



Recommendation ITU-R BS.1196-6
(12/2017)

Audio coding for digital broadcasting

BS Series
Broadcasting service (sound)



International
Telecommunication
Union
Documents
Document Sharing Hub

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Annex 1 of Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from <http://www.itu.int/ITU-R/go/patents/en> where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Recommendations

(Also available online at <http://www.itu.int/publ/R-REC/en>)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	Broadcasting service (television)
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management
SNG	Satellite news gathering
TF	Time signals and frequency standards emissions
V	Vocabulary and related subjects

Note: This ITU-R Recommendation was approved in English under the procedure detailed in Resolution ITU-R 1.

Electronic Publication
Geneva, 2017

© ITU 2017

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without written permission of ITU.

RECOMMENDATION ITU-R BS.1196-6*

Audio coding for digital broadcasting

(Question ITU-R 19-1/6)

(1995-2001-2010-2012-02/2015-10/2015-2017)

Scope

This Recommendation specifies audio source coding systems applicable for digital sound and television broadcasting. It further specifies a system applicable for the backward compatible multichannel enhancement of digital sound and television broadcasting systems.

Keywords

Audio, Audio Coding, Broadcast, Digital, Broadcasting, Sound, Television, Codec

The ITU Radiocommunication Assembly,

considering

- a) that user requirements for audio coding systems for digital broadcasting are specified in Recommendation ITU-R BS.1548;
- b) that multi-channel sound system with and without accompanying picture is the subject of Recommendation ITU-R BS.775 and that a high-quality, multi-channel sound system using efficient bit rate reduction is essential in a digital broadcasting system;
- c) that the advanced sound system specified in Recommendation ITU-R BS.2051 consists of three-dimensional channel configurations and uses either static or dynamic metadata to control object-based, scene-based, and channel-based signals;
- d) that subjective assessment of audio systems with small impairments, including multi-channel sound systems is the subject of Recommendation ITU-R BS.1116;
- e) that subjective assessment of audio systems of intermediate audio quality is subject of Recommendation ITU-R BS.1534 (MUSHRA);
- f) that low bit-rate coding for high quality audio has been tested by the ITU Radiocommunication Sector;
- g) that commonality in audio source coding methods among different services may provide increased system flexibility and lower receiver costs;
- h) that many broadcast services already use or have specified the use of audio codecs from the families of MPEG-1, MPEG-2, MPEG-4, AC-3 and E-AC-3;
- i) that Recommendation ITU-R BS.1548 lists codecs that have been shown to meet the broadcaster's requirements for contribution, distribution and emission;
- j) that those broadcasters which have not yet started services should be able to choose the system which is best suited to their application;
- k) that broadcasters may need to consider compatibility with legacy broadcasting systems and equipment when selecting a system;

* This Recommendation should be brought to the attention of the International Standardization Organization (ISO) and the International Electrotechnical Commission (IEC).

- l) that when introducing a multi-channel sound system existing mono and stereo receivers should be considered;
- m) that a backward compatible multi-channel extension to an existing audio coding system can provide better bit rate efficiency than simulcast;
- n) that an audio coding system should preferably be able to encode both speech and music with equally high fidelity,

recommends

1 that for new applications of digital sound or television broadcasting emission, where compatibility with legacy transmissions and equipment is not required, one of the following low bit-rate audio coding systems should be employed:

- Extended HE AAC as specified in ISO/IEC 23003-3:2012;
- E-AC-3 as specified in ETSI TS 102 366 (2014-08);
- AC-4 as specified in ETSI TS 103 190-1 v1.1.1 (2015-06) and ETSI TS 103 190-2 v1.2.1 (2015-09);
- MPEG-H 3D Audio LC Profile as specified in ISO/IEC 23008-3:2015/Amd 3:2017.

NOTE 1 – Extended HE AAC is a more flexible superset of MPEG-4 HE AAC v2, HE AAC and AAC LC, and includes MPEG-D Unified Speech and Audio Coding (USAC).

