International Telecommunication Union

ITU-R

Radiocommunication Sector of ITU

Recommendation ITU-R BS.1196-2

(03/2010)

Audio coding for digital broadcasting

BS Series

Broadcasting service (sound)

International Telecommunication Union

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, 2010

RECOMMENDATION ITU-R BS.1196-2*, **

Audio coding for digital broadcasting

(Question ITU-R 19/6)

(1995-2001-2010)

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.

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 subjective assessment of audio systems with small impairments, including multi-channel sound systems is the subject of Recommendation ITU-R BS.1116;

d) that subjective assessment of audio systems of intermediate audio quality is subject of Recommendation ITU-R BS.1534 (MUSHRA);

e) that low bit-rate coding for high quality audio has been tested by the ITU Radiocommunication Sector;

f) that commonality in audio source coding methods among different services may provide increased system flexibility and lower receiver costs;

g) that several 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;

h) 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;

l) that when introducing a multi-channel sound system existing mono and stereo receivers should be considered;

---

* Radiocommunication Study Group 6 made editorial amendments to this Recommendation in 2003 in accordance with Resolution ITU-R 44.

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

m) that a backward compatible multi-channel extension to an existing audio coding system can provide better bit rate efficiency than simulcast,

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:

– MPEG-4 HE AAC v2 as specified in ISO/IEC 14496-3:2009;

– E-AC-3 as specified in ETSI TS 102 366 (2008-08);

NOTE 1 – MPEG-4 HE AAC v2 and E-AC-3 are more flexible supersets of MPEG-4 AAC-LC and AC-3.

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 (2008-08);

NOTE 1 – ISO/IEC 11172-3 may sometimes be referred to as 13818-3 as this specification includes 11172-3 by reference.

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 1 – 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;

8 that compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g. interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words shall in no way be construed to imply partial or total compliance with this Recommendation, *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.

NOTE 1 – Information about the codecs included in this Recommendation may be found in Appendices 1 to 4.

Appendix 1

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

1 Encoding

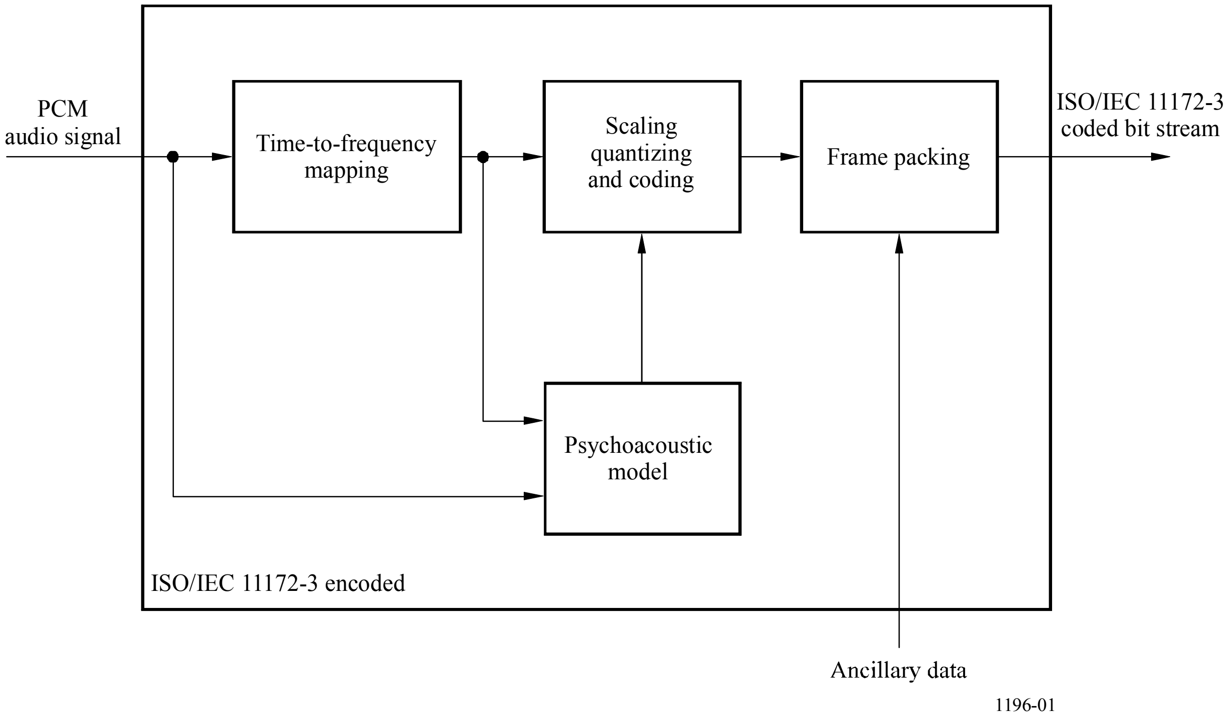
The encoder processes the digital audio signal and produces the compressed bit stream. The encoder algorithm is not standardized and may use various means for encoding, such as estimation of the auditory masking threshold, quantization, and scaling (following Note 1). However, the encoder output must 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



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);

