# INTERNATIONAL STANDARD

# ISO/IEC 14496-3

Second edition
2001-12-15

**AMENDMENT 3**
2005-04-01

## Information technology — Coding of audio-visual objects —

## Part 3: **Audio**

## AMENDMENT 3: MPEG-1/2 Audio in MPEG-4

*Technologies de l'information — Codage des objets audiovisuels —*

*Partie 3: Codage audio*

*AMENDEMENT 3: Audio MPEG-1/2 dans MPEG-4*



Reference number
ISO/IEC 14496-3:2001/Amd.3:2005(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

# Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Amendment 3 to ISO/IEC 14496-3:2001 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

# Information technology — Coding of audio-visual objects —

## Part 3: **Audio**

## AMENDMENT 3: MPEG-1/2 Audio in MPEG-4

*In subclause 1.2 (Normative references), add:*

ISO/IEC 14496-1, *Information technology - Coding of audio-visual objects - Part 1: Systems*

ISO/IEC 14496-11, *Information technology - Coding of audio-visual objects - Part 11: Scene description and application engine*

*In subclause 1.5.1.1 (Audio object type definition), replace the following line in Table 1.1:*

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (reserved) | | | | | | | | | | | | | | | | | | | | | 30 |
| (reserved) | | | | | | | | | | | | | | | | | | | | | 31 |

*with:*

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (reserved) | | | | | | | | | | | | | | | | | | | | | 30 |
| (escape) | | | | | | | | | | | | | | | | | | | | | 31 |
| Layer-1 | | | | | | | | | | | | | | | | | | | | | 32 |
| Layer-2 | | | | | | | | | | | | | | | | | | | | | 33 |
| Layer-3 | | | | | | | | | | | | | | | | | | | | | 34 |
| (reserved) | | | | | | | | | | | | | | | | | | | | | 35 - 95 |

*In subclause 1.5.1.1(Audio object type definition), Table 1.1 (Audio Object Type definition), add three tools as additional columns:*

- Layer-1, enter X in row for AOT 32
- Layer-2, enter X in row for AOT 33
- Layer-3, enter X in row for AOT 34

*Add the following subclauses:*

1.5.1.2.26 Layer-1 Audio object type

The Layer-1 object is the counterpart of the audio coding scheme Layer-1 specified in ISO/IEC 11172-3 and 13818-3.

1.5.1.2.27 Layer-2 Audio object type

The Layer-2 object is the counterpart of the audio coding scheme Layer-2 specified in ISO/IEC 11172-3 and ISO/IEC 13818-3.

1.5.1.2.28 Layer-3 Audio object type

The Layer-3 object is very similar to the audio coding scheme Layer-3 specified in ISO/IEC 11172-3 and ISO/IEC 13818-3. However, the use of Layer 3 encoded data as defined in ISO/IEC 13818-3 is limited to the “Lower Sampling Frequencies” case, i.e. the Layer 3 multi-channel syntax defined in ISO/IEC 13818-3 is not permitted in this scope. Furthermore, additional sampling rates have been specified.

*In subclause 1.6.2.1 (AudioSpecificConfig), Table 1.8 (Syntax of AudioSpecificConfig), replace (as often as it appears):*

| | | |
|---|---|---|
| **audioObjectType;** | **5** | **bslbf** |

*with:*

| |
|---|
| audioObjectType = GetAudioObjectType(); |

*In subclause 1.6.2.1 (AudioSpecificConfig), Table 1.8 (Syntax of AudioSpecificConfig), replace (as often as it appears):*

| | | |
|---|---|---|
| **extensionAudioObjectType;** | **5** | **bslbf** |

*with:*

| |
|---|
| extensionAudioObjectType = GetAudioObjectType(); |

*In subclause 1.6.2.1 (AudioSpecificConfig), add below Table 1.8 (Syntax of AudioSpecificConfig):*

**Table 1.8a – Syntax of GetAudioObjectType()**

| Syntax | No. of bits | Mnemonic |
|---|---|---|
| GetAudioObjectType() | | |
| { | | |
| **audioObjectType;** | **5** | **uimsbf** |
| if (audioObjectType == 31) { | | |
| audioObjectType = 32 + **audioObjectTypeExt**; | **6** | **uimsbf** |
| } | | |
| return audioObjectType; | | |
| } | | |

*In subclause 1.6.2.1 (AudioSpecificConfig), Table 1.8 (Syntax of AudioSpecificConfig), add above of the data element epConfig:*

```
if (  audioObjectType == 32 || audioObjectType == 33 ||
      audioObjectType == 34 )
      MPEG_1_2_SpecificConfig();
```

*Add the following subclause:*

1.6.2.1.10 MPEG_1_2_SpecificConfig

Defined in ISO/IEC 14496-3 subpart 9.