NOTE 2 – E-AC-3 is a more flexible superset of AC-3.

NOTE 3 – The AC-4 and MPEG-H 3D Audio LC profile specifications include capabilities that are able to support the advanced sound system specified in Recommendation ITU-R BS.2051 and users should refer to Recommendation ITU-R BS.1548 for codec compliance;

2 that for applications of digital sound or television broadcasting emission, where compatibility with legacy transmissions and equipment is required, one of the following low bit-rate coding systems should be employed:

- MPEG-1 Layer II as specified in ISO/IEC 11172-3:1993;
- MPEG-2 Layer II half sample rate as specified in ISO/IEC 13818-3:1998;
- MPEG-2 AAC-LC or MPEG-2 AAC-LC with SBR as specified in ISO/IEC 13818-7:2006;
- MPEG-4 AAC-LC as specified in ISO/IEC 14496-3:2009;
- MPEG-4 HE AAC v2 as specified in ISO/IEC 14496-3:2009;
- AC-3 as specified in ETSI TS 102 366 (2014-08);

NOTE 4 – ISO/IEC 11172-3 may sometimes be referred to as 13818-3 as this specification includes 11172-3 by reference.

NOTE 5 – It is encouraged to support Extended HE AAC as specified in ISO/IEC 23003-3:2012. It includes all of the above mentioned AAC versions, thus guaranteeing compatibility with new future as well as legacy broadcast systems worldwide with the same single decoder implementation;

3 that for backward compatible multi-channel extension of digital television and sound broadcasting systems, the multichannel audio extensions described in ISO/IEC 23003-1:2007 should be used;

NOTE 6 – Since the MPEG Surround technology described in ISO/IEC 23003-1:2007 is independent of the compression technology (core coder) used for transmission of the backward compatible signal, the described multi-channel enhancement tools can be used in combination with any of the coding systems recommended under *recommends* 1 and 2;

- 4 that for distribution and contribution links, ISO/IEC 11172-3 Layer II coding may be used at a bit rate of at least 180 kbit/s per audio signal (i.e. per mono signal, or per component of an independently coded stereo signal) excluding ancillary data;
- 5 that for commentary links, ISO/IEC 11172-3 Layer III coding may be used at a bit rate of at least 60 kbit/s excluding ancillary data for mono signals, and at least 120 kbit/s excluding ancillary data for stereo signals, using joint stereo coding;
- 6 that for high quality applications the sampling frequency should be 48 kHz;
- 7 that the input signal to the low bit rate audio encoder should be emphasis-free and no emphasis should be applied by the encoder,

further recommends

- 1 that Recommendation ITU-R BS.1548 should be referred to for information about coding system configurations that have been demonstrated to meet quality and other user requirements for contribution, distribution, and emission;
- 2 that further studies of the requirements for the advanced sound system specified in Recommendation ITU-R BS.2051 are needed and that this Recommendation should be updated when these studies are completed.

NOTE – Information about the codecs included in this Recommendation may be found in Annexes 1 to 8.

Annex 1 (informative)

