INTERNATIONAL STANDARD

ISO/IEC 11172-4

First edition
1995-03-15

Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s —

Part 4:
Compliance testing

Technologies de l'information — Codage de l'image animée et du son associé pour les supports de stockage numérique jusqu'à environ 1,5 Mbit/s —

Partie 4: Essais de conformité

Reference number
ISO/IEC 11172-4:1995(E)

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for world-wide standardization. National Bodies that are members of ISO and 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.

ISO/IEC 11172 consists of the following parts, under the general title *Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s:*

- *Part 1: Systems*
- *Part 2: Video*
- *Part 3: Audio*
- *Part 4: Compliance testing*

Annex A forms an integral part of this part of ISO/IEC 11172. Annex B is for information only.

Introduction

This International Standard was prepared by ISO/IEC JTC1/SC29/WG11 also known as MPEG (Moving Pictures Expert Group). MPEG was formed in 1988 to establish an International Standard for the coded representation of moving pictures and associated audio stored on digital storage media. Parts 1, 2 and 3 of this International Standard were unanimously approved by the participating National Bodies in November 1992.

This International Standard is published in four 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 synchronized 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 reconstruct audio. Part 4 - compliance testing - specifies procedures to determine characteristics of coded bitstreams and to test compliance of bitstreams and decoders with the requirements specified in Parts 1, 2 and 3.

Parts 1, 2 and 3 of ISO/IEC 11172 specify a multiplex structure and coded representations of audiovisual information. Parts 1, 2 and 3 of ISO/IEC 11172 allow for large flexibility, achieving suitability of this International Standard for many different applications. The flexibility is obtained by including parameters in the bitstream that define the characteristics of coded bitstreams. Examples are the audio sampling frequency, picture size, picture rate and bitrate parameters.

This part of ISO/IEC 11172 specifies how tests can be designed to verify whether bitstreams and decoders meet the requirements as specified in parts 1, 2 and 3 of ISO/IEC 11172. These tests can be used for various purposes such as:

- manufacturers of encoders, and their customers, can use the tests to verify whether the encoder produces valid bitstreams.
- manufacturers of decoders and their customers can use the tests to verify whether the decoder meets the requirements specified in parts 1, 2 and 3 of ISO/IEC 11172 for the claimed decoder capabilities.
- applications can use the tests to verify whether the characteristics of a given bitstream meet the application requirements, for example whether the size of the coded picture does not exceed the maximum value allowed for the application.

Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s —

Part 4:
Compliance testing

Section 1: General

1.1 Scope

This part of ISO/IEC 11172 specifies how tests can be designed to verify whether bitstreams and decoders meet requirements specified in parts 1, 2 and 3 of ISO/IEC 11172. In this part of ISO/IEC 11172, encoders are not addressed specifically. An encoder is entitled to be an ISO/IEC 11172 encoder if it generates bitstreams compliant with the syntactic and semantic bitstream requirements specified in parts 1, 2 and 3 of ISO/IEC 11172.

Characteristics of coded bitstreams and decoders are defined for parts 1, 2 and 3 of ISO/IEC 11172. The characteristics of a bitstream define the subset of the standard that is exploited in the bitstream. Examples are the applied values or range of the picture size and bitrate parameters. Decoder characteristics define the properties and capabilities of the applied decoding process. An example of a property is the applied arithmetic accuracy. The capabilities of a decoder specify which coded bitstreams the decoder can decode and reconstruct, by defining the subset of the standard that may be exploited in decodable bitstreams. A bitstream can be decoded by a decoder if the characteristics of the coded bitstream are within the subset of the standard specified by the decoder capabilities.

Procedures are descibed for testing compliance of bitstreams and decoders to the requirements defined in parts 1, 2 and 3 of ISO/IEC 11172. Given the set of characteristics claimed, the requirements that must be met are fully determined by parts 1, 2 and 3 of ISO/IEC 11172. This part of ISO/IEC 11172 summarizes the requirements, cross references them to characteristics, and defines how compliance with them can be tested. Guidelines are given how to construct tests and determine their outcome. Some actual tests are defined only for audio.

1.2 Normative references

The following International Standards contain provisions which, through reference in this text, constitute provisions of this part of ISO/IEC 11172. 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 11172 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 11172-1:1993 *Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s - Part 1: Systems.*

ISO/IEC 11172-2:1993 *Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s - Part 2: Video.*

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 601-2 *Encoding parameters of digital television for studios.*

CCIR Report 624-4 *Characteristics of systems for monochrome and colour television.*

CCIR Recommendation 648 *Recording of audio signals.*

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

CCITT Recommendation J.17 *Pre-emphasis used on Sound-Programme Circuits.*

IEEE Draft Standard P1180/D2 1990 *Specification for the implementation of 8x8 inverse discrete cosine transform".*

IEC publication 908:1987 *CD Digital Audio System.*

Section 2: Technical elements

2.1 Definitions

For the purposes of this part of ISO/IEC 11172, the following definitions apply. If the definition is specific to a part, this is noted in square brackets.

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

2.1.2 access unit [system]: In the case of compressed audio an access unit is an audio access unit. In the case of compressed video an access unit is the coded representation of a picture.

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

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

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

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

2.1.8 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.

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

2.1.10 audio sequence [audio]: A non-interrupted series of audio frames in which the following parameters are not changed:

- ID
- Layer
- Sampling Frequency
- For Layer I and II: Bitrate index

2.1.11 backward motion vector [video]: A motion vector that is used for motion compensation from a reference picture at a later time in display order.

2.1.12 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.

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

2.1.14 bitrate: The rate at which the compressed bitstream is delivered from the storage medium to the input of a decoder.

2.1.15 bitstream characteristics [compliance]: The subset of the standard that is exploited by the encoder in generating the bitstream. For example, an encoder may apply syntactic and semantic constraints, such as restricted ranges of parameters, to produce a bitstream that exploits a subset of the capabilities supported by parts 1, 2 and 3 of ISO/IEC 11172. Examples are the applied values or range of the picture size and bitrate parameters in video bitstreams.

2.1.16 bitstream compliance [compliance]: A bitstream is compliant, if the bitstream meets the syntactic and semantic bitstream requirements, specified in the normative clauses of parts 1, 2 and 3 of ISO/IEC 11172.

2.1.17 bitstream requirements [compliance]: Requirements for bitstreams defined in the normative clauses of parts 1, 2 and 3 of ISO/IEC 11172.

2.1.18 block companding [audio]: Normalizing of the digital representation of an audio signal within a certain time period.

2.1.19 block [video]: An 8-row by 8-column orthogonal block of pels.

2.1.20 bound [audio]: The lowest subband in which intensity stereo coding is used.

2.1.21 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.

2.1.22 byte: Sequence of 8-bits.

2.1.23 channel: A digital medium that stores or transports an ISO/IEC 11172 stream.

2.1.24 channel [audio]: The left and right channels of a stereo signal

2.1.25 chrominance (component) [video]: A matrix, block or single pel representing one of the two colour difference signals related to the primary colours in the manner defined in CCIR Rec 601. The symbols used for the colour difference signals are Cr and Cb.

2.1.26 coded audio bitstream [audio]: A coded representation of an audio signal as specified in ISO/IEC 11172-3.

2.1.27 coded video bitstream [video]: A coded representation of a series of one or more pictures as specified in ISO/IEC 11172-2.

2.1.28 coded order [video]: The order in which the pictures are stored and decoded. This order is not necessarily the same as the display order.

2.1.29 coded representation: A data element as represented in its encoded form.

2.1.30 coding parameters [video]: The set of user-definable parameters that characterize a coded video bitstream. Bitstreams are characterized by coding parameters. Decoders are characterized by the bitstreams that they are capable of decoding.

2.1.31 component [video]: A matrix, block or single pel from one of the three matrices (luminance and two chrominance) that make up a picture.

2.1.32 compression: Reduction in the number of bits used to represent an item of data.

2.1.33 constant bitrate coded video [video]: A compressed video bitstream with a constant average bitrate.

2.1.34 constant bitrate: Operation where the bitrate is constant from start to finish of the compressed bitstream.

2.1.35 constrained parameters [video]: The values of the set of coding parameters defined in 2.4.3.2 of ISO/IEC 11172-2.

2.1.36 constrained system parameter stream (CSPS) [system]: An ISO/IEC 11172 multiplexed stream for which the constraints defined in 2.4.6 of this part of ISO/IEC 11172 apply.

2.1.37 CRC: Cyclic redundancy code.

2.1.38 critical band rate [audio]: Psychoacoustic function of frequency. At a given audible frequency it is proportional to the number of critical bands below that frequency. The units of the critical band rate scale are Barks.

2.1.39 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.

2.1.40 data element: An item of data as represented before encoding and after decoding.

2.1.41 dc-coefficient [video]: The DCT coefficient for which the frequency is zero in both dimensions.

2.1.42 dc-coded picture; D-picture [video]: A picture that is coded using only information from itself. Of the DCT coefficients in the coded representation, only the dc-coefficients are present.

2.1.43 DCT coefficient: The amplitude of a specific cosine basis function.

2.1.44 decoded stream: The decoded reconstruction of a compressed bitstream.

2.1.45 decoder characteristics [compliance]: The properties and capabilities of the decoding process applied in the decoder.

2.1.46 decoder compliance [compliance]: A decoder is compliant, if the decoder meets the decoder requirements, specified in the normative clauses of parts 1, 2 and 3 of ISO/IEC 11172, to decode compliant bitstreams within the subset of the standard defined by the specified capabilities of the decoder.

2.1.47 decoder input buffer [video]: The first-in first-out (FIFO) buffer specified in the video buffering verifier.

2.1.48 decoder input rate [video]: The data rate specified in the video buffering verifier and encoded in the coded video bitstream.

2.1.49 decoder: An embodiment of a decoding process.

2.1.50 decoding (process): The process defined in ISO/IEC 11172 that reads an input coded bitstream and produces decoded pictures or audio samples.

2.1.51 decoder requirements [compliance]: Requirements for decoders defined in the normative clauses of parts 1, 2 and 3 of ISO/IEC 11172.

2.1.52 decoding time-stamp; DTS [system]: A field that may be present in a packet header that indicates the time that an access unit is decoded in the system target decoder.

2.1.53 de-emphasis [audio]: Filtering applied to an audio signal after storage or transmission to undo a linear distortion due to emphasis.

2.1.54 dequantization [video]: The process of rescaling the quantized DCT coefficients after their representation in the bitstream has been decoded and before they are presented to the inverse DCT.

2.1.55 digital storage media; DSM: A digital storage or transmission device or system.

2.1.56 discrete cosine transform; DCT [video]: Either the forward discrete cosine transform or the inverse discrete cosine transform. The DCT is an invertible, discrete orthogonal transformation. The inverse DCT is defined in annex A of ISO/IEC 11172-2.

2.1.57 display order [video]: The order in which the decoded pictures should be displayed. Normally this is the same order in which they were presented at the input of the encoder.

2.1.58 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.

2.1.59 editing: The process by which one or more compressed bitstreams are manipulated to produce a new compressed bitstream. Conforming edited bitstreams must meet the requirements defined in ISO/IEC 11172.

2.1.60 elementary stream [system]: A generic term for one of the coded video, coded audio or other coded bitstreams.

2.1.61 emphasis [audio]: Filtering applied to an audio signal before storage or transmission to improve the signal-to-noise ratio at high frequencies.

2.1.62 encoder: An embodiment of an encoding process.

2.1.63 encoding (process): A process, not specified in ISO/IEC 11172, that reads a stream of input pictures or audio samples and produces a valid coded bitstream as defined in ISO/IEC 11172.

2.1.64 entropy coding: Variable length lossless coding of the digital representation of a signal to reduce redundancy.

2.1.65 fast forward playback [video]: The process of displaying a sequence, or parts of a sequence, of pictures in display-order faster than real-time.