*In subclause 1.6.2.2.1 (Overview), add the following lines to Table 1.9:*

| | | | |
|---|---|---|---|
| Layer-1 | 32 | ISO/IEC 14496-3 subpart 9 | |
| Layer-2 | 33 | ISO/IEC 14496-3 subpart 9 | |
| Layer-3 | 34 | ISO/IEC 14496-3 subpart 9 | |

*Add after subclause 1.6.3.1 (AudioObjectType):*

1.6.3.1a AudioObjectTypeExt

This data element extends the range of audio object types.

*Create a new subpart, with the following content:*

# Subpart 9: MPEG-1/2 Audio in MPEG-4

## 9.1 Scope

The MPEG-1/2 Audio in MPEG-4 subpart of MPEG-4 Audio specifies the usage of MPEG-1/2 Layer-1, 2 or 3 in an MPEG-4 oriented way, i.e. such that signalling and access unit handling on Systems level is identical to the other MPEG-4 audio object types.

In order to be carried in MPEG-4, the MPEG-1/2 Layer 1, 2 or 3 bitstream frames are re-formatted such that they become self-contained MPEG-4 access units. This facilitates transport over packet based networks, random access, and editability. Those self-contained access units, as used in an MPEG-4 Systems compliant transport or storage format, can be re-converted to MPEG-1/2 compliant bitstreames, and then decoded with any MPEG-1/2 compliant decoder. Several methods of re-conversion are given in an informative annex.

The MPEG-4 Audio syntax is further extended to allow multi-channel configurations based on ISO/IEC 11172-3 and ISO/IEC 13818-3 Layer 3. The multi-channel configurations are similar to the configurations defined for the other multi-channel capable MPEG-4 audio object types. Note that for MPEG-1/2 Layer 1 and 2 the format is not extended. The multi-channel format for these layers is described in ISO/IEC 13818-3.

Furthermore, the permitted sampling frequencies for Layer-3 are extended.

Assistance is furthermore provided in the use of decSpecificInfo and accessUnit as to utilize MPEG-1/2 Layer 1, 2 or 3 in the MPEG-4 world by means of the legacy MPEG-4 Systems interface using ObjectTypeIndication 0x69 or 0x6b.

## 9.2 MPEG_1_2_SpecificConfig

**Table 9.1 – MPEG_1_2_SpecificConfig()**

| Syntax | No. of bits | Mnemonic |
|---|---|---|
| MPEG_1_2_SpecificConfig() | | |
| { | | |
| **extension;** | **1** | **bslbf** |
| } | | |

**extension** shall be zero.

## 9.3 Channel Mapping

The following rules apply:

- single_channel_element()'s and lfe_element()'s are represented by mono audio frames.
- channel_pair_element()'s are represented by stereo audio frames.
- For Layer-1 and Layer-2, not more than one mono audio frame representing a single_channel_element() or one stereo audio frame representing a channel_pair_element() is permitted.

## 9.4 Access Unit Format

### 9.4.1 Layer 1 and 2

One audio frame maps directly to one access unit.

### 9.4.2 Layer 3

One access unit consists of one or more mp3_channel_elements. An mp3_channel_element equals a Layer 3 audio frame with the following modifications compared to its definition in ISO/IEC 11172-3 or ISO/IEC 13818-3:

| | |
|---|---|
| syncword (12 bit) | signals the total length in bytes of the mp3_channel_element (consisting of header, error_check, side info and main data). |
| main_data_begin (9/8 bit) | is either set to the correct value of the corresponding MPEG-1/2 Layer 3 bitstream or to zero. |
| main_data() | is generally stored after the side information. |

All other data elements shall be set according to their specification in ISO/IEC 11172-3 or ISO/IEC 13818-3. All settings in the header shall correspond to the settings in the AudioSpecificConfig().

