American National Standard

INCITS/ISO/IEC 13818-3:1998 (R2017)

(ISO/IEC 13818-3:1998, IDT)

Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 3: Audio

Developed by

incits SM

Where IT all begins

Approved American National Standard

ANSI

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

Adopted by INCITS (InterNational Committee for Information Technology Standards) as an American National Standard.

Date of ANSI Approval: 12/19/2017

Published by American National Standards Institute,
25 West 43rd Street, New York, New York 10036

INTERNATIONAL STANDARD ISO/IEC 13818-3

Second edition
1998-04-15

Information technology — Generic coding of moving pictures and associated audio information —

Part 3:
Audio

Technologies de l'information — Codage générique des images animées et des informations sonores associées —

Partie 3: Son

Reference number
ISO/IEC 13818-3:1998(E)

Page

ISO/IEC Copyright Office • Case Postale 56 • CH1211 Genève 20 • Switzerland

Printed in Switzerland.

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

International Standard ISO/IEC 13818-3 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information.*

This second edition cancels and replaces the first edition (ISO/IEC 13818-3:1995), which has been technically revised.

ISO/IEC 13818 consists of the following parts, under the general title *Information technology — Generic coding of moving pictures and associated audio information*:

– *Part 1: Systems*

– *Part 2: Video*

– *Part 3: Audio*

– *Part 4: Compliance testing*

– *Part 5: Software simulation*

– *Part 6: Extensions for DSM-CC*

– *Part 7: Advanced Audio Coding (AAC)*

– *Part 9: Extension for real time interface for systems decoders*

– *Part 10: Conformance extensions for DSM-CC*

Annexes A and B form an integral part of this part of ISO/IEC 13818. Annexes C to I are for information only.

ISO/IEC 13818 was prepared by SC29/WG11, also known as MPEG (Moving Pictures Expert Group). MPEG was formed in 1988 to establish a standard for the coded representation of moving pictures and associated audio stored on digital storage media.

ISO/IEC 13818 is published in three parts. Part 1 - systems - specifies the system coding layer of the standard. It defines a multiplexed structure for combining audio and video data and means of representing the timing information needed to replay synchronised sequences in real-time. Part 2 - video - specifies the coded representation of video data and the decoding process required to reconstruct pictures. Part 3 - audio - specifies the coded representation of audio data and the decoding process required to decode audio signals.

The technical changes in this 2nd edition compared to the first publication of ISO/IEC 13818-3 (1995) are:

1. In the first publication, certain combinations of dynamic crosstalk and prediction were not prohibited but not practically implementable. In this 2nd revision, these combinations are explicitly prohibited.
2. In the first publication, a low-pass filter was to be applied to the monophonic surround signal in matrix mode 2 (analogue surround mode). This filter is omitted in this edition, greatly simplifying the decoder.
3. The description of the syntax of the LFE channel was ambiguous. This description has been clarified.

Next to these technical changes, many editorial changes have been made, improving readability and clarity.

0.1 Extension of ISO/IEC 11172-3 Audio Coding to Lower Sampling Frequencies

In order to achieve better audio quality at very low bit rates (<64 kbit/s per audio channel), in particular if compared with ITU-T (formerly CCITT) Recommendation G.722 performance, three additional sampling frequencies are provided for ISO/IEC 11172-3 layers I, II and III. The additional sampling frequencies (Fs) are 16 kHz, 22,05 kHz and 24 kHz. This allows corresponding audio bandwidths of approximately 7,5 kHz, 10,3 kHz and 11,25 kHz. The syntax, semantics, and coding techniques of ISO/IEC 11172-3 are maintained except for a new definition of the sampling frequency field, the bitrate index field, and the bit allocation tables. These new definitions are valid if the ID bit in the ISO/IEC 11172-3 header equals zero. To obtain the best audio performance, the parameters of the psychoacoustic model used in the encoder have to be changed accordingly.

With these sampling frequencies, the duration of the audio frame corresponds to:

Layer	Sampling Frequency in kHz		
	16	22,05	24
I	24 ms	17,41.. ms	16 ms
II	72 ms	52,24.. ms	48 ms
III	36 ms	26,12.. ms	24 ms

0.2 Low bitrate coding of multichannel audio

0.2.1 Universal multichannel audio system

A standard on low bit rate coding for mono or stereo audio signals was established by MPEG-1 Audio in ISO/IEC 11172-3. This standard is applicable for carrying of high quality digital audio signals associated with or without picture information on storage media or transmission channels with limited capacity.