2.1.66 FFT: Fast Fourier Transformation. A fast algorithm for performing a discrete Fourier transform (an orthogonal transform).

2.1.67 filterbank [audio]: A set of band-pass filters covering the entire audio frequency range.

2.1.68 fixed segmentation [audio]: A subdivision of the digital representation of an audio signal into fixed segments of time.

2.1.69 forbidden: The term "forbidden" when used in the clauses defining the coded bitstream indicates that the value shall never be used. This is usually to avoid emulation of start codes.

2.1.70 forced updating [video]: The process by which macroblocks are intra-coded from time-to-time to ensure that mismatch errors between the inverse DCT processes in encoders and decoders cannot build up excessively.

2.1.71 forward motion vector [video]: A motion vector that is used for motion compensation from a reference picture at an earlier time in display order.

2.1.72 frame [audio]: A part of the audio signal that corresponds to audio PCM samples from an audio access unit.

2.1.73 free format [audio]: Any bitrate other than the defined bitrates that is less than the maximum valid bitrate for each layer.

2.1.74 future reference picture [video]: The future reference picture is the reference picture that occurs at a later time than the current picture in display order.

2.1.75 granules [Layer II] [audio]: The set of 3 consecutive subband samples from all 32 subbands that are considered together before quantization. They correspond to 96 PCM samples.

2.1.76 granules [Layer III] [audio]: 576 frequency lines that carry their own side information.

2.1.77 group of pictures [video]: A series of one or more coded pictures intended to assist random access. The group of pictures is one of the layers in the coding syntax defined in ISO/IEC 11172-2.

2.1.78 Hann window [audio]: A time function applied sample-by-sample to a block of audio samples before Fourier transformation.

2.1.79 Huffman coding: A specific method for entropy coding.

2.1.80 hybrid filterbank [audio]: A serial combination of subband filterbank and MDCT.

2.1.81 IMDCT [audio]: Inverse Modified Discrete Cosine Transform.

2.1.82 intensity stereo [audio]: A method of exploiting stereo irrelevance or redundancy in stereophonic audio programmes based on retaining at high frequencies only the energy envelope of the right and left channels.

2.1.83 interlace [video]: The property of conventional television pictures where alternating lines of the picture represent different instances in time.

2.1.84 intra coding [video]: Coding of a macroblock or picture that uses information only from that macroblock or picture.

2.1.85 intra-coded picture; I-picture [video]: A picture coded using information only from itself.

2.1.85a ISO/IEC 11172-1 decoder [compliance]: An embodiment of a decoding process for an ISO/IEC 11172-1 bitstream. MPEG-system decoder is a synonym.

2.1.85b ISO/IEC 11172-2 decoder [compliance]: An embodiment of a decoding process for an ISO/IEC 11172-2 bitstream. MPEG-video decoder is a synonym.

2.1.85c ISO/IEC 11172-3 decoder [compliance]: An embodiment of a decoding process for an ISO/IEC 11172-3 bitstream. MPEG-audio decoder is a synonym.

2.1.86 ISO/IEC 11172 (multiplexed) stream [system]: A bitstream composed of zero or more elementary streams combined in the manner defined in this part of ISO/IEC 11172.

2.1.87 joint stereo coding [audio]: Any method that exploits stereophonic irrelevance or stereophonic redundancy.

2.1.88 joint stereo mode [audio]: A mode of the audio coding algorithm using joint stereo coding.

2.1.89 layer [audio]: One of the levels in the coding hierarchy of the audio system defined in ISO/IEC 11172-3.

2.1.90 layer [video and systems]: One of the levels in the data hierarchy of the video and system specifications defined in this part of ISO/IEC 11172 and ISO/IEC 11172-2.

2.1.91 luminance (component) [video]: A matrix, block or single pel representing a monochrome representation of the signal and related to the primary colours in the manner defined in CCIR Rec 601. The symbol used for luminance is Y.

2.1.92 macroblock [video]: The four 8 by 8 blocks of luminance data and the two corresponding 8 by 8 blocks of chrominance data coming from a 16 by 16 section of the luminance component of the picture. Macroblock is sometimes used to refer to the pel data and sometimes to the coded representation of the pel values and other data elements defined in the macroblock layer of the syntax defined in ISO/IEC 11172-2. The usage is clear from the context.

2.1.93 mapping [audio]: Conversion of an audio signal from time to frequency domain by subband filtering and/or by MDCT.

2.1.94 masking [audio]: A property of the human auditory system by which an audio signal cannot be perceived in the presence of another audio signal .

2.1.95 masking threshold [audio]: A function in frequency and time below which an audio signal cannot be perceived by the human auditory system.

2.1.96 MDCT [audio]: Modified Discrete Cosine Transform.

2.1.97 motion compensation [video]: The use of motion vectors to improve the efficiency of the prediction of pel values. The prediction uses motion vectors to provide offsets into the past and/or future reference pictures containing previously decoded pel values that are used to form the prediction error signal.

2.1.98 motion estimation [video]: The process of estimating motion vectors during the encoding process.

2.1.99 motion vector [video]: A two-dimensional vector used for motion compensation that provides an offset from the coordinate position in the current picture to the coordinates in a reference picture.

2.1.100 MS stereo [audio]: A method of exploiting stereo irrelevance or redundancy in stereophonic audio programmes based on coding the sum and difference signal instead of the left and right channels.

2.1.101 non-intra coding [video]: Coding of a macroblock or picture that uses information both from itself and from macroblocks and pictures occurring at other times.

2.1.102 non-tonal component [audio]: A noise-like component of an audio signal.

2.1.103 Nyquist sampling: Sampling at or above twice the maximum bandwidth of a signal.

2.1.104 pack [system]: A pack consists of a pack header followed by one or more packets. It is a layer in the system coding syntax described in this part of ISO/IEC 11172.

2.1.105 packet data [system]: Contiguous bytes of data from an elementary stream present in a packet.

2.1.106 packet header [system]: The data structure used to convey information about the elementary stream data contained in the packet data.

2.1.107 packet [system]: A packet consists of a header followed by a number of contiguous bytes from an elementary data stream. It is a layer in the system coding syntax described in this part of ISO/IEC 11172.

2.1.108 padding [audio]: A method to adjust the average length in time of an audio frame to the duration of the corresponding PCM samples, by conditionally adding a slot to the audio frame.

2.1.109 past reference picture [video]: The past reference picture is the reference picture that occurs at an earlier time than the current picture in display order.

2.1.110 pel aspect ratio [video]: The ratio of the nominal vertical height of pel on the display to its nominal horizontal width.

2.1.111 pel [video]: Picture element.

2.1.112 picture period [video]: The reciprocal of the picture rate.

2.1.113 picture rate [video]: The nominal rate at which pictures should be output from the decoding process.

2.1.114 picture [video]: Source, coded or reconstructed image data. A source or reconstructed picture consists of three rectangular matrices of 8-bit numbers representing the luminance and two chrominance signals. The Picture layer is one of the layers in the coding syntax defined in ISO/IEC 11172-2. Note that the term "picture" is always used in ISO/IEC 11172 in preference to the terms field or frame.

2.1.115 polyphase filterbank [audio]: A set of equal bandwidth filters with special phase interrelationships, allowing for an efficient implementation of the filterbank.

2.1.116 prediction [video]: The use of a predictor to provide an estimate of the pel value or data element currently being decoded.

2.1.117 predictive-coded picture; P-picture [video]: A picture that is coded using motion compensated prediction from the past reference picture.

2.1.118 prediction error [video]: The difference between the actual value of a pel or data element and its predictor.

2.1.119 predictor [video]: A linear combination of previously decoded pel values or data elements.

2.1.120 presentation time-stamp; PTS [system]: A field that may be present in a packet header that indicates the time that a presentation unit is presented in the system target decoder.

2.1.121 presentation unit; PU [system]: A decoded audio access unit or a decoded picture.

2.1.122 psychoacoustic model [audio]: A mathematical model of the masking behaviour of the human auditory system.

2.1.123 quantization matrix [video]: A set of sixty-four 8-bit values used by the dequantizer.

2.1.124 quantized DCT coefficients [video]: DCT coefficients before dequantization. A variable length coded representation of quantized DCT coefficients is stored as part of the compressed video bitstream.

2.1.125 quantizer scalefactor [video]: A data element represented in the bitstream and used by the decoding process to scale the dequantization.

2.1.126 random access: The process of beginning to read and decode the coded bitstream at an arbitrary point.

2.1.127 reference picture [video]: Reference pictures are the nearest adjacent I- or P-pictures to the current picture in display order.

2.1.128 reorder buffer [video]: A buffer in the system target decoder for storage of a reconstructed I-picture or a reconstructed P-picture.

2.1.129 requantization [audio]: Decoding of coded subband samples in order to recover the original quantized values.

2.1.130 reserved: The term "reserved" when used in the clauses defining the coded bitstream indicates that the value may be used in the future for ISO/IEC defined extensions.

2.1.131 reverse playback [video]: The process of displaying the picture sequence in the reverse of display order.

2.1.132 scalefactor band [audio]: A set of frequency lines in Layer III which are scaled by one scalefactor.

2.1.133 scalefactor index [audio]: A numerical code for a scalefactor.

2.1.134 scalefactor [audio]: Factor by which a set of values is scaled before quantization.

2.1.135 sequence header [video]: A block of data in the coded bitstream containing the coded representation of a number of data elements.

2.1.136 side information: Information in the bitstream necessary for controlling the decoder.

2.1.137 skipped macroblock [video]: A macroblock for which no data are stored.

2.1.138 slice [video]: A series of macroblocks. It is one of the layers of the coding syntax defined in ISO/IEC 11172-2.

2.1.139 slot [audio]: A slot is an elementary part in the bitstream. In Layer I a slot equals four bytes, in Layers II and III one byte.

2.1.140 source stream: A single non-multiplexed stream of samples before compression coding.

2.1.141 spreading function [audio]: A function that describes the frequency spread of masking.

2.1.142 start codes [system and video]: 32-bit codes embedded in that coded bitstream that are unique. They are used for several purposes including identifying some of the layers in the coding syntax.

2.1.143 STD input buffer [system]: A first-in first-out buffer at the input of the system target decoder for storage of compressed data from elementary streams before decoding.

2.1.144 stereo mode [audio]: Mode, where two audio channels which form a stereo pair (left and right) are encoded within one bitstream. The coding process is the same as for the dual channel mode.

2.1.145 stuffing (bits); stuffing (bytes) : Code-words that may be inserted into the compressed bitstream that are discarded in the decoding process. Their purpose is to increase the bitrate of the stream.

2.1.146 subband [audio]: Subdivision of the audio frequency band.

2.1.147 subband filterbank [audio]: A set of band filters covering the entire audio frequency range. In ISO/IEC 11172-3 the subband filterbank is a polyphase filterbank.

2.1.148 subband samples [audio]: The subband filterbank within the audio encoder creates a filtered and subsampled representation of the input audio stream. The filtered samples are called subband samples. From 384 time-consecutive input audio samples, 12 time-consecutive subband samples are generated within each of the 32 subbands.

2.1.149 syncword [audio]: A 12-bit code embedded in the audio bitstream that identifies the start of a frame.

2.1.150 synthesis filterbank [audio]: Filterbank in the decoder that reconstructs a PCM audio signal from subband samples.

2.1.151 system header [system]: The system header is a data structure defined in this part of ISO/IEC 11172 that carries information summarising the system characteristics of the ISO/IEC 11172 multiplexed stream.

2.1.152 system target decoder; STD [system]: A hypothetical reference model of a decoding process used to describe the semantics of an ISO/IEC 11172 multiplexed bitstream.

2.1.153 test procedure [compliance]: a method to verify compliance of a bitstream or a decoder.

2.1.154 time-stamp [system]: A term that indicates the time of an event.

2.1.155 triplet [audio]: A set of 3 consecutive subband samples from one subband. A triplet from each of the 32 subbands forms a granule.

