RECOMMENDATION ITU-R BS.1115-1*, **

Low bit-rate audio coding

(Question ITU-R 19/6)

(1994-2004)

The ITU Radiocommunication Assembly,

considering

a) that high quality digital audio services are available to the consumer via various media such as: compact disc (CD), magnetic tape and broadcasting;

b) that the basis of a digital sound-broadcasting system is contained in Recommendations ITU-R BS.774 and ITU-R BO.789;

c) that user requirements for audio coding systems for digital broadcasting are specified in Recommendation ITU-R BS.1548;

d) that the basic audio and stereo image quality required for digital sound broadcasting is to be virtually indistinguishable from that of CD;

e) that the performance of low bit-rate coding systems used in programme connections (contribution, distribution and commentary links) feeding digital sound-broadcasting systems must be such as to be able to deliver audio signals of the highest quality to the digital sound-broadcasting emitter;

f) that, in the case of contribution and distribution links, this implies a level of performance exceeding that of CD because of the requirements for processing and cascading margins and (in the case of contribution links) of overload margin or "headroom";

g) that, in the case of commentary links, this implies a level of performance which is capable of delivering speech of excellent quality to the listener, but in the event that these links carry musical programme material, a reduced level of performance is to be expected;

h) that low bit-rate coding generally incurs a delay of several tens of milliseconds that may have consequences for operational practices such as the provision of off-air cue feeds;

j) that low bit-rate coding for high quality audio has been tested by the ITU Radiocommunication Sector, and satisfactory performance obtained in a number of applications (see Appendix 6 and Recommendation ITU-R BS.1548);

k) that conventional objective methods (e.g. for measuring signal-to-noise ratio and distortion) may no longer be adequate in assessing the quality of systems which use low bit-rate audio coding schemes, and that the objective perceptual measurement method given in Recommendation ITU-R BS.1387 may eventually complement or supersede conventional objective assessment methods, and may eventually complement subjective assessment methods,

---

* 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 Telecommunication Standardization Study Groups 9 and 15, the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) (JTC 1/SC 29/WG 11), the European Broadcasting Union (EBU) and the Audio Engineering Society (AES).

recommends

1 that for digital sound-broadcasting applications listed in *recommends* 4 to 6, low bit-rate audio coding described in ISO/IEC 11172-3: 1993 (see Note 1), ISO/IEC 13818-7:2003 or AC-3 as per Annex 2 to Recommendation ITU-R BS.1196, should be used (short descriptions of these standards are given in Appendices 1, 3 and 4);

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

2 that for all applications the sampling frequency should be either 48 kHz or 32 kHz;

3 that the input signal to the low bit rate audio encoder should be emphasis-free and no emphasis should be applied by the encoder;

4 that for digital sound-broadcasting emission (see Note 1), the following codecs should be used:

– ISO/IEC 11172-3 Layer II coding at a bit rate of 128 kbit/s for a mono signal and 256 kbit/s (i.e. 2 × 128 kbit/s) for a stereo signal, with independent coding of the left and right components of the stereo signal (see Note 2),

– ISO/IEC 13818-7 coding at a bit rate of at least 144 kbit/s for a stereo signal,

– AC-3 coding at a bit rate of at least 192 kbit/s for a stereo signal;

NOTE 1 – Appendix 1 to Annex 2 of Recommendation ITU-R BS.1548 contains information about coding systems that have been demonstrated to meet quality, and other user requirements for emission.

NOTE 2 – It is anticipated that broadcasters will wish to have some capacity for ancillary data within the emitted signal. They should, however, be made aware that reducing the audio bit rate generally affects the audio quality.

5 that for distribution and contribution links, ISO/IEC 11172-3 Layer II coding should 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 (see Appendix 2 and Note 2), but that for the special case of a single distribution link without further cascading (see Note 3), a bit rate of at least 120 kbit/s per audio signal may be used;

NOTE 1 – Appendix 1 to Annex 1 of Recommendation ITU-R BS.1548 contains information about coding systems that have been demonstrated to meet quality, and other user requirements for contribution and distribution.

NOTE 2 – It should be taken into account when applying this Recommendation that some telecommunication networks currently do not permit unrestricted use of 64 kbit/s channels.

NOTE 3 – When, e.g., an analogue transmitter is fed via a single distribution link, or when a signal coded to the final emission standard at the studio is sent to the transmitter for emission without any further decoding and encoding.

