International Telecommunication Union

ITU-R

Radiocommunication Sector of ITU

Report ITU-R BS.2434-0

(10/2018)

Loudness in Internet delivery of broadcast-originated soundtracks

BS Series

Broadcasting service (sound)

International Telecommunication Union

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from http://www.itu.int/ITU-R/go/patents/en where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Reports

(Also available online at http://www.itu.int/publ/R-REP/en)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	**Broadcasting service (sound)**
BT	Broadcasting service (television)
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management

Note: *This ITU-R Report was approved in English by the Study Group under the procedure detailed in Resolution ITU-R 1.*

Electronic Publication
Geneva, 2018

REPORT ITU-R BS.2434-0

Loudness in Internet delivery of broadcast-originated soundtracks

(2018)

1 Introduction

The ITU-R has three Recommendations and three Reports relating to measurement of the audio levels and loudness of broadcast programme soundtracks:

– Recommendation ITU-R BS.1770 – Algorithms to measure audio programme loudness and true-peak audio level

– Recommendation ITU-R BS.1771 – Requirements for loudness and true-peak indicating metres

– Recommendation ITU-R BS.1864 – Operational practices for loudness in the international exchange of digital television programmes

– Report ITU-R BS.2054 – Audio levels and loudness

– Report ITU-R BS.2103 – Short-term loudness metering

– Report ITU-R BS.2217 – Compliance material for Recommendation ITU-R BS.1770

Recommendation ITU-R BS.1864 specifies the target loudness level of digital television programmes as −24 LKFS (full programme-mix or normal dialogue) in international exchange. ITU-R has not set any target loudness levels for distribution and emission of broadcast programmes. The target loudness level of Internet delivery services is determined by each Internet service provider. As the Internet is becoming increasingly widespread, the opportunities for broadcasters to provide their content to Internet service providers and for viewers/listeners to receive broadcast-originated soundtracks via the Internet are increasing. This Report addresses issues involved with the Internet delivery of broadcast-originated soundtracks and describes examples of operational practices in different countries.

2 Current state of loudness in Internet delivered services

It is acknowledged that target loudness levels as high as −11 LKFS and as low as −27 LKFS are used in Internet delivery of broadcast originated soundtracks. Content processed to a higher target loudness level will, by default, have a lower peak to loudness ratio (PLR) which may alter the audio quality.

There is no common target loudness level in use amongst Internet service providers. A few of the larger content providers have their own in-house targets but these are not always enforced. Two target loudness levels are widely used among non-broadcast content providers: −16.5 and −13 LKFS.

The Audio Engineering Society (AES) technical document AES TD1004.1.15-10 [1] recommends that the target loudness level for audio streaming and network file playback should be set within the range from −16 LKFS to −20 LKFS.

The AES Standards Committee has published AES71-2018 [2] that recommends that the target loudness level for over-the-top television and online video distribution services should follow the appropriate broadcast regional content delivery and exchange Recommendations (−24 LKFS internationally and −23 LUFS in Europe) when there is no prior arrangement between parties about content delivery or distribution. The guidelines separate the use-cases between systems with and without metadata (or uncertain capability), devices with limited dynamic range and whether there has been a prior agreement between the parties regarding programme distribution. The use of loudness and dynamic range control metadata that matches the programme is encouraged when systems support it. If there is no prior agreement between the parties regarding programme distribution, it is recommended that the various regional recommendations are followed. These are summarized in the following Tables.