All mp3_channel_elements belonging to the same timestamp are stored sequentially into one access unit according to the order given in subpart 1, subclause 1.6.3.4, Table 1.11 (Channel Configuration). An example of a 5.1 channel configuration is given in Figure 9.1.

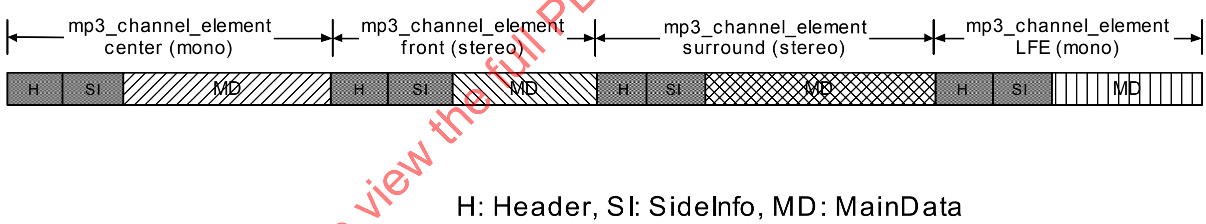


**Figure 9.1 – Access Unit containing mp3_channel_elements for a 5.1 channel configuration**

## 9.5 Sampling rate extension for Layer 3

This subsection provides specifications to allow the usage of Layer 3 with sampling rates not specified in ISO/IEC 11172-3 or ISO/IEC 13818-3.

The bitstream syntax and description for the extension towards sampling frequencies lower than those specified in ISO/IEC 13818-3 are in accordance of ISO/IEC 13818-3 (one frame covers 576 samples).

The subsequent subclauses outline the necessary extensions.

### 9.5.1 Bitrates

Table 9.1 specifies the bitrate depending on the bitrate_index and sampling frequency.

**Table 9.1 – Bitrate depending on bitrate_index and sampling frequency**

| bitrate_index | bitrate specified (kbit/s) | | |
|---|---|---|---|
| | 8, 11.025, 12 kHz | 16, 22.05, 24 kHz (see ISO/IEC 13818-3) | 32, 44.1, 48 kHz (see ISO/IEC 1aa72-3) |
| '0000' | free | free | free |
| '0001' | 8 | 8 | 32 |
| '0010' | 16 | 16 | 40 |
| '0011' | 24 | 24 | 48 |
| '0100' | 32 | 32 | 56 |
| '0101' | 40 | 40 | 64 |
| '0110' | 48 | 48 | 80 |
| '0111' | 56 | 56 | 96 |
| '1000' | 64 | 64 | 112 |
| '1001' | forbidden | 80 | 128 |
| '1010' | forbidden | 96 | 160 |
| '1011' | forbidden | 112 | 192 |
| '1100' | forbidden | 128 | 224 |
| '1101' | forbidden | 144 | 256 |
| '1110' | forbidden | 160 | 320 |
| '1111' | forbidden | forbidden | forbidden |



### 9.5.2 Samping frequency

Depending on the sampling frequency signaled in the AudioSpecificConfig, the data element sampling_frequency in the header has to be set as specified in Table 9.2.

**Table 9.1 – Setting of the data element sampling_frequency depending on the sampling frequency specified in the AudioSpecificconfig()**

| sampling_frequency | sampling frequency |
|---|---|
| 00 | 11.025 kHz and multiples thereof |
| 01 | 12 kHz and multiples thereof |
| 10 | 8 kHz and multiples thereof |
| 11 | reserved |

### 9.5.3 Padding

Padding is necessary with a sampling frequency of 11.025 kHz and multiples thereof.

### 9.5.4 Scalefactor bands

The subdivision of the spectrum into scalefactor bands is fixed for every block length and sampling frequency and stored in tables in the coder and decoder. The tables for the sampling frequencies not specified in ISO/IEC 11172-3:1993 or ISO/IEC 13818-3:1998 are specified in Annex 9.A. In accordance with ISO/IEC 11172-3 or ISO/IEC 13818-3, the scale factor for frequency lines above the highest line in the tables is zero, which means that the actual multiplication factor is 1.0.