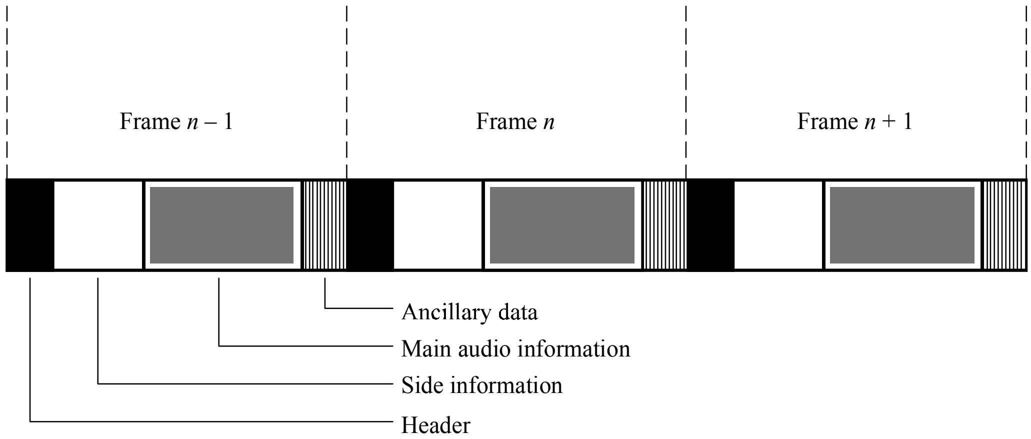
– joint stereo (left and right signals of a stereo pair coded within one bit stream with the stereo irrelevancy and redundancy exploited). The joint stereo mode can be used to increase the audio quality at low bit rates and/or to reduce the bit rate for stereophonic signals.

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



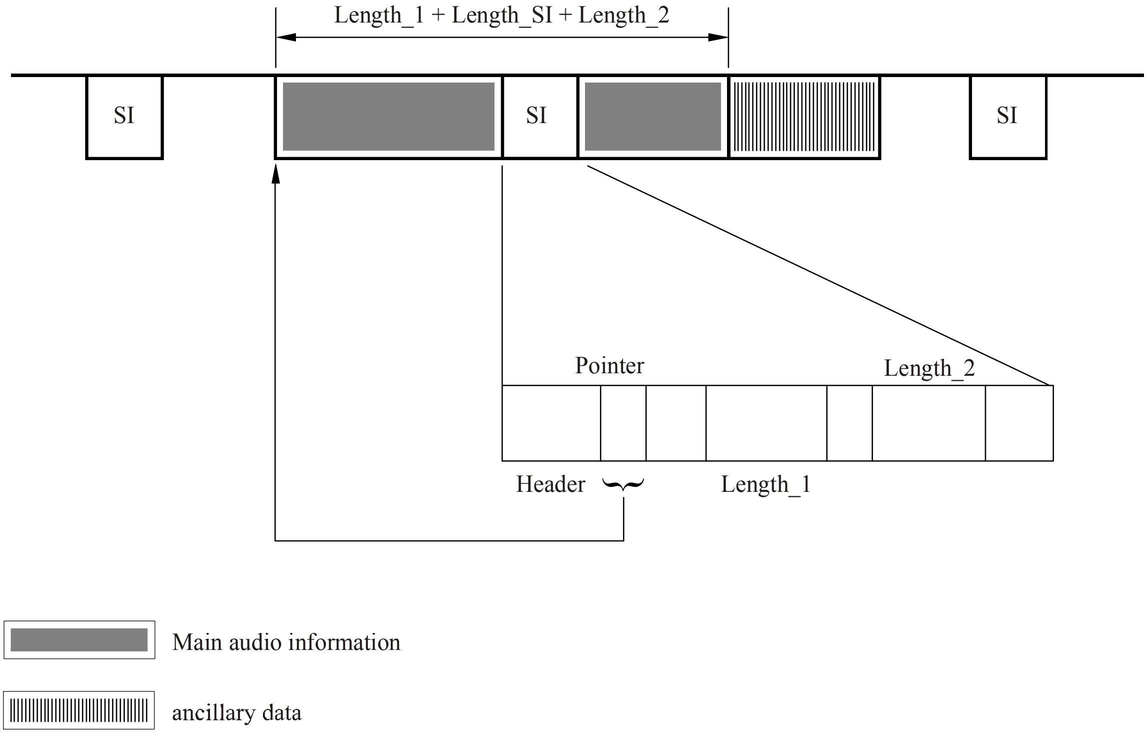
Layer II:

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

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.

1196-03

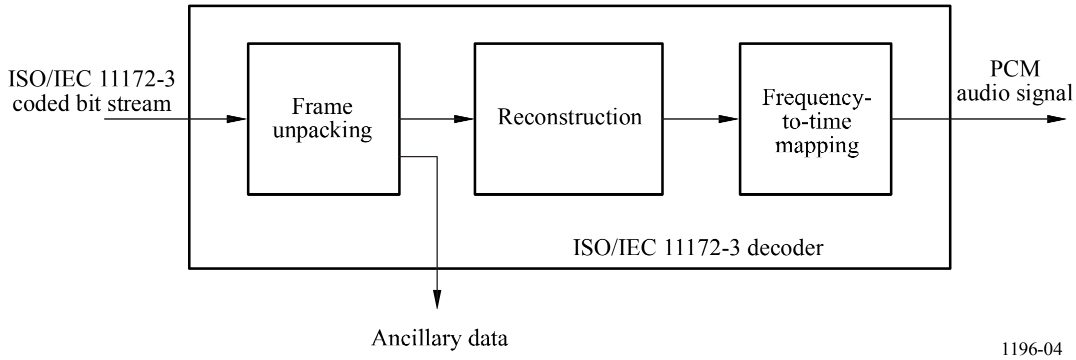
4 Decoding

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



Appendix 2

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

Number of speakers	**Audio syntactic elements, listed in order received**	**Default element to speaker mapping**
1	single_channel_element	Centre front speaker
2	channel_pair_element	Left and right front speakers
3	single_channel_element()	Centre front speaker
	channel_pair_element()	Left and right front speakers
4	single_channel_element()	Centre front speaker
	channel_pair_element(),	Left and right front speakers
	single_channel_element()	Rear surround speaker
5	single_channel_element()	Centre front speaker
	channel_pair_element()	Left and right front speakers
	channel_pair_element()	Left surround and right surround rear speakers
5 + 1	single_channel_element()	Centre front speaker
	channel_pair_element()	Left and right front speakers
	channel_pair_element()	Left surround and right surround rear speakers
	Lfe_element()	Low frequency effects speaker