2.1.156 tonal component [audio]: A sinusoid-like component of an audio signal.

2.1.157 variable bitrate: Operation where the bitrate varies with time during the decoding of a compressed bitstream.

2.1.158 variable length coding; VLC: A reversible procedure for coding that assigns shorter code-words to frequent events and longer code-words to less frequent events.

2.1.159 video buffering verifier; VBV [video]: A hypothetical decoder that is conceptually connected to the output of the encoder. Its purpose is to provide a constraint on the variability of the data rate that an encoder or editing process may produce.

2.1.160 video sequence [video]: A series of one or more groups of pictures. It is one of the layers of the coding syntax defined in ISO/IEC 11172-2.

2.1.161 zig-zag scanning order [video]: A specific sequential ordering of the DCT coefficients from (approximately) the lowest spatial frequency to the highest.

2.2 Symbols and abbreviations

The mathematical operators used to describe this International Standard are similar to those used in the C programming language. However, integer division with truncation and rounding are specifically defined. The bitwise operators are defined assuming twos-complement representation of integers. Numbering and counting loops generally begin from zero.

2.2.1 Arithmetic operators

+	Addition.
-	Subtraction (as a binary operator) or negation (as a unary operator).
++	Increment.
- -	Decrement.
*	Multiplication.
^	Power.
/	Integer division with truncation of the result toward zero. For example, 7/4 and -7/-4 are truncated to 1 and -7/4 and 7/-4 are truncated to -1.
//	Integer division with rounding to the nearest integer. Half-integer values are rounded away from zero unless otherwise specified. For example 3//2 is rounded to 2, and -3//2 is rounded to -2.
DIV	Integer division with truncation of the result towards-∞.
| |	Absolute value. $\lvert x \rvert = x$ when $x > 0$; $\lvert x \rvert = 0$ when $x == 0$; $\lvert x \rvert = -x$ when $x < 0$
%	Modulus operator. Defined only for positive numbers.
Sign( )	Sign(x) = 1 x > 0; 0 x == 0; -1 x < 0
NINT ( )	Nearest integer operator. Returns the nearest integer value to the real-valued argument. Half-integer values are rounded away from zero.
sin	Sine.
cos	Cosine.
exp	Exponential.
$\sqrt{}$	Square root.
$\log_{10}$	Logarithm to base ten.
$\log_e$	Logarithm to base e.
$\log_2$	Logarithm to base 2.

2.2.2 Logical operators

||	Logical OR.
&&	Logical AND.

! Logical NOT.

2.2.3 Relational operators

> Greater than.

>= Greater than or equal to.

< Less than.

<= Less than or equal to.

== Equal to.

!= Not equal to.

max [,...,] the maximum value in the argument list.

min [,...,] the minimum value in the argument list.

2.2.4 Bitwise operators

A twos complement number representation is assumed where the bitwise operators are used.

& AND.

| OR.

>> Shift right with sign extension.

<< Shift left with zero fill.

2.2.5 Assignment

= Assignment operator.

2.2.6 Mnemonics

The following mnemonics are defined to describe the different data types used in the coded bit-stream.

bslbf Bit string, left bit first, where "left" is the order in which bit strings are written in ISO/IEC 11172. Bit strings are written as a string of 1s and 0s within single quote marks, e.g. '1000 0001'. Blanks within a bit string are for ease of reading and have no significance.

ch Channel. If ch has the value 0, the left channel of a stereo signal or the first of two independent signals is indicated. (Audio)

nch Number of channels; equal to 1 for single_channel mode, 2 in other modes. (Audio)

gr Granule of 3 * 32 subband samples in audio Layer II, 18 * 32 sub-band samples in audio Layer III. (Audio)

main_data The main_data portion of the bitstream contains the scalefactors, Huffman encoded data, and ancillary information. (Audio)

main_data_beg The location in the bitstream of the beginning of the main_data for the frame. The location is equal to the ending location of the previous frame's main_data plus one bit. It is calculated from the main_data_end value of the previous frame. (Audio)

part2_length The number of main_data bits used for scalefactors. (Audio)

rpchof	Remainder polynomial coefficients, highest order first. (Audio)
sb	Subband. (Audio)
sblimit	The number of the lowest sub-band for which no bits are allocated. (Audio)
scfsi	Scalefactor selection information. (Audio)
switch_point_l	Number of scalefactor band (long block scalefactor band) from which point on window switching is used. (Audio)
switch_point_s	Number of scalefactor band (short block scalefactor band) from which point on window switching is used. (Audio)
uimsbf	Unsigned integer, most significant bit first.
vlclbf	Variable length code, left bit first, where "left" refers to the order in which the VLC codes are written.
window	Number of the actual time slot in case of block_type==2, $0 \le \text{window} \le 2$. (Audio)

The byte order of multi-byte words is most significant byte first.

2.2.7 Constants

π	3,14159265358...
e	2,71828182845...

2.3 Bitstream characteristics

Bitstream characteristics specify the constraints that are applied by the encoder in generating the bitstream. These syntactic and semantic constraints may for example restrict the range or the values of parameters that are encoded directly or indirectly in the bitstream. The constraints applied to a given bitstream may or may not be known *a priori* .

2.3.1 System bitstreams

System encoders may apply restrictions to the following parameters of system bitstreams (see ISO/IEC 11172-1):

a) mux_rate
b) rate_bound
c) STD_buffer_size
d) STD_buffer_size_bound
e) delay caused by system target decoder input buffering
f) difference between two SCRs in successive packs
g) length of a pack
h) length of a packet
i) number of packets in a pack
j) presence of time stamps in packet headers (DTS, PTS)
k) CSPS_flag
l) use of private streams
m) packet rate
n) fixed or variable bitrate operation (fixed_flag parameter)
o) number of multiplexed audio streams (audio_bound parameter)
p) number of multiplexed video streams (video_bound parameter)
q) locking of audio sampling frequency and frequency of system clock (system_audio_lock_flag parameter)
r) locking of video picture rate and frequency of system clock (system_video_lock_flag parameter)

2.3.2 Video bitstreams

A requirement for MPEG video encoders is that the arithmetic precision in the decoder process used in the encoder to produce the coded bitstream shall have the full accuracy specified in ISO/IEC 11172-2.

Video encoders may apply restrictions to the following parameters of video bitstreams (see ISO/IEC 111722):

a) horizontal_size
b) vertical_size
c) pel_aspect_ratio
d) picture_rate
e) bit_rate
f) VBV_buffer_size
g) constrained_parameter_flag
h) forward_f_code
i) backward_f_code
j) total number of macroblocks.
k) The number of macroblocks per second.
l) range and accuracy (half or integer pixel) of motion vectors.
m) use of a non-default quantizer matrix for intra coded blocks
n) use of a non-default quantizer matrix for non-intra coded blocks
o) slice structure, that is the definition of where slices start and end within the picture.
p) IBPD structure. That is the picture coding types and sequences of different picture coding types, such as for example the number of consecutive B frames, may be restricted.
q) fixed and/or variable bitrate operation (encoded in the bit_rate field and in the vbv_delay field)

r) the occurrence and specification of user_data

2.3.3 Audio bitstreams

Audio encoders may apply restrictions to the following parameters of audio bitstreams (see ISO/IEC 11172-3):

a) layer
b) bitrate_index
c) sampling_frequency
d) mode
e) mode_extension
f) emphasis
g) generation of crc_check
h) value of fixed bitrate when coding in free format mode.
i) generation of ancillary data

2.4 Decoder characteristics

The characteristics of a decoder specify the properties and capabilities of the decoding process applied in the decoder. An example of a property is the arithmetic accuracy that is applied. The capabilities of a decoder specify which coded bitstreams the decoder can reconstruct, by defining the subset of the standard that may be exploited in decodable bitstreams. A bitstream can be decoded by a decoder if the characteristics of the coded bitstream are within the subset of the standard defined by the decoder capabilities. Compliance to ISO/IEC 11172 by a decoder requires that the capabilities of the decoder are specified. That is, each constraint on the subset of the standard that may be exploited in bitstreams decodable by the decoder shall be specified. Compliant decoders are required to decode all bitstreams that are compliant with the defined subset without relying on private data, user data, ancillary data or other information.

2.4.1 System decoders

An ISO/IEC 11172-1 decoder may support specific values only, or a specific range of the following parameters in system bitstreams. These parameters are encoded directly or indirectly in the bitstream.

a) mux_rate
b) STD_buffer_size
c) packet rate

Furthermore, a decoder may constrain the support of fixed and/or variable bitrate operation (see definition of the fixed_flag field in 2.4.4.2 of ISO/IEC 11172-1), and may require locking between the 90 kHz system clock and the audio sampling frequency and/or the video picture rate (see the system_audio_lock_flag and the system_video_lock_flag fields in 2.4.4.2 of ISO/IEC 11172-1). Decoders shall specify how private streams are handled.

2.4.2 Video decoders

An ISO/IEC 11172-2 video decoder may support specific values only, or a specific range, or a specific combination of values or ranges of the following parameters in video bitstreams. These parameters are encoded directly or indirectly in the bitstream.

a) horizontal_size
b) vertical_size
c) pel_aspect_ratio
d) picture_rate
e) bit_rate
f) VBV_buffer_size
g) The total number of macroblocks in a picture.
h) The total number of macroblocks per second.
i) range of motion vectors (encoded in the various fields related to motion vectors)

Furthermore, an ISO/IEC 11172-2 video decoder may constrain:

a) IBPD structure. That is the picture coding types and sequences of different picture coding types that are supported, such as for example the number of consecutive B frames, may be restricted.
b) the support of fixed and/or variable bitrate operation.

An ISO/IEC 11172-2 video decoder shall specify how sequence_error_codes and user data are handled.

For a video decoder to be compliant to ISO/IEC 11172-2 a statement of whether or not computation is carried out with the full accuracy specified in ISO/IEC 11172-2 is required. If the full arithmetic precision is not implemented, the accuracy shall be specified. In the case of the inverse DCT, compliance testing requires performing the tests described in IEEE standard P1180/D2, as indicated in annex A of ISO/IEC 11172-2, and a statement of the numerical results for peak error and mean square error.

2.4.2.1 Constrained parameter decoders

Constrained parameter bitstreams are defined to provide a strong guideline to industry on a widely useful operating point. An ISO/IEC 11172-2 video decoder may specify that all constrained parameter streams can be decoded. A constrained parameter stream has the following characteristics:

horizontal_size <= 768 pels
vertical_size <= 576 pels
((horizontal_size+15)/16)*((vertical_size+15)/16) <= 396
((horizontal_size+15)/16)*((vertical_size+15)/16)*picture_rate <= 396*25
picture_rate <= 30 picture/s
forward_f_code <= 4 (see 2.4.3.4 of ISO/IEC 11172-2)
backward_f_code <= 4 (see 2.4.3.4 of ISO/IEC 11172-2)
VBV buffer size less than or equal to 327 680 bits.
bitrate less than or equal to 1 856 000 bits/s.

2.4.3 Audio decoders

An ISO/IEC 11172-3 audio decoder may support only specific values, or a specific range, or a specific combination of values or ranges of the following parameters in audio bitstreams. These parameters are encoded directly or indirectly in the bitstream.

a) layer
b) bitrate_index
c) sampling_frequency
d) mode
e) mode_extension
f) emphasis

Furthermore, an ISO/IEC 11172-3 audio decoder may constrain the support of free format mode. For an ISO/IEC 11172-3 audio decoder the handling of ancillary data and error protection (crc_check) shall be specified, as shall the single channel performance (single channel output at one or at both output channels).

Compliance of an audio decoder to ISO/IEC 11172-3 requires that the output signal of the decoder is reconstructed accurately. For actual tests see 2.6.3.