### 9.5.5 Intensity stereo mode

Step 3 of the intensity stereo mode decoding (see ISO/IEC 11172-3, subclause 2.4.3.4.9.3) is clarified as follows:

- if only the uppermost scalefactor band is in intensity stereo mode, then is_ratio(20) = 1 for long blocks

is_ratio(11) = 1 for short blocks

- if at least the upper two scalefactor bands are in intensity stereo mode, then

is_ratio(20) = is_ratio(19) for long blocks

is_ratio(11) = is_ratio(10) for short blocks

# Annex 9.A
(normative)

## Scalefactor band tables

**Table 9.A.1 – 8 kHz sampling rate, long blocks, number of lines 576**

| scalefactor band | width of band | index_of_start | index_of_end |
|---|---|---|---|
| 0 | 12 | 0 | 11 |
| 1 | 12 | 12 | 23 |
| 2 | 12 | 24 | 35 |
| 3 | 12 | 36 | 47 |
| 4 | 12 | 48 | 59 |
| 5 | 12 | 60 | 71 |
| 6 | 16 | 72 | 87 |
| 7 | 20 | 88 | 107 |
| 8 | 24 | 108 | 131 |
| 9 | 28 | 132 | 159 |
| 10 | 32 | 160 | 191 |
| 11 | 40 | 192 | 231 |
| 12 | 48 | 232 | 279 |
| 13 | 56 | 280 | 335 |
| 14 | 64 | 336 | 399 |
| 15 | 76 | 400 | 475 |
| 16 | 90 | 476 | 565 |
| 17 | 2 | 566 | 567 |
| 18 | 2 | 568 | 569 |
| 19 | 2 | 570 | 571 |
| 20 | 2 | 572 | 573 |

**Table 9.A.2 – 8 kHz sampling rate, short blocks, number of lines 192**

| scalefactor band | width of band | index_of_start | index_of_end |
|---|---|---|---|
| 0 | 8 | 0 | 7 |
| 1 | 8 | 8 | 15 |
| 2 | 8 | 16 | 23 |
| 3 | 12 | 24 | 35 |
| 4 | 16 | 36 | 51 |
| 5 | 20 | 52 | 71 |
| 6 | 24 | 72 | 95 |
| 7 | 28 | 96 | 123 |
| 8 | 36 | 124 | 159 |
| 9 | 2 | 160 | 161 |
| 10 | 2 | 162 | 163 |
| 11 | 2 | 164 | 165 |

**Table 9.A.3 – 11.025 kHz sampling rate, long blocks, number of lines 576**

| scalefactor band | width of band | index_of_start | index_of_end |
|---|---|---|---|
| 0 | 6 | 0 | 5 |
| 1 | 6 | 6 | 11 |
| 2 | 6 | 12 | 17 |
| 3 | 6 | 18 | 23 |
| 4 | 6 | 24 | 29 |
| 5 | 6 | 30 | 35 |
| 6 | 8 | 36 | 43 |
| 7 | 10 | 44 | 53 |
| 8 | 12 | 54 | 65 |
| 9 | 14 | 66 | 79 |
| 10 | 16 | 80 | 95 |
| 11 | 20 | 96 | 115 |
| 12 | 24 | 116 | 139 |
| 13 | 28 | 140 | 167 |
| 14 | 32 | 168 | 199 |
| 15 | 38 | 200 | 237 |
| 16 | 46 | 238 | 283 |
| 17 | 52 | 284 | 335 |
| 18 | 60 | 336 | 395 |
| 19 | 68 | 396 | 463 |
| 20 | 58 | 464 | 521 |

**Table 9.A.4 – 11.025 kHz sampling rate, short blocks, number of lines 192**

| scalefactor band | width of band | index_of_start | index_of_end |
|---|---|---|---|
| 0 | 4 | 0 | 3 |
| 1 | 4 | 4 | 7 |
| 2 | 4 | 8 | 11 |
| 3 | 6 | 12 | 17 |
| 4 | 8 | 18 | 25 |
| 5 | 10 | 26 | 35 |
| 6 | 12 | 36 | 47 |
| 7 | 14 | 48 | 61 |
| 8 | 18 | 62 | 79 |
| 9 | 24 | 80 | 103 |
| 10 | 30 | 104 | 133 |
| 11 | 40 | 134 | 173 |