TABLE 2 (*end*)

Number of speakers	Audio syntactic elements, listed in order received	Default element to speaker mapping
7 + 1	single_channel_element()	Centre front speaker
	channel_pair_element()	Left and right centre front speakers
	channel_pair_element()	Left and right outside front speakers
	channel_pair_element()	Left surround and right surround rear speakers
	lfe_element()	Low frequency effects speaker

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.

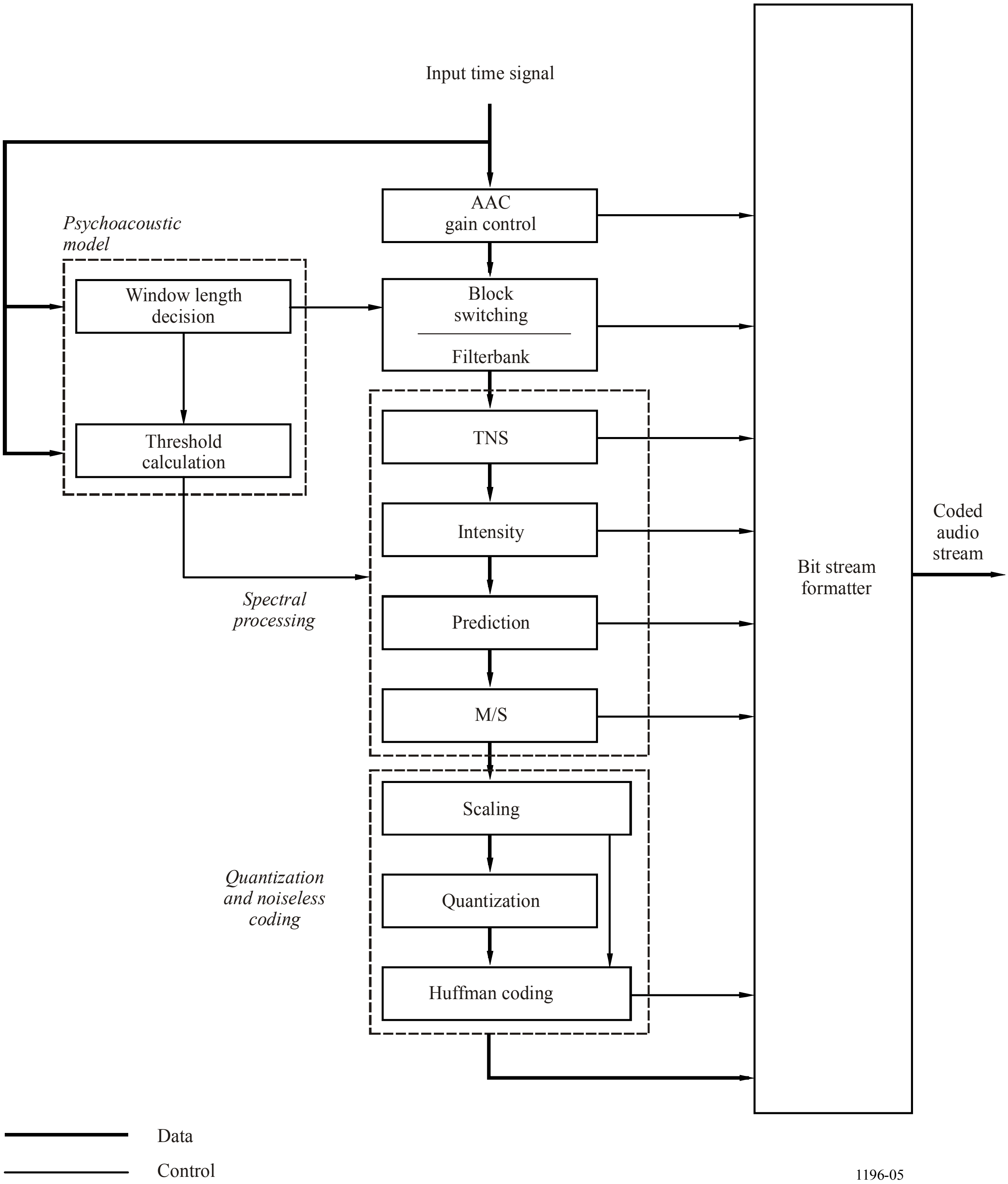
– *Quantization and noiseless coding*: In the quantization tool, a non-uniform quantizer is used with a step size of 1.5 dB. Huffman coding is applied for quantized spectrum, the different scale factors, and directional information.

– *Bit-stream formatter*: Finally a bit-stream formatter is used to multiplex the bit stream, which consists of the quantized and coded spectral coefficients and some additional information from each tool.

– *Psychoacoustic model*: The current masking threshold is computed using a psychoacoustic model from the input signal. A psychoacoustic model similar to ISO/IEC 11172-3 psychoacoustic model 2 is employed. A signal-to-mask ratio, which is derived from the masking threshold and input signal level, is used during the quantization process in order to minimize the audible quantization noise and additionally for the selection of adequate coding tool.