2.5 Procedures to test bitstream compliance

A bitstream is compliant if the bitstream meets each syntactic and semantic bitstream requirement specified in the normative clauses of parts 1, 2 and 3 of ISO/IEC 11172. Any ISO/IEC 11172 compliant bitstream is entitled to be an ISO/IEC 11172 bitstream regardless of the subset of the standard exploited in the bitstream. In this clause, guidelines are described to test bitstream compliance. From these guidelines explicit tests can be derived. This clause also includes guidelines on testing whether a bitstream meets specific bitstream characteristics. Such tests can be used to verify whether a bitstream can be decoded by a decoder with matching capabilities or whether the bitstream meets some specific application requirements.

To test a bitstream for compliance, compliance with all of the syntactic requirements of ISO/IEC 11172 shall be tested. Such a syntax test can be constructed directly from the syntactic specification contained in parts 1, 2 and 3 of ISO/IEC 11172. Therefore no explicit guidelines are given to test the syntax of the bitstream. One of the syntactic requirements is that syntactic elements reserved for future use shall not occur in the bitstream.

Semantic tests can likewise be derived directly from the requirements in parts 1, 2 and 3 of ISO/IEC 11172. One of the semantic requirements is that parameter fields shall not contain values reserved for future use by ISO/IEC. This clause contains guidelines for specific semantic tests, including timing requirements and consistency among coded parameters and values in a bitstream. The guidelines are given for parts 1, 2 and 3 of ISO/IEC 11172 in separate sub-clauses. Compliance with the timing requirements on the bitstream can be tested by examining the bitstream itself, without dependency on the behaviour of a delivery mechanism. There is a semantic constraint on the relative timing encoded in the complete bitstream, which is in addition to the explicit requirements of parts 1, 2 and 3 of ISO/IEC 11172. This requirement is detailed in 2.5.1.4.

2.5.1 System bitstream tests

2.5.1.1 Tests on the pack level

system_clock_reference: in successive packs, the system_clock_reference field contains encoded values which are samples of a nominal 90 kHz system clock. The maximum interval between system_clock_reference fields is limited by the difference between encoded values in successive packs; this difference shall not exceed 0,7*90 000, as specified in 2.4.5.2 of ISO/IEC 11172-1.

mux_rate (1): the value encoded in the mux_rate field shall be sufficently large that, if all bytes in the pack are transmitted at that rate, they are delivered to the system target decoder before the time the first byte of the subsequent pack is delivered. The time that the first byte of the subsequent pack is delivered may be derived from the system_clock_reference and mux_rate fields in that subsequent pack.

mux_rate (2): the mux_rate field shall not be encoded with the value zero.

packet rate: if the CSPS_flag in the system header is set to "1" and the rate specified in the mux_rate field is less than 5 Mbit/s then the rate at which packets arrive at the input of the system target decoder shall be less than 300 packets/s. If CSPS_flag is set to "1" and the rate specified in the mux_rate field is greater than 5 Mbit/s, the packet rate is bounded by a linear relation to the value encoded in the mux_rate field, as specified in 2.4.6 in ISO/IEC 11172-1.

length of a pack: the length of a pack may be determined by counting the bytes between successive pack start codes.

length of a packet: the length of a packet may be determined by counting the bytes between successive packet start codes.

2.5.1.2 Tests on the system header

general test (1): the first pack of an ISO/IEC 11172 stream shall contain a system header.

general test (2): a system header, if present in a pack, shall immediately follow the pack header.

general test (3): if an ISO/IEC 11172 stream contains more than one system header, then the values encoded in all the headers shall be identical.

header_length: the header_length field shall be encoded with a value equal to the number of bytes in the system header that follow the header_length field.

rate_bound: the rate_bound field shall denote a bitrate which is greater than or equal to the maximum bitrate value encoded in any mux_rate field in the same ISO/IEC 11172 stream.

audio_bound (1): the value encoded in the audio_bound field shall be greater than or equal to the maximum number of simultaneously active ISO/IEC 11172-3 audio streams in the ISO/IEC 11172 stream. For the purpose of this clause, an ISO/IEC 11172-3 audio stream is active if :

a) the input buffer of the system target decoder of that ISO/IEC 11172-3 audio stream is not empty, or if
b) one of the presentation units decoded from that ISO/IEC 11172-3 audio stream is being presented.

audio_bound (2): the value encoded in the audio_bound field shall be less than or equal to 32.

fixed_flag: if the fixed_flag is set to "1", then the values encoded in all system_clock_ reference fields shall satisfy 2.4.4.2 in ISO/IEC 11172-1.

CSPS_flag: if the CSPS_flag is set to "1", then the packet rate and the system target decoder buffer size shall satisfy 2.4.6 in ISO/IEC 11172-1.

system_audio_lock_flag: if the system_audio_lock_flag is set to "1", then the difference between the values encoded in any two presentation_time_stamp fields in audio packets of the same ISO/IEC 11172-3 audio stream shall correspond to the duration of the decoded audio access units in that ISO/IEC 11172-3 audio stream. For this purpose the duration (in terms of units of the system clock frequency) shall be derived from the number of samples and the sampling frequency of the decoded access units and the ratio SCASR as specified in 2.4.4.2 of ISO/IEC 11172-1. This assumes that no discontinuities occurred in the ISO/IEC 11172-3 audio stream in the presentation of the access units during the presentation period defined by both fields. See 2.4.5.4 of ISO/IEC 11172-1 for the definition of discontinuities.

system_video_lock_flag: if the system_video_lock_flag is set to "1", then the difference between the values encoded in any two presentation_time_stamp fields in video packets of the same ISO/IEC 11172-2 video stream shall correspond to the duration of the decoded pictures in that ISO/IEC 11172-2 video stream. For this purpose the duration (in terms of units of the system clock frequency) shall be derived from the picture rate of the decoded pictures and the ratio SCPR as specified in 2.4.4.2 of ISO/IEC 11172-1. This assumes that no discontinuities occurred in the ISO/IEC 11172-2 video stream in the presentation of the access units during the presentation period defined by both fields. See 2.4.5.4 of ISO/IEC 11172-1 for the definition of discontinuities.

video_bound (1): the value encoded in the video_bound field shall be greater than or equal to the maximum number of simultaneously active ISO/IEC 11172-2 video streams in the ISO/IEC 11172 stream. For the purpose of this clause, an ISO/IEC 11172-2 video stream is active if:

a) the input buffer of the system target decoder of that ISO/IEC 11172-2 video stream is not empty, or if
b) the reorder buffer of the system target decoder of that ISO/IEC 11172-2 video stream is not empty, or if
c) one of the presentation units decoded from that ISO/IEC 11172-2 video stream is being presented.

video_bound (2): the value encoded in the audio_bound field shall be less than or equal to 16.

stream_id (1): the value encoded in the stream_id field shall be one of the values permitted by table 1 (stream_id) in 2.4.4.2 in ISO/IEC 11172-1.

stream_id (2): the stream_id mechanism refers exactly once to each elementary stream in the multiplex.

STD_buffer_bound_scale: if the stream_id shall refer to an ISO/IEC 11172-3 audio stream, the STD_buffer_bound_scale shall be set to "0". If the stream_id refers to an ISO/IEC 11172-2 video stream, the STD_buffer_bound_scale shall be set to "1".

STD_buffer_size_bound: in the STD_buffer_size_bound field a value shall be encoded greater than or equal to the maximum value encoded in any of the STD_buffer_size fields in packets of the same elementary stream.

2.5.1.3 Tests on the packet level

stream_id: the value encoded in the stream_id field shall be one of the values permitted by table 1 (stream_id) in 2.4.4.2 in ISO/IEC 11172-1.

header_length: the header_length field shall be encoded with a value equal to the number of bytes in the packet that follow the header_length field.

stuffing bytes: a packet header shall contain no more than sixteen stuffing bytes.

STD_buffer_scale: if the stream_id refers to an ISO/IEC 11172-3 audio stream, the STD_buffer_scale shall be set to "0". If the stream_id refers to an ISO/IEC 11172-2 video stream, the STD_buffer_scale shall be set to "1",

STD_buffer_size (1): the STD_buffer_size for each elementary stream shall be encoded in the first packet of each elementary stream.

STD_buffer_size (2): if the CSPS_flag is set to "1" in the system header, and if the stream_id refers to an ISO/IEC 11172-3 audio stream, then the value encoded in the STD_buffer_size field shall define a buffer size value less than or equal to 4 096 Bytes, as specified in 2.4.6 in ISO/IEC 11172-1.

STD_buffer_size (3): if the CSPS_flag is set to "1" in the system header, and if the stream_id refers to an ISO/IEC 11172-2 video stream, then the value encoded in the STD_buffer_size field shall satisfy 2.4.6 in ISO/IEC 11172-1.

presentation_time_stamp (1): the presentation_time_stamp field may be encoded in a packet containing ISO/IEC 11172-3 audio data only if an audio access unit (an audio frame) commences in that packet.

presentation_time_stamp (2): the presentation_time_stamp field may be encoded in a packet containing ISO/IEC 11172-2 video data only if the first byte of a video picture start code commences in that packet.

presentation_time_stamp (3): the difference between the encoded values in any two successive presentation_time_stamps fields in packets of the same elementary stream, shall satisfy 2.4.5.3 in ISO/IEC 11172-1.

presentation_time_stamp (4): the presentation_time_stamp field shall satisfy 2.4.5.4 in ISO/IEC 11172-1.

decoding_time_stamp: the encoding of the decoding_time_stamp field shall satisfy 2.4.5.4 in ISO/IEC 11172-1.

buffer management (1): the access unit containing a data byte from an ISO/IEC 11172-3 audio or ISO/IEC 11172-2 video packet shall be decoded within one second after the data byte is received by the system target decoder in the STD model.

buffer management (2): the STD buffer shall neither overflow nor underflow; see 2.5.1.5.

2.5.1.4 Tests on timing accuracy

The entire ISO/IEC 11172 stream shall be constructed so that, in addition to meeting all other requirements, when the values and semantics of the presentation_time_stamp fields are compared with the audio and video samples to which they refer and their nominal sample rates, the calculated tolerance betweeen the actual sample frequency and the system_clock_frequency shall not exceed 100 parts per million. The timing relationships can be verified in practice solely by examining the bitstream. For example, to verify that the relationship between the audio sample rate and the system clock frequency meets this constraint, the following test can be performed:

a) Determine the nominal audio sample frequency as specified in the audio bitstream.

b) Find a PTS (PTS_1) associated with audio and the sample it refers to.

c) Count the number of audio samples until any other PTS (PTS_2) is found that refers to audio in the same stream, as long as no discontinuities occur between these two points.

d) Find the value of that second PTS (PTS_2), and the difference between PTS_2 and PTS_1.

e) Calculate the effective audio sample rate by dividing the number of samples passed by the difference in time as indicated by the PTSs (PTS_2 minus PTS_1) times 1/90 000).

f) If the calculated sample rate differs from the nominal by more than 100 ppm (allowing for integer arithmetic rounding effects) then this bitstream is not compliant.

2.5.1.5 Buffer overflow/underflow test

ISO/IEC 11172-1 specifies the requirement for the bitstream to be decodable by a System Target Decoder (STD). This means that, using the hypothetical STD model, when all audio and video streams are decoded and presented with exact synchronisation as specified by the respective Presentation Time Stamps, the STD buffers never underflow nor overflow. In other words, all bytes required for decoding audio and video are present in the STD buffers when they are needed, and the STD buffers are never filled beyond the capacity of the STD buffer sizes specified. The STD model is defined and described in 2.4.2 of ISO/IEC 11172-1.

In an ISO/IEC 11172 multiplexed bitstream, the buffer control mechanism specified for ISO/IEC 11172-2 video, using the vbv_delay field and the VBV model, is not directly applicable. In this case it is superseded by the STD model. The STD model allows for direct decoding of elementary streams from the multiplexed stream without the need to reconstruct the elementary streams first. However, it is possible to reconstruct elementary streams from the multiplex. In the case that an ISO/IEC 11172-2 video stream is reconstructed from the multiplexed stream, the VBV model again becomes applicable. For buffer overflow/underflow tests in cases where the VBV model is applicable, see 2.5.2.3.