**Table 9.A.5 – 12 kHz sampling rate, long blocks, number of lines 576**

| scalefactor band | width of band | index_of_start | index_of_end |
|---|---|---|---|
| 0 | 6 | 0 | 5 |
| 1 | 6 | 6 | 11 |
| 2 | 6 | 12 | 17 |
| 3 | 6 | 18 | 23 |
| 4 | 6 | 24 | 29 |
| 5 | 6 | 30 | 35 |
| 6 | 8 | 36 | 43 |
| 7 | 10 | 44 | 53 |
| 8 | 12 | 54 | 65 |
| 9 | 14 | 66 | 79 |
| 10 | 16 | 80 | 95 |
| 11 | 20 | 96 | 115 |
| 12 | 24 | 116 | 139 |
| 13 | 28 | 140 | 167 |
| 14 | 32 | 168 | 199 |
| 15 | 38 | 200 | 237 |
| 16 | 46 | 238 | 283 |
| 17 | 52 | 284 | 335 |
| 18 | 60 | 336 | 395 |
| 19 | 68 | 396 | 463 |
| 20 | 58 | 464 | 521 |

**Table 9.A.6 – 12 kHz sampling rate, short blocks, number of lines 192**

| scalefactor band | width of band | index_of_start | index_of_end |
|---|---|---|---|
| 0 | 4 | 0 | 3 |
| 1 | 4 | 4 | 7 |
| 2 | 4 | 8 | 11 |
| 3 | 6 | 12 | 17 |
| 4 | 8 | 18 | 25 |
| 5 | 10 | 26 | 35 |
| 6 | 12 | 36 | 47 |
| 7 | 14 | 48 | 61 |
| 8 | 18 | 62 | 79 |
| 9 | 24 | 80 | 103 |
| 10 | 30 | 104 | 133 |
| 11 | 40 | 134 | 173 |

# Annex 9.B
(informative)

## Converting MPEG-1/2 Layer 3 bitstreams into mp3_channel_elements

The use of the bit-reservoir usually causes the start of the main_data() to appear in a past bitstream frame. This needs to be modified by moving the main_data() immediately adjacent to its side information. This is illustrated for one Layer 3 bitstream (mono or stereo) in Figure 9.B.1. Each resulting mp3_channel_element is mapped directly into an access unit. Resulting header and side info are indicated by H' and SI' respectively.

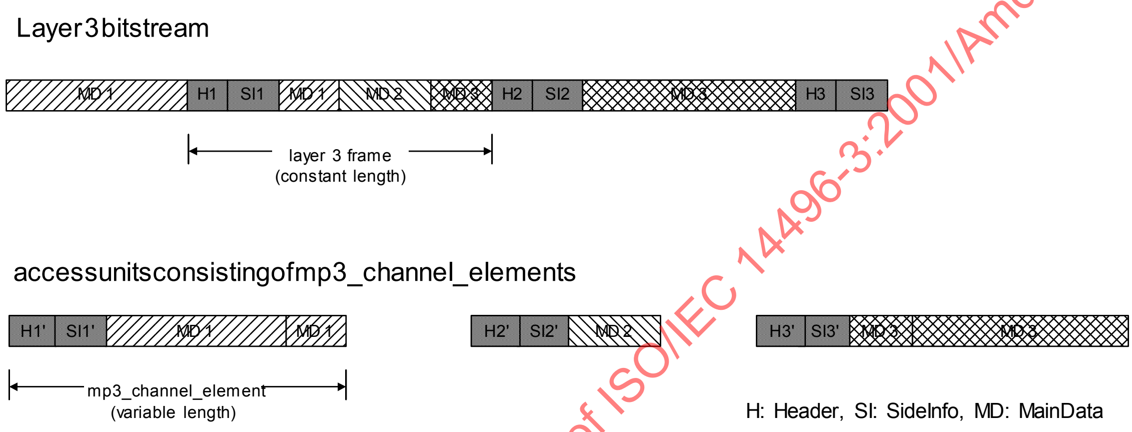


**Figure 9.B.1 – Converting an MPEG-1/2 Layer 3 bitstream into mp3_channel_elements**

All data elements of the header() shall be preserved. The data element main_data_begin might be set to zero. In that case the CRC has to be recalculated.