TABLE 1

Recommendations for short form content for any conditions

Broadcast region	Standard	Integrated loudness	Maximum short term loudness	Maximum true peak	Anchor element measurement	Full program mix measurement
North America	A/85	−24 ± 2 LKFS	N/A	−2 dB TP	Not permitted	Recommended
Europe	R128	−23 ± 0.5 LUFS	−18 LUFS and + 5 LU relative to IL	−1 dB TP	Not permitted	Recommended
Japan	TR-B32	−24 ± 1 LKFS	N/A	−1 dB TP	Not permitted	Recommended
Australia	OP-59	−24 ± 1 LKFS	N/A	−2 dB TP	Not permitted	Recommended

TABLE 2

Recommendations for long form content for any conditions

Broadcast region	Standard	Integrated loudness	Maximum short term loudness	Maximum true peak	Anchor element measurement	Full program mix measurement
North America	A/85	−24 ± 2 LKFS	N/A	−2 dB TP	Recommended	Conditionally permitted (1)
Europe	R128	−23 ± 0.5 LUFS	N/A	−1 dB TP	Conditionally permitted (2)	Recommended
Japan	TR-B32	−24 ± 1 LKFS	N/A	−1 dB TP	Not permitted	Recommended
Australia	OP-59	−24 ± 1 LKFS	N/A	−2 dB TP	Recommended	Permitted

(1) If the Anchor Element cannot be isolated and measured (per ATSC A/85).

(2) Permitted for wide dynamic range content (per EBU R128).

TABLE 3

Recommendations for delivery or distribution of content where prior arrangements exist

	Integrated loudness	Maximum true peak
Short form content	Up to −16 ± 1 LKFS/LUFS (1) or Table 1	−1 dB TP
Long form content	Up to −16 ± 1 LKFS/LUFS (1) or Table 2	−1 dB TP

(1) Not applicable for Japan.

A report on music streaming [3] recommends target loudness levels for the loudest track on an album of −14 LKFS[1] for portable devices and −18 to −20 LKFS for stationary devices, with overall loudness for an album up to 6 LU lower.

Video games are one of the major content types accessed over the Internet. A working document [5] reports that the typical loudness level of video game software is −24±2 LKFS for game consoles and −18±2 LKFS for portable game devices.

The target loudness level of non-broadcast content is not fixed to a single value. Moreover, a substantial amount of audio content from Internet delivery services has a lower dynamic range and higher average loudness level and consequently a lower peak to loudness ratio (PLR) than those of broadcast content. In particular, a higher target loudness level may be required for portable devices used in noisy environments than for devices used in the home. When broadcasters provide their contents to Internet service providers, the broadcasters should consider the difference in target loudness levels between broadcasting and Internet delivery services.

3 Considerations for Internet delivery of broadcast-originated soundtracks

3.1 Variety of reproduction devices and viewing/listening conditions

Consumers of Internet-delivered broadcast-originated soundtracks use an increasing number of receiving platforms, including traditional TV receivers, hybrid TV receivers, home theatre systems, personal computers, car-borne receivers, tablet computers and smartphones. Screen size, aspect ratio and picture resolution of these devices cover a wide range from screens of less than 100 mm to over 1 m with resolution from QVGA (320 × 240) to UHD (4K, 8K).

Similarly the number of audio channels, sampling frequency, bandwidth and bit depth also cover a wide range of channel configurations from mono to 9+10+3 (system H in Recommendation ITU-R BS.2051).

Some portable viewing devices are only capable of a modest maximum sound level, as they have very small loudspeakers, often less than 25 mm diameter. In addition, hearing protection legislation limits the maximum sound levels from devices in some regions. Compliance is currently often effected by the device design having a limited amount of amplification. Viewing/listening environments can cover a wide range, from quiet, private spaces to noisy, public spaces. In most circumstances it would be possible to achieve comfortable viewing/listening by adjusting the volume control of the playback device provided the device had sufficient gain and system headroom to allow this without distortion.

The optimal loudness level and dynamic range will differ with the type of content, for instance, pop music and feature film soundtracks [4]. As well as tailoring the picture to the capabilities of the receiving device, the sound may also need to be adjusted to suit the capabilities and constraints of the listening environment. However, it is not possible to infer the consumer’s listening conditions from the device type. For example, a Smartphone may be being used in conjunction with a Bluetooth loudspeaker or as an HDMI source to multi-channel home theatre system.

3.2 Compatibility of broadcast originated soundtracks with Internet delivery

It is possible that broadcast content and broadcast-originated Internet content will be viewed using the same device. Therefore, it is desirable that the target loudness level is the same regardless of how the content is delivered.