The ISO/IEC 11172-3 audio coding standard can be used together with both MPEG-1 and MPEG-2 Video as long as only two-channel stereo is required. MPEG-2 Audio (ISO/IEC 13818-3) provides the extension up to 3/2 multichannel audio and an optional low frequency enhancement channel (LFE).

This part of ISO/IEC 13818 describes an audio subband coding system called ISO/MPEG-Audio Multichannel, which can be used to transfer high quality digital multichannel and/or multilingual audio information on storage media or transmission channels with limited capacity. One of the basic features is the backwards compatibility to ISO/IEC 11172-3 coded mono, stereo or dual channel audio programmes. It is designed for use in different applications as considered by the ISO/MPEG audio group and the specialist groups TG 10/1, 10/2 and 10/3 of the ITU-R (previously CCIR).

Multichannel audio systems provide enhanced stereo performance compared to conventional two channel audio systems. It is recognised that improved presentation performance is desirable not only for applications with

system applicable to satellite or terrestrial television broadcasting, digital audio broadcasting (terrestrial and satellite), as well as other non-broadcasting media, e.g.,

CATV	Cable TV Distribution
CDAD	Cable Digital Audio Distribution
DAB	Digital Audio Broadcast
DVD	Digital Versatile Disc
ENG	Electronic News Gathering (including Satellite News Gathering)
HDTV	High Definition Television
IPC	Interpersonal Communications (video conference, videophone, etc.)
ISM	Interactive Storage Media (optical disks, etc.)
NDB	Network Database Services (via ATM, etc.)
DSM	Digital Storage Media (digital VTR, etc.)
EC	Electronic Cinema
HTT	Home Television Theatre
ISDN	Integrated Services Digital Network

seems to be very attractive to the manufacturer, producer and consumer.

0.2.2 Representation of multichannel audio

0.2.2.1 The 3/2-stereo plus LFE format

Regarding stereophonic presentation, specialist groups of ITU-R, SMPTE, and EBU recommend the use of an additional centre loudspeaker channel C and two surround loudspeaker channels LS and RS, augmenting the front left and right loudspeaker channels L and R. This reference audio format is referred to as "3/2-stereo" (3 front / 2 surround loudspeaker channels) and requires the transmission of five appropriately formatted audio signals.

For audio accompanying picture applications (e.g. HDTV), the three front loudspeaker channels ensure sufficient directional stability and clarity of the picture related frontal images, according to the common practice in the cinema. The dominant benefit is the "stable centre", which is guaranteed at any location of the listener and important for most of the dialogue.

Additionally, for audio-only applications, the 3/2-stereo format has been found to be an improvement over two-channel stereophony. The addition of one pair of surround loudspeaker channels allows improved realism of auditory ambience.

A low frequency enhancement channel (in this part of ISO/IEC 13818 called LFE channel) can, optionally, be added to any of these configurations. The purpose of this channel is to enable listeners to extend the low frequency content of the reproduced programme in terms of both frequency and level. In this way it is the same as the LFE channel proposed by the film industry for their digital sound systems.

The LFE channel should not be used for the entire low frequency content of the multichannel sound presentation. The LFE channel is optional at the receiver, and thus should only carry low frequency sound effects, which may have a high level. The LFE channel is not included in any dematrixing operation in the decoder. The sampling frequency of the LFE channel corresponds to the sampling frequency of the main channels, divided by a factor of 96. This provides 12 LFE samples within one audio frame. The LFE channel is capable of handling signals in the range from 15 Hz to 120 Hz.

0.2.2.2 Compatibility

Extension from 2/0-stereo towards multichannel sound.

As a result of the widespread use of conventional two-channel stereo (2/0-stereo) reproduction, compatibility with existing 2/0-stereo sound reproduction systems or with existing matrixed surround sound receivers has to be maintained. This means that for many applications a basic stereo signal which contains an appropriate downmix of the audio information of the multichannel programme has to be transmitted together with the multichannel audio information. Appropriate downmix equations are given by equation pairs (1,2), (3,4), (5,6) and (7,8).

$$Ro = R + ½\sqrt{2} * C + ½\sqrt{2} * RS \quad (2)$$

or

$$Lo = L + ½\sqrt{2} * C + ½ * LS \quad (3)$$

$$Ro = R + ½\sqrt{2} * C + ½ * RS \quad (4)$$

or

$$Lo = L \quad (5)$$

$$Ro = R \quad (6)$$

or

$$Lo = L + ½\sqrt{2} * C - ½\sqrt{2} * jS \quad (7)$$

$$Ro = R + ½\sqrt{2} * C + ½\sqrt{2} * jS \quad (8)$$

where jS is derived from LS and RS by calculation of the mono component. Then, a dynamic range compression and 90 degrees phase shift are applied to this component. The downmix (7,8) is suitable for existing matrixed surround decoders.

