In analogue systems the alignment level is equivalent to the digital alignment level of  $-20$  dBFS. In an analogue transmission system, audio peaks should nominally not exceed a level of 8 dB above the alignment level.

The television station alignment level of  $-20$  dBFS<sup>3</sup> (at a frequency of 400 Hz for transmitters with pre-emphasis) will be the level that causes reference modulation in the station's analogue and digital transmitters under test conditions (transmitter processing bypassed). It will also ensure optimum operation and headroom in the station's analogue and digital recording equipment.

Audio-processing techniques employed during the production of audio material must not produce passages of audio which are strident or excessively loud. Broadcasters should process the transmission audio in the analogue and digital streams to maintain the dynamic range within a range of sound control settings that are comfortable for the listening audience.

---

---

<sup>3</sup> Normal alignment level in a studio domain is normally set using a 1 kHz tone. Analogue TV transmitters employ pre-emphasis which results in a 0.5 dB gain. Transmitter alignment level must therefore be set using a  $-20$  dBFS tone at a frequency of 400 Hz.