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

Part 3:
Audio
(standards.iteh.ai)

*Technologies de l'information — Codage générique des images animées
et des informations sonores associées —*
<https://standards.iteh.ai/standards/iso-iec-13818-3-1998>
Partie 3: Son

Contents

	Page
Section 1: General	1
1.1 Scope	1
1.2 Normative references	1
Section 2: Technical elements	2
2.1 Definitions	2
2.2 Symbols and abbreviations	13
2.3 Method of describing bit stream syntax	16
2.4 Requirements for Extension of ISO/IEC 11172-3 to Lower Sampling Frequencies	18
2.5 Requirements for Extension of ISO/IEC 11172-3 to Multichannel Audio	24
2.6 Registration of Copyright Identifiers	65
 Annexes	
A. Diagrams	66
B. Tables	69
C. The encoding process	73
D. Psychoacoustic models	84
E. Ancillary Data Use	108
F. List of patent holders	110
G. Registration Procedure	112
H. Registration Application Form	114
I. Registration Authority	115

© ISO/IEC 1998

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the publisher.

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

Printed in Switzerland.

Foreword

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.

Introduction

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 (F_s) 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

accompanying picture but also for audio-only applications. A universal and compatible multichannel audio 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).

$$L_o = L + \frac{1}{2}\sqrt{2} * C + \frac{1}{2}\sqrt{2} * LS \quad (1)$$

$$R_o = R + \frac{1}{2}\sqrt{2} * C + \frac{1}{2}\sqrt{2} * RS \quad (2)$$

or

$$L_o = L + \frac{1}{2}\sqrt{2} * C + \frac{1}{2} * LS \quad (3)$$

$$R_o = R + \frac{1}{2}\sqrt{2} * C + \frac{1}{2} * RS \quad (4)$$

or

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

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

or

$$L_o = L + \frac{1}{2}\sqrt{2} * C - \frac{1}{2}\sqrt{2} * jS \quad (7)$$

$$R_o = R + \frac{1}{2}\sqrt{2} * C + \frac{1}{2}\sqrt{2} * jS \quad (8)$$

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

drawback of this solution is the additional bit rate necessary for transmitting 7 audio channels instead of 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'0/R'0, and the multichannel decoder produces the complete 3/2-stereo presentation L'0/C'0/R'0/LS'0/RS'0 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.

iTeh STANDARD PREVIEW (standards.iteh.ai)

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).

- Audio data, for Layer II consisting of bit allocation (BAL), scalefactor select information (SCFSI), 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 1172-3 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

- f) Three channels using the 2/1 configuration
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

the binaural processor of the human auditory system. Therefore, stereo-irrelevant components of any stereo 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 II 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.	

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

Part 3:

Audio

Section 1: General

1.1 Scope

This part of ISO/IEC 13818 specifies the extension of ISO/IEC 11172-3 to lower sampling frequencies, the coded representation of multichannel and multilingual high quality audio for broadcasting, transmission and storage media, and the method for decoding of multichannel and multilingual high quality audio signals. The input of the encoder and the output of the decoder are compatible with existing PCM standards.

1.2 Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this part of ISO/IEC 13818. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this part of ISO/IEC 13818 are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO/IEC 11172-3: 1993, *Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s - Part 3: Audio*.

CCIR Recommendation 601-1: 1990, *Encoding parameters of digital television for studios*.

CCIR Recommendation 648: 1986, *Recording of audio signals*.

CCIR Recommendation 775: 1992, *Multichannel stereophonic sound system with and without accompanying picture*.

CCIR Report 955-2: 1990, *Sound broadcasting by satellite for portable and mobile receivers, including Annex IV Summary description of Advanced Digital System II*.

IEC 908: 1987, *Compact disc digital audio system*.

IEE Draft Standard P1180/D2: 1990, *Specification for the implementation of 8x 8 inverse discrete cosine transform*.

ITU-T Recommendation G.722: 1988, *7 kHz audio coding within 64 kbit/s*.

European Telecommunication Standard pr ETS 300 401: 1995, *Radio Broadcasting system; Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers*.

ITU-T Recommendation J.52: 1995, *Digital Transmission of High Quality Sound Programme Signals Using One, Two or Three 64 kbit/s Channels per Mono Signal (and up to Six per Stereo Signal)*.

Section 2: Technical elements

2.1 Definitions

For the purposes of this part of ISO/IEC 13818, the following definitions apply. If specific to a part, this is noted in square brackets.

2.1.1 16x8 prediction [video]: A prediction mode similar to field-based prediction but where the predicted block size is 16x8 luminance samples.