Underflow and overflow of the STD buffer may be tested by constructing an STD model decoder. The test need not be performed in real time. For example, to test the behaviour of the STD video buffer, a test may be performed as follows:

a) The STD test model has a 90kHz time base (the system_clock_frequency) which is used to increment a counter.

b) Begin parsing the ISO/IEC 11172 stream starting at the first byte.

c) When a pack header is encountered, the system clock reference (SCR) is extracted, converted to its 33 bit value, and loaded onto a counter variable called System Time Clock which is incremented by the 90kHz time base.

d) Also from the pack header, extract the mux_rate field.

e) Calculate the number of bytes per unit of increments of the 90kHz time base as an accurate ratio of mux_rate to 90kHz (not an integer ratio). Note that mux_rate is coded in units of 50 bytes/s.

f) Read data from the source into the input buffer of the system target decoder at the rate indicated by the ratio of mux_rate to 90 kHz, calculated in step e). Note that accuracy must be maintained over the course of the bitstream, and the average number of bytes per system clock frequency increment is in general not an integer.

g) When packets of the video stream under test are encountered, the STD_buffer_scale and STD_buffer_size are extracted and converted to the STD buffer size in bytes. This is subsequently used as the upper limit on the STD buffer fullness. Presentation_time_ stamps and decoding_time_stamps are extracted and stored temporarily.

h) All packet_data bytes from the video stream under test are put into a fifo buffer in the order that they are received. This buffer is at least as large as the STD_buffer_size as specified for this stream in the packet header.

i) When a PTS or DTS is extracted, the PTS or DTS is saved temporarily until a video picture_start_code is encountered, and then the PTS or DTS is associated with the coded picture that immediately follows that picture_start_code. Note that if only the PTS is coded, the value of the DTS equals the value of the PTS.

j) When the System Time Clock value equals the DTS value of the picture that has been in the STD buffer the longest, the data of that coded picture is removed instantaneously from the STD buffer as specified in 2.4.2 in ISO/IEC 11172-1.

k) For a coded picture which has no DTS associated with it, the value of the DTS is derived from:
   - the DTS of the last coded picture to which a PTS or DTS is associated; this coded picture is referred to as coded picture (n);
   - the number (k) of coded pictures between coded picture (n) and the current coded picture;
   - the picture rate as encoded in the picture_rate field in the video sequence_header for the coded pictures between coded picture (n) and the current picture;
   - the derived DTS value is the sum of the previous coded DTS, which may have been implied by a coded PTS value, and the product of the number of succeeding pictures (k) and the picture period in units of (1/90 000) seconds.

l) Compliance requires that the STD buffer neither overflows nor underflows.

2.5.2 Video bitstream tests

2.5.2.1 Tests on the sequence header

horizontal_size (1): the value encoded in the horizontal_size field shall be at least 1.

horizontal_size (2): if the constrained_parameters_flag is "1", then the horizontal_size field shall be less than or equal to 768.

horizontal_size (3): within a video sequence, all horizontal_size fields shall be encoded with the same value.

horizontal_size and vertical_size: if the constrained_parameters_flag is set to "1", then the total number of macroblocks shall be less than or equal to 396.

vertical_size (1): the value encoded in the vertical_size field shall be at least 1.

vertical_size (2): if the constrained_parameters_flag is "1", then the vertical_size field shall be less than or equal to 576.

vertical_size (3): within a video sequence, all vertical_size fields shall be encoded with the same value.

horizontal_size, vertical_size and picture_rate: if the constrained_parameters_flag is set to "1", then the total number of macroblocks coded per second shall be less than or equal to 396*25.

pel_aspect_ratio (1): the binary value encoded in the pel_aspect_ratio field shall be in the range from 0001 up to 1110.

pel_aspect_ratio (2): within a video sequence, all pel_aspect_ratio fields shall be encoded with the same value.

picture_rate (1): the binary value encoded in the picture_rate field shall be in the range from 0001 up to 1000.

picture_rate (2): if the constrained_parameters_flag is set to "1", then the binary value encoded in the picture_rate field shall be in the range from 0001 up to 0101.

picture_rate (3): within a video sequence, all picture_rate fields shall be encoded with the same value.

bit_rate (1): the value encoded in the bit_rate field shall not be equal to zero.

bit_rate (2): within a video sequence, all bit_rate fields shall be encoded with the same value.

bit_rate (3): if the constrained_parameters_flag is set to "1", then the value encoded in the bit_rate field shall be less than or equal to 1 856 000 bits/s.

bit_rate (4): if the constrained_parameters_flag is set to "1", then the hexadecimal value encoded in the bit_rate field shall not be equal to 3FFFF (variable bitrate operation).

vbv_buffer_size (1): within a video sequence, all vbv_buffer_size fields shall be encoded with the same value.

vbv_buffer_size (2): if the constrained_parameters_flag is set to "1", then the value encoded in the vbv_buffer_size field shall be less than or equal to 40*1 024 bytes.

constrained_parameter_flag: within a video sequence, all constrained_parameter_flag fields shall be encoded with the same value.

intra_quantizer_matrix (1): the intra_quantizer_matrix field shall contain no values of intra_quant[i][j] equal to zero.

intra_quantizer matrix (2): the value of intra_quant[0][0] shall be equal to 8.

non_intra_quantizer_matrix (1): the non_intra_quantizer_matrix field shall contain no values of non_intra_quant[i][j] equal to zero.

user_data: user data shall not contain a string of 23 or more zero bits.

2.5.2.2 Tests on the group of pictures layer

In this section, a number of parameters are used. The values of these parameters are obtained from the values encoded in the time_code fields at the Group of Pictures layer and in the picture_rate field in the Sequence Header as follows:

> M is the value encoded in the time_code_minutes field;
>
> S is the value encoded in the time_code_seconds field;
>
> P is the value encoded in the picture_rate field in the sequence header, rounded to the nearest integer number.

drop_frame_flag: the drop_frame_flag may only be set to "1", if the picture_rate field in the Sequence Header indicates a picture rate of 29,97 Hz.

time_code_hours: the value encoded in the time_code_hours field shall be in the range from zero up to 23.

time_code_minutes: the value encoded in the time_code_minutes field shall be in the range from zero up to 59.

time_code_seconds: the value encoded in the time_code_seconds field shall be in the range from zero up to 59.

time_code_pictures: the value encoded in the time_code_pictures field shall be in the range from F0 up to (P-1), where

F0 = 2 , if the drop_frame_flag is set to "1" and S = 0 and M % 10 $\neq$ 0
= 0 otherwise.

time_code: the time_code field shall be encoded in compliance with 2.4.3.3 in ISO/IEC 11172-2.

user_data: user data shall not contain a string of 23 or more zero bits.

2.5.2.3 Tests on the picture layer

temporal_reference (1): for the first picture, in display order, within a Group of Pictures the temporal_reference field shall be encoded with the value zero.

temporal_reference (2): the value encoded in the temporal_reference field shall be the display order number modulo 1 024 of the picture within the Group of Pictures.

picture_coding_type (1): the binary value encoded in the picture_coding_type field shall be in the range from 001 up to 100.

picture_coding_type (2): the first picture, in coded order, of a Group of Pictures shall be an I- or a D-picture (see also the test on picture_coding_type(6) below).

picture_coding_type (3): the last picture, in display order, of a Group of Pictures shall be an I-, P- or D-picture (see also the test on picture_coding_type(6) below).

picture_coding_type (4): if a Group of Pictures contains at least two pictures of type I or P, the last two of which have coding order numbers m and n, then all B-pictures with coding order numbers m+1, ..., n-1 shall also be included in the Group of Pictures.

picture_coding_type (5): a D-type picture shall not appear in the same video sequence as a picture of any other type.

picture_coding_type (6): Let m and n be the natural display order of any I- or P-pictures and j and k be the decoding order. If m>n, then j>k and vice-versa.

vbv_delay (1): if the bit_rate field in the Sequence Header indicates variable bitrate operation, then the vbv_delay field shall be encoded with the hexadecimal value FFFF.

vbv_delay (2): if the bit_rate field in the Sequence Header indicates constant bitrate operation, then the vbv_delay field shall be encoded such that neither overflow nor underflow occurs in the VBV buffer. Overflow and underflow do not occur, if the following conditions apply:

a) all bytes of picture (i) shall arrive before picture (i) is to be decoded.

$$N(i) / R \leq V(i) / 90\,000,$$

where

i indicates the number of the picture in coding order

N(i) is the number of bits after the final byte of the picture start code of picture (i) that are removed from the VBV buffer at decoding of picture (i) (see also ISO/IEC 11172-2, annex C)

R is the bitrate in bits/s, derived from the bit_rate field as encoded in the sequence header

V(i) is the value encoded in the vbv_delay field of picture (i) in units of the 90 kHz system clock

b) the bitrate at which all picture data are delivered shall be constant.

$N(i)/T(i) = R(i)$

where

i as specified in a)

N(i) as specified in a)

T(i) is the elapsed time in seconds between the arrival of the final bytes of the two successive picture start codes of pictures (i) and (i+1), as derived from :

$T(i) = (V(i)/90\ 000) + (1/P) - (V(i+1)/90\ 000)$,

where

V(i) is the value of the vbv_delay of picture (i), as encoded in the vbv_delay field of picture i in units of the 90 kHz system clock

P is the picture rate

R(i) is the bitrate in bits/s during delivery of the coded data of picture (i) applied by the encoder. The encoder used constant bitrate operation. The value of R(i) found from this formula should therefore be constant for each picture (i) within the accuracy constraints of the parameters. If underflow occurs for picture (i), a higher value of R(i) is found. R(i) rounded upwards equals R, the value encoded in the Sequence Header in units of 400 bits/s.

c) the number of bits in the VBV buffer immediately before picture (i) is decoded shall be less than the size of the VBV buffer.

$H(i) + V(i)*R(i)/90\ 000 \leq B(vbv)$,

where

i as specified in a)

H(i) is the number of bits of the picture start code and any preceding (header) data that is removed instantaneously at decoding of picture (i); see annex C of ISO/IEC 11172-2.

V(i) as specified in a)

R(i) as specified in b)

B(vbv) is the value of the vbv_buffer_size in bits

forward_f_code (1): the forward_f_code field shall not be encoded with the value zero

forward_f_code (2): if the constrained_parameters_flag is set to "1", then the forward_f_code field shall be encoded with a value in the range from 1 up to 4

backward_f_code (1): the backward_f_code field shall not be encoded with the value zero

backward_f_code (2): if the constrained_parameters_flag is set to "1", then the value encoded in the backward_f_code field shall be in the range from 1 up to 4

user_data: user shall not contain a string of 23 or more zero bits.

2.5.2.4 Tests on the slice layer

slice_vertical_position (1): the value encoded in the slice_vertical_position field shall be in the range from 1 to 175.

slice_vertical_position (2): the value encoded in the slice_vertical_position shall not be greater than the vertical picture size in units of 16 lines, rounded upwards. The vertical picture size is indicated in the Sequence Header.

quantizer_scale: the value encoded in the quantizer_scale field shall not be equal to zero.

