RECOMMENDATION ITU-R BS.1548

User requirements for audio coding systems for digital broadcasting

(Question ITU-R 208/10)

(2001)

The ITU Radiocommunication Assembly,

considering

a)   that the multichannel sound system, with or without accompanying picture, is the subject of Recommendation ITU-R BS.775;

b)   that audio coding systems for digital terrestrial television broadcasting are the subject of Recommendation ITU-R BS.1196;

c)   that low bit-rate audio coding is the subject of Recommendation ITU-R BS.1115;

d)   that the coding system recommended in Recommendation ITU-R BS.1115 offers monophonic and two-channel stereophonic coding modes;

e)   that the basic audio and stereo image quality required for sound systems for television and sound broadcasting is to be the highest possible, generally indistinguishable from the source material;

f)   that interoperability and network operation involving programme connections such as contribution and distribution links should be carefully considered;

g)   that interoperability with existing consumer multichannel audio equipment, such as matrix surround decoders and discrete multichannel decoders, should be carefully considered;

h)   that when introducing a multichannel sound system in an existing broadcasting service using Recommendation ITU-R BS.1115, compatibility with existing receivers to maintain the service must be considered;

j)   that more generally, in view of the many applications of such systems, all technical, quality and operational requirements should be clearly specified;

k)   that the performance of multichannel audio coding systems according to Recommendation ITU-R BS.1196 is widely dependent on the configuration under which the system is operated (bit rate, use of pre-matrixing, use of composite coding, etc.);

l)   that several broadcast services already use or have specified the use of the systems recommended in Recommendation ITU-R BS.1196;

m)   that, consequently, the broadcasters have an urgent need of all information necessary to set up all the available coding parameters of the systems recommended in Recommendation ITU-R BS.1196;

n)   that the introduction of incompatible systems with similar performance characteristics is highly undesirable;

o)   that those broadcasters which have not yet started services should be able to choose the system which is best suited to their application and which is the most cost-effective,

recommends

1   that the audio coding systems for digital television and sound broadcasting for contribution and distribution applications shall fulfil the requirements listed in Annex 1;

2   that the audio coding systems for digital television and sound broadcasting for emission applications shall fulfil the requirements listed in Annex 2.

ANNEX 1

Requirements for contribution and distribution

The audio coding systems for digital television and sound broadcasting for contribution and distribution applications shall fulfil the requirements listed below.

1   Service requirements

1.1   Channel configurations

For audio services the following channel configurations should be supported according to the needs of applications (see Recommendation ITU-R BS.775 – Multichannel sterophonic sound system with and without accompanying picture):

1 channel	1/0	Mono
2 channels	2/0	Left, right
3 channels	3/0 2/1	Left, right, centre Left, right/surround
4 channels	3/1 2/2	Left, right, centre/surround Left, right/surround left, surround right
5 channels	3/2	Left, right, centre/surround left, surround right

together with an optional low frequency enhancement (LFE) channel.

For contribution, in addition, it could be necessary to convey programmes produced in other formats than those listed above, e.g. 3/4, thus the coding system should allow for accommodation of additional high quality channels.

1.2 Flexible allocation of channels

A bit stream shall provide identification data for signalling and controlling of sound configurations. It must be possible in the transmission system to dynamically switch among the channel configurations listed in § 1.1.

1.3 Ancillary data

The audio coding system shall provide for the possibility of transmission of ancillary data. The ancillary data can convey various types of information, including dynamic range control, loudness control, user data, and any metadata required by the emission encoder that will encode the final audio for delivery to the consumer.

2 Performance requirements

2.1 Audio quality

2.1.1 Basic audio quality

The quality of sound reproduced after reference contribution/distribution cascade (five contribution codecs and three distribution codecs in tandem) should be subjectively indistinguishable from the source for most types of audio programme material. Using the triple stimuli double blind with hidden reference test, described in Recommendation ITU-R BS.1116 – Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems, this requires mean scores generally higher than 4.5 in the impairment 5-grade scale, for listeners at the reference listening position. The worst rated item should not be graded lower than 4 (Recommendation ITU-R BS.775).