---

[1] LUFS (Loudness unit relative to Full Scale) is used in the references [3] and [4] instead of LKFS (Loudness, K-weighted, relative to nominal Full Scale). Both units indicate the same quantity of loudness level.

In the case of live production, a wider peak margin is required compared with that for the production of packaged media. If a higher target loudness level were used, the production process may not provide sufficient peak margin to prevent accidental clipping. For this reason, secondary formats with higher loudness levels might be generated downstream, rather than in the production phase.

Broadcast originated programmes are usually produced in one or two formats such as HD and SD services. Internet service provider originated programmes can be in many formats from a low-quality and low-bit-rate format to a high-quality and high-bit-rate format for the same presentation.

Internet delivery service providers may offer different audio versions for various device classes.

Internet service contracts may however forbid alterations to the audio level. In this case it is the broadcaster's responsibility to provide audio at alternate levels with satisfactory audio quality.

In some cases, users may be allowed to freely choose their preferred audio format. More usually the provider's service will make a choice for the user, based on the device capabilities and connection bandwidth.

3.3 Codec-based solutions

Some audio codecs such as AC-3, E-AC-3, DTS-HD, AAC, HE-AAC, and xHE-AAC [15] allow support (of varying degrees depending on implementations) of multiple device profiles. Newer NGA codecs such as AC-4 [6], DTS-UHD [13] and MPEG-H 3D Audio [14] support the transmission of multiple device profiles. In the case of AC-4, each profile specifies a target loudness level and a target amount of dynamic range reduction allowing support for varying playback device characteristics. In the case of DTS-UHD, loudness and dynamic range control metadata enables consistent content presentation in diverse listening environments and playback devices. In the case of MPEG-H 3D Audio, multiple device profiles including metadata for loudness and dynamic range control are enabled based on the codec-agnostic MPEG-D DRC standard [16]. MPEG-D DRC is also used with xHE-ACC to enable similar features. In all cases these parameters can be set by the broadcaster. Each client device can be set by the manufacturer, or perhaps by the user, to choose a specific device profile that best suits the usage conditions and capabilities of the device. This allows a single audio stream to be used for several device types without the need for extra processing and multiple streams by the broadcaster. This is a very good solution to the diversity problem, but it can only be used on newer receiving devices that have this capability.

4 Examples of operational practices for loudness in Internet delivery services of broadcast-originated soundtracks

This Report proposes that a device agnostic solution is required which uses a single harmonized target loudness level of −24 LKFS (−23 LUFS for Europe) for the Internet delivery of broadcast originated soundtracks. Furthermore, any adjustments that may be required to optimize the viewing/listening experience should be made in the receiving /reproduction device at the point of consumption if possible.

4.1 Example of Australia

In the case of portable devices, it is considered that the primary criterion for the quality of the audio service is intelligibility [7]. All other factors are secondary. Intelligibility on portable devices will be best achieved in most listening environments by reducing the dynamic range and raising the target loudness level to −16 LKFS. This level harmonises with the upper limits of AES TD1004.1.15-10 [1] and TD1005.1.16-09 [2].

Most of these portable viewing platforms also receive content from non-broadcast sources. Much of the content from these alternate service providers has a low dynamic range and/or is user-generated, and the average loudness level is often much higher than broadcast loudness level.

In a viewing environment with reduced dynamic range, a higher average loudness level has the advantage that it can greatly improve speech intelligibility [8]. This is desirable for broadcast content as well as for non-broadcast content.

The disparity between broadcast loudness levels and non-broadcast loudness levels can also be an annoying for the portable device viewer and makes programme switching a frustrating experience.

To make the portable device viewing and listening experience more consistent in audible quality, it is therefore desirable for the loudness level of broadcast content delivered to these portable devices to be adjusted to be closer to the levels used by non-broadcast content providers.