6 that for commentary links, ISO/IEC 11172-3 Layer III coding should be used at a bit rate of at least 60 kbit/s excluding ancillary data for mono signals (see Note 1), and at least 120 kbit/s excluding ancillary data for stereo signals, using joint stereo coding (see Appendix 2 and Note 2);

NOTE 1 – Results obtained in tests, in 1993, using headphone listening to assess the performance of a 60 kbit/s Layer III mono commentary codec indicated that this did not meet the required quality with a speech signal. This Note will be retained in the Recommendation until the performance of the commentary codec fulfils the requirements.

NOTE 2 – It should be taken into account when applying this Recommendation that some telecommunication networks currently do not permit unrestricted use of 64 kbit/s channels.

7 that delay should be minimized in the implementation of encoders and decoders.

NOTE 1 – ISO/IEC 11172-3 (MPEG-1 audio) and 13818-7 (MPEG-2 AAC) standards are available in electronic version at the following address: http://www.iso.org/itu

Appendix 1

Description of ISO/IEC 11172-3 coding method
Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s – Part 3: Audio

1 Encoding

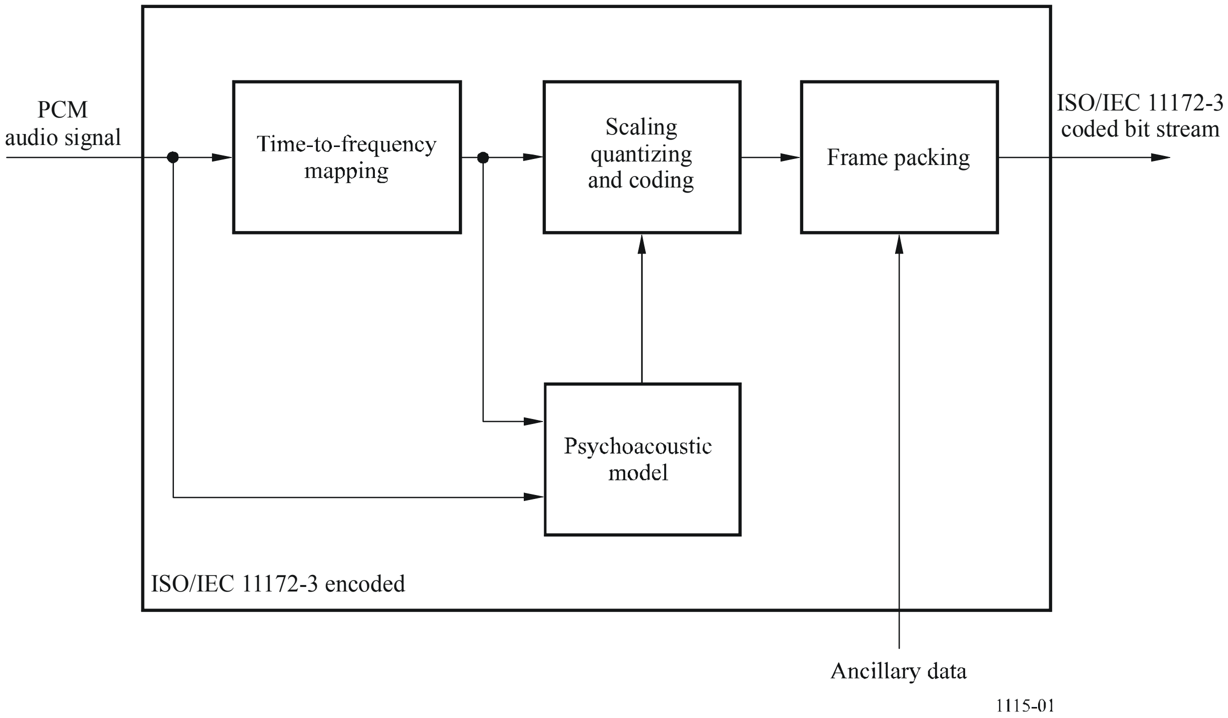
The encoder processes the digital audio signal and produces the compressed bit stream. The encoder algorithm is not standardized and may use various means for encoding, such as estimation of the auditory masking threshold, quantization, and scaling (Note 1). However, the encoder output must be such that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