MPEG-1 and MPEG-2, layer II and III audio

1 Encoding

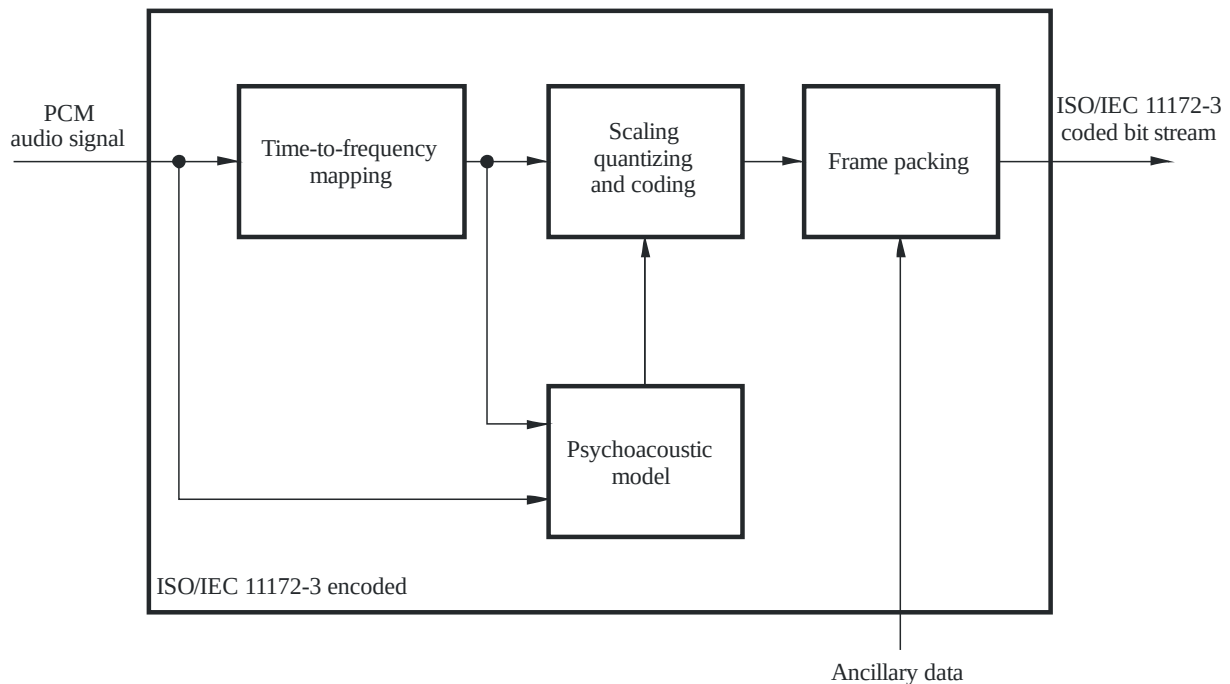
The encoder processes the digital audio signal and produces the compressed bit stream. The encoder algorithm is not standardized and may use various means for encoding, such as estimation of the auditory masking threshold, quantization, and scaling (following Note 1). However, the encoder output must be such that a decoder conforming to this Recommendation will produce an audio signal suitable for the intended application.

NOTE 1 – An encoder complying with the description given in Annexes C and D to ISO/IEC 11172-3, 1993 will give a satisfactory minimum standard of performance.

The following description is of a typical encoder, as shown in Fig. 1. Input audio samples are fed into the encoder. The time-to-frequency mapping creates a filtered and sub-sampled representation of the input audio stream. The mapped samples may be either sub-band samples (as in Layer I or II, see below) or transformed sub-band samples (as in Layer III). A psycho-acoustic model, using a fast Fourier transform, operating in parallel with the time-to-frequency mapping of the audio signal creates a set of data to control the quantizing and coding. These data are different depending on the actual coder implementation. One possibility is to use an estimation of the masking threshold to control the quantizer. The scaling, quantizing and coding block creates a set of coded symbols from the mapped input samples. Again, the transfer function of this block can depend on the implementation of the encoding system. The block “frame packing” assembles the actual bit stream

for the chosen layer from the output data of the other blocks (e.g. bit allocation data, scale factors, coded sub-band samples) and adds other information in the ancillary data field (e.g. error protection), if necessary.

FIGURE 1
Block diagram of a typical encoder



BS.1196-01

2 Layers

Depending on the application, different layers of the coding system with increasing complexity and performance can be used.

Layer I: This layer contains the basic mapping of the digital audio input into 32 sub-bands, fixed segmentation to format the data into blocks, a psycho-acoustic model to determine the adaptive bit allocation, and quantization using block companding and formatting. One Layer I frame represents 384 samples per channel.

Layer II: This layer provides additional coding of bit allocation, scale factors, and samples. One Layer II frame represents $3 \times 384 = 1\,152$ samples per channel.