The format of an ISO/IEC 13818-3 bit stream is such that an ISO/IEC 11172-3 audio decoder properly decodes the basic stereo information according to one of the sets of downmix equations above (see 0.2.3.1). Compatibility with existing surround sound decoders by use of equations (7) and (8) has not been verified at the time of printing of this part of ISO/IEC 13818.

In the case of this part of ISO/IEC 13818, three different possibilities can be identified to provide to the user a basic stereo downmix together with the multichannel audio information:

1. Transmitting the 2/0-stereo sound inherently with the multichannel information in one bit stream in a backwards compatible way with ISO/IEC 11172-3, thus avoiding simulcast. This allows for the most efficient use of bit rate required for both, the 2/0-stereo and the multichannel audio signal. Additional advantages are that both programmes are strictly synchronized on a PCM audio sample basis, and that audio programme associated data carried in the ancillary data field of the MPEG-Audio bit stream have to be transmitted only once. The stereo downmix from the multichannel audio signal is handled by the ISO/IEC 13818-3 encoder. For this downmix, a number of matrix options according to equations (1) and (2) and equations (3) and (4) are provided by this part of ISO/IEC 13818 (see 2.5.2.13).
2. Simulcast of the multichannel audio signal, coded according to this part of ISO/IEC 13818, together with the 2/0-stereo signal coded according to ISO/IEC 11172-3. This solution requires two independent bit streams which can be multiplexed and transmitted by ISO/IEC 13818-1. The programme provider has to make provisions if a synchronization of both bit streams is required. Further, the simulcast option requires a significantly higher bit rate because instead of 5 channels in the case of 3/2 multichannel sound, altogether 7 audio channels have to be transmitted. However, the simulcast option allows for an individual, i.e. dynamic downmix to 2/0-stereo sound which can be controlled by a sound engineer.
3. Transmitting only the multichannel signal, by using the non-matrixed mode (downmix equation (5,6)). Each stereo decoder has then to be able to decode all the five channels, and to make a stereo downmix. Although the downmix can be applied before the filtering operation in the decoder, and the filter only needs to be done on two channels, this complicates the decoder significantly.

If compatibility with existing matrixed surround sound decoders is required, this part of ISO/IEC 13818 again provides three solutions:

1. To ensure a high efficiency regarding the bit rate required for both, the 3/2-multichannel and the matrixed surround signal, this surround signal can be transmitted in the backwards compatible stereo channel. The matrix-option '10' according to equations (7) and (8) provides an appropriate compatible signal which is transmitted in the basic stereo channels. A matrixed surround signal, suitable for existing matrixed surround decoders, can be obtained at the receiver by using an ISO/IEC 11172-3 two-channel decoder. The corresponding 3/2-channel output can be derived by using an ISO/IEC 13818-3 decoder.
2. A higher bit rate is necessary for simulcast of a matrixed surround signal using ISO/IEC 11172-3 and a 3/2-multichannel audio signal using this part of ISO/IEC 13818. This simulcast option allows for an independent mix of the matrixed surround signal which can be controlled by a sound engineer. The

only five channels if matrix-option '10' (see 2.5.2.13) is used.

3. Transmitting only the multichannel signal, by using the non-matrixed mode. Each stereo decoder has then to be able to decode all the five channels, and to make the downmix according to equation (7,8). Although the downmix can be applied before the filtering operation in the decoder, and the filter only needs to be done on two channels, this complicates the decoder significantly.

Downwards compatibility.

A hierarchy of audio formats providing a lower number of loudspeaker channels and reduced presentation performance (down to 2/0-stereo or even mono) and a corresponding set of downwards mixing equations are recommended in ITU-R Recommendation 775: "Multichannel stereophonic audio system with and without accompanying picture", November 1992. Alternative lower level audio formats which may be used in circumstances where economic or channel capacity constraints apply, are 3/1, 3/0, 2/2, 2/1, 2/0, and 1/0. Corresponding loudspeaker arrangements are 3/2, 3/1, 3/0, 2/2, 2/1, 2/0, and 1/0.

Backwards compatibility.

For several applications, the intention is to extend the existing 2/0-stereo sound system by transmitting additional audio channels (centre, surround) without making use of simulcast operation. This provision of backwards compatibility with existing receivers implies the use of compatibility matrices: the decoder of the previous generation must reproduce the two conventional basic stereo signals L´o/R´o, and the multichannel decoder produces the complete 3/2-stereo presentation L´/C´/R´/LS´/RS´ from the basic stereo signal and the extension signals.