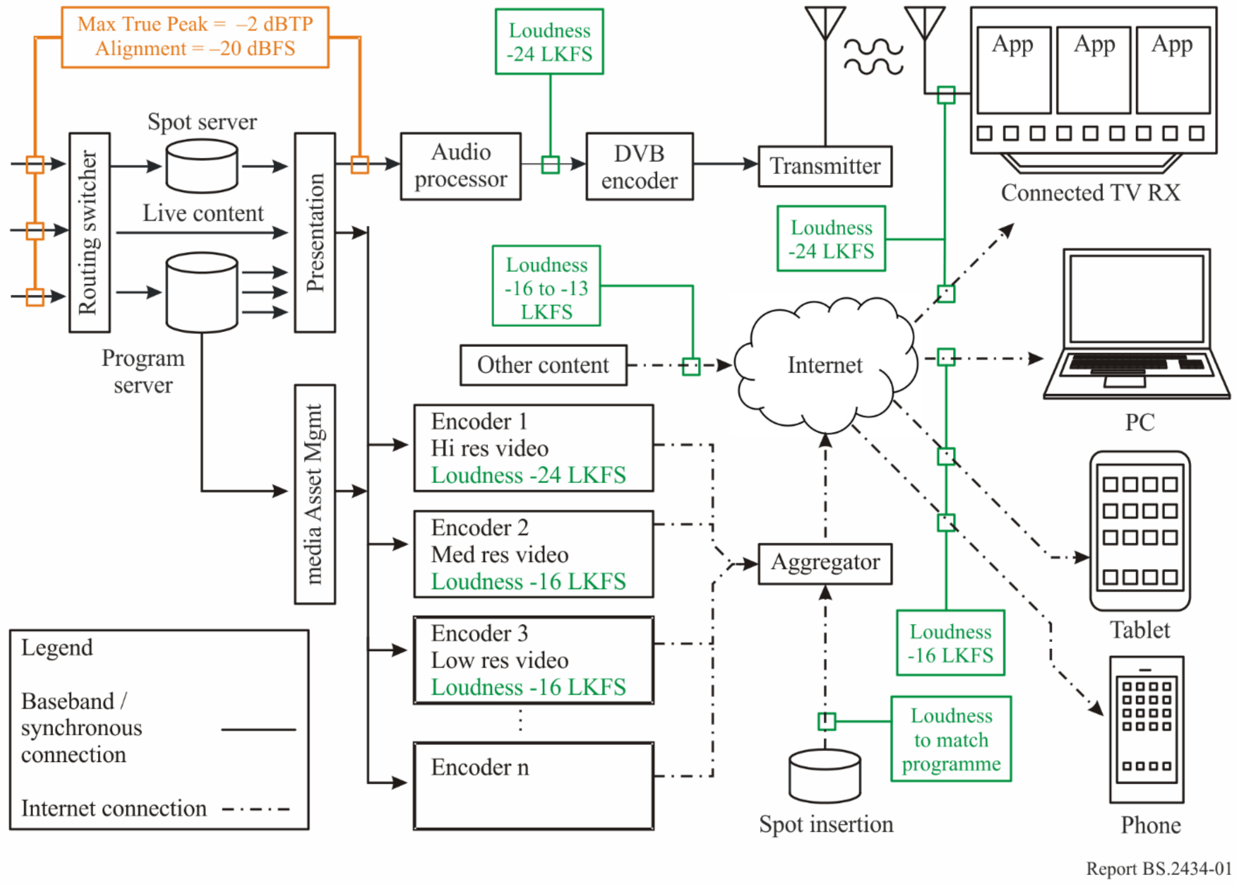
It is noted that two reference loudness levels are widely used among non-broadcast content providers: −16.5 LKFS and −13 LKFS. Of these, −16.5 LKFS is closer to broadcast loudness level, and is therefore less problematic for loudness level adjustment of broadcast content to portable devices. It is therefore suggested that broadcasters should consider raising the loudness level of their programmes to −16 LKFS for transmission to older portable devices if it is possible to do so without seriously impairing the audio quality of the soundtrack. Newer portable devices may be able to achieve the same result with a standard broadcast level using advanced codec features (see § 3.3).