FIGURE 5

MPEG-2 AAC encoder block diagram

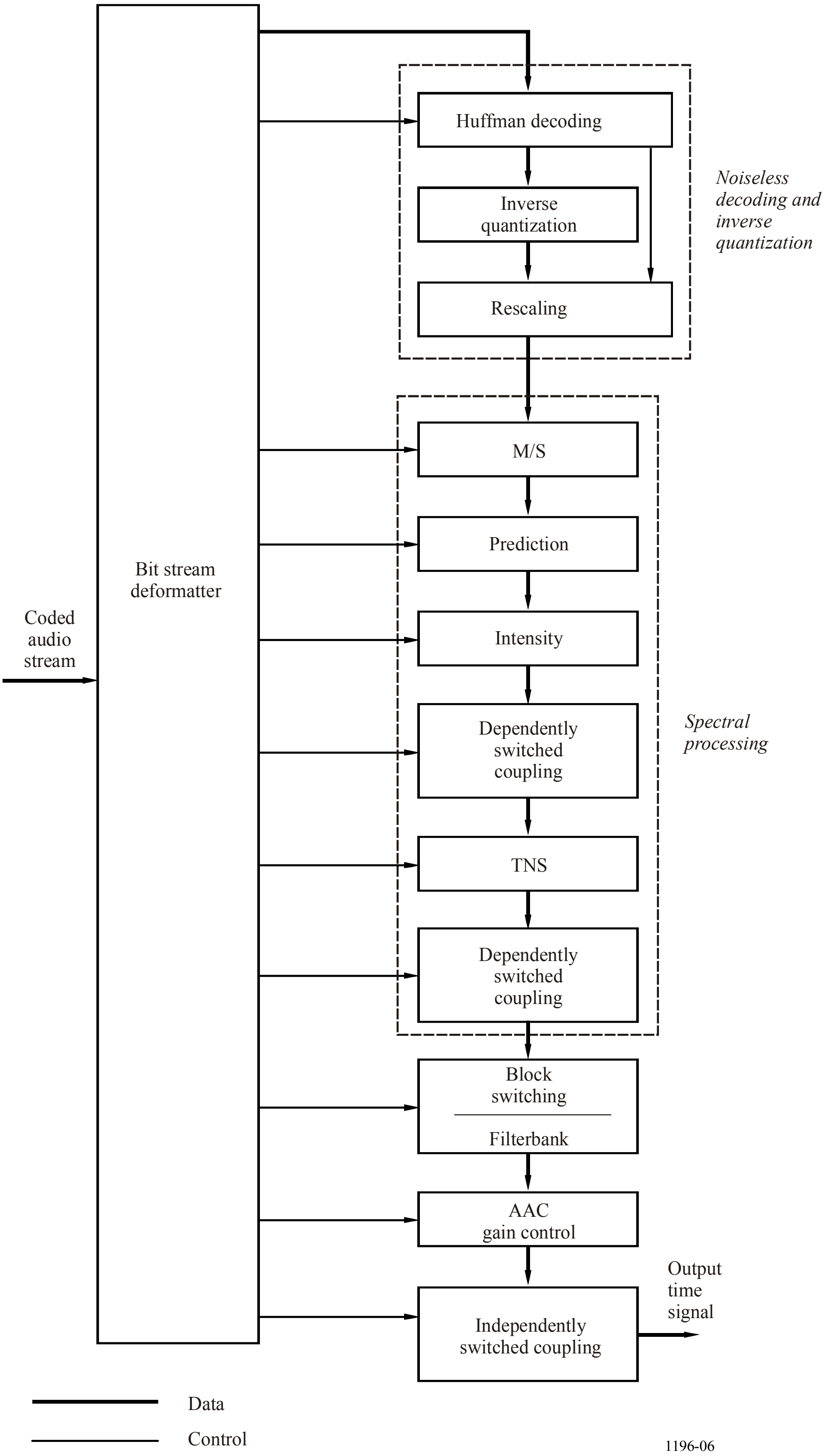


3 Decoding

The basic structure of the MPEG-2 AAC decoder is shown in Fig. 6. The decoding process is basically the inverse of the encoding process.

FIGURE 6

MPEG-2 AAC decoder block diagram

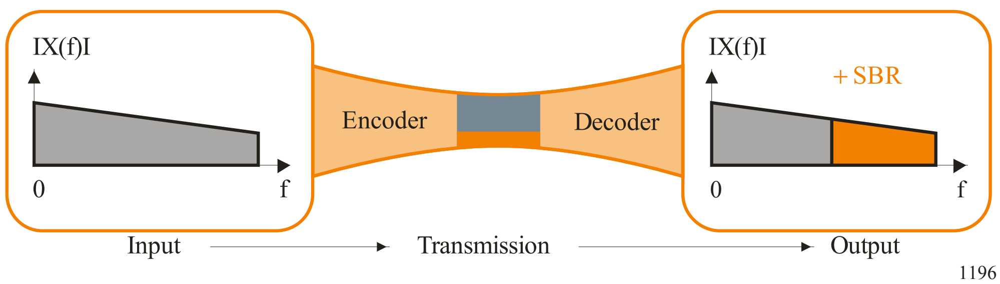


The functions of the decoder are to find the description of the quantized audio spectra in the bit stream, decode the quantized values and other reconstruction information, reconstruct the quantized spectra, process the reconstructed spectra through whatever tools are active in the bit stream in order to arrive at the actual signal spectra as described by the input bit stream, and finally convert the frequency domain spectra to the time domain, with or without an optional gain control tool. Following the initial reconstruction and scaling of the spectrum reconstruction, there are many optional tools that modify one or more of the spectra in order to provide more efficient coding. For each of the optional tools that operate in the spectral domain, the option to "pass through" is retained, and in all cases where a spectral operation is omitted, the spectra at its input are passed directly through the tool without modification.