Layer III: This layer introduces increased frequency resolution based on a hybrid filter bank (a 32 sub-band filter bank with variable length modified discrete cosine transform). It adds a non-uniform quantizer, adaptive segmentation, and entropy coding of the quantized values. One Layer III frame represents 1 152 samples per channel.

There are four different modes possible for any of the layers:

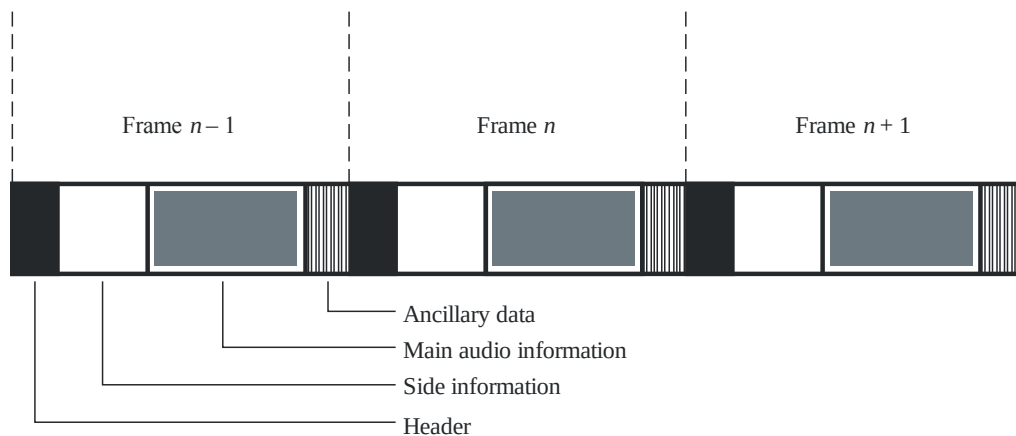
- single channel;
- dual channel (two independent audio signals coded within one bit stream, e.g. bilingual application);
- stereo (left and right signals of a stereo pair coded within one bit stream);

- joint stereo (left and right signals of a stereo pair coded within one bit stream with the stereo irrelevancy and redundancy exploited). The joint stereo mode can be used to increase the audio quality at low bit rates and/or to reduce the bit rate for stereophonic signals.

3 Coded bit stream format

An overview of the ISO/IEC 11172-3 bit stream is given in Fig. 2 for Layer II and Fig. 3 for Layer III. A coded bit stream consists of consecutive frames. Depending on the layer, a frame includes the following fields:

FIGURE 2
ISO/IEC 11172-3 Layer II bit stream format

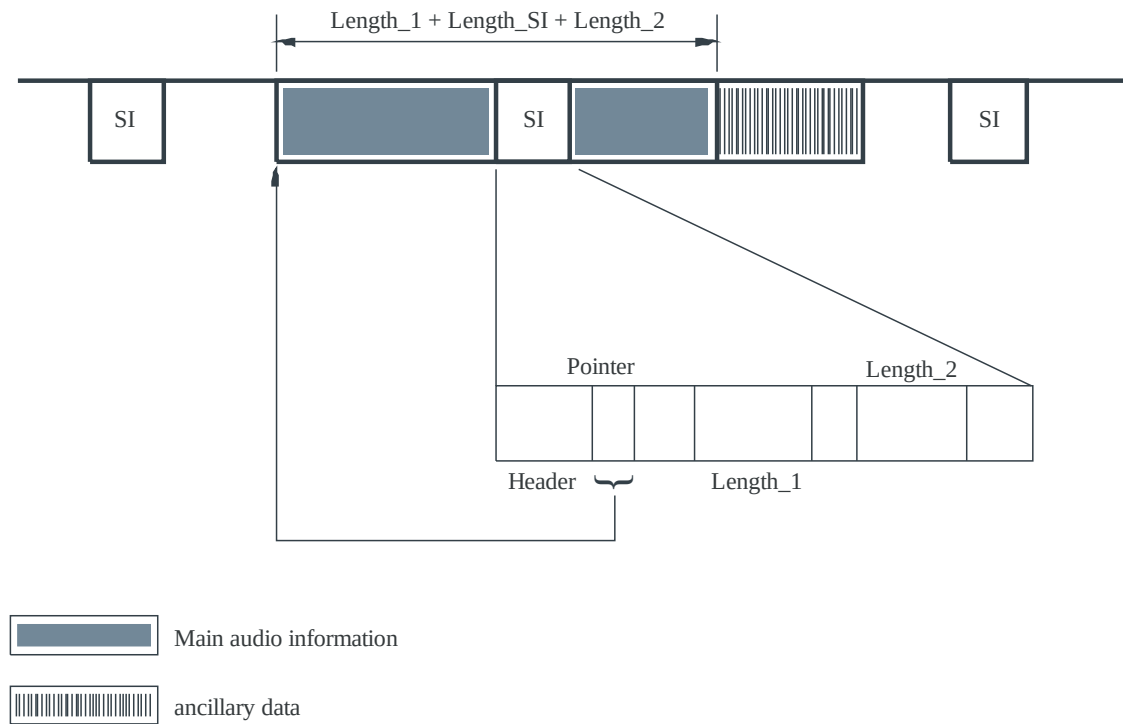


Layer II:

Header:	part of the bit stream containing synchronization and status information
Side information:	part of the bit stream containing bit allocation and scale factor information
Main audio information:	part of the bit stream containing encoded sub-band samples
Ancillary data:	part of the bit stream containing user definable data

BS.1196-02

FIGURE 3
ISO/IEC 11172-3 Layer III bit stream format



Layer III:

Side information (SI):	part of the bit stream containing header, pointer, length_1 and length_2, scale factor information, etc.;
Header:	part of the bit stream containing synchronization and status information;
Pointer:	pointing to beginning of main audio information;
Length_1:	length of first part of main audio information;
Length_2:	length of second part of main audio information;
Main audio information:	part of the bit stream containing encoded audio;
Ancillary data:	part of the bit stream containing user definable data.

BS.1196-03

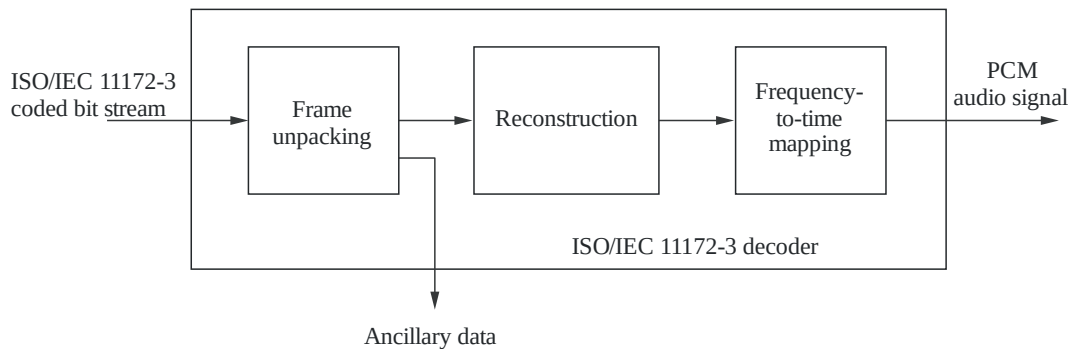
4 Decoding

The decoder accepts coded audio bit streams in the syntax defined in ISO/IEC 11172-3, decodes the data elements, and uses the information to produce digital audio output.

The coded audio bit stream is fed into the decoder. The bit stream unpacking and decoding process optionally performs error detection if error-check is applied in the encoder. The bit stream is unpacked to recover the various pieces of information, such as audio frame header, bit allocation, scale factors, mapped samples, and, optionally, ancillary data. The reconstruction process reconstructs the quantized version of the set of mapped samples. The frequency-to-time mapping transforms these mapped samples back into linear PCM audio samples.