Figure 1 below shows how this could be done in a typical emission system. The servers are normally able to detect the type of device they are serving, and can choose the appropriate version of the soundtrack for that device, just as they choose the appropriate video resolution for the screen and the appropriate codec bit rate for the programme carrier.

In the case of the hybrid TV, called connected TV in Fig. 1, it is anticipated that the viewing environment will support a wider audio dynamic range, and therefore the broadcast loudness level of −24 LKFS is preserved. This also maintains consistency with the loudness level of the conventionally received programme.

FIGURE 1

Example of emission system for multiple viewing platforms, showing dual audio level implementation



4.2 Example of Japan

4.2.1 Characteristics of broadcast programme soundtracks

Many television broadcasters in Japan have established common alignment levels, −20 dBFS and −18 dBFS, in accordance with SMPTE RP155 and EBU R68-2000, respectively. Broadcast programmes have a loudness level of −24 LKFS regardless of the alignment level and are broadcast with a loudness level of −24 LKFS in accordance with ARIB TR-B32. A single target loudness level of −24 LKFS was set to harmonise with signal levels used in traditional broadcasting production, and this value has been widely used in Japanese broadcasting circuits even if programmes are provided to other service providers to avoid parallel production work flows.

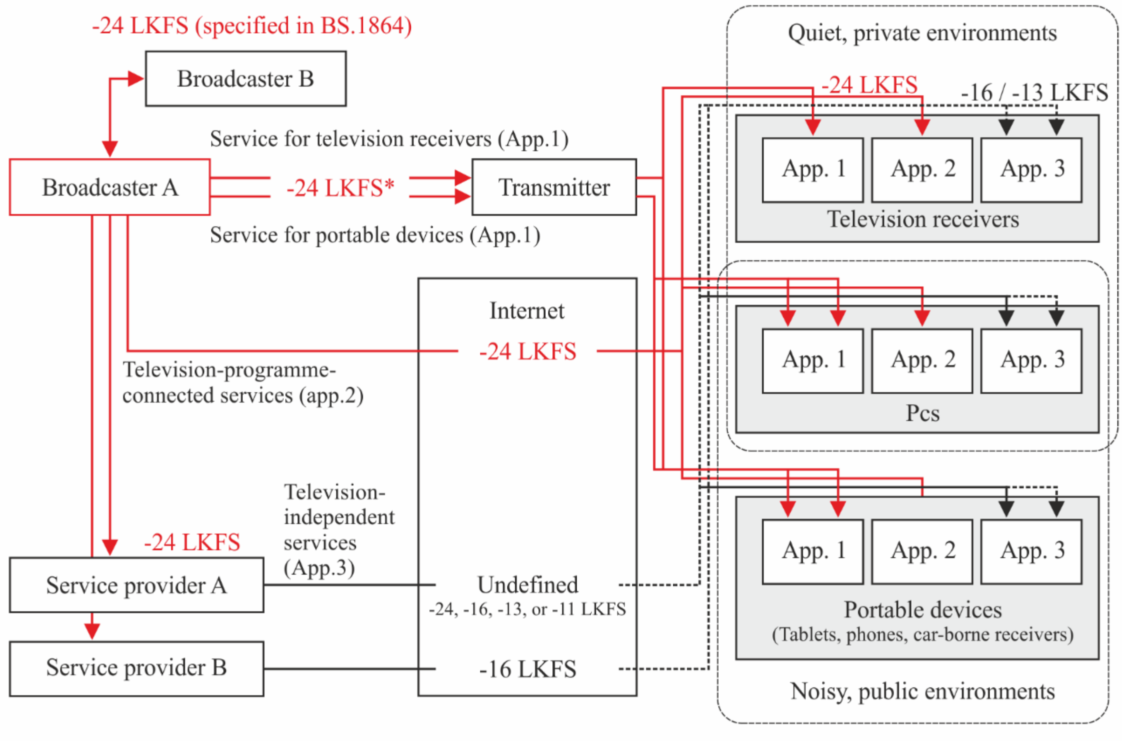
Recommendation ITU-R BS.1770 recommends that the true peak level be calculated to avoid clipping when audio signals are reproduced. Many television broadcasters require a wider peak margin than that of package media productions to produce live broadcast programmes. Although broadcasters confirm the peak margin of audio signals, a common compression method has not been specified and audio signals with a loudness level of −24 LKFS are maintained after the final mix.