4 High efficiency AAC and spectral band replication

High Efficiency AAC (HE AAC) introduces spectral band replication (SBR). SBR is a method for highly efficient coding of high frequencies in audio compression algorithms. It offers improved performance of low bit rate audio and speech codecs by either increasing the audio bandwidth at a given bit rate or by improving coding efficiency at a given quality level.

Only the lower part of the spectrum is encoded and transmitted. This is the part of the spectrum to which the human ear is most sensitive. Instead of transmitting the higher part of the spectrum, SBR is used as a post-decoding process to reconstruct the higher frequencies based on an analysis of the transmitted lower frequencies. Accurate reconstruction is ensured by transmitting SBR-related parameters in the encoded bit stream at a very low data rate.

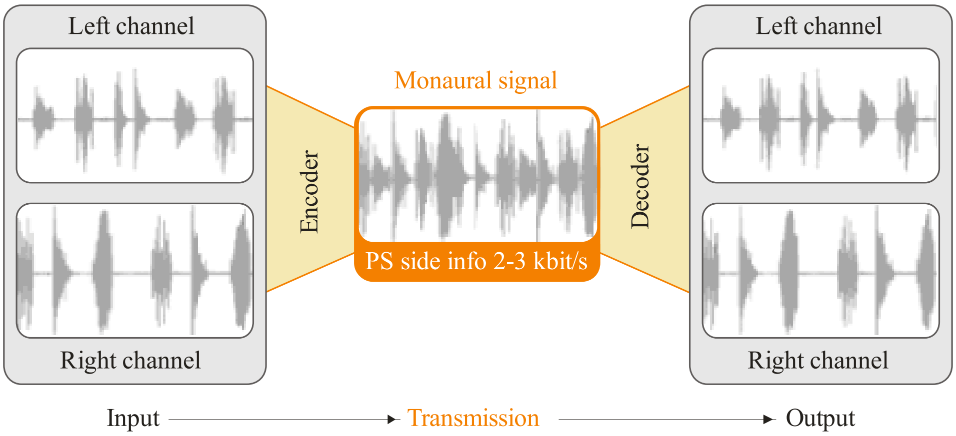


The HE AAC bit stream is an enhancement of the AAC audio bit stream. The additional SBR data is embedded in the AAC fill element, thus guaranteeing compatibility with the AAC standard. The HE AAC technology is a dual-rate system. The backward compatible plain AAC audio bit stream is run at half the sample rate of the SBR enhancement, thus an AAC decoder, which is not capable of decoding the SBR enhancement data, will produce an output time-signal at half the sampling rate than the one produced by an HE AAC decoder.

5 High efficiency AAC version 2 and parametric stereo

HE AAC v2 is an extension to HE AAC and introduces parametric stereo (PS) to enhance the efficiency of audio compression for low bit rate stereo signals.

The encoder analyses the stereo audio signal and constructs a parametric representation of the stereo image. There is now no need to transmit both channels and only a mono-aural representation of the original stereo signal is encoded. This signal is transmitted together with parameters required for the reconstruction of the stereo image.



As a result, the perceived audio quality of a low bit rate audio bit stream (for example, 24 kbit/s) incorporating parametric stereo is significantly higher compared to the quality of a similar bit stream without parametric stereo.

The HE AAC v2 bit stream is built on the HE AAC bit stream. The additional parametric stereo data is embedded in the SBR extension element of a mono HE AAC stream, thus guaranteeing compatibility with HE AAC as well as with AAC.

An HE AAC decoder, which is not capable of decoding the parametric stereo enhancement, produces a mono output signal at the full bandwidth. A plain AAC decoder, which is not capable of decoding the SBR enhancement data, produces a mono output time-signal at half the sampling rate.