2.1.2 AC coefficient [video]: Any DCT coefficient for which the frequency in one or both dimensions is non-zero.

2.1.3 access unit [system]: A coded representation of a presentation unit. In the case of audio, an access unit is the coded representation of an audio frame.

In the case of video, an access unit includes all the coded data for a picture, and any stuffing that follows it, up to but not including the start of the next access unit. If a picture is not preceded by a group_start_code or a sequence_header_code, the access unit begins with the picture start code. If a picture is preceded by a group_start_code and/or a sequence_header_code, the access unit begins with the first byte of the first of these start codes. If it is the last picture preceding a sequence_end_code in the bitstream all bytes between the last byte of the coded picture and the sequence_end_code (including the sequence_end_code) belong to the access unit.

2.1.4 adaptive bit allocation [audio]: The assignment of bits to subbands in a time and frequency varying fashion according to a psychoacoustic model.

2.1.5 adaptive multichannel prediction [audio]: A method of multichannel data reduction exploiting statistical inter-channel dependencies.

2.1.6 adaptive noise allocation [audio]: The assignment of coding noise to frequency bands in a time and frequency varying fashion according to a psychoacoustic model.

2.1.7 adaptive segmentation [audio]: A subdivision of the digital representation of an audio signal in variable segments of time.

2.1.8 alias [audio]: Mirrored signal component resulting from sub-Nyquist sampling.

2.1.9 analysis filterbank [audio]: Filterbank in the encoder that transforms a broadband PCM audio signal into a set of subsampled subband samples.

2.1.10 ancillary data [audio]: part of the bitstream that might be used for transmission of ancillary data.

2.1.11 audio access unit [audio]: For Layers I and II, an audio access unit is defined as the smallest part of the encoded bitstream which can be decoded by itself, where decoded means "fully reconstructed sound". For Layer III, an audio access unit is part of the bitstream that is decodable with the use of previously acquired main information.

2.1.12 audio buffer [audio]: A buffer in the system target decoder for storage of compressed audio data.

2.1.13 audio sequence [audio]: A non-interrupted series of audio frames (base frames plus optional extension frames) in which the following parameters are not changed:

- ID
- Layer
- Sampling Frequency

For Layer I and II, a decoder is not required to support a continuously variable bitrate (change in the bitrate index) of the base stream. Such a relaxation of requirements does not apply to the extension stream.

2.1.14 B-field picture [video]: A field structure B-Picture.

2.1.15 B-frame picture [video]: A frame structure B-Picture.

2.1.16 B-picture; bidirectionally predictive-coded picture [video]: A picture that is coded using motion compensated prediction from past and/or future reference fields or frames.

2.1.17 Bark [audio]: Unit of critical band rate. The Bark scale is a non-linear mapping of the frequency scale over the audio range closely corresponding with the frequency selectivity of the human ear across the band.