FIGURE 4

Block diagram of the decoder



BS.1196-04

Annex 2 (informative)

MPEG-2 and MPEG-4 AAC audio

1 Introduction

ISO/IEC 13818-7 describes the MPEG-2 audio non-backwards compatible standards called MPEG-2 Advanced Audio Coding (AAC), a higher quality multichannel standard than achievable while requiring MPEG-1 backwards compatibility.

The AAC system consists of three profiles in order to allow a trade-off between the required memory and processing power, and audio quality:

– *Main profile*

Main profile provides the highest audio quality at any given data rate. All tools except the gain control may be used to provide high audio quality. The required memory and processing power are higher than the LC profile. A main profile decoder can decode an LC-profile encoded bit stream.

– *Low complexity (LC) profile*

The required processing power and memory of the LC profile are smaller than the main profile, while the quality performance keeps high. The LC profile is without predictor and the gain control tool, but with temporal noise shaping (TNS) order limited.

– *Scalable sampling rate (SSR) profile*

The SSR profile can provide a frequency scalable signal with gain control tool. It can choose frequency bands to decode, so the decoder requires less hardware. To decode the only lowest frequency band at the 48 kHz sampling frequency, for instance, the decoder can reproduce 6 kHz bandwidth audio signal with minimum decoding complexity.

AAC system supports 12 types of sampling frequencies ranging from 8 to 96 kHz, as shown in Table 1, and up to 48 audio channels. Table 2 shows default channel configurations, which include mono, two-channel, five-channel (three front/two rear channels) and five-channel plus low-frequency effects (LFE) channel (bandwidth < 200 Hz), etc. In addition to the default configurations, it is possible to specify the number of loudspeakers at each position (front, side, and back), allowing

flexible multichannel loudspeaker arrangement. Down-mix capability is also supported. The user can designate a coefficient to down-mix multichannel audio signals into two-channel. Sound quality can therefore be controlled using a playback device with only two channels.

TABLE 1
Supported sampling frequencies

Sampling frequency (Hz)
96 000
88 200
64 000
48 000
44 100
32 000
24 000
22 050
16 000
12 000
11 025
8 000

TABLE 2
Default channel configurations (NOTE 1)

Value ^{*1}	Number of speakers	Audio syntactic elements, listed in order received	Default element to speaker mapping (Note 2)	Channel name specified in Recommendation ITU-R BS.775 or BS.2051 (Note 3)
1	1	single_channel_element	M+000	Mono
2	2	channel_pair_element	M+030, M-030	Left, right
3	3	single_channel_element()	M+000	Centre
		channel_pair_element()	M+030, M-030	Left, right
4	4	single_channel_element()	M+000	Centre
		channel_pair_element()	M+030, M-030	Left, right
		single_channel_element()	M+180	Mono surround
5	5	single_channel_element()	M+000	Centre
		channel_pair_element()	M+030, M-030	Left, right
		channel_pair_element()	M+110, M-110	Left surround, right surround

TABLE 2 (cont)