Appendix 3

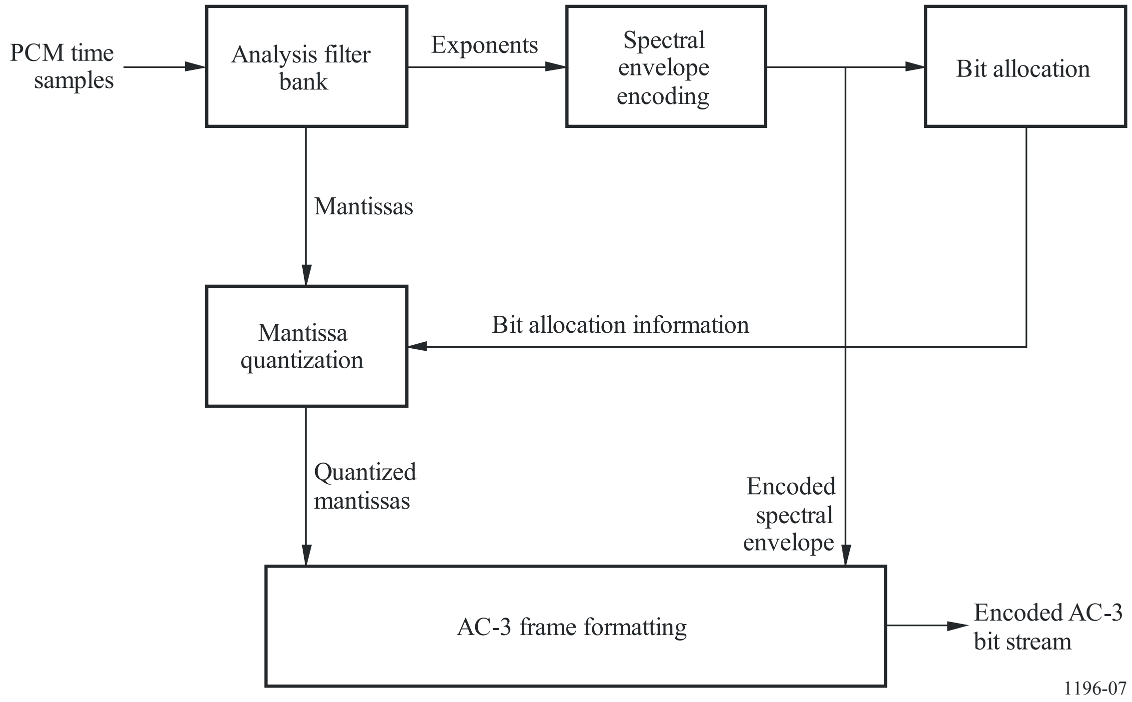
AC-3 and E-AC-3 audio

1 Encoding

The AC-3 digital compression algorithm can encode from 1 to 5.1 channels of source audio from a PCM representation into a serial bit stream at data rates ranging from 32 kbit/s to 640 kbit/s. The AC-3 algorithm achieves high coding gain (the ratio of the input bit rate to the output bit rate) by coarsely quantizing a frequency domain representation of the audio signal. A block diagram of this process is shown in Fig. 7. The first step in the encoding process is to transform the representation of audio from a sequence of PCM time samples into a sequence of blocks of frequency coefficients. This is done in the analysis filter bank. Overlapping blocks of 512 time-samples are multiplied by a time window and transformed into the frequency domain. Due to the overlapping blocks, each PCM input sample is represented in two sequential transformed blocks. The frequency domain representation may then be decimated by a factor of two so that each block contains 256 frequency coefficients. The individual frequency coefficients are represented in binary exponential notation as a binary exponent and a mantissa. The set of exponents is encoded into a coarse representation of the signal spectrum which is referred to as the spectral envelope. This spectral envelope is used by the core bit allocation routine which determines how many bits to use to encode each individual mantissa. The spectral envelope and the coarsely quantized mantissas for 6 audio blocks (1 536 audio samples) are formatted into an AC-3 frame. The AC-3 bit stream is a sequence of AC-3 frames.

FIGURE 7

The AC- encoder



The actual AC-3 encoder is more complex than indicated in Fig. 7. The following functions not shown above are also included:

– a frame header is attached which contains information (bit rate, sample rate, number of encoded channels, etc.) required to synchronize and decode the encoded bit stream;

– error detection codes are inserted in order to allow the decoder to verify that a received frame of data is error free;

– the analysis filter bank spectral resolution may be dynamically altered so as to better match the time/frequency characteristic of each audio block;

– the spectral envelope may be encoded with variable time/frequency resolution;

– a more complex bit allocation may be performed, and parameters of the core bit allocation routine modified so as to produce a more optimum bit allocation;

– the channels may be coupled together at high frequencies in order to achieve higher coding gain for operation at lower bit rates;

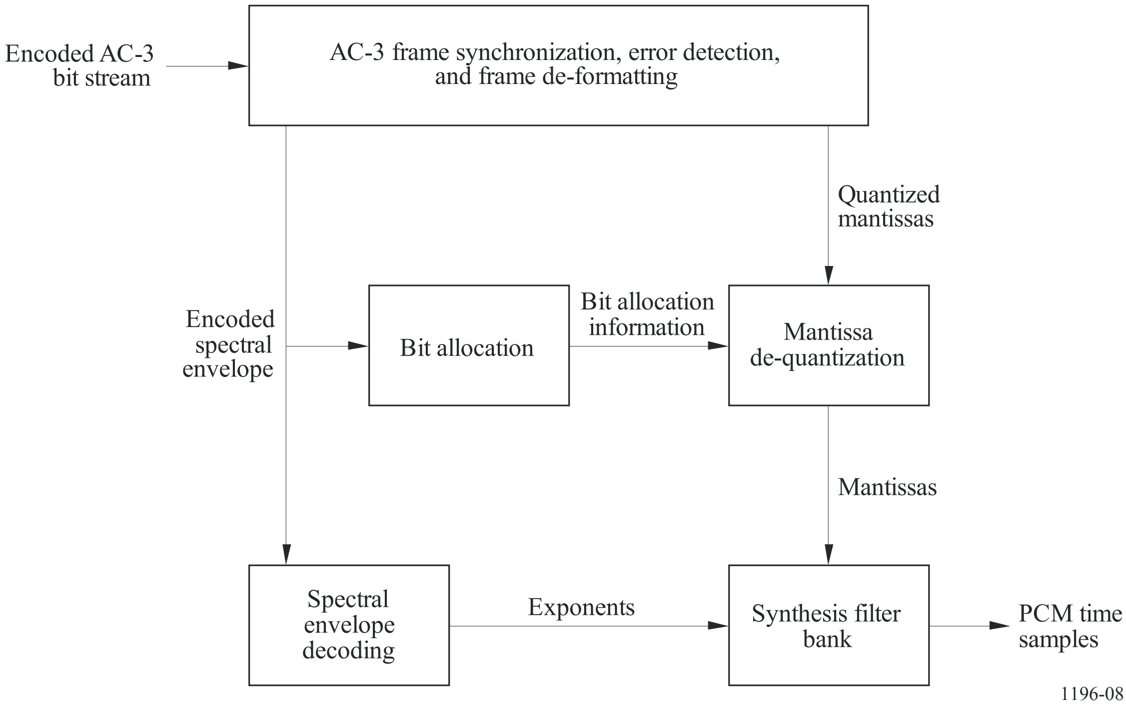
– in the two-channel mode a rematrixing process may be selectively performed in order to provide additional coding gain, and to allow improved results to be obtained in the event that the two-channel signal is decoded with a matrix surround decoder.