It is recognised that backward compatibility may not be required for all applications of MPEG-2 Audio. Therefore, nonbackward compatible (NBC) audio coding systems free of the constraints of backwards compatibility are being evaluated for optional use with this part of ISO/IEC 13818.

0.2.2.3 Multilingual capability

Particularly for HDTV applications, multichannel stereo performance and bilingual programmes or multilingual commentaries are required. This part of ISO/IEC 13818 provides for alternative audio channel configurations in the five-channel sound system, for example a bilingual 2/0 stereo programme or one 2/0, 3/0 stereo sound plus accompanying services (e.g. "clean dialogue" for the hard-of-hearing, commentary for the visually impaired, multilingual commentary etc.). An important configuration is the reproduction of commentary dialogue (e.g. via centre loudspeaker) together with the common music/effect stereo downmix (examples are documentation film, sport reports).

0.2.3 Basic Parameters of the Multichannel Audio Coding System

The transmission of the five audio signals of a 3/2 sound system requires five transmission channels (although, in the context of bitrate reduced signals, these are not necessarily independent). In order that two of the transmitted signals can provide a stereo service on their own, the source sound signals are generally combined in a linear matrix prior to encoding. These combined signals (and their transmission channels) are identified by the notation T0, T1, T2, T3 and T4.

0.2.3.1 Compatibility with ISO/IEC 11172-3

The ISO/MPEG-Audio Multichannel system provides full compatibility with ISO/IEC 11172-3. For a multichannel audio bit stream, backwards compatibility means, that an ISO/IEC 11172-3 audio decoder properly decodes the basic stereo information (see 0.2.2.2). Forwards compatibility means that an MPEG-2 multichannel audio decoder is able to decode properly an ISO/IEC 11172-3 audio bit stream.

The backwards compatibility is realised by coding the basic stereo information in conformance with ISO/IEC 11172-3 and exploiting the ancillary data field of the ISO/IEC 11172-3 audio frame (base frame, in the context of this part of ISO/IEC 13818) plus an optional extension frame for the multichannel extension.

The complete ISO/IEC 11172-3 audio frame incorporates four different types of information:

- Header information within the first 32 bits of the ISO/IEC 11172-3 audio frame.
- Cyclic Redundancy Check (CRC), consisting of 16 bits, just after the header information (optional).

scalefactors (SCF), and the subband samples.

- Ancillary data. Due to the large number of different applications which will use this part of ISO/IEC 13818, the length and usage of this field are not specified.

The variable length of the ancillary data field enables packing the complete extension information of the channels T2/T3/T4 into the first part of the ancillary data field. If the MC encoder does not use all of the ancillary data field for the multichannel extension information, the remaining part of the field can be used for other ancillary data.

The bit rate required for the multichannel extension information may vary on a frame by frame basis, depending on the sound signals. The overall bit rate may be increased above that provided for in ISO/IEC 11172-3 by the use of an optional extension bit stream. The maximum bit rate, including the extension bit stream, is given by the following table:

Sampling Frequency	Layer	Maximum Total Bit Rate
32 kHz	I	903 kbit/s
32 kHz	II	839 kbit/s
32 kHz	III	775 kbit/s
44.1 kHz	I	1075 kbit/s
44.1 kHz	II	1011 kbit/s
44.1 kHz	III	947 kbit/s
48 kHz	I	1130 kbit/s
48 kHz	II	1066 kbit/s
48 kHz	III	1002 kbit/s

This part of ISO/IEC 13818 describes the combinations of the basic Lo, Ro stereo of Layer I, II and III and the multichannel extension of Layer II mc and Layer III mc. The following combinations are possible:

Basic Lo, Ro Stereo	Multichannel Extension
Layer II	Layer II mc
Layer III	Layer III mc
Layer I	Layer II mc

0.2.3.2 Audio Input/Output Format

Sampling frequencies: 48, 44.1 or 32 kHz
Quantisation: up to 24 bits/sample PCM resolution

The following combinations of audio channels can be applied as inputs to the audio encoder:

a) Five channels, using the 3/2 configuration
L, C, R plus two surround channels LS, RS

b) Four channels, using the 3/1 configuration
L, C, R plus single surround channel S

c) Three channels using the 3/0 configuration
L, C, R without surround

d) Five channels, using the 3/0 + 2/0 configuration
L, C, R of first programme plus L2, R2 of second programme

e) Four channels, using the 2/2 configuration
L, R plus two surround channels LS, RS