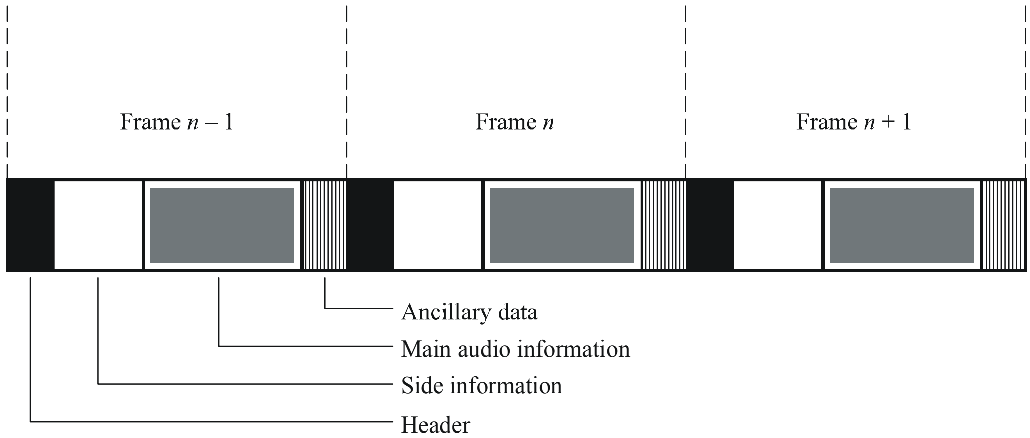
– single channel;
– dual channel (two independent audio signals coded within one bit stream, e.g. bilingual application);
– stereo (left and right signals of a stereo pair coded within one bit stream); and
– joint stereo (left and right signals of a stereo pair coded within one bit stream with the stereo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



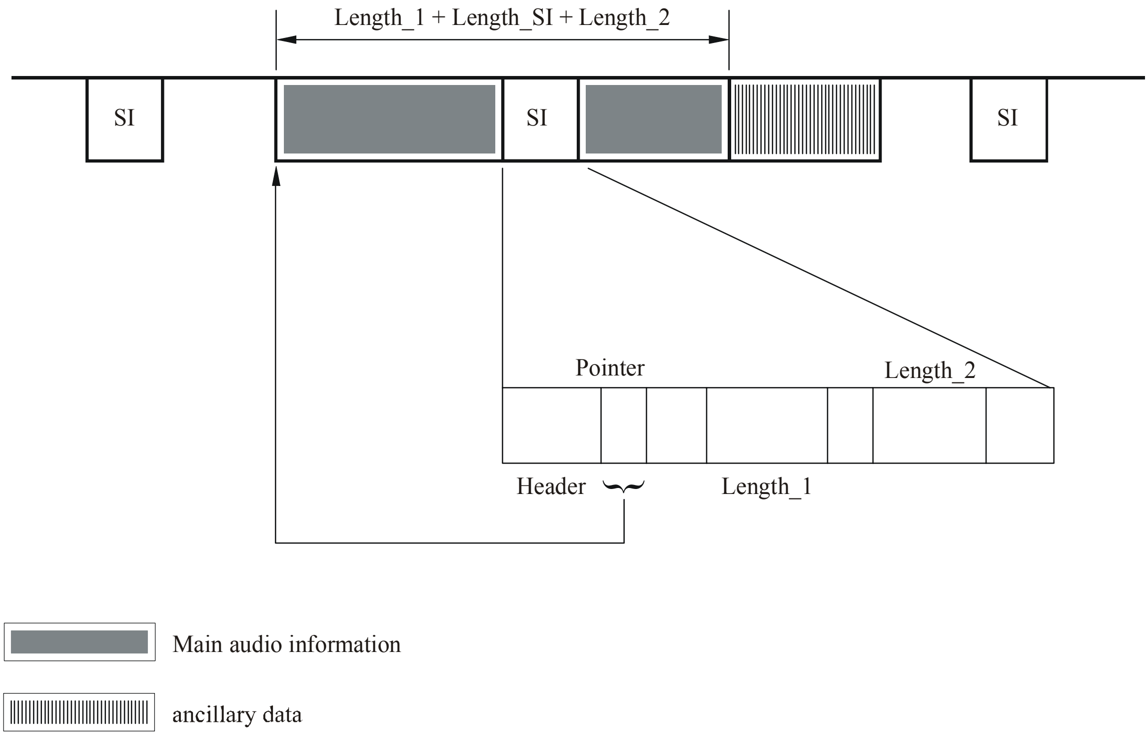
Layer II:

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

1115-02

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.