2 Decoding

The decoding process is basically the inverse of the encoding process. The decoder, shown in Fig. 8, must synchronize to the encoded bit stream, check for errors, and de-format the various types of data such as the encoded spectral envelope and the quantized mantissas. The bit allocation routine is run and the results used to unpack and de-quantize the mantissas. The spectral envelope is decoded to produce the exponents. The exponents and mantissas are transformed back into the time domain to produce the decoded PCM time samples.

FIGURE 8

The AC-3 decoder



The actual AC-3 decoder is more complex than indicated in Fig. 8. The following functions not shown above are included:

– error concealment or muting may be applied in case a data error is detected;

– channels which have had their high-frequency content coupled together must be de-coupled;

– de-matrixing must be applied (in the 2-channel mode) whenever the channels have been re-matrixed;

– the synthesis filter bank resolution must be dynamically altered in the same manner as the encoder analysis filter bank had been during the encoding process.

3 E-AC-3

Enhanced AC-3 (E-AC-3) adds several additional coding tools and features to the basic AC-3 codec described above. The additional coding tools provide improved coding efficiency allowing operation at lower bit rates, while the additional features provide additional application flexibility.

Additional coding tools:

– Adaptive Hybrid Transform – Additional layer applied in the analysis/synthesis filter bank to provide finer (1/6 of AC-3) spectral resolution.

– Transient pre-noise processing – Additional tool to reduce transient pre-noise.

– Spectral extension – Decoder synthesis of highest frequency components based on side information created by encoder.

– Enhanced coupling – Treats phase as well as amplitude in channel coupling.

Additional features:

– Finer data rate granularity.
– Higher maximum data rate (3 Mbit/s).
– Sub-streams can carry additional audio channels, e.g. 7.1 chs, or commentary tracks.

Appendix 4

MPEG Surround

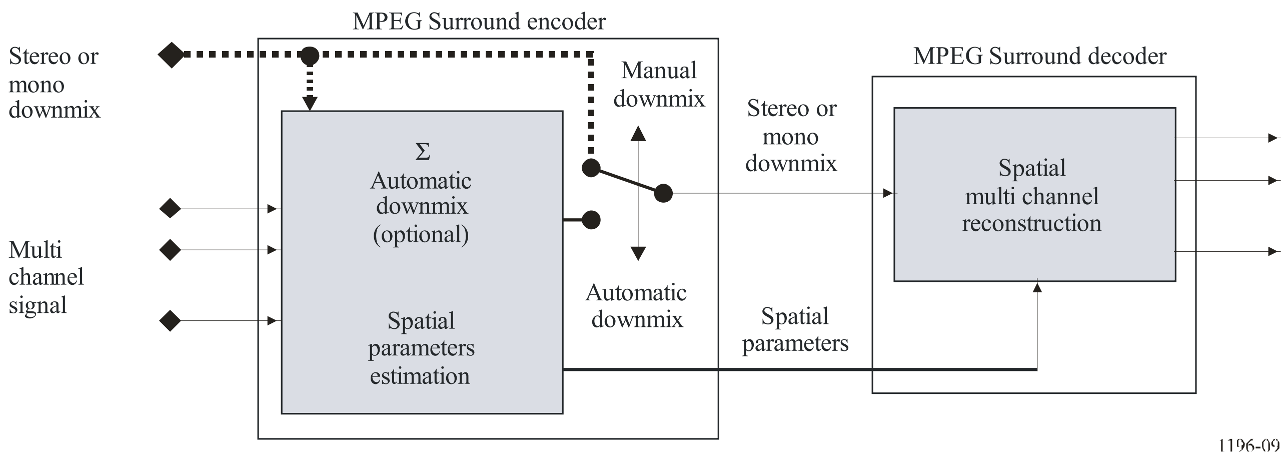
1 Introduction

ISO/IEC 23003-1 or MPEG Surround technology provides an extremely efficient method for coding of multi-channel sound and allows the transmission of surround sound at bit-rates that have been commonly used for coding of mono or stereo sound. It is capable of representing an N channel multi-channel audio signal based on an M<N channel downmix and additional control data. In the preferred operating modes, an MPEG Surround encoder creates either a mono or stereo downmix from the multi-channel audio input signal. This downmix is encoded using a standard core audio codec, e.g. one of the coding systems recommended under *recommends* 1 and 2. In addition to the downmix, MPEG Surround generates a spatial image parameter description of the multi-channel audio that is added as an ancillary data stream to the core audio codec in a backwards compatible fashion. Legacy mono or stereo decoders will ignore the ancillary data and playback the stereo or mono downmix audio signal. MPEG Surround capable decoders will first decode the mono or stereo downmix and then use the spatial image parameters extracted from the ancillary data stream to generate a high quality multi-channel audio signal.

Figure 9 illustrates the principle of MPEG Surround.

FIGURE 9

Principle of MPEG Surround, the downmix is coded using a core audio codec



By using MPEG Surround, existing services can easily be upgraded to provide for surround sound in a backward compatible fashion. While a stereo decoder in an existing legacy consumer device ignores the MPEG Surround data and plays back the stereo signal without any quality degradation, an MPEG Surround-enabled decoder will deliver high quality multi-channel audio.

2 Encoding

The aim of the MPEG Surround encoder is to represent a multi-channel input signal as a backward compatible mono or stereo signal, combined with spatial parameters that enable reconstruction of a multi-channel output that resembles the original multi-channel input signals from a perceptual point of view. Other than the automatically generated downmix an externally created downmix ("artistic downmix") can be used. The downmix shall preserve the spatial characteristics of the input sound.