NOTE 1 – The confidence interval (error bar) associated with the single mean score for a codec and item shows the range above and below the stated mean score in which the true score may fall, with some degree of certainty, usually 95%. The true score for a codec and item may be as poor as the lower limit of the confidence interval about the stated score. In order to make a meaningful evaluation of the expected performance of cascaded codecs, the confidence interval associated with the reported mean scores for the individual codecs must be approximately equal to or less than the difference between the scores being compared.

NOTE 2 – The contribution/distribution cascade, when placed in tandem with the emission codec, should not cause a significant reduction in quality compared to the basic audio quality of the emission codec. Precise specification requires further study.

NOTE 3 – The objective audio quality parameters for contribution/distribution can be incorporated later, conforming to Recommendation ITU-R BS.1387.

2.1.2 Quantization resolution

The required resolution should be at least 18 bits for distribution and 20 bits or greater is preferable for contribution.

2.1.3 Sampling frequency

In agreement with Recommendation ITU-R BS.646 – Source encoding for digital sound signals in broadcasting studios, the sampling frequency shall be 48 kHz.

2.1.4 Bandwidth

Main audio channels: 20-20 000 Hz.

LFE channel: 15-120 Hz.

2.1.5   Emphasis

The audio coding system should be emphasis free.

2.1.6   Tandem capability

The tandem capability required depends on the application according to the following table:

Distribution	3 codecs in cascade
Contribution	5 codecs in cascade

These figures have been taken from previous experiments done to evaluate two-channel sound broadcasting systems (see Recommendation ITU-R BS.1115 – Low bit-rate audio coding) and may not be representative of the practical radio and television broadcasting operational situations. More information is required to better specify this aspect.

2.1.7   Post-processing capability

The post-processing capability required is strongly dependent on the application. For distribution crossfades can be applied together with dynamic range control. It is more difficult to determine the signal processing that can take place between contribution links.

2.2   Coding delay

Coding delay for all channels in a programme must be identical. The coding delay should be as low as possible, considering the coding performance (i.e. amount of bit rate reduction) required. In case of television sound, the delay of audio must be matched with the delay of video. It is desirable that the audio coder produces encoded audio frames (access units) that correspond exactly to the time period of the matching video frame.

2.3   Error resilience

A mechanism must be provided in the audio bit stream to allow the decoder to identify residual channel errors and to adopt proper concealment methods.

2.4   Recovery time

The recovery time should be as low as possible. In case of audio access unit (AAU) applied, the recovery time should be within a few AAU, and preferably the audio should resume upon receipt of the first error free AAU.

3   Functional and operational requirements

3.1   Bit rate and coding scheme

For distribution and contribution links, Recommendation ITU-R BS.1115 recommends International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) IS 11172-3 Layer II at a bit rate of 180 kbit/s per channel or above. For several reasons the system may be applied at higher bit rate or other systems may be employed.

These reasons may include the following:

–   additional coding margin to support signal processing that may be inserted between coding generations (this was not tested or verified in the development of Recommendation ITU-R BS.1115);

–   to obtain a lower bit rate in the distribution and contribution link;

–   to obtain a higher quality;

–   suitability of synchronization and switching with accompanying video signals.

3.2   Composite coding

A two-channel or multichannel programme material often contains some inter-channel statistical correlation. Composite coding can be an effective way to reduce the inter-channel irrelevance or redundancy, thus increasing the coding efficiency. Some coding systems use perceptual criteria to eliminate part of the inter-channel irrelevance by joining together two or more channels in frequency regions where the ability of the human ear to discriminate the direction of the source is rather poor. The disadvantage of this technique is that it is generally not possible to correctly reposition the sound information in the original channels at a later stage. For contribution and many distribution applications such composite coding schemes should not be used.

ANNEX 2

Requirements for emission

The audio coding systems for digital television and sound broadcasting for emission applications shall fulfil the requirements listed below.

1   Service requirements

1.1   Channel configurations

For audio services the following channel configurations should be supported according to the needs of applications (see Recommendation ITU-R BS.775):

1 channel	1/0	Mono
2 channels	2/0	Left, right
3 channels	3/0 2/1	Left, right, centre Left, right/surround
4 channels	3/1 2/2	Left, right, centre/surround Left, right/surround left, surround right
5 channels	3/2	Left, right, centre/surround left, surround right

together with an optional LFE channel.