- 2.1.18 backward compatibility:** A newer coding standard is backward compatible with an older coding standard if decoders designed to operate with the older coding standard are able to continue to operate by decoding all or part of a bitstream produced according to the newer coding standard.
- 2.1.19 backward motion vector [video]:** A motion vector that is used for motion compensation from a reference frame or reference field at a later time in display order.
- 2.1.20 backward prediction [video]:** Prediction from the future reference frame (field).
- 2.1.21 base bit stream [audio]:** Information contained in a bit stream which consists of continuous base frames. This bit stream is decodable by an ISO/IEC 11172-3 and an ISO/IEC 13818-3 decoder. An ISO/IEC 13818-3 bit stream shall always consist of the base bit stream and optionally of an extension bit stream.
- 2.1.22 base frame [audio]:** The part of the ISO/IEC 13818-3 encoded audio frame which can be decoded by an ISO/IEC 11172-3 decoder and contains the basic stereo signal.
- 2.1.23 base layer [video]:** First, independently decodable layer of a scalable hierarchy.
- 2.1.24 big picture [video]:** A coded picture that would cause VBV buffer underflow as defined in C.7 Annex C of ISO/IEC 13818-2. Big pictures can only occur in sequences where low_delay is equal to 1. “Skipped picture” is a term that is sometimes used to describe the same concept.
- 2.1.25 bitrate:** The rate at which the compressed bitstream is delivered to the input of a decoder.
- 2.1.26 bitstream; stream:** An ordered series of bits that forms the coded representation of the data.
- 2.1.27 bitstream verifier [video]:** A process by which it is possible to test and verify that all the requirements specified in ISO/IEC 13818-2 are met by the bitstream.
- 2.1.28 block [video]:** An 8-row by 8-column matrix of samples, or 64 DCT coefficients (source, quantised or dequantised).
- 2.1.29 block companding [audio]:** Normalising of the digital representation of an audio signal within a certain time period.
- 2.1.30 bottom field [video]:** One of two fields that comprise a frame. Each line of a bottom field is spatially located immediately below the corresponding line of the top field.
- 2.1.31 bound [audio]:** The lowest subband in which intensity stereo coding is used.
- 2.1.32 byte aligned:** A bit in a coded bitstream is byte-aligned if its position is a multiple of 8-bits from the first bit in the stream.
- 2.1.33 byte:** Sequence of 8-bits.
- 2.1.34 centre channel [audio]:** An audio presentation channel used to stabilise the central component of the frontal stereo image.
- 2.1.35 channel [audio]:** A sequence of data representing an audio signal being transported.
- 2.1.36 chroma simulcast [video]:** A type of scalability (which is a subset of SNR scalability) where the enhancement layer (s) contain only coded refinement data for the DC coefficients, and all the data for the AC coefficients, of the chrominance components.
- 2.1.37 chrominance format [video]:** Defines the number of chrominance blocks in a macroblock.
- 2.1.38 chrominance component [video]:** A matrix, block or single sample representing one of the two colour difference signals related to the primary colours in the manner defined in the bitstream. The symbols used for the chrominance signals are Cr and Cb.
- 2.1.39 coded audio bitstream [audio]:** A coded representation of an audio signal as specified in this part of ISO/IEC 13818.
- 2.1.40 coded B-frame [video]:** A B-frame picture or a pair of B-field pictures.
- 2.1.41 coded frame [video]:** A coded frame is a coded I-frame, a coded P-frame or a coded B-frame.
- 2.1.42 coded I-frame [video]:** An I-frame picture or a pair of field pictures, where the first field picture is an I-picture and the second field picture is an I-picture or a P-picture.

- 2.1.43 coded order [video]:** The order in which the pictures are transmitted and decoded. This order is not necessarily the same as the display order.
- 2.1.44 coded P-frame [video]:** A P-frame picture or a pair of P-field pictures.
- 2.1.45 coded picture [video]:** A coded picture is made of a picture header, the optional extensions immediately following it, and the following picture data. A coded picture may be a coded frame or a coded field.
- 2.1.46 coded representation:** A data element as represented in its encoded form.
- 2.1.47 coded video bitstream [video]:** A coded representation of a series of one or more pictures as defined in ISO/IEC 13818-2.
- 2.1.48 coding parameters [video]:** The set of user-definable parameters that characterise a coded bitstream. Bitstreams are characterised by coding parameters. Decoders are characterised by the bitstreams that they are capable of decoding.
- 2.1.49 component [video]:** A matrix, block or single sample from one of the three matrices (luminance and two chrominance) that make up a picture.
- 2.1.50 compression:** Reduction in the number of bits used to represent an item of data.
- 2.1.51 constant bitrate:** Operation where the bitrate is constant from start to finish of the coded bitstream.
- 2.1.52 constrained parameters [video]:** The values of the set of coding parameters defined in 2.4.3.2 of ISO/IEC 11172-2.
- 2.1.53 constrained system parameter stream; CSPTS [system]:** A Program Stream for which the constraints defined in 2.7.9 of ISO/IEC 13818-1 apply.
- 2.1.54 CRC:** The Cyclic Redundancy Check to verify the correctness of data.
- 2.1.55 critical band [audio]:** Psychoacoustic measure in the spectral domain which corresponds to the frequency selectivity of the human ear. This selectivity is expressed in Bark.
- 2.1.56 critical band rate [audio]:** Psychoacoustic function of frequency. At a given audible frequency, it is equal to the number of critical bands below that frequency. The units of the critical band rate scale are Barks.
- 2.1.57 data element:** An item of data as represented before encoding and after decoding.
- 2.1.58 data partitioning [video]:** A method for dividing a bitstream into two separate bitstreams for error resilience purposes. The two bitstreams have to be recombined before decoding.
- 2.1.59 DC coefficient [video]:** The DCT coefficient for which the frequency is zero in both dimensions.
- 2.1.60 DCT coefficient [video]:** The amplitude of a specific cosine basis function.
- 2.1.61 de-emphasis [audio]:** Filtering applied to an audio signal after storage or transmission to undo a linear distortion due to emphasis.
- 2.1.62 decoded stream:** The decoded reconstruction of a compressed bitstream.
- 2.1.63 decoder input buffer [video]:** The first-in first-out (FIFO) buffer specified in the video buffering verifier.
- 2.1.64 decoder:** An embodiment of a decoding process.
- 2.1.65 decoder sub-loop [video]:** Stages within encoder which produce numerically identical results to the decode process described in ISO/IEC 13818-2, clause 7. Encoders capable of producing more than just I-pictures embed a decoder sub-loop to create temporal predictions and to model the behaviour of downstream decoders.
- 2.1.66 decoding (process):** The process defined in ISO/IEC 13818 parts 1, 2 and 3 that reads an input coded bitstream and outputs decoded pictures or audio samples.
- 2.1.67 decoding time-stamp; DTS [system]:** A field that may be present in a PES packet header that indicates the time that an access unit is decoded in the system target decoder.
- 2.1.68 dequantisation [video]:** The process of rescaling the quantised DCT coefficients after their representation in the bitstream has been decoded and before they are presented to the inverse DCT.
- 2.1.69 digital storage media; DSM:** A digital storage or transmission device or system.