MPEG Surround builds upon the parametric stereo technology that has been combined with HE- AAC, resulting in the HE AAC v2 standard specification. By combining multiple parametric stereo modules and other newly developed modules, various structures supporting different combinations of number of output and downmix channels have been defined. As an example, for a 5.1 multi-channel input signal, three different configurations are available; one configuration for stereo downmix based systems (525 configuration.), and two different configurations for the mono downmix based systems (a 515_1 and 515_2 configuration that employ a different concatenation of boxes).

MPEG Surround incorporates a number of tools enabling features that allow for broad application of the standard. A key feature of MPEG Surround is the ability to scale the spatial image quality gradually from very low spatial overhead towards transparency. Another key feature is that the decoder input can be made compatible to existing matrixed surround technologies.

These and other features are realized by the following prominent encoding tools:

– Residual coding: In addition to the spatial parameters, also residual signals can be conveyed using a hybrid coding technique. These signals substitute part of the decorrelated signals (that are part of the Parametric stereo boxes). Residual signals are coded by transforming the QMF domain signals to the MDCT domain after which the MDCT coefficients are coded using AAC.
– Matrix compatibility: Optionally, the stereo downmix can be pre-processed to be compatible to legacy matrix surround technologies to ensure backward-compatibility with decoders that can only decode the stereo bit-stream but are equipped with a matrix-surround decoder.
– Arbitrary downmix signals: The MPEG Surround system is capable of handling not just encoder-generated downmixes but also artistic downmixes supplied to the encoder in addition to the multi-channel original signal.
– MPEG Surround over PCM: Typically, the MPEG Surround spatial parameters are carried in the ancillary data portion of the underlying audio compression scheme. For applications where the downmix is transmitted as PCM, MPEG Surround also supports a method that allows the spatial parameters to be carried over uncompressed audio channels. The underlying technology is referred to as buried data.

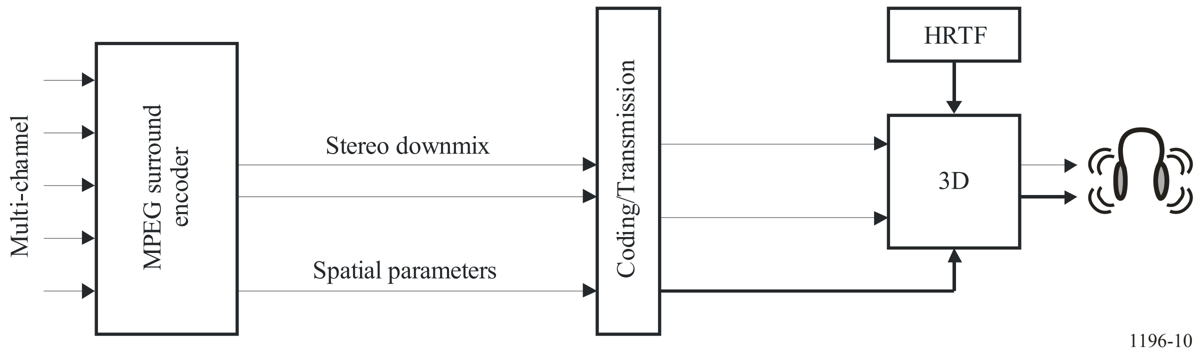
3 Decoding

Next to rendering to a multi-channel output, an MPEG Surround decoder also supports rendering to alternative output configurations:

– Virtual Surround: The MPEG Surround system can exploit the spatial parameters to render the downmix to a stereo virtual surround output for playback over legacy headphones. The standard does not specify the Head Related Transfer Function (HRTF), but merely the interface to these HRTF allowing freedom in implementation depending on the use case. The virtual surround processing can be applied in both the decoder as well as in the encoder, the latter providing the possibility for a virtual surround experience on the downmix, not requiring an MPEG Surround decoder. An MPEG Surround decoder can however undo the virtual surround processing on the downmix and reapply an alternative virtual surround. The basic principle is outlined in the Fig. 10.

FIGURE 10

Virtual Surround decoding of MPEG Surround



– Enhanced Matrix Mode: In the case of legacy stereo content, where no spatial side information is present, the MPEG Surround is capable of estimating the spatial side information from the downmix and thus creates the multi-channel output yet offering a quality which is beyond conventional matrix-surround systems.

– Pruning: As a result of the underlying structure, an MPEG Surround decoder can render its output to channel configurations where the number of channels is lower than the number of channels in the multi-channel input of the encoder.

4 Profiles and levels

The MPEG Surround decoder can be implemented as a high quality version and a low power version. Both versions operate on the same data stream, albeit with different output signals.

The MPEG Surround Baseline Profile defines six different hierarchical levels which allow for different numbers of input and output channels, for different ranges of sampling rates, and for a different bandwidth of the residual signal decoding. The level of the decoder must be equal to, or larger than the level of the bit stream in order to ensure proper decoding. In addition, decoders of Level 1 and 2 are capable of decoding all bit streams of Level 2 and 3, though at a possibly slightly reduced quality due to the limitations of the decoder. The quality and format of the output of an MPEG Surround decoder furthermore depends on the specific decoder configuration. However, decoder configuration aspects are completely orthogonal to the different levels of this profile.

5 Interconnection with audio codecs

MPEG Surround operates as a pre- and post-processing extension on top of legacy audio coding schemes. It is therefore equipped with means to accommodate virtually any core audio coder. The framing in MPEG Surround is highly flexible to ensure synchrony with a wide range of coders and means to optimize the connection with coders that already use parametric tools (e.g. spectral band replication) are provided.