INTERNATIONAL STANDARD ISO/IEC 23003-1:2007
TECHNICAL CORRIGENDUM 2

Published 2009-10-01

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION • МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ ПО СТАНДАРТИЗАЦИИ • ORGANISATION INTERNATIONALE DE NORMALISATION
INTERNATIONAL ELECTROTECHNICAL COMMISSION • МЕЖДУНАРОДНАЯ ЭЛЕКТРОТЕХНИЧЕСКАЯ КОМИССИЯ • COMMISSION ÉLECTROTECHNIQUE INTERNATIONALE

Information technology — MPEG audio technologies —

Part 1:
MPEG Surround

TECHNICAL CORRIGENDUM 2

Technologies de l'information — Technologies audio MPEG —

Partie 1: Ambiance MPEG

RECTIFICATIF TECHNIQUE 2

Technical Corrigendum 2 to ISO/IEC 23003-1:2007 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

Throughout this Technical Corrigendum, changes to existing text and tables are highlighted using a grey background.

In the first paragraph of 3.5, replace:

$\mathbf{a}^m(l)$ aliasing condition vector defined for every parameter time slot l and all QMF subbands m that are the last subband (highest in frequency) within a parameter band.

with:

$\mathbf{a}^m(l)$ aliasing condition vector defined for every parameter set l and all QMF subbands m that are the last subband (highest in frequency) within a parameter band.

ICS 35.040 **Ref. No. ISO/IEC 23003-1:2007/Cor.2:2009(E)**

In the first paragraph of 3.5, replace:

$\mathbf{r}^m(l)$ weighted correlation sum based on the input downmix signal, defined for every parameter time slot l and all QMF subbands m that have an adjoining parameter border, used for Low Power MPEG surround.

with:

$\mathbf{r}^m(l)$ weighted correlation sum based on the input downmix signal, defined for every parameter set l and all QMF subbands m that have an adjoining parameter border, used for Low Power MPEG surround.

In 4.3.2, replace the title of Table 2:

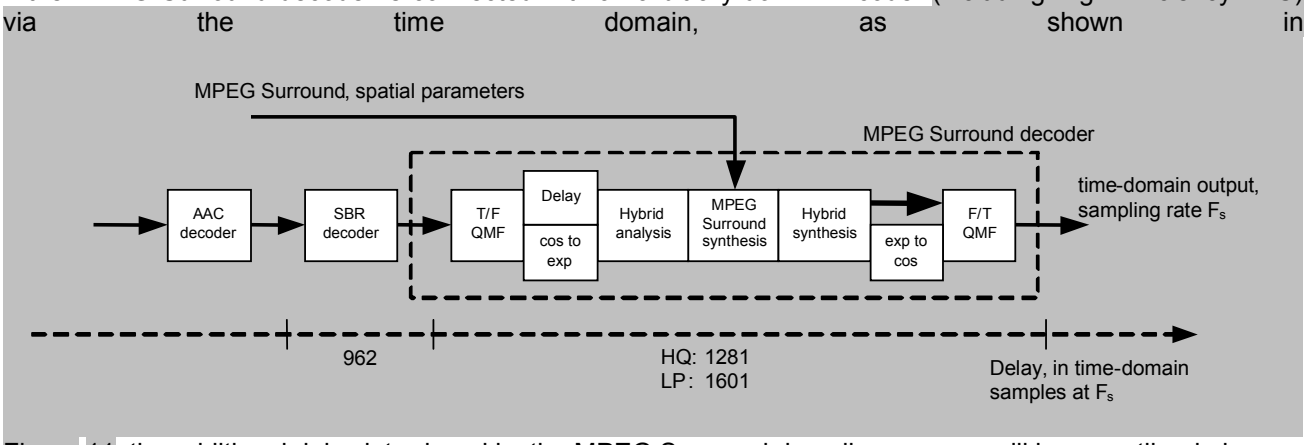
Table 2 — Outline of difference between the High Quality and the Low Power MEG Surround system

with:

Table 2 — Outline of difference between the High Quality and the Low Power MPEG Surround system

In 4.5, replace the third paragraph and the caption of Figure 11:

If the MPEG Surround decoder is connected with an arbitrary downmix coder (including High Efficiency AAC) via the time domain, as shown in Figure 11, the additional delay introduced by the MPEG Surround decoding process will be as outlined above.