- 2.1.70 discrete cosine transform; DCT:** Either the forward discrete cosine transform or the inverse discrete cosine transform. The DCT is an invertible, discrete orthogonal transformation.
- 2.1.71 display aspect ratio [video]:** The ratio height/width (in SI units) of the intended display.
- 2.1.72 display order [video]:** The order in which the decoded pictures are displayed. Normally this is the same order in which they were presented at the input of the encoder.
- 2.1.73 display process [video]:** The (non-normative) process by which reconstructed frames are displayed.
- 2.1.74 downmix [audio]:** A matrixing of n channels to obtain less than n channels.
- 2.1.75 drift [video]:** Accumulation of mismatch between the reconstructed output produced by the hypothetical decoder sub-loop embedded within an encoder (see definition of "decoder sub-loop") and the reconstructed outputs produced by a (downstream) decoder.
- 2.1.76 DSM-CC:** digital storage media command and control.
- 2.1.77 dual channel mode [audio]:** A mode, where two audio channels with independent programme contents (e.g. bilingual) are encoded within one bitstream. The coding process is the same as for the stereo mode.
- 2.1.78 dual-prime prediction [video]:** A prediction mode in which two forward field-based predictions are averaged. The predicted block size is 16x16 luminance samples. Dual-prime prediction is only used in interlaced P-pictures.
- 2.1.79 dynamic crosstalk [audio]:** A method of multichannel data reduction in which stereo-irrelevant signal components are copied to another channel.
- 2.1.80 dynamic transmission channel switching [audio]:** A method of multichannel data reduction by allocating the most orthogonal signal components to the transmission channels.
- 2.1.81 editing:** The process by which one or more coded bitstreams are manipulated to produce a new coded bitstream. Conforming edited bitstreams must meet the requirements defined in parts 1, 2, and 3 of ISO/IEC 13818.
- 2.1.82 Elementary Stream Clock Reference; ESCR [system]:** A time stamp in the PES Stream from which decoders of PES streams may derive timing. <https://standards.iteh.ai/catalog/standards/sist/c5bfc71-1430-4339-9e1e-4aef04b7c630/iso-iec-13818-3-1998>
- 2.1.83 elementary stream; ES [system]:** A generic term for one of the coded video, coded audio or other coded bitstreams in PES packets. One elementary stream is carried in a sequence of PES packets with one and only one stream_id.
- 2.1.84 emphasis [audio]:** Filtering applied to an audio signal before storage or transmission to improve the signal-to-noise ratio at high frequencies.
- 2.1.85 encoder:** An embodiment of an encoding process.
- 2.1.86 encoding (process):** A process, not specified in ISO/IEC 13818, that reads a stream of input pictures or audio samples and produces a valid coded bitstream as defined in parts 1, 2, and 3 of ISO/IEC 13818.
- 2.1.87 enhancement layer [video]:** A relative reference to a layer (above the base layer) in a scalable hierarchy. For all forms of scalability, its decoding process can be described by reference to the lower layer decoding process and the appropriate additional decoding process for the enhancement layer itself.
- 2.1.88 entitlement control message; ECM [systems]:** Entitlement Control Messages are private conditional access information which specify control words and possibly other, typically stream-specific, scrambling and/or control parameters.
- 2.1.89 entitlement management message; EMM [systems]:** Entitlement Management Messages are private conditional access information which specify the authorisation levels or the services of specific decoders. They may be addressed to single decoders or groups of decoders.
- 2.1.90 entropy coding:** Variable length lossless coding of the digital representation of a signal to reduce redundancy.
- 2.1.91 event [systems]:** An event is defined as a collection of elementary streams with a common time base, an associated start time, and an associated end time.
- 2.1.92 evil bitstreams:** bitstreams orthogonal to reality.