1.2   Audio services

Together with a main audio service, the following associated audio services can be provided according to the needs of applications: a multilingual service, audio services for the hearing and visually impaired and some capacity for ancillary data (Recommendation ITU-R BS.775) [ATSC, 1992]. The multilingual service, consisting of one or more independent channels, can be used to distribute a programme with commentary in several languages. The service for the visually impaired usually contains a vocal description of the picture content, while the service for the hearing impaired would contain the clean dialogue without, or with a lower level of, music and special effects so as to improve the intelligibility of the speech. The ancillary data can convey various types of information, including dynamic range control, loudness control and user data.

The various services can be grouped as:

–   *Main service* (every channel of a main service is assigned to the same programme, including the optional LFE channel).

–   *Extended service(s)*, which could be:

–   *Independent services* (for additional programmes which are independent of the main service programme, such as commentary, or other services containing two or more channels; channel configurations can be chosen according to the table in § 1.1).

–   *Alternative services* (for programmes which are intended to replace one or more of the main service channels, such as multilingual, hearing impaired).

–   *Additional services* (containing channels to be added to channels of the main service, such as commentary, or additional channels for enhanced sound systems as 3D TV).

As any transmission system should include a system layer able to perform multiplexing operations, it is not required that all the audio services listed above be conveyed by a single bit stream.

1.3   Flexible allocation of channels

A bit stream shall provide identification data for signalling and controlling of the sound configurations. The transmission system must provide the ability to dynamically switch among any of the channel configurations listed in § 1.1.

1.4   Ancillary data

The audio coding system shall provide for the possibility of transmission of ancillary data. The ancillary data can convey various types of information, including dynamic range control, loudness control and user data.

2   Performance requirements

2.1   Audio quality

2.1.1   Basic audio quality

The broadcaster typically has the ability to trade off the audio quality versus the bit rate applied to audio. Ideally, the quality of the sound reproduced after decoding will be subjectively similar to the original signal for most types of audio programme material. Using the triple stimuli double blind

with hidden reference test, described in Recommendation ITU-R BS.1116, this requires mean values consistently higher than 4 on the Recommendation ITU-R BS.1116 impairment 5 grade scale at the reference listening position. In practice, commercial requirements sometimes lead to operation with bit rates lower than that necessary to achieve this level of quality. However, the system should offer the broadcaster the option to operate at this level of quality.

NOTE 1 – The objective audio quality parameters for contribution/distribution can be incorporated later, conforming to Recommendation ITU-R BS.1387.

2.1.2   Stereo image quality

In the case of stereophonic (two-channel or multichannel) configurations, the quality of the sound image of source material should be preserved. For the configurations including a centre channel (3/0, 3/1, 3/2) the directional stability of the frontal sound image shall be maintained within reasonable limits over a listening area larger than that provided by conventional two-channel stereophony. For the configurations including surround (2/1, 2/2, 3/1, 3/2) the sensation of spatial reality (ambience) shall be significantly enhanced over that provided by conventional two-channel stereophony (Recommendation ITU-R BS.775).

2.1.3   Quantization resolution

The required resolution should be at least 16 bits.

2.1.4   Sampling frequency

In agreement with Recommendation ITU-R BS.646, the sampling frequency shall be of 48 kHz.

2.1.5   Bandwidth

Main audio channels: 20-20 000 Hz.

LFE channel: 15-120 Hz.

2.1.6   Emphasis

The audio coding system should not employ emphasis.

2.1.7   Post-processing capability

The post-processing capability required is strongly dependent on the application. For emission links, it can be restricted to equalization and dynamic range adjustment (e.g. to match the dynamic range of the programme material to that of the listening environment).

2.2   Coding delay

Coding delay for all channels in a programme must be identical. In case of television sound, the delay of audio must be matched with the delay of video.

2.3   Error resilience

A mechanism must be provided in the audio bit stream to allow the decoder to identify residual channel errors and to adopt proper concealment methods.

