

RECOMMENDATION ITU-R BS.1688

Baseband sound system and audio source-coding at delivery interfaces of large-screen digital imagery applications

(Question ITU-R 15/6)

(2004)

The ITU Radiocommunication Assembly,

considering

- a) that Recommendation ITU-R BT.1662 – General reference chain and management of post-processing headroom for programme essence in large screen digital imagery applications, specifies a general reference chain for typical large screen digital imagery (LSDI) applications, and its principles equally apply to image essence and to sound essence;
- b) that an important feature of the general reference chain is the identification of the main functional blocks in a generic chain, and of the interfaces between them;
- c) that, for the delivery of LSDI applications, it is necessary to identify the reference baseband sound system and to also identify the source-coding¹ method used for the audio programme delivery:
 - at the interface between the LSDI programme distribution master² and the delivery channel;
 - at the interface between the delivery channel and the audio system in the presentation venue;
- d) that Recommendation ITU-R BT.1666 – User requirements for large screen digital imagery applications intended for presentation in a theatrical environment, identifies the subjective overall picture and sound quality to be provided at the highest performance level of the family of applications intended for LSDI presentation in a theatrical environment;
- e) that Recommendation ITU-R BS.775 – Multichannel stereophonic sound system with and without accompanying picture, specifies a 5-1 reference sound system with five channels³ plus an optional low-frequency effects channel, as the highest level in a hierarchy of multichannel sound systems that range from 1/0 (monophonic) up to 3/2;

¹ ITU defines “source coding” as the encoding of the original digital signal (respectively video, audio or data) in bit-rate reduced (BRR) representation before protection is applied against bit errors in the distribution channel.

² The term “distribution master” is used here to denote the programme master that is obtained from the original finished master of the programme after it has been tailored (in content and in quality) to the envisaged distribution medium.

³ The five main channels carry front left, front centre and front right sound, plus left surround and right surround sound.

- f) that the 5-1 reference sound system specified in Recommendation ITU-R BS.775 can provide the sound quality identified in Recommendation ITU-R BT.1666;
- g) that in some cases it is important to optimize the sound programme (e.g. 2 channel stereo, 2-channel matrix surround, or 5-1 channels) for presentation in a specific target venue (e.g. small venue or large venue);
- h) that for the delivery of LSDI content over bandwidth constrained channels (e.g. satellite), a reduction in audio bit rate will reduce transmission cost;
- j) that ITU-R has performed subjective audio tests, has evaluated subjective audio tests conducted by other organizations and has documented, in Recommendation ITU-R BS.1548 – User requirements for audio coding systems for digital broadcasting, the quality that can be provided by various source-coding algorithms at certain bit rates;
- k) that LSDI can benefit from the use of equipment and systems that have been developed to support digital television;
- l) that digital audio as practiced in broadcast and other professional areas employs a sampling rate of 48 kHz without application of frequency dependent emphasis,

recommends

- 1 that, for LSDI applications, the reference digital baseband sound system at the interface between the programme distribution master and the delivery channel and at the interface between the delivery channel and the audio system at the presentation venue, should be based on the hierarchical reference sound system specified in Recommendation ITU-R BS.775⁴ which specifies a hierarchy ranging from monophonic, through 2-channel stereophonic and up to 5-1 channels of sound;
- 2 that each channel in the reference digital sound system should use a PCM signal representation with a sampling rate of 48 kHz and a minimum of 16 bits/sample without emphasis as specified in Recommendation ITU-R BS.1548, unless there is prior agreement between the programme provider and destination venue to employ a different sampling rate or signal representation;
- 3 that, in those cases when the LSDI presentation venue receives a programme with a number of sound channels larger than the number that it is equipped to present, it may down-mix based on the specifications given in Annex 4 to Recommendation ITU-R BS.775;

⁴ While in some cases the “interface” may be physical, in many cases this interface may be a conceptual point buried within equipment with no physical interface presented.

4 that in those cases when the LSDI presentation venue receives a programme with more than two channels, and the equipment in the venue can only accept two channels at the physical interface into an audio system that contains a matrix surround sound decoder⁵, the audio may be down-mixed based on the equations in Annex 1;

5 that in those cases when the LSDI presentation venue receives a programme with a number of sound channels smaller than the number that it is equipped to present, it may perform an upwards conversion based on the specifications given in Annex 5 to Recommendation ITU-R BS.775;

6 that in the case when the LSDI presentation venue receives a programme with two channels, the audio may be upwards converted with a matrix surround sound decoder, and that programme providers should be aware that two channel programmes are likely to be reproduced in this manner;

7 that programme providers should ensure that the delivered audio programme is appropriate for the anticipated venue(s)⁶, and should strongly consider the possible need to deliver other versions of the programme, perhaps containing different numbers of channels, optimized for other anticipated venues;

8 that programme providers should be aware that the most common available physical interface in theatres is for a two channel LtRt matrix surround encoded signal, and that provision of this signal in parallel with any 5-1 delivery should be strongly considered unless it is known that all venues can satisfactorily handle a 5-1 programme that has been prepared for satisfactory results when reproduced in the anticipated venues,

further recommends

1 that for delivery to the theatrical environment, the 5-1 audio format is preferred as it can yield the best result within the hierarchy specified in Recommendation ITU-R BS.775;

2 that the audio may be delivered in the original baseband PCM representation, unless otherwise prohibited by specifications that cover specific delivery media;

⁵ A “matrix surround sound decoder” is found in most cinemas today. This decoder can up-mix, or “decode”, a two channel audio programme into loudspeaker feeds for left, centre, right, and surround channels.

⁶ If the anticipated venue is a theatrical environment, it is important that the programme provider monitor the mix in a large room; mixes created and optimized in small rooms with near field monitor speakers may not be appropriate for presentation in a large room.

3 that, in the cases when audio source coding is used in the delivery channel to the theatrical environment, the system used for audio source-coding should be AC-3, as per Annex 2 to Recommendation ITU-R BS.1196⁷, unless there is prior agreement between the programme provider and the destination venue to employ another source coding system, or another source coding system is used in parallel with the system specified above;

4 that, in those cases when audio source coding is used in the delivery channel of an LSDI application where bit-rate efficiency is of primary importance, the system used for audio source-coding should be ISO/IEC IS 13818-7 (AAC);

5 that if future contributions to ITU-R show that sound systems using additional channels or more advanced technology can provide a significant benefit over the basic 5-1 channel sound system, or of the source coding systems specified in this Recommendation, then modifications to this Recommendation should be considered.

NOTE 1 – ISO/IEC IS 13818-7 Standard is available in electronic version at the following address: <http://www.iso.org/itu>.

Annex 1

Down-mix into a 3/1 matrix surround sound encoded LtRt signal

The following equations specify how to encode a 3/1 format sound signal into a stereo compatible matrix encoded surround signal. These equations have been in use in the cinema and home video fields since 1980.

$$L_t = \text{Left} + 0.707 \text{ Centre} - 0.707 J (\text{Left surround} + \text{Right surround})$$

$$R_t = \text{Right} + 0.707 \text{ Centre} + 0.707 J (\text{Left surround} + \text{Right surround})$$

where J represents a 90° phase shift.

NOTE 1 – Scaling or other means must be used to prevent overload if this down-mix is performed in the digital domain.

⁷ According to Recommendation ITU-R BS.1548, the AC-3 system provides broadcast quality at a bit rate of 96 kbit/s per channel, and the ISO/IEC IS 13818-7 system provides broadcast quality at a bit rate of 72 kbit/s per channel.