2.5.2.5 Tests on the macroblock layer

macroblock_address_increment (1): the indicated macroblock shall be within the boundaries of the picture.

macroblock_address_increment (2): the first slice of the picture shall start with the first macroblock in the picture.

macroblock_address_increment (3): the first macroblock of a slice shall not be skipped.

macroblock_address_increment (4): the horizontal position of the first macroblock of a slice shall be less than or equal to the width of the picture in units of macroblocks.

macroblock_address_increment (5): the last macroblock of a slice shall not be skipped.

macroblock_address_increment (6): the last slice of the picture shall end with the last macroblock of the picture.

macroblock_address_increment (7): slices shall not overlap.

macroblock_address_increment (8): there shall be no gaps between slices in the same picture.

macroblock_address_increment (9): within I- and D-pictures skipped macroblocks shall not occur.

macroblock_type (1): in a B-picture, a skipped macroblock shall not follow an intra-coded macroblock.

macroblock_type (2): each macroblock shall be intra-coded at least once per every 132 times it is coded in a P-picture without an intervening I-picture.

macroblock_type (3): if the macroblock_type field indicates macroblock_motion_forward and in case the picture is a B-picture that is preceded by exactly one I-picture in the same Group of Pictures, then the closed_gop flag shall be set to "0".

macroblock_type (4): if the macroblock_type field indicates macroblock_motion_forward and in case the picture is a B-picture that is preceded by exactly one I-picture in the same Group of Pictures, and if that Group of Pictures is the first Group of Pictures in the sequence, then the broken_link flag shall be set to "1".

reconstruction of motion vectors: both right_little and down_little shall not be equal to forward_f*16 or backward_f*16, whichever is appropriate.

reconstructed motion vectors: each reconstructed motion vector shall refer to a macroblock that is fully within the boundaries of the picture.

2.5.2.6 Tests on the block layer

number of DCT coefficients: for each transformed block the block indices of the coefficients shall be in the range from zero to sixty-three. The block index defines the position of the coefficient in the array of quantized DCT coefficients in zig-zag scanning order (see 2.4.3.7 in ISO/IEC 11172-2).

2.5.3 Audio bitstream tests

2.5.3.1 General tests

ID: the ID flag shall be set to "1".

layer: the Layer field shall not be encoded with the binary value 00.

bitrate: the bitrate field shall not be encoded with the binary value 1111.

sampling_frequency: the sampling frequency field shall not be encoded with the binary value 11.

padding: padding shall be applied such that the accumulated length of the coded audio frames, after a certain number of audio frames, shall not deviate more than (+0,-1) slot from the value specified in 2.4.2.3 of ISO/IEC 11172-3. This shall apply only if the layer, the bitrate and the sampling frequency do not change in the course of the considered audio frames.

emphasis: the emphasis field shall not be encoded with the binary value 10.

protection: if the protection bit is set to "0", then the correct CRC16 value shall be in the crc_check field

2.5.3.2 Tests on Layer I

allocation: the allocation[sb] or allocation[ch][sb] field shall not be encoded with the binary value 1111.

scalefactor: the scalefactor[sb] or scalefactor[ch][sb] field shall not refer to index 63.

samples: for the coded representation of subband samples the valid range is from zero up to (nlevels -1), where nlevels equals the number of levels used for quantization of that sample, that is the coded representation of a sample shall not consist of a bitstring with only "1"s.

frame length (1): the bit allocation shall be such that the total number of bits for a frame does not exceed the frame length for Layer I.

frame length (2): for Layer 1 the frame length shall equal the number of slots times the slot size for Layer I.

2.5.3.3 Tests on Layer II

bitrate: only allowed combinations of bitrate and mode shall be encoded in the bitrate field.

scalefactor: the scalefactor[sb][p] or scalefactor[ch][sb][p] field shall not refer to index 63

samples: for un-grouped samples the coded representation of subband samples the valid range is from zero up to (nlevels -1), where nlevels equals the number of levels used for quantization of that sample, that is the coded representation of a sample shall not consist of a bitstring with only "1"s. For grouped samples the range shall be from zero up to 26 if nlevels equals 3, from zero up to 124 if nlevels equals 5, and from zero up to 728 if nlevels equals 9.

frame length (1): the bit allocation and the scalefactor select information shall be such that the total number of bits for a frame does not exceed the frame length for Layer II.

frame length (2): for Layer II the frame length shall equal the number of slots times the slot size for Layer II.

2.5.3.4 Tests on Layer III

part2_3_length: the value encoded in the part2_3_length[gr] or part2_3_length[gr][ch] field shall correspond to the total length of scalefactors and Huffman encoded data.

table_select: the table_select[region][gr] or table_select[region][gr][ch] fields shall be encoded correctly.

frame_length (1): the Huffman code data shall be such that the total number of bits for a frame does not exceed the frame length for Layer III.

frame length (2): for Layer III the frame length shall equal the number of slots times the slot size for Layer III.

buffer control: the value of main_data_begin shall comply with the buffer considerations specified in 2.4.3.4 of ISO/IEC 11172-3.

2.6 Procedures to test decoder compliance

This clause describes procedures to verify compliance of system, video and audio decoders. All tests are performed using error free bit streams. For correct interpretation of syntax and semantics, test sequences covering a wide range of parameters shall be supplied to the decoder under test and its output sequence shall be compared with the output of a reference decoder. The comparison can be done, for example, by performing subjective tests, by evaluation of the difference signal and by comparing the timing performance. Each compliant decoder shall be able to decode all compliant ISO/IEC 11172 streams within the subset of the standard defined by the specified capabilities of the decoder. The procedures to test decoder compliance are given for parts 1, 2 and 3 of ISO/IEC 11172 in separate sub-clauses.

2.6.1 System decoder tests

A system decoder shall ensure that the coded data buffers do not overflow or underflow. Specifically this requires implementation of a combination of effective synchronization using the timing information and sufficient buffers and compensation to allow for actual decoding performance and inaccuracies in the actual decoder synchronization. In this manner, direct specification of the decoder synchronization accuracy is replaced with a simple binary test. Decoders need to synchronize properly because if they do not they will tend to have buffer problems. In addition, a decoder should synchronize audio and video accurately, in terms of PTS values, to ensure subjective quality.

The decoder can assume that bitstreams meet the constraint that when the values and semantics of the PTS time stamps are compared with the audio and video samples to which they refer and their nominal sample rates, the calculated difference between actual sample frequency and the frequency of the system clock does not exceed 100 parts per million. To test decoder characteristics with regards to relative and absolute timing accuracy of the input bitstream, bitstreams can be used that meet various constraints, for example:

a) relative timing tolerance limited to 100 ppm;

b) system_audio_lock_flag and/or system_video_lock_flag are set to "1";

c) relative timing tolerance varies within tighter bounds than 100 ppm;

d) absolute data delivery timing, measured by the SCRs, varies within certain timing bounds; this is an application specification beyond the scope of ISO/IEC 11172, but it is suggested that manufacturers of decoders specify the timing tolerances the decoder can accept, and it is recommended that manufacturers of data sources specify the timing tolerance that they keep, and whether these tolerances are maintained with variable rate data delivery.

2.6.2 Video decoder tests

Testing of a video decoder requires testing of the reconstructed pictures. In particular accumulation of errors in the decoded pictures needs to be tested. Furthermore testing is needed of the timing of the decoding process with the associated buffer performance. When the timing of the decoding process is not synchronized properly with the coded video input stream, then the decoder will tend to have buffer underflow or overflow problems.

A method is as follows :

a) Select a representative set of video sequences for the application. The video material should be selected so that any decoding errors are made easily visible. It is recommended for example that the source video has low noise, that is a high SNR, and contains some large areas of uniform colour. Some of the video material should contain motion vectors near the limit set by the motion vector limit of the decoder characteristics. Some of the video material should contain abrupt transitions of VBV buffer occupancy, to test the decoder buffer behaviour as it approaches the almost-empty and almost-full states.

b) Encode the video material using a range of parameters to cover the decoder characteristics to be tested. Parameters should be tested at normal values as well as close to the maximum and minimum values permitted by the decoder characteristics. Horizontal and vertical picture sizes should be tested that are not multiples of 16. The bitstreams should contain a mix of I-, B- and P-pictures, as appropriate for the characteristics of the decoder under test. To test IDCT mismatch, bitstreams should contain series of 132 successive P-pictures.

c) Decode the video bitstream using a reference decoder. The reference decoder could for example apply the reference inverse DCT specified in the IEEE draft standard P1180/D2 (see annex A of ISO/IEC 11172-2). The reference decoder may be an accurate software implementation.

d) Perform a set of visual checks to compare the decoded video from step c) with the decoded pictures from the decoder under test. A compliant decoder exhibits no visible difference in the decoded pictures. The decoded pictures from the reference decoder should be given the same postprocessing as the pictures decoded by the decoder under test. Both decoded video sequences should then be displayed on identical displays under good viewing conditions.

e) Measure the pixel values of the decoded pictures from the decoder under test and compare them with the pixel values of the decoded pictures from the reference decoder. It may not be possible to perform this test with a production decoder. In that case this test should be performed by the manufacturer during the design and development phase.

On arithmetic precision, compliance of an ISO/IEC 11172-2 video decoder requires a statement of whether or not computation is carried out with the full accuracy specified in ISO/IEC 11172-2. If the computation is not fully accurate, the accuracy shall be specified. In the case of the IDCT, this includes performing the tests described in IEEE standard P1180/D2 as indicated in annex A of ISO/IEC 11172-2, and stating the numerical results for peak error and mean square error. To be called a limited accuracy ISO/IEC 11172-2 video decoder, the numerical results for peak error and mean square error for the decoder shall not be greater than exactly twice the values specified in the IEEE standard P1180/D2. To be ISO/IEC 11172-2 compliant, a video decoder shall meet the accuracy requirements for limited accuracy ISO/IEC 11172-2 video decoders. To be called an ISO/IEC 11172-2 video decoder, the peak error and mean square error of the decoder shall meet the full accuracy requirements from IEEE standard P1180/D2.

2.6.3 Audio decoder tests

An ISO/IEC 11172-3 compliant decoder that is able to support at least one but not all combinations of the options defined in 2.4.3 such as bit rates, sampling rates and modes, will be designated as an ISO/IEC 11172-3 Layer "N" audio decoder. Decoders that support all combinations are designated as Full Layer "N" ISO/IEC 11172-3 audio decoders, where "N" indicates I, II or III.

To test audio decoders, ISO/IEC JTC1 SC29/WG11 supplies a number of test sequences. Supplied sequences cover only Full Layer "N" Decoders and so not all decoders will be able to decode all the supplied test sequences. For a supplied test sequence, testing can be done by comparing the output of a decoder under test with a reference output also supplied by ISO/IEC JTC1 SC29/WG11. Measurements are carried out relative to full scale where the output signals of the decoders are normalized to be in the range between -1 and +1.

To be called an ISO/IEC 11172-3 audio decoder, the decoder shall provide an output such that the rms level of the difference signal between the output of the decoder under test and the supplied reference output is less than 2^{-15}/sqrt(12) for the supplied sine sweep (20Hz-10kHz) with an amplitude of -20dB relative to full scale. In addition, the difference signal shall have a maximum absolute value of at most 2^{-14} relative to full-scale.

To be called a limited accuracy ISO/IEC 11172-3 audio decoder, the decoder shall provide an output for a provided test sequence such that the rms level of the difference signal between the output of the decoder under test and the supplied reference output is less than 2^{-11} /(sqrt(12) for the supplied sine sweep (20Hz-10kHz) with an amplitude of -20dB relative to full scale.

The above two tests only verify the computational accuracy of an implementation.

Annex A

(normative)

Definition of audio decoder tests

A.1. Calculation for RMS

All measurements are carried out relative to full scale where the output signals of the decoder and supplied test sequences are normalized to be in the range between -1,0 and +1,0. The supplied sine sweep with an amplitude of -20dB relative to full scale has an absolute amplitude of +-0,1. This test sequence has a precision (P) of 24 bit. i.e. the MSB represents the value of -1, the MSB-1 bit represents the value of +1/2, etc.

MSB	-1/2^0	= -1
MSB-1	1/2^1	= 1/2
MSB-2	1/2^2	= 1/4
MSB-23	1/2^23	= 1/8 388 608.

