EESTI STANDARD **EVS-EN ISO/IEC 13818-3:1999**

Infotehnoloogia. Videosignaali ja sellega seotud heliinformatsiooni üldkodeerimine. Osa 3: Helisignaal

Information technology - Generic of moving coding pictures associated audio information - Part 3: Audio

EVS **EESTI STANDARDIKESKUS**

EESTI STANDARDI EESSÕNA

NATIONAL FOREWORD

Käesolev Eesti standard EVS-EN ISO/IEC 13818-3:1999 sisaldab Euroopa standardi EN ISO/IEC 13818-3:1996 ingliskeelset teksti. Käesolev dokument on jõustatud 12.12.1999 ja selle kohta on avaldatud teade Eesti standardiorganisatsiooni ametlikus väljaandes. Standard on kättesaadav Eesti standardiorganisatsioonist.	This Estonian standard EVS-EN ISO/IEC 13818-3:1999 consists of the English text of the European standard EN ISO/IEC 13818-3:1996. This document is endorsed on 12.12.1999 with the notification being published in the official publication of the Estonian national standardisation organisation. The standard is available from Estonian standardisation organisation.

Käsitlusala: Käesolev standard spetsifitseerib standardi ISO/IEC 11172-3 laienduse madalamatele diskretiseerimissagedustele ja kodeeritud kõrgekvaliteetse mitmekanalilise ning mitmekeelse helisignaali esituseks ringhäälingule, andmeedastusele ja andmekandjale, ning meetodi kõrgekvaliteetsete mitmekanaliliste ja mitmekeelsete helisignaalide dekodeerimiseks. Kooderi sisend ja dekoodri väljund ühilduvad kehtivate PCM standarditega.	**Scope:**

ICS 35.040

Võtmesõnad: andmeteisendus, andmetöötlus, audioandmed, filmid, kodeerimine (andmeteisendus), video salvestamine

EUROPEAN STANDARD
NORME EUROPÉENNE
EUROPÄISCHE NORM

EN ISO/IEC 13818-3

Mai 1996

ICS 35.040

Descriptors: data processing, moving pictures, audio data, video recording, data converting, coding (data conversion)

English version

Information technology

Generic of coding pictures and associated audio information
Part 3: Audio (ISO/IEC 13818-3 : 1995)

Technologies de l'information — Codage générique des images animées et des informations sonores associées — Partie 3: Son (ISO/IEC 13818-3 : 1995)

Informationstechnik — Codierung von bewegten Bildern und damit verbundenen Toninformationen — Teil 3: Audio (ISO/IEC 13818-3 : 1995)

This European Standard was approved by CEN on 1996-04-18.

CEN members are bound to comply with the CEN/CENELEC Internal Regulations which stipulate the conditions for giving this European Standard the status of a national standard without any alteration.

Up-to-date lists and bibliographical references concerning such national standards may be obtained on application to the Central Secretariat or to any CEN member.

This European Standard exists in three official versions (English, French, German). A version in any other language made by translation under the responsibility of a CEN member into its own language and notified to the Central Secretariat has the same status as the official versions.

CEN members are the national standards bodies of Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and United Kingdom.

CEN

EUROPEAN COMMITTEE FOR STANDARDIZATION
Comité Européen de Normalisation
Europäisches Komitee für Normung

Central Secretariat: rue de Stassart 36, B-1050 Brussels

Ref. No. EN ISO/IEC 13818-3 : 1996 E

Foreword

This European Standard was taken over by CEN from the work of ISO/IEC/JTC 1 "Information technology" of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) has been taken as a European Standard by the CEN Technical Board.

This European Standard shall be given the status of a national standard, either by publication of an identical text or by endorsement, at the latest by 1996-11, and conflicting national standards shall be withdrawn at the latest by 1996-11.

According to the CEN/CENELEC Internal Regulations, the national standards organizations of the following countries are bound to announce this European Standard:

Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom.

Endorsement notice

The text of the International Standard ISO/IEC 13818-3 : 1995 was approved by CEN as a European Standard without any modification.

INTERNATIONAL STANDARD

ISO/IEC 13818-3

First edition
1995-05-15

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

Contents

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*.

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*

Annexes A and B form an integral part of this part of ISO/IEC 13818. Annexes C to E 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.

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 CCITT Standard G-722 performance, three additional sampling frequencies are provided for ISO/IEC 11172-3 layers I, II and III. The additional sampling frequencies 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 to 3/2 multichannel audio and an optional low frequency enhancement channel (LFE).

Multichannel audio systems provide enhanced stereophonic 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
ENG   Electronic News Gathering (including Satellite News Gathering)
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.

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 TG10/1, 10/2 and 10/3 of the ITU-R (previously CCIR).

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

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 Lo/Ro, 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 backward compatibility are being evaluated for optional use with the standard.

0.2.2.3 Multilingual capability

Particularly for HDTV applications, multichannel stereo performance and bilingual programmes or multilingual commentaries are required. This standard 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

Backwards and forwards compatibility with an ISO/IEC 11172-3 decoder is provided.

For a multichannel audio bit stream, backwards compatibility means, that an ISO/IEC 11172-3 audio decoder properly decodes the basic stereo information. The basic stereo information consists of a left and right channel that constitute an appropriate downmix of the audio information in all channels, or, optionally, the basic stereo information may consist only of the left and right channel of the multichannel audio configuration. Appropriate downmix equations are given by equation pairs (1) and (2), (3) and (4), and (5) and (6).

$$Lo = L + x * C + y * LS \qquad (1)$$
$$Ro = R + x * C + z * RS \qquad (2)$$

or

$$Lo = L \qquad (3)$$
$$Ro = R \qquad (4)$$

or

$$Lo = L + x * C - y * jS \qquad (5)$$
$$Ro = R + x * C + y * jS \qquad (6)$$

where jS is derived from LS and RS by calculation of the mono component, bandwidth limitation to the range 100-7000 Hz, half Dolby®[1] B-type encoding, and 90 degrees phase shifting (Prologic®[1] surround matrixing).

---

[1]Dolby and Prologic are registered trademarks of Dolby Laboratories Licensing Corp.

Compatibility with existing surround sound decoders by use of equations (5) and (6) has not been verified at the time of printing of this part of ISO/IEC 13818.

Forwards compatibility means that an MPEG 2 multichannel audio decoder is able to decode properly an ISO/IEC 11172-3 audio bit stream.

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

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 ISO/MPEG-Audio Multichannel system provides full compatibility with the ISO Standard 11172-3. This 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 and an optional extension bit stream for the multichannel extension.

The complete ISO/IEC 11172-3 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 the ISO/IEC 11172-3 Standard, 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

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

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

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

e) Four channels, using the 2/0 + 2/0 configuration
L, R of first programme plus L2, R2 of second programme

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

g) Three channels using the 2/1 configuration
L, R with single surround channel S

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

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

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) channels 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) channels plus centre channel (C)
Surround: Mono surround (S)

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

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

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

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.

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. ISO/IEC 13818-3 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, 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, 3/0 + 2/0, 3/1, 2/0 + 2/0, 3/0, 2/2, 2/1, 2/0, 1/0
> 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

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 13818are 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.*

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