4.2.2 Target loudness level for Internet

Figure 2 shows examples of services using broadcast-originated soundtracks in Japan. ARIB TR-B32 specifies the target loudness level of −24 LKFS in the production, exchange, distribution and emission of digital television programmes (Application 1) but does not specify the target loudness level for sound-only broadcasting services or Internet delivery services.

FIGURE 2

Examples of broadcast-originated services in Japan



Report BS.2434-02

* Target level of emission or distribution is specified in ARIB TR-B32, not in ITU-R BS.1864.

Some broadcasters provide low-bit-rate broadcasting services for portable devices in addition to normal high-bit-rate broadcasting services; portable devices can also receive normal high-bit-rate broadcasting services (Application 1). The target loudness of −24 LKFS is commonly used for low- and high-bit-rate services because these services are switched automatically and seamlessly by portable devices such as car-borne receivers, depending on the receiving conditions (available bit rate, etc.).

Some broadcasters provide Internet delivery services (Television-programme-connected services in Fig. 2), including IBB and VOD services (Application 2). The main target is viewers who watch broadcast-originated content using digital television receivers at home, in which case the target loudness of −24 LKFS is used to provide services with the same sound quality as normal broadcast content. A higher target loudness level such as −16 LKFS, −13 LKFS or −11 LKFS will impair the sound quality, even if the average loudness level is readjusted to −24 LKFS by the receivers, because these higher target loudness levels typically require dynamic range compression.

On the other hand, some Internet delivery services operated by non-broadcasting companies deliver broadcast-originated content where the target loudness of −24 LKFS is not used (Television-independent services in Fig. 2, Application 3). The content, however, is produced with the target loudness of −24 LKFS by the broadcaster and provided to Internet delivery service companies. The Internet delivery service providers adjust the loudness levels of broadcast-originated content to harmonise with other content in their services under their local rules.

4.3 Example of United Kingdom

4.3.1 Audio characteristics of broadcast programmes

The UK uses a common alignment level of −18 dBFS for both radio and television programmes in accordance with EBU R68-2000. Television programmes are made to a loudness target of −23 LUFS and −1 dBTP in accordance with EBU R128-2014. Radio programmes are made to a peak level target of −10 dBFS. Currently there is no defined loudness target for radio programmes, however there is a desire to adopt the EBU R128-2014 recommendations.

Production values should be treated as absolute, programmes are mixed such that dialogue audibility and intelligibility are ensured. No further dynamic range or loudness processing is required to achieve speech audibility or intelligibility in normal listening conditions.

For radio transmission, loudness and dynamic range processing is applied at levels appropriate for the transmission platform (AM, FM, DAB and DVB) and for the content genre (pop music, light music, classical music, speech and drama). AM and FM transmissions have the heaviest processing, DAB and DVB distributions have the lightest.

4.3.2 Target loudness for Internet

Currently there is no defined target loudness for radio programmes on the Internet however there is a desire to comply with AES TD1004.1.15-10 at or below −18 LUFS. Content produced using −10 dBFS peak normalisation can easily be adapted to −18 LUFS with a 5 dB gain increase. The small amount of audio above −10 dBFS comprises transient peaks that are managed through peak limiting. Using higher loudness targets result in the need to use more aggressive dynamic range compression.

Figure 3 shows examples of broadcast originated television services in the United Kingdom. Programmes are produced, transmitted and distributed at −23 LUFS. Broadcasters use the production target loudness of −23 LUFS as defined by EBU R128-2014 for the creation and publishing of online content.