The output signal of the decoder under test requires to be in the same format. In the case that the output of the decoder has a precision of P' bits and if P' is smaller than 24, then the values for the bits between the positions P'-1 and 24 shall be set to zero.

MSB	-1/2^0	= -1
MSB-1	1/2^1	= 1/2
...		
MSB-(P'-1)	1/2^P'	= 1/2^P'
MSB-P'	0	= 0
...		
MSB-23	0	= 0

In the next step the difference (diff) of the samples of these signals has to be calculated. If two channels are present (in case of stereo, joint-stereo and dual-channel) both channels shall be tested. The total number of samples for each channel is N.

diff(n) = 'output signal of decoder under test(n)' - 'supplied sine sweep(n)'
for n = 1 to N

The values of all difference samples shall be squared, summed, divided by N and then the square-root shall be calculated. This calculation finally gives the rms level.

rms := sqrt(1/N*sum(diff^2))

The decoder under test may be called an ISO/IEC 11172-3 audio decoder, if the rms is less than 1/(2^15 * 12^0,5) and if the maximum absolute value is less than or equal to 1/2^14.

The decoder under test may be called a limited accuracy ISO/IEC 11172-3 audio decoder, if the rms is less than 1/(2^11 * 12^0,5).

Annex B

(informative)

Descriptions of the ISO/IEC 11172 (MPEG) audio test bitstreams

The ISO/IEC 11172-3 (MPEG/Audio) Layer I and Layer II test bitreams have a different format from the Layer III bitstreams. For Layer I and II both compressed and decompressed bitstreams are provided. For Layer III, only compressed bitstreams are provided. A reference Layer III decoder is included so that the decompressed bitstreams can be generated. Detailed descriptions of the bitstreams are furnished below.

B.1 Layer I.

For the input, compressed-data, files, there are one slot (4 bytes in Layer I) per line. Each 4 bits are translated into one hex character, i.e. 8 characters per line.

For the output files, there are one output sample per line. The samples are represented with 24 bits accuracy (6 characters per line). For stereo output, the left and right output samples are interleaved, so the left samples are on the even lines (starting to count with line 0), the right samples are on the odd lines. File number 8 contains a signal that can be used to do the accuracy check as described in CD11172-4. The provided output signals are rounded to 24 bits.

The number of frames for the files FL1 through FL6, and FL8 is forty-nine. Forty-nine is chosen because the padding sequence in Layer I in case of 44,1 kHz repeats itself every forty-nine frames. This makes it possible to loop the data infinitely. The scalefactor test (FL7) is a special one because it tests all sixty-three scalefactors, one scalefactor per frame.

Filenames	FL1.MPG (compressed), FL1.PCM (decompressed)
Layer	I
Fs	32 kHz
bit rate	384 kbit/s
CRC	yes
mode	intensity stereo
emphasis	none
private bit	not set
copyright bit	not set
original bit	set
signal	-20 dB sweep in left and right
#frames	49 (18 816 samples/ch in decoded signal)

Filenames	FL2.MPG (compressed), FL2.PCM (decompressed)
Layer	I
Fs	44,1 kHz
bit rate	384 kbit/s
CRC	yes
mode	intensity stereo
emphasis	none
private bit	not set
copyright bit	not set
original bit	set
signal	-20 dB sweep in left and right
#frames	49 (18 816 samples/ch in decoded signal)

Filenames	FL3.MPG (compressed), FL3.PCM (decompressed)
Layer	I
Fs	48 kHz
bit rate	384 kbit/s
CRC	yes
mode	intensity stereo

emphasis	none
private bit	not set
copyright bit	not set
original bit	set
signal	-20 dB sweep in left and right
#frames	49 (18 816 samples/ch in decoded signal)

Filenames	FL4.MPG (compressed), FL4.PCM (decompressed)
Layer	I
Fs	32 kHz
bit rate	32 kbit/s
CRC	no
mode	single channel
emphasis	none
private bit	not set
copyright bit	not set
original bit	set
signal	-20 dB sweep
#frames	49 (18 816 samples/ch in decoded signal)

Filenames	FL5.MPG (compressed), FL5.PCM (decompressed)
Layer	I
Fs	48 kHz
bit rate	448 kbit/s
CRC	yes
mode	dual channel
emphasis	none
private bit	not set
copyright bit	not set
original bit	set
signal	-1 dB, 1 kHz sine wave in left, -1 dB, 2 kHz sine wave in right
#frames	49 (18 816 samples/ch in decoded signal)

Filenames	FL6.MPG (compressed), FL6.PCM (decompressed)
Layer	I
Fs	44,1 kHz
bit rate	384 kbit/s
CRC	yes
mode	stereo
emphasis	none
private bit	not set
copyright bit	not set
original bit	set
signal	noise in left and right, all subbands get all possible allocations
#frames	49 (18 816 samples/ch in decoded signal)

Filenames	FL7.MPG (compressed), FL7.PCM (decompressed)
Layer	I
Fs	44,1 kHz
bit rate	384 kbit/s
CRC	yes
mode	stereo
emphasis	none
private bit	not set
copyright bit	not set
original bit	set
signal	pecial, all scalefactors occur
#frames	63 (24 192 samples/ch in decoded signal)

Filenames	FL8.MPG (compressed), FL8.PCM (decompressed)
Layer	I

Fs	44,1 kHz
bit rate	384 kbit/s
CRC	no
mode	stereo
emphasis	none
private bit	not set
copyright bit	not set
original bit	set
signal	-20 dB sweep in left and right, from 0 to 10 kHz,
#frames	49 (18 816 samples/ch in decoded signal)

B.2 Layer II.

Filenames	FL_INP10 (compressed), FL_PCM10 (decompressed)
Layer	II
Fs	32 kHz
bit rate	192 kbit/s
CRC	yes
mode	stereo
emphasis	none
private bit	not set
copyright bit	not set
original bit	not set
signal	sine sweep
#frames	49 (56448 samples/ch in decoded signal)

Filenames	FL_INP11 (compressed), FL_PCM11 (decompressed)
Layer	II
Fs	44,1 kHz
bit rate	192 kbit/s
CRC	yes
mode	stereo
emphasis	none
private bit	not set
copyright bit	not set
original bit	not set
signal	sine sweep
#frames	49 (56448 samples/ch in decoded signal)

Filenames	FL_INP12 (compressed), FL_PCM12 (decompressed)
Layer	II
Fs	48 kHz
bit rate	192 kbit/s
CRC	yes
mode	stereo
emphasis	none
private bit	not set
copyright bit	not set
original bit	not set
signal	sine sweep
#frames	49 (56448 samples/ch in decoded signal)

Filenames	FL_INP13 (compressed), FL_PCM13 (decompressed)
Layer	II
Fs	32 kHz
bit rate	32 kbit/s
CRC	no
mode	mono
emphasis	none
private bit	not set
copyright bit	not set
original bit	not set

signal	sine sweep
#frames	49 (56448 samples/ch in decoded signal)

Filenames	FL_INP14 (compressed), FL_PCM14 (decompressed)
Layer	II
Fs	48 kHz
bit rate	384 kbit/s
CRC	yes
mode	dual channel
emphasis	none
private bit	not set
copyright bit	not set
original bit	not set
signal	1 kHz sinewave (left ch), 2 kHz sinewave (right ch)
#frames	16 (18432 samples/ch in decoded signal)

Filenames	FL_INP15 (compressed), FL_PCM15 (decompressed)
Layer	II
Fs	48 kHz
bit rate	384 kbit/s
CRC	yes
mode	stereo
emphasis	none
private bit	not set
copyright bit	not set
original bit	not set
signal	noise
#frames	16 (18432 samples/ch in decoded signal)

Filenames	FL_INP16 (compressed), FL_PCM16 (decompressed)
Layer	II
Fs	48 kHz
bit rate	256 kbit/s
CRC	yes
mode	stereo
emphasis	none
private bit	not set
copyright bit	not set
original bit	not set
signal	noise
#frames	16 (18432 samples/ch in decoded signal)

B.3 Layer III.

The package consists of:

a) Bitstreams which are generated synthetically (i.e. only a bitstream encoder was used) to test as many bitstream parameters as possible. Therefore decoding these bitstreams will give strange audio outputs..

hecommon.bit	tests less necessary header bits
he_32khz.bit	all bitrates at 32 kHz
he_44khz.bit	all bitrates at 44,1 kHz (padding included)
he_48khz.bi	all bitrates at 48 kHz
he_free.bit	120 kbit 44,1 kHz free format bitstream
he_mode.bit	tests mono/stereo/MS-stereo/intensity-stereo
si.bit	tests main_data_begin, part2_3_length, bigvalues, scalefactors
si_block.bit	tests the block-modes (long,start,short/mixed,stop)
si_huff.bit	tests the huffman side information and the whole huffman code

b) A bitstream which should produce an audible output:

sin1k0db.bit	A 1kHz 0dB sinusoid. The Decoder should generate full scaled output with this bitstream.

c) The newest version of the 'l3dec' simulation decoder. It should be possible to decode all bitstreams, including 'he_free.bit', with this decoder. Using the command line switches of 'l3dec' you can watch the decoding process at several points. This package includes the SUN-executable 'l3dec' and the PC-executable 'l3dec.exe'.

d) The compliance accuracy test bitstream (see part 4 of ISO/IEC 11172-3):

compl.bit	the test bitstream (10Hz-10kHz/-20dB sine sweep) mono, 48 kHz
compl.hex	reference output in 24bit ascii hexadecimal format (mono)

B.3.1 Detailed Bitstream Descriptions

Filename hecommon.bit (compressed)
Layer III
Fs 44,1 kHz
bit rate 128 kbit/s
CRC see signal description
mode stereo
emphasis see signal description
private bit see signal description
copyright bit see signal description
original bit see signal description
scalefac_compress = 0
blocksplit_flag = 0
three bigvalue regions, count1 bits
#frames 29 (33408 samples/ch in decoded signal)
signal tonal

Frame	Action
1-5	crc_check = 1 (= no crc_check)
6	crc_check = 0 (= switched on (for all following blocks))
7	private_bit = 1
8	private_bit = 0
	copyright = 1
9	copyright = 0
	original_home = 1
10	original_home = 0
	emphasis = 1
11	emphasis = 2
12	emphasis = 3
13-29	emphasis = 0

Filename he_32khz.bit (compressed)
Layer III
Fs 32 kHz
bit rate see signal description
CRC none
mode mono
emphasis none
private bit not set
copyright bit not set
original bit not set
scalefac_compress = 0
blocksplit_flag = 0
three bigvalue regions, count1 bits
#frames 149 (171648 samples/ch in decoded signal)
signal tonal

Frame	Action
1-10	bitrate = 32 kbit
11-20	bitrate = 40 kbit

21-30	bitrate = 48 kbit
31-40	bitrate = 56 kbit
41-50	bitrate = 64 kbit
51-60	bitrate = 80 kbit
61-70	bitrate = 96 kbit
71-80	bitrate = 112 kbit
81-90	bitrate = 128 kbit
91-100	bitrate = 160 kbit
101-110	bitrate = 192 kbit
111-120	bitrate = 224 kbit
121-130	bitrate = 256 kbit
131-149.	bitrate = 320 kbit

Filename he_44khz.bit (compressed)
Layer III
Fs 44,1 kHz
bit rate see signal description
CRC none
mode mono
emphasis none
private bit not set
copyright bit not set
original bit not set
scalefac_compress = 0
blocksplit_flag = 0
three bigvalue regions, count1 bits
#frames 409 (471168 samples/ch in decoded signal)
signal tonal, each bitrate test lasts 30 frames to test the padding function

Frame	Action
1-30	bitrate = 32 kbit
31-60	bitrate = 40 kbit
61-90	bitrate = 48 kbit
91-120	bitrate = 56 kbit
121-150	bitrate = 64 kbit
151-180	bitrate = 80 kbit
181-210	bitrate = 96 kbit
211-240	bitrate = 112 kbit
241-270	bitrate = 128 kbit
271-300	bitrate = 160 kbit
301-330	bitrate = 192 kbit
331-360	bitrate = 224 kbit
361-390	bitrate = 256 kbit
391-409	bitrate = 320 kbit