1115-03

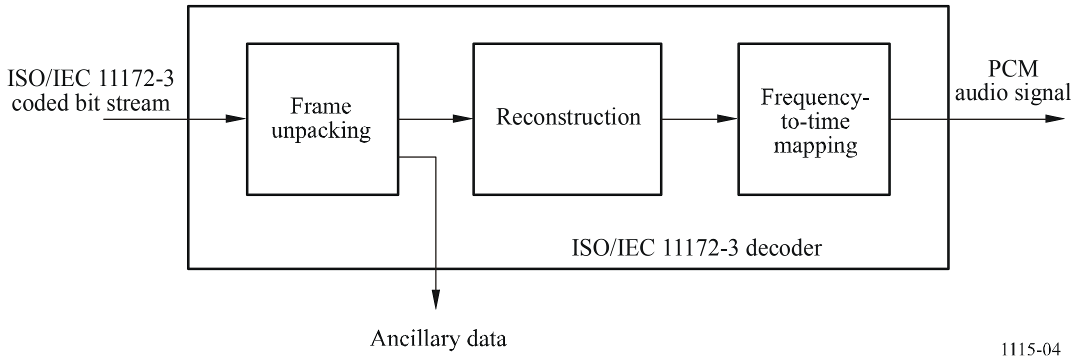
4 Decoding

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

Guidelines for operating ISO/IEC 11172-3 codecs at specific bit rates

This Appendix is intended to provide some guidelines for applications where it might be necessary to operate an ISO/IEC 11172-3 codec at a specific bit rate not explicitly listed in the bit rate table of the ISO/IEC standard (Table 1). These specific bit rates might have to be chosen, taking into account the available bit rate of a transmission channel and the appropriate minimum bit rate given in the Recommendation.

A more complete discussion is available in ITU-T Recommendation J.52 on Digital Transmission of High-Quality Sound-Programme Signals using One, Two or Three 64 kbit/s Channels per Mono Signal (and up to 6 per stereo signal).

The following bit rates are explicitly listed in the ISO/IEC standard:

TABLE 1

Layer II (kbit/s)	**Layer III (kbit/s)**
32	32
48	40
56	48
64	56
80	64
96	80
112	96
128	112
160	128
192	160
224	192
256	224
320	256
384	320

The figures given in Table 1 indicate the total bit rate per audio programme, irrespective of the mode. There is an additional format, the free format, which may be used for fixed bit rates which are not in the table, with a maximum of 384 kbit/s for Layer II and 320 kbit/s for Layer III.

An ISO/IEC 11172-3 Layer II or Layer III audio frame always corresponds to 1 152 audio PCM input samples per channel. The duration of the frame is 24 ms with a sampling rate of 48 kHz, and 36 ms with a sampling rate of 32 kHz. A frame always consists of an integer number of bytes. The number of bytes can be calculated by frame duration multiplied by bit rate and divided by 8. The bit rate and the sampling frequency are indicated in the header in the beginning of the audio frame.

The following three methods may be used to realize an audio bit rate not listed in the table:

1 Free format

The free format condition has to be indicated in the header. In this case the decoder measures at start-up the distance between consecutive sync words, after which a flywheel synchronization procedure can be used as in the case of a defined bit rate. The length of the frame in bytes can be calculated by the formula given above. This figure has to be accounted for by the encoder.

For example, if a bit rate of 240 kbit/s for stereo programme is desired with a sampling frequency of 48 kHz, the length of the frame will be $0.024 \times 240\,000/8 = 720$ bytes. For bit rates which would result in a non-integer number of bytes in a frame, the required bit rate can be implemented by using padding. Details for this are given in the sub-clause 2.4.2.3 of ISO/IEC 11172-3.

This method may be used for both Layers II and III.

2 Use of the ancillary data field (for Layer II)

One of the listed bit rates (higher or equal than the required bit rate) can be chosen. By constraining the bit allocation the encoder can be set to reserve a certain amount of bits for ancillary data, such that the actual audio bit rate is equal to the required bit rate. The ancillary data for Layer II are at the very end of each audio frame.

The ancillary data bits may be removed or overwritten before transmission. If these bits are removed, a corresponding number of dummy bits have to be inserted again before decoding.

For example, a bit rate of 240 kbit/s for a stereo programme is desired with a sampling frequency of 48 kHz, a bit rate of 256 kbit/s can be chosen and according to the ISO/IEC standard, indicated in the header of the frame. The encoder has to be set to reserve 48 bytes per frame for ancillary data, corresponding to 16 kbit/s ancillary data capacity.

3 Dynamic bit rate switching (for Layer III)

In Layer III, dynamic bit rate switching can be used to obtain bit rates not specified in Table 1. If, for instance, a bit rate of 120 kbit/s is required, then a Layer III encoder can operate in a mode where bit rates explicitly defined in ISO/IEC standard of 128 kbit/s and 112 kbit/s are selected in an alternating sequence. The encoder operating in this mode must be set to a bit allocation number derived from the average of these alternating bit rates. The Layer III bit reservoir technique maintains a constant bit rate available for the coding process. With this method, the distance in bits between sync words alternates between two different values.