Content is exchanged with third-party Internet service providers at −23 LUFS as defined by EBU R128-2014. Those third party providers may choose to change the target loudness value and apply dynamic range compression based on their own requirements and the target devices.

4.3.3 Encoding characteristics

Modern A/V encoders are capable of supporting multiple output profiles from a single input. This feature allows adaptive bitrate adaption sets to be produced that enables the playback client to select the most appropriate rendition based on device capabilities and available bandwidth. Whilst it is possible to produce multiple audio renditions it is common for only a single audio rendition to be made as maintaining the audio quality helps improve the perceived quality at lower video bitrates. Studies have shown [9] [10] [11] that audio quality plays an important role in the perceived quality of experience of audio-visual content.

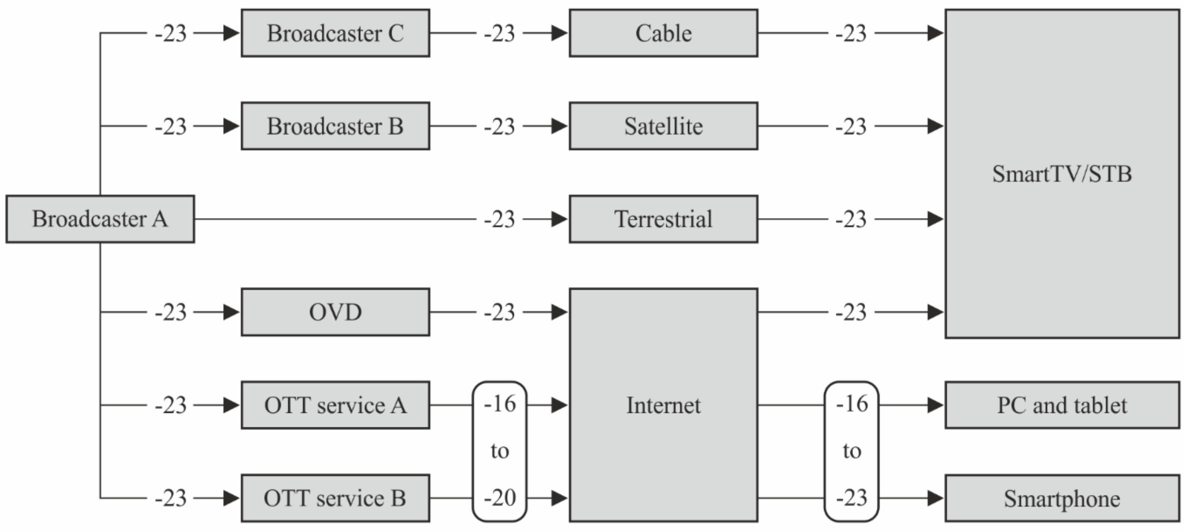
4.3.4 Playback characteristics

It is not possible to infer listening capabilities and environmental conditions from the type of device being used for playback; there are too many exceptions and variables. The only valid arbiters of the actual listening conditions are the listener and possibly their specific device. To try and make assumptions at the encoding and distribution stage is presumptuous and will tend to provide the lowest common quality to a particular device type thus depriving many consumers of a better viewing/listening experience. The onus of audibility and intelligibility lies primarily in the editorial intent and with expertise of the content production team. Where the listener requires changes to the loudness level and/or dynamic range of the content this should be made at the point of consumption.

It is noted that current EU hearing protection legislation limits the maximum listening levels on Portable Media Players (PMP). However there is revised legislation in progress to move away from the maximum SPL method to an individual dose based system [12].

FIGURE 3

Example of broadcast originated services in United Kingdom



Report BS.2434-03

5 Future work

Internet service providers can upload content with multiple target loudness levels aimed at specific devices. However, this strategy may not be completely effective as the device type gives no consideration to possible reproducer capabilities and listening environments. Different delivery services and applications currently use different local target loudness levels for their delivery. Harmonising these levels to a single common target loudness level, −24 LKFS, would require no changes to the system architecture of the playback device. This approach however requires the features of a codec that supports metadata (see § 3.3) and the receiving devices reliably use it to support this facility.