L, R with single surround channel S

g) Two channels, using the 2/0 (or 1/0+1/0) configuration
Stereo (or dual channel mode) as in ISO/IEC 11172-3

h) Four channels, using the 2/0 + 2/0 (or 1/0+1/0+ 2/0) configuration
L, R (or channel I and channel II) of first programme plus L2, R2 of second programme

i) One channel, using the 1/0 configuration
Single channel mode (as in ISO/IEC 11172-3)

j) Three channels, using the 1/0 + 2/0 configuration
Single channel mode (as in ISO/IEC 11172-3) plus L2, R2 of second programme

The different combinations of audio input signals are encoded and transmitted within the up to five available transmission channels T0, T1, T2, T3 and T4, of which channels T0 and T1 are the two basic channels of ISO/IEC 11172-3 and convey the backwards compatible signals Lo and Ro. Transmission channels T2, T3 and T4 together form the multichannel extension information, which is compatibly transmitted within the ISO/IEC 11172-3 ancillary data field and an optional extension bit stream.

After multichannel decoding, the up to five audio channels are recovered and can then be presented in any convenient format at the choice of the listeners:

a) Five channels, using the 3/2 configuration
Front: Left (L) and right (R) channel plus centre channel (C)
Surround: Left surround (LS) and right surround (RS)

b) Four channels, using the 3/1 configuration
Front: Left (L) and right (R) channel plus centre channel (C)
Surround: Mono surround (S)

c) Three channels using the 3/0 configuration
Front: Left (L) and right (R) channel plus centre channel (C)
Surround: No surround

d) Four channels, using the 2/2 configuration
Front: Left (L) and right (R) channel
Surround: Left surround (LS) and right surround (RS)

e) Three channels, using the 2/1 configuration
Front: Left (L) and right (R) channel
Surround: Mono surround (S)

f) Two channels, using the 2/0 configuration
Front: Left (L) and right channel (R)
Surround: No surround

g) One channel output, using the 1/0 configuration
Front: Mono channel (Mo)
Surround: No surround

A low frequency enhancement channel can, optionally, be added to any of these configurations, except for the 1/0 configuration.

Outputs may be required to provide discrete signals, or may be combined in accordance with downward mixing, or upwards conversion equations, as defined in ITU-R Recommendation 775.

0.2.3.3 Composite Coding Modes

Dynamic Transmission Channel Switching

In order to provide a better orthogonality between the two compatible signals T0 and T1, and the three additionally transmitted signals T2, T3 and T4, it is necessary to have flexibility in the choice of the channels T2, T3 and T4. This part of ISO/IEC 13818 allows, independently for a number of frequency regions, the selection of a number of combinations of three out of the five signals L, C, R, LS, RS to be transmitted in T2, T3 and T4.

Dynamic Crosstalk

According to a binaural hearing model, it is possible to determine the portion of the stereophonic signal which is irrelevant with respect to the spatial perception of the stereophonic presentation. The stereo-irrelevant signal components are not masked, but they do not contribute to the localisation of sound sources. They are ignored in

signal (L, C, R, LS or RS) may be reproduced via any loudspeaker, or via several loudspeakers of the arrangement, without affecting the stereophonic impression. This can be done independently for a number of frequency regions.

Adaptive Multichannel Prediction

In order to make use of the statistical inter-channel dependencies, adaptive multichannel prediction is used for redundancy reduction. Instead of transmitting the actual signals in the transmission channels T2, T3, T4, the corresponding prediction error signals are transmitted. A predictor of up to 2nd order with delay compensation is used.

Phantom Coding of Centre

Due to the fact that the human auditory system uses only intensity cues of the audio signal for localisation at higher frequencies, it is possible to transmit the high frequency part of the centre channel in the front left and right channels, constituting a phantom source at the location of the centre loudspeaker.

0.2.3.4 Encoder and Decoder Parameters

Encoding and decoding:	similar to ISO/IEC 11172-3.	
Coding modes:	3/2, 3/1, 3/0 (+ 2/0), 2/2, 2/1, 2/0 (+ 2/0), 1/0+1/0 (+ 2/0), 1/0 (+ 2/0) second stereo programme, up to 7 additional multilingual or commentary channels, associated services.	
Subband filter transforms:	Number of subbands:	32
	Sampling frequency:	Fs/32
	Bandwidth of subbands:	Fs/64
Additional decomposition by MDCT (Layer III only):		
	Frequency Resolution:	6 or 18 components per subband
LFE channel filter transform:	Number of LFE channels:	1
	Sampling frequency:	Fs/96
	Bandwidth of LFE channel:	125 Hz
Dynamic range:	more than 20 bits.	