Appendix 3

Description of ISO/IEC 13818-7 coding method Generic coding of moving pictures and associated audio information – Part 7: Advanced Audio Coding (AAC)

See ISO/IEC 13818-7.

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 2, and up to 48 audio channels. Table 3 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 2

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 3

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
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.

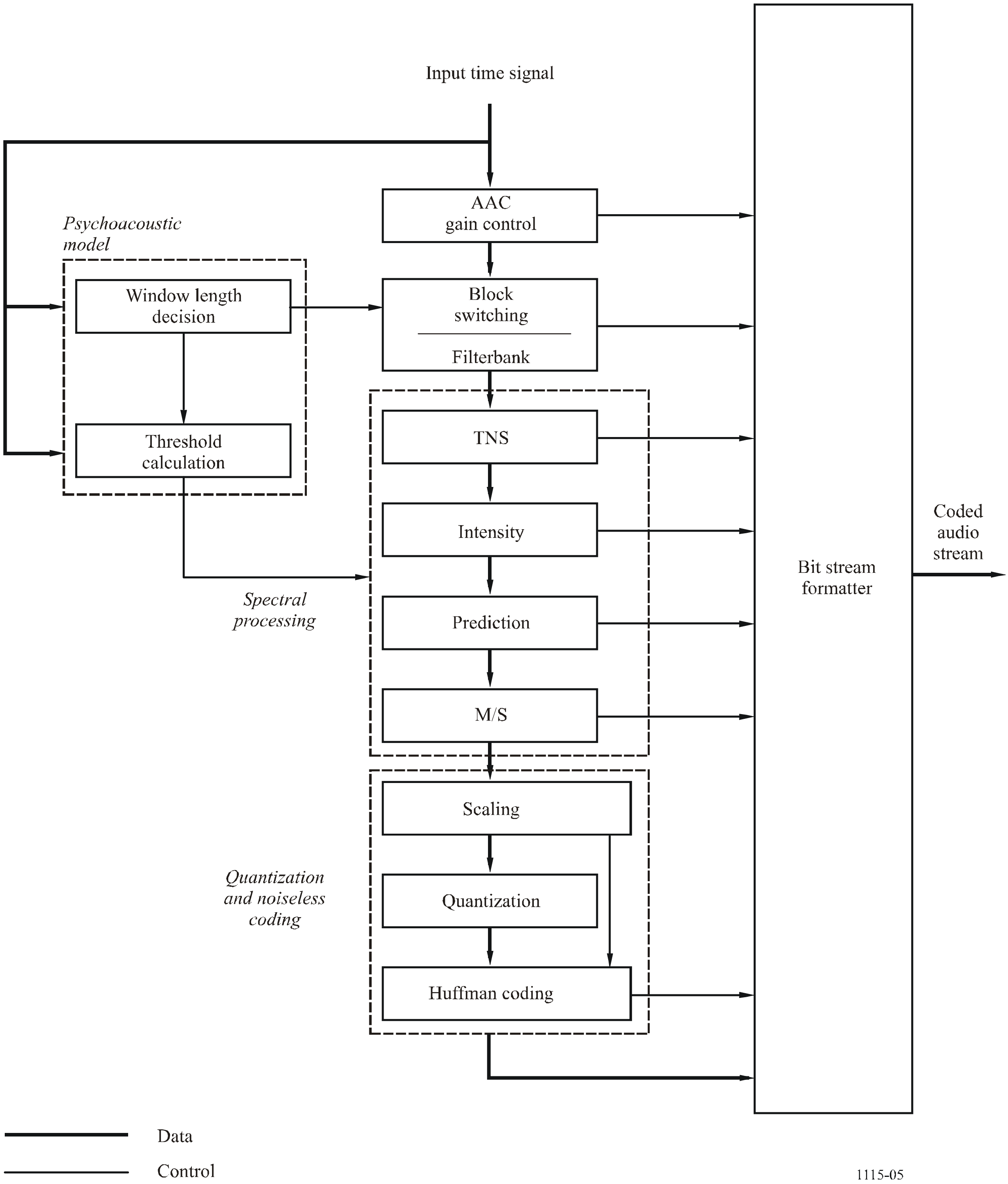
– *Quantization and noiseless coding*: In the quantization tool, a non-uniform quantizer is used with a step size of 1.5 dB. Huffman coding is applied for quantized spectrum, the different scale factors, and directional information.