Value ^{*1}	Number of speakers	Audio syntactic elements, listed in order received	Default element to speaker mapping (Note 2)	Channel name specified in Recommendation ITU-R BS.775 or BS.2051 (Note 3)
6	5 + 1	single_channel_element()	M+000	Centre
		channel_pair_element()	M+030, M-030	Left, right
		channel_pair_element()	M+110, M-110	Left surround, right surround
		lfe_element()	LFE1	Low frequency effects
7	7 + 1 Front	single_channel_element()	M+000	N/A ^{*2}
		channel_pair_element()	M+030, M-030	
		channel_pair_element()	M+045, M-045	
		channel_pair_element()	M+110, M-110	
		lfe_element()	LFE1	
8-10	-	—	reserved	-
11	6 + 1	single_channel_element()	M+000	N/A
		channel_pair_element()	M+030, M-030	
		channel_pair_element()	M+110, M-110	
		single_channel_element()	M+180	
		lfe_element()	LFE1	
12	7 + 1 Back	single_channel_element()	M+000	Centre
		channel_pair_element()	M+030, M-030	Left, right
		channel_pair_element()	M+090, M-090	Left side surround, right side surround
		channel_pair_element()	M+135, M-135	Left rear surround, right rear surround
		lfe_element()	LFE1	Low frequency effect
13	22 + 2	single_channel_element()	M+000	Front centre
		channel_pair_element()	M+030, M-030	Front left centre, front right centre
		channel_pair_element()	M+060, M-060	Front left, front right
		channel_pair_element()	M+090, M-090	Side left, side right
		channel_pair_element()	M+135, M-135	Back left, back right
		single_channel_element()	M+180	Back centre
		lfe_element()	LFE1	Low frequency effects-1
		lfe_element()	LFE2	Low frequency effects-2
		single_channel_element()	U+000	Top front centre
		channel_pair_element()	U+045, U-045	Top front left, top front right
		channel_pair_element()	U+090, U-090	Top side left, top side right
		single_channel_element()	T+000	Top centre
		channel_pair_element()	U+135, U-135	Top back left, top back right
		single_channel_element()	U+180	Top back centre
		single_channel_element()	B+000	Bottom front centre
		channel_pair_element()	B+045, U-045	Bottom front left, bottom front right

TABLE 2 (*end*)

Value ^{*1}	Number of speakers	Audio syntactic elements, listed in order received	Default element to speaker mapping (Note 2)	Channel name specified in Recommendation ITU-R BS.775 or BS.2051 (Note 3)
14	7 + 1 Top	single_channel_element()	M+000	Centre
		channel_pair_element()	M+030, M-030	Left, right
		channel_pair_element()	M+110, M-110	Left surround, right surround
		lfe_element()	LFE1	Low frequency effects
		channel_pair_element()	U+030, U-030	Left top front, right top front
15	—	—	reserved	—

^{*1} The audio output channel configuration is indicated by a four-bit field that carries a channel configuration value as defined in ISO/IEC 23001-8:2013, “Coding Independent Code Points”. MPEG-2 is applicable to channel configuration values of up to 7. MPEG-4 AAC is applicable to channel configuration values of up to 15.

^{*2} N/A: not applicable; channel configuration not available in Recommendation ITU-R BS.2051 and Recommendation ITU-R BS.775.

NOTE 1 – The list is quoted from Table 1.19 of ISO/IEC 14496-3:2009/Amd.4:2013.

NOTE 2 – Identification of the speakers by labels according to Recommendation ITU-R BS.2051.

NOTE 3 – Please note that the channel labels and names depend on the actual channel configuration.

2 Encoding

The basic structure of the MPEG-2 AAC encoder is shown in Fig. 5. The AAC system consists of the following coding tools:

- Gain control: A gain control splits the input signal into four equally spaced frequency bands. The gain control is used for SSR profile.
- Filter bank: A filter bank modified discrete cosine transform (MDCT) decomposes the input signal into sub-sampled spectral components with frequency resolution of 23 Hz and time resolution of 21.3 ms (128 spectral components) or with frequency resolution of 187 Hz and time resolution of 2.6 ms (1 024 spectral components) at 48 kHz sampling. The window shape is selected between two alternative window shapes.
- Temporal noise shaping (TNS): After the analysis filter bank, TNS operation is performed. The TNS technique permits the encoder to have control over the temporal fine structure of the quantization noise.
- Mid/side (M/S) stereo coding and intensity stereo coding: For multichannel audio signals, intensity stereo coding and M/S stereo coding may be applied. In intensity stereo coding only the energy envelope is transmitted to reduce the transmitted directional information. In M/S stereo coding, the normalized sum (M as in middle) and difference signals (S as in side) may be transmitted instead of transmitting the original left and right signals.
- Prediction: To reduce the redundancy for stationary signals, the time-domain prediction between sub-sampled spectral components of subsequent frames is performed.
- Quantization and noiseless coding: In the quantization tool, a non-uniform quantizer is used with a step size of 1.5 dB. Huffman coding is applied for quantized spectrum, the different scale factors, and directional information.