2.4   Recovery time

The recovery time should be as low as possible. For systems that provide AAUs, the recovery time should be within a few AAU, and ideally within a single AAU.

3   Functional and operational requirements

3.1   Compatibility

3.1.1   Downward compatibility (Recommendation ITU-R BS.775)

A multichannel bit stream format must be decodable by classes of decoders of varying complexity. It must be possible in the decoder to arrange a presentation with a number of channels lower than the number of transmitted channels, according to the user reproduction capabilities, without impairment other than the loss of the stereo or multichannel localization effect.

Two methods have been identified which provide downward compatibility with low receiver complexity. The first requires the use of the matrix process. A low-cost receiver then only requires the A- and B-channels as in the case of the 2/0 system, i.e. a system which does not use a backwards compatibility matrix. The second method is applicable to the discrete 3/2 delivery system. The delivered signals are digitally combined using the equations, which enable the required number of signals to be provided. In the case of low bit rate source coded signals, the downward mixing of the 3/2 signals may be performed prior to the synthesis portion of the decoding process (where the bulk of the complexity lies).

3.1.2   Backward compatibility with Recommendation ITU-R BS.1115

This requirement applies in situations where an existing mono/stereo application based on Recommendation ITU-R BS.1115 must be upgraded to multichannel sound and services to existing receivers must be maintained. Recommendation ITU-R BS.1115 recommends (for applications that only require mono and stereo low bit rate audio coding) the ISO/IEC IS 11172-3 Layer II system. In systems that already employ this type of audio coding for mono or stereo, backward compatibility for multichannel low bit rate coding means that an ISO/IEC IS 11172-3 decoder shall properly decode basic stereo information, constituted by an appropriate down mix of the audio information from all source channels. To fulfil this requirement either the simulcast method or the matrixing method may be applied.

Simulcast method

One method is to continue providing the existing Recommendation ITU-R BS.1115 channel service and to add the new 3/2 channel service. This approach is referred to as a simulcasting operation. The advantage of this approach is that the existing Recommendation ITU-R BS.1115 service could be discontinued at some point in the future, and the 2/0 and 3/2 programme mixes may be independently optimized.

Matrixing method

Another method is the use of compatibility matrices. The matrix equations may be used to provide compatibility with existing receivers. In this case, the existing left and right emission channels are used to convey the compatible A and B matrix signals. Additional emission channels are used to

convey the T, Q_1, and Q_2 matrix signals. The advantage of this approach may be that less additional data capacity is required to add the new service.

3.1.3 Forward compatibility with Recommendation ITU-R BS.1115

For applications where the new multichannel system must coexist with the mono/stereo system defined in Recommendation ITU-R BS.1115, it may be required that decoders are able to decode an ISO/IEC IS 11172-3 audio bit stream.

3.2 Bit rate

The ISO/IEC IS 11172-3 Layer II system [ISO/IEC, 1993] is recommended, at a bit rate of 128 kbit/s per channel (Recommendation ITU-R BS.1115). Thus 2×128 kbit/s can be considered an upper limit for the two-channel service. Thus 5×128 kbit/s can be considered an upper limit for the multichannel service in case that backward compatibility (see § 3.1.2) is not required. As the composite coding techniques should provide for additional coding gain, a bit rate not above 512 kbit/s should be achievable by the new multichannel coding system for the five-channel main service.

The coding system should provide the audio quality defined in § 2.1 at a bit rate comparable to or lower than the bit rate that would be required by the system recommended in Recommendation ITU-R BS.1115.

3.3 Decoder complexity

The decoder for the audio programme should be of not unduly high complexity so that the decoder cost may be kept low. In the case where a fewer number of channels, *M*, are to be reproduced from an audio programme containing *N* channels, the decoder complexity should be smaller than the complexity of the complete *N* channel decoder.

Bibliography

ATSC [1992] Digital audio and ancillary data services for an advanced television service. ATSC Document T3/186. Advanced Television Systems Committee.

ISO/IEC [1993] Coding of moving pictures and associated audio for digital storage media at up to 1.5 Mbit/s. ISO/IEC IS 11172-3. International Standards Organization/International Electrotechnical Commission.