– *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



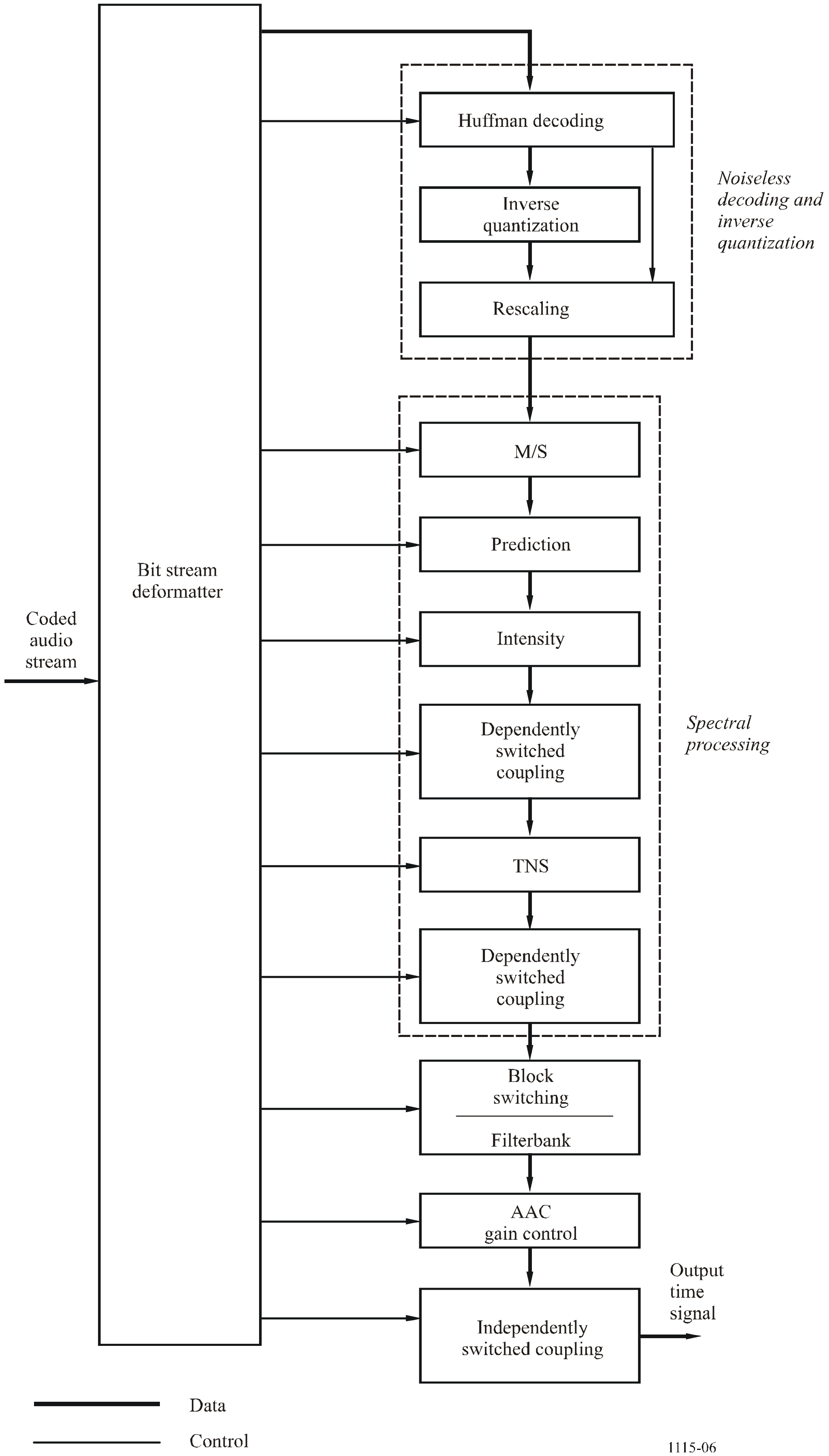
3 Decoding

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.

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.