Changing the target loudness level may improve the audibility of reproduced sound where the dynamic range is also managed effectively. The optimal adjustment of the target loudness level depends on the specifications of the playback device, the reproduction device and the ambient noise level. Where portable devices have the capability for dynamic range control, devices may adjust the playback loudness level to a suitable value depending on the listening conditions, independent of the target loudness level.

Ideally, loudness management should be implemented at the point of consumption as it is not possible to determine the consumers listening conditions earlier in the distribution chain. This however requires either the receiving device or the consumer to understand the factors at play and make intelligent choices about listening level and dynamic range. It also requires the receiver to have the required control of loudness and dynamic range. Even if the consumer understands these factors, the receiver may not have the required degree of control for the consumer to achieve the desired result. Broadcasters may therefore need to assist in this task by providing alternative versions of the soundtrack where necessary.

It is desirable to develop high-quality methods to adjust the loudness level within devices.

User control of the dynamic range and target loudness level is currently available in some broadcast codecs such as those defined by DTS-UHD, AC-4, and MPEG-H 3D Audio.

ITU-R has also published Recommendation ITU-R BS.2076 – The Audio Definition Model, which specifies loudness metadata including the integrated loudness level (full-programme mix and normal dialogue). The advanced sound systems specified in Recommendation ITU-R BS.2051 – Advanced sound system for programme production, require a rendering processer in each receiver and reproduction device. Portable devices employing the advanced sound systems will be able to control their reproduction loudness level and dynamic range even for channel-based sound systems including stereo and 5.1 formats.

6 References

[1] AES technical document, AES TD1004.1.15-10, "Recommendation for loudness of Audio Streaming and Network File Playback", October 2015.

[2] AES standard, AES71-2018, "Recommended Practice Loudness Guidelines for Over the Top Television and Online Video Distribution", July 2018.

[3] Eelco Grimm, HKU University of the Arts Utrecht, "Recommendation for loudness normalization by Music Steaming Services", version 1.2, June 2017.

[4] White paper of the Music Loudness Alliance, "loudness Normalization: The Future of File-Based Playback", version 1.0, July 2012.

[5] Sony Worldwide Studios Audio Standards Working Group (ASWG), Recommendation ASWG-R001 v1.10 "Average loudness and peak levels of audio content on Sony computer entertainment platforms", August 2013.

[6] ETSI TS 103 190 Digital Audio Compression (AC-4) Standard, version 1.1.1, April 2014.

[7] Shannon, Claude E. "A Mathematical Theory of Communication". Bell System Technical Journal. 27 (3): 379–423, July 1948.

[8] Davis, D. & Davis, C., Sound System Engineering, Howard Sams / Prentice Hall 1994.

[9] Neuman, W. Russell; Crigler, Ann N.; Bove, V. Michael, AES conference paper 9-019 "Television Sound and Viewer Perceptions" 1991.

[10] Hollier M P, Voelker R; AES convention paper 4590 "Objective performance assessment: Video quality as an influence on audio perception" 1997.

[11] Belmudez B, et al; IEEE International Workshop paper "Audio and Video channel impact on perceived audio-visual quality in different interactive contexts" 2009.

[12] Bertin C, WHO and ITU Make listening safe initiative. "Standards for safe listening devices, situation analysis" 2015.

[13] ETSI TS 103 491 "DTS-UHD Audio Format; Delivery of Channels, Objects and Ambisonic Sound Fields", 2017.

[14] ISO/IEC 23008-3:2015/Amd 3:2017, "Information technology - High efficiency coding and media delivery in heterogeneous environments – Part 3: 3D audio", MPEG-H 3D Audio LC Profile, 2017.

[15] ISO/IEC 23003-3:2012, "Information technology – MPEG audio technologies - Part 3: Unified speech and audio coding", 2012.

[16] ISO/IEC 23003-4:2015, "Information technology – MPEG audio technologies - Part 3: Dynamic Range Control", 2015.