Filename he_48khz.bit (compressed)
Layer III
Fs 48 kHz
bit rate see signal description
CRC none
mode mono
emphasis none
private bit not set
copyright bit not set
original bit not set
scalefac_compress = 0
blocksplit_flag = 0
three bigvalue regions, count1 bits
#frames 149 (171648 samples/ch in decoded signal)
signal tonal

Frame	Action
1-10	bitrate = 32 kbit
11-20	bitrate = 40 kbit

21-30	bitrate = 48 kbit
31-40	bitrate = 56 kbit
41-50	bitrate = 64 kbit
51-60	bitrate = 80 kbit
61-70	bitrate = 96 kbit
71-80	bitrate = 112 kbit
81-90	bitrate = 128 kbit
91-100	bitrate = 160 kbit
101-110	bitrate = 192 kbit
111-120	bitrate = 224 kbit
121-130	bitrate = 256 kbit
131-149	bitrate = 320 kbit

Filename he_free.bit (compressed)
Layer III
Fs 44,1 kHz
bit rate free format, 120 kbit/s
CRC none
mode stereo
emphasis none
private bit not set
copyright bit not set
original bit not set
scalefac_compress = 0
blocksplit_flag = 0
three bigvalue regions, count1 bits
signal tonal
Frame Action
1-end free format bitstream 120 kbit

Filename he_mode.bit (compressed)
Layer III
Fs 44,1 kHz
bit rate 128 kbit/s
CRC none
mode see signal description
emphasis none
private bit not set
copyright bit not set
original bit not set
scalefac_compress = 0
blocksplit_flag = 0 (unless otherwise noted below)
three bigvalue regions(two, if window_switching_flag == 1), count1 bits
#frames 127 (146304 samples/ch in decoded signal)
signal tonal

Frame	Action
1 - 10	single channel mode
11 - 20	dual channel mode
21 - 30	stereo mode
31 - 40	joint stereo mode - mode extension: normal stereo
41 - 50	joint stereo mode - mode extension: MS stereo only
51 - 60	joint stereo mode - mode extension: intensity only
61 - 70	joint stereo mode - mode extension: intensity & MS stereo
71	Gr 0: intensity only Gr 1: start block
72 - 80	short blocks & intensity only
81 - 89	short blocks & intensity & MS stereo
90	Gr 1: stop_block
91	Gr 0: long_block intensity only Gr 1: start_block (mixed blocks) intensity only
92 - 100	mixed blocks intensity only
101 - 109	mixed blocks intensity & MS stereo
110	Gr 1: stop_block (mixed blocks)

111 - 127	single channel mode

Filename	si.bit (compressed)
Layer	III
Fs	44,1 kHz
bit rate	64 kbit/s
CRC	none
mode	single channel
emphasis	none
private bit	not set
copyright bit	not set
original bit	not set

scalefac_compress = 0 (unless otherwise noted below) blocksplit_flag = 0

#frames	117 (134784 samples/ch in decoded signal)
signal	tonal, test of side information values

Frame Action

1 - 6 main_data_begin = 0

6 - 21 main_data_begin increases up to the limit 511 22 main_data_begin = 0 suddenly

23 - 27 main_data_begin increases slowly

27 - 31 main_data_begin increases fast up to the limit 511

32 part2_3_length of GR 0 is maximal

33 - 36 part2_3_length changes

37 big_values is maximal

38 - 42 global_gain = 255 with only one spectral line set to '1', the corresponding scalefactor maximal (=15) and scalefac_scale=1

43 - 47 global_gain = 200, scalefactor = 0

48 - 52 global_gain = 100, only one spectral value set to '1000'

53 - 57 global_gain = 50, only one spectral value set to '8191'

58 - 62 global_gain = 0, dto.

64	GR 0: scalefactors:		slen1 = 0	slen2 = 1	
	GR 1: scalefactors:		slen1 = 0	slen2 = 2	
65	GR 0:	dto	0	3	
	GR 1:	dto	3	0	
66	GR 0:	dto	1	1	
	GR 1:	dto	1	2	
67	GR 0:	dto	1	3	
	GR 1:	dto	2	1	
68	GR 0:	dto	2	2	
	GR 1:	dto	2	3	
69	GR 0:	dto	3	1	
	GR 1:	dto	3	2	
70	GR 0:	dto	3	3	
	GR 1:	dto	4	2	
71	GR 0:	dto	4	3	
	GR 1:	dto	4	3	
72	GR 0: scalefactors:		slen1 = 0	slen2 = 1	scfsi = 0001
	GR 1: scalefactors:		slen1 = 0	slen2 = 2	
73	GR 0:	dto	0	3	scfsi = 0010
	GR 1:	dto	3	0	
74	GR 0:	dto	1	1	scfsi = 0100
	GR 1:	dto	1	2	
75	GR 0:	dto	1	3	scfsi = 1000
	GR 1:	dto	2	1	
76	GR 0:	dto	2	2	scfsi = 1001
	GR 1:	dto	2	3	
77	GR 0:	dto	3	1	scfsi = 1011
	GR 1:	dto	3	2	
78	GR 0:	dto	3	3	scfsi = 1111
	GR 1:	dto	4	2	
79	GR 0:	dto	4	3	scfsi = 0000
	GR 1:	dto	4	3	

80 scalefactor_scale = 1

81	scalefactor_scale = 0	preflag = 1
82-117	preflag = 0	

Filename	si_block.bit (compressed)
Layer	III
Fs	44,1 kHz
bit rate	64 kbit/s
CRC	none
mode	single channel
emphasis	none
private bit	not set
copyright bit	not set
original bit	not set

scalefac_compress = 0 (unless noted otherwise below)
three bigvalue regions(two if window_switching_flag==1), count1 bits (unless noted otherwise below)

#frames	63 (72576 samples/ch in decoded signal)
signal	tonal, test of the block modes.

Frame	Action
1 - 4	long blocks
5	GR 0: long block GR 1: start block
6	GR 0: short blocks GR 1: stop block
7	GR 0: long block GR 1: start block
8	GR 0: short blocks GR 1: stop block
9	GR 0: long block GR 1: start block
10	short blocks GR 0: subblock_gain = 0,1,2 GR 1: subblock_gain = 3,4,5
11	short blocks GR 0: subblock_gain = 6,7,0 GR 1: subblock_gain = 7,7,7
12 - 13	short blocks: several scalefactor & scalefac_scale combinations
14	short blocks: scalefactors = 0
15	GR 0: short blocks GR 1: stop block
16	long blocks
17	GR 0: long block GR 1: start block (mixed type)
18	GR 0: mixed blocks GR 1: stop block (mixed type)
19	GR 0: long block GR 1: start block (mixed type)
20	GR 0: mixed blocks GR 1: stop block (mixed type)
21	GR 0: long block GR 1: start block (mixed type)
22	mixed blocks GR 0: subblock_gain = 0,1,2 GR 1: subblock_gain = 3,4,5
23	mixed blocks GR 0: subblock_gain = 6,7,0 GR 1: subblock_gain = 7,7,7
24 - 25	mixed blocks: several scalefactor & scalefac_scale combinations
26	mixed blocks: scalefactors = 0
27	GR 0: mixed blocks GR 1: stop block (mixed type)
28-63	long block

Filename	si_huf.bit (compressed)
Layer	III
Fs	44,1 kHz
bit rate	64 kbit/s

CRC	none
mode	single channel
emphasis	none
private bit	not set
copyright bit	not set
original bit	not set
scalefac_compress	= 0
blocksplit_flag	= 0
#frames	74 (85248 samples/ch in decoded signal)
signal	tonal, test of huffman side info and the huffman code.
Frame	Action
2	test of every huffcode pair table (two values for each table)
	GR 0: table no 0,1,2
	GR 1: 3,5
3	GR 0: 6,7,8
	GR 1: 9,10,11
4	GR 0: 12,13,15
	GR 1: 16,17,18
5	GR 0: 19,20,21
	GR 1: 22,23,24
6	GR 0: 25,26,27
	GR 1: 28,29,30
7	GR 0: 31, 0, 1
	GR 1: 2, 3, 5
8	test of every huffcode count1 table
	GR 0: table no 0
	GR 1: table no 1
9	Test of region_adresses and pairs in regions:
	big_values = 0 (= 0 pairs, 0 regions)
	region pairs0,1,2 region0,1_count
	GR 0: 0, 0, 0 0, 0
	big_values = 1 (= 1 pair, 1 region)
	region pairs0,1,2 region0,1_count
	GR 1: 1, 0, 0 0, 0
10	big_values = 3 (= 3 pairs, 2 regions)
	region pairs0,1,2 region0,1_count
	GR 0: 2, 1, 0 0, 0
	GR 1: 3, 0, 0 1, 0
11	big_values = 5 (= 5 pairs, 3 regions)
	region pairs0,1,2 region0,1_count
	GR 0: 2, 2, 1 0, 0
	GR 1: 4, 1, 0 1, 0
12	GR 0: 5, 0, 0 2, 0
	GR 1: 2, 3, 0 0, 1
13	big_values = 7 (= 7 pairs, 4 regions)
	region pairs0,1,2 region0,1_count
	GR 0: 2, 2, 3 0, 0
	GR 1: 2, 2, 3 0, 0
14	GR 0: 2, 4, 1 0, 1
	GR 1: 2, 5, 0 0, 2
15	GR 0: 4, 3, 0 1, 1
	GR 1: 6, 1, 0 2, 0
16	GR 0: 7, 0, 0 3, 0
	GR 1: 4, 2, 1 1, 0
17	big_values = 13 (= 13 pairs, 7 regions)
	region pairs0,1,2 region0,1_count
	GR 0: 13, 0, 0 6, 0
	GR 1: 2, 11, 0 0, 5
18	GR 0: 2, 2, 9 0, 0
	GR 1: 8, 5, 0 3, 2
19	GR 0: 8, 2, 3 3, 0
	GR 1: 2, 6, 5 0, 2
20	GR 0: 4, 4, 5 1, 1

	GR 1: 4, 4, 5 1, 1
21	big_values = 288 (= 288 pairs = whole spec., 22 regions)
	region pairs0,1,2 region0,1_count
	GR 0: 81,207, 0 15, 5
	GR 1: 81, 17,190 15, 0
22	GR 0: 55,233, 0 13, 7
	GR 1: 18, 49,221 7, 6
23	GR 0: table 0 (no huffcode values)
	GR 1: table 1 all huffcode values
24	GR 0: table 2 all huffcode values
	GR 1: table 3 all huffcode values
25	GR 0: table 5 all huffcode values
	GR 1: table 6 all huffcode values
26	GR 0: table 7 all huffcode values
	GR 1: table 8 all huffcode values
27	GR 0: table 9 all huffcode values
	GR 1: table 10 all huffcode values
28	GR 0: table 11 all huffcode values
	GR 1: table 12 all huffcode values
29 - 30	table 13 all huffcode values
31 - 32	table 15 all huffcode values
33 - 34	table 16 all huffcode values
35	GR 0: table 20 (== table 16 + more linbits) several linbit values
	GR 1: table 23 (== table 16 + more linbits) several linbit values
36 - 37	table 24 all huffcode values
38	GR 0: table 28 (== table 24 + more linbits) several linbit values
	GR 1: table 31 (== table 24 + more linbits) several linbit values
39-74	GR 0: count1 table 0 all values
	GR 1: count1 table 1 all values

This page intentionally left blank

ICS 35.040.00

Descriptors: data processing, moving pictures, video data, audio data, video recording, data storage, digital storage, coded representation, coding (data conversion), digital encoders, conformity tests.

Price based on 42 pages