FIGURE 6

MPEG-2 AAC decoder block diagram



Appendix 4

Description of AC-3 coding method

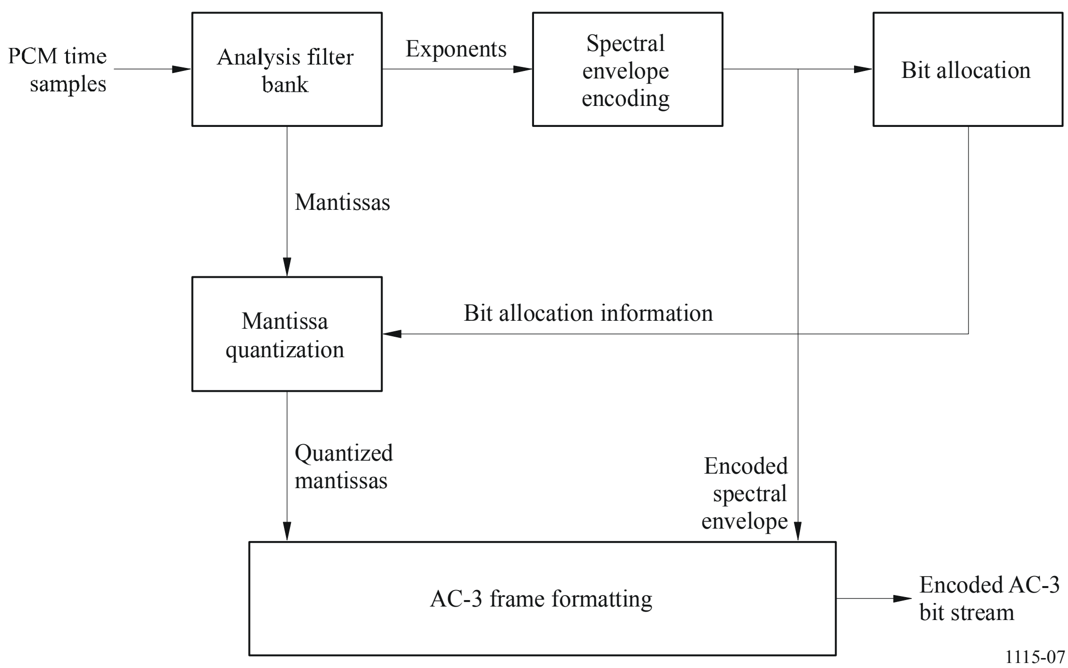
See Annex 2 to Recommendation ITU-R BS.1196 and Annex B to ATSC Standard A/53B.

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-3 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;

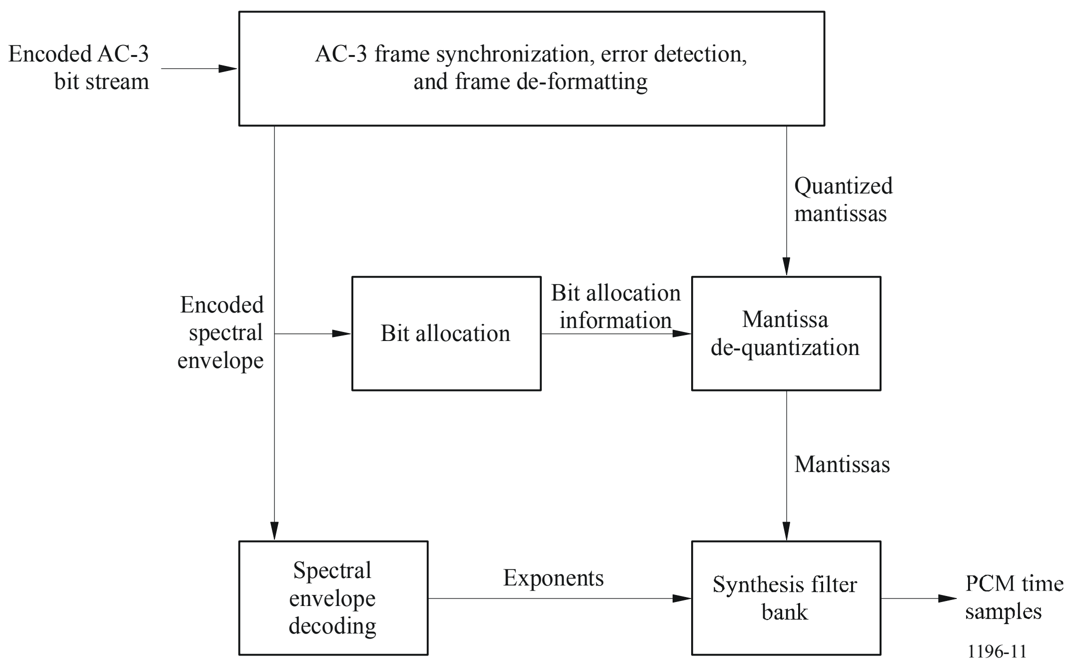
– in the two-channel mode a rematrixing process may be selectively performed in order to provide additional coding gain, and to allow improved results to be obtained in the event that the two-channel signal is decoded with a matrix surround decoder.

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.

Appendix 5

Perceptual objective methods

Perceptual objective methods have been used experimentally to measure codec performance, and some published results show good correlation with the results of subjective tests.