# Annex 9.C
(informative)

# Converting mp3_channel_elements into MPEG-1/2 Layer 3 bitstreams

## 9.C.1 Overview

mp3_channel_elements extracted from an access unit have to undergo the following conversion operations in order to obtain MPEG-1/2 Layer 3 audio bitstreams compliant to ISO/IEC 11172-3 or ISO/IEC 13818-3:

- for each mp3_channel_element per access unit open a decoder instance or output stream
- for each mp3_channel_element in every access unit do
- restore syncword and IDex
- correct bitrate_index
- adjust main_data_begin
- recalculate crc_word
- reconstruct framing

## 9.C.2 Sampling rate signaling

The last bit of the syncword shall be used in a backwards compatible way to allow the signalling of sampling rates not specified in ISO/IEC 11172-3 or ISO/IEC 13818-3. This leads to the following syntax modification:

| Syntax | No. of bits | Mnemic |
|---|---|---|
| header() | | |
| { | | |
| **syncword** | **11** | **bslbf** |
| **IDex** | **1** | **bslbf** |
| ... | | |

**syncword** The bit string '1111 1111 111'.

**IDex** One bit to indicate the extended ID of the algorithm. Has the value '0' for sampling rates not specified in ISO/IEC 11172-3 or ISO/IEC 13818-3.

The following table specifies the sampling rate depending on the values for IDex and ID:

| IDex | ID | sampling rate |
|---|---|---|
| 0 | 0 | 8, 11.025, 12 kHz |
| 1 | 0 | 16, 22.05, 24 kHz (see ISO/IEC 13818-3) |
| 1 | 1 | 32, 44.1, 48 kHz (see ISO/IEC 11172-3) |

### 9.C.3 Reconstruction instructions

This reconstruction process offers some degrees of freedom:

a) bitrate_index (stuffing might be required to adjust the bitstream frame length according to the new bitrate_index, sampling frequency and padding_bit settings)

   1) set to the maximum value allowed (signalling the maximum allowed bitstream frame length).

   2) set to nearest higher value that matches the mp3_channel_element length.

   3) set to nearest higher value that matches the mp3_channel_element length minus main_data_begin of the current audio frame.

b) main_data_begin

   1) set to zero.

   2) set to the value pointing back to the end of main_data of the previous audio frame.

   3) set to the correct value of the corresponding MPEG-1/2 Layer 3 bitstream.

c) location of stuffing

   1) at end of main_data: preserves ancillary data written in forward direction, starting after the last Huffman code word.

   2) at the end of the last Huffman codeword (location can be calculated using the part_2_3_length): preserves ancillary data written in reverse direction starting before the main_data of the next frame.

   3) no stuffing required: preserves any ancillary data.

Depending on bitrate requirements and ancillary data handling, these possibilities can be combined in several ways:

The simplest method sets the bitrate to the maximum size. This is the preferred method when feeding existing MPEG-1/2 Layer 3 decoders. main_data_begin is set to zero. Stuffing bits are added either before or after ancillary data (see Figure 9.C.1, example A and B).

A more advanced method can be derived from this simple method by setting the bitrate_index to the nearest higher value that corresponds to the length of the mp3_channel_element. With this modification, the bitrate can be significantly reduced (see Figure 9.C.1, example C and D).

To avoid the necessity of stuffing, main_data_begin is set to the value pointing back to the end of main_data of the previous frame. The bitrate_index is now set to the nearest higher value that matches the mp3_channel_element length minus main_data_begin of the current audio frame (see Figure 9.C.1, example E in Figure 9.C.1). Only if main_data_begin would exceed the allowed value, stuffing has to be performed.

The original Layer 3 bitstream can be reconstructed perfectly, if the correct main_data_begin value of the corresponding MPEG-1/2 Layer 3 bitstream was preserved (see Figure 9.C.1, example F).