The following are methods which have been used experimentally within the Radiocommunication Sector:

– noise to mask ratio (NMR);
– PERCeptual EVALuation model (PERCEVAL);
– perceptual audio quality measure (PAQM);
– perceptual objective model (POM 620).

Recommendation ITU-R BS.1387 specifies a method for objective measurements of perceived audio quality.

Appendix 6

Selection methodology of the recommended low bit rate audio codecs in the tests conducted in 1992-1993

Subjective tests were performed for various single channel and conventional 2-channel applications and the results were compared to the requirements. Further studies are being conducted on multichannel coding systems.

For the three applications, emission, distribution, and contribution, the basic audio quality of the codecs reproduced after decoding has to be equivalent to compact disc quality, that is, subjectively indistinguishable, for most types of audio programme material. The quality of the stereo sound image should be maintained.

For emission, the most critical material for the codecs must be such that the degradation may be "perceptible but not annoying" (grade 4). (The subjective tests were performed using the five point impairment scale (Recommendation ITU-R BS.562).) In tests conducted in 1992, the following systems fulfilled these requirements:

2 × 128 kbit/s independent channels (Note 1):

- ISO/IEC 11172-3 Layer II;
- ISO/IEC 11172-3 Layer III;
- Dolby AC-2.

192 kbit/s joint stereo:

- ISO/IEC 11172-3 Layer II;
- ISO/IEC 11172-3 Layer III.

For distribution, the signal, after passing through three codecs in tandem at 120 kbit/s per independent channel, should be graded not more than 0.5 lower than the original source signal (Note 1). Re-population of the binary codes was performed by applying a 0.1 dB gain reduction in the linear PCM domain. The ISO/IEC 11172-3 Layer II codec was the only system that fulfilled the requirements and is therefore the basis of the Recommendation for this application.

NOTE 1 – At this bit rate, these codecs were graded 4.5 or higher for the most critical items. As test methodologies and codecs continue to evolve, future tests may yield different grades.

For contribution, the signal, after passing through five codecs in tandem at 180 kbit/s per independent channel, shall be subjectively indistinguishable from the original source signal. The tandeming was performed using a resolution of 18-bit linear PCM. Re-population of the binary codes was performed by applying a 0.1 dB gain reduction in the linear PCM domain. The ISO/IEC 11172-3 Layer II codec was the only system that fulfilled the requirements and is therefore the basis of the Recommendation for this application.

For commentary, the basic audio quality of speech signals reproduced after decoding should be equivalent to that of a 14-bit linear PCM original. In the actual tests, a 16-bit linear PCM format was used. Perceptible levels of impairment are tolerated for music sequences of compact disc quality. In tests conducted in 1992 with loudspeakers only, the ISO/IEC 11172-3 Layer III codec alone performed satisfactorily on speech, and it was, on average, the best for music signals. On speech signals, the grades were always higher than 4.0 (perceptible but not annoying) in both mono (60 kbit/s) and stereo (120 kbit/s joint stereo) modes. This codec is therefore the basis of the Recommendation for this application. In subsequent tests, in 1993, the results obtained using headphone listening to assess the performance of a 60 kbit/s Layer III mono commentary codec indicated, with a grade below 4.0, that this did not meet the required quality with a speech signal. Further studies at 60 kbit/s are required.

The complexity evaluation showed that the ISO/IEC 11172-3 Layer II and the Dolby AC-2 decoders proved to be of the lowest complexity. Based on the following considerations:

- low complexity decoder;
- commonality with the distribution and contribution applications;
- flexibility to improve the encoder;
- codec tested at both bit rates.

The ISO/IEC 11172-3 Layer II format and decoder were chosen to be the basis of the Recommendation for the emission application at 2 × 128 kbit/s. In tests conducted in 1992, the ISO/IEC 11172-3 Layer II only marginally fulfilled the requirements at 192 kbit/s, and it was decided that improvements were needed at that bit rate in order to meet clearly the basic audio quality requirements (as was done by the ISO/IEC 11172-3 Layer III). Further tests conducted in 1993 did not show improvements at this stage. Further studies at 192 kbit/s using joint stereo mode would be required for the use of this bit rate to be reconsidered. Administrations are requested to contribute on this subject.

Network verification tests were conducted in 1993 with a complete broadcasting chain, including five contribution codecs at 180 kbit/s in tandem, three distribution codecs at 120 kbit/s in tandem, and one emission codec. The results of these tests showed that the basic audio quality at the extremity of such a chain is not satisfactory.

Tests were also conducted with eight codecs at 180 kbit/s in tandem which proved to be a satisfactory configuration for maintaining an acceptable quality.