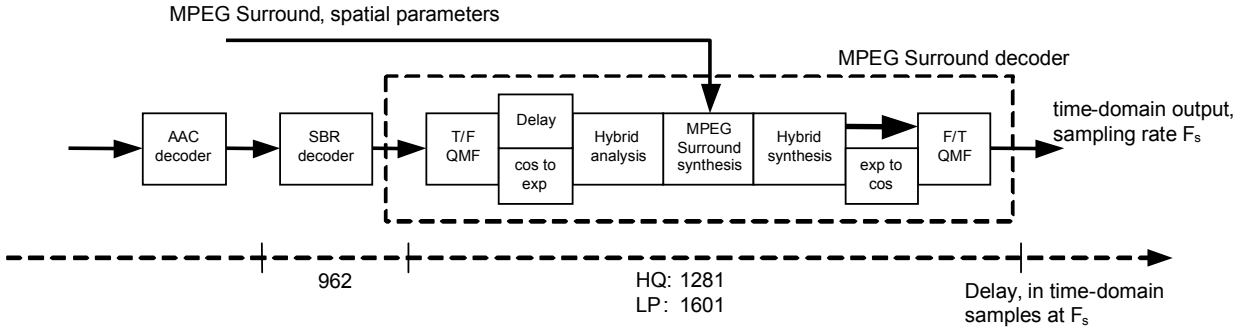
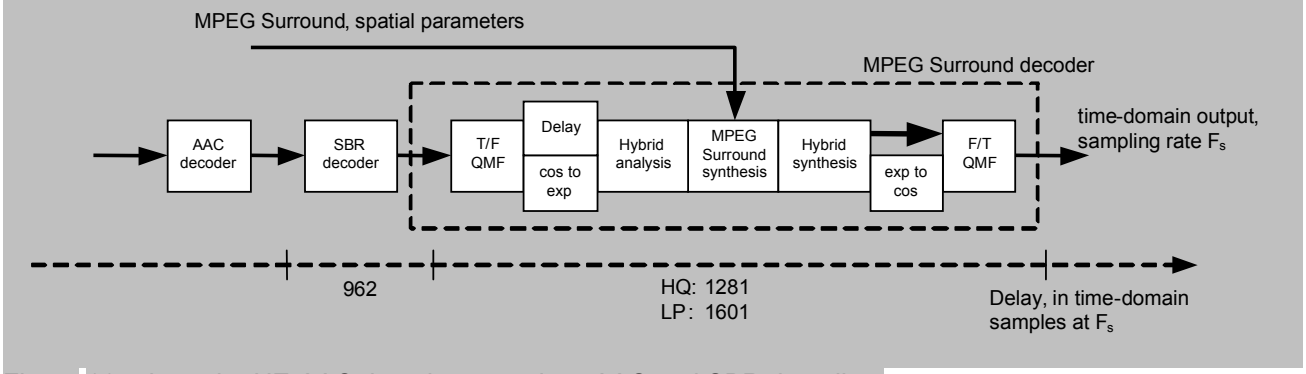


Figure 11 — Delay when connecting MPEG Surround in the time-domain for arbitrary core codec (including HE-AAC)

with:

If the MPEG Surround decoder is connected with an arbitrary downmix coder, the additional delay introduced by the MPEG Surround decoding process will be as outlined above. The connection of MPEG Surround with a High Efficiency AAC downmix coder in the time domain is shown in Figure 11, where the HE-AAC decoder comprises AAC and SBR decoding.



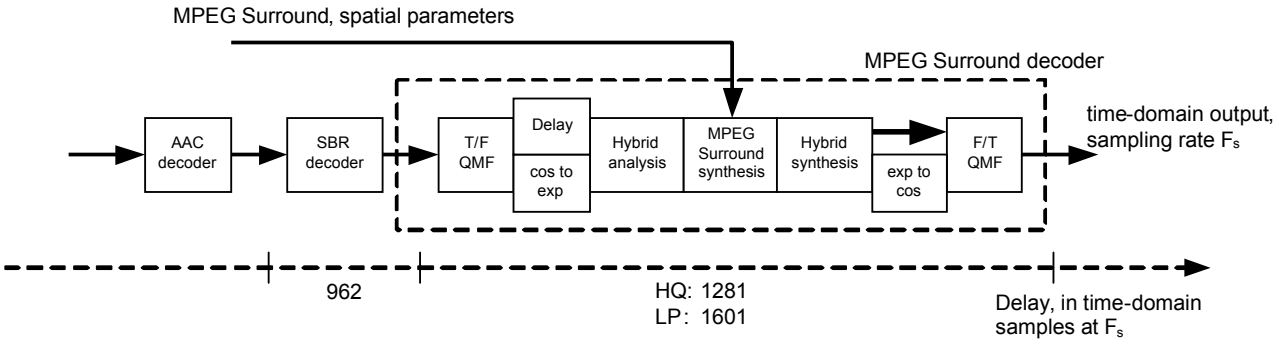


Figure 11 — Delay when connecting MPEG Surround with HE-AAC in the time domain

In 4.5, replace the title of Figure 12:

Figure 12 — Delay when connecting MPEG Surround with HE-AAC in the QMF-domain

with:

Figure 12 — Delay when connecting MPEG Surround with HE-AAC in the QMF domain

In 4.5, replace the fifth paragraph:

Transmission of MPEG Surround side information with respect to transmission of the coded downmix signal is done in such a manner that there is no need to delay the downmix signal before it is processed by the MPEG Surround decoder. This means that MPEG Surround data is conveyed such that it is available when needed by the MPEG Surround decoding process. The temporal relationship between downmix data and spatial data is defined in Clause 6. Note that special consideration is required if an MPEG Surround decoder and an HE-AAC decoder are connected in the time domain while a connection in the QMF domain would have been possible according to subclause 4.4. In this case, the spatial parameters have to be delayed by 961 time samples, which is the sum of 257 samples for HE-AAC QMF synthesis and 704 samples for MPEG Surround QMF and Nyquist analysis.

with:

Transmission of MPEG Surround side information with respect to transmission of the coded downmix signal is preferably done in such a manner that there is no need to delay the downmix signal before it is processed by the MPEG Surround decoder. This means that MPEG Surround data is preferably conveyed such that it is available when needed by the MPEG Surround decoding process. The temporal relationship between downmix data and spatial data is defined in Clause 7. For the Baseline MPEG Surround Profile defined in 4.7.2, restrictions for the temporal relationship are specified. Note that special consideration is required if an MPEG Surround decoder and an HE-AAC decoder are connected in the time domain while a connection in the QMF domain would have been possible according to 4.4. In this case, the spatial parameters have to be delayed by 961 time samples, which is the sum of 257 samples for HE-AAC QMF synthesis and 704 samples for MPEG Surround QMF and Nyquist analysis.

In 4.7.2, Note 3 at the bottom of Table 4, replace:

Note 3: A low power decoder utilizes only residual coding data for the first 8 QMF bands, corresponding to approximately 2.7 kHz bandwidth.

with:

Note 3: A low power decoder utilizes only residual coding data for the first 7 QMF bands, corresponding to approximately 2.4 kHz bandwidth at a sampling frequency of 44.1 kHz.

At the end of 4.7.2, add the following new paragraphs and table:

If a Baseline MPEG Surround Profile decoder is used in combination with one of the MPEG downmix coders listed in Table 4A, the following additional restrictions apply. The MPEG Surround frame length must be an integer multiple of the downmix coder frame length, i.e. bsFrameLength must have one of the allowed values listed in Table 4A (see 6.3.3 for details of downsampled or upsampled operation of MPEG Surround). Furthermore, in the case that MPEG Surround data is embedded in a downmix bitstream (as defined in 7.2.3 and 7.2.4), the temporal relationship between downmix data and spatial data must be such that the sacTimeAlign parameter (defined in 7.2.5) has the value 0. In the case that MPEG Surround data and downmix data are conveyed in separate streams (as defined in 7.2.2), the temporal relationship must be such that the time stamp of an MPEG Surround access unit must be the same as the time stamp of an downmix access unit.

Table 4A — Allowed values for *bsFrameLength* in the Baseline MPEG Surround Profile when used in combination with MPEG downmix coders

Downmix coder	Downmix coder frame length (QMF samples)	Allowed values for *bsFrameLength*
AAC 1024	16	15, 31, 47, 63
- with downsampled MPEG Surround	32	31, 63
- with upsampled MPEG Surround	8	7, 15, 23, 31, 39, 47, 55, 63, 71
AAC 960	15	14, 29, 44, 59
- with downsampled MPEG Surround	30	29, 59
- with upsampled MPEG Surround	7.5	14, 29, 44, 59
HE-AAC 1024/2048	32	31, 63
- with downsampled MPEG Surround	64	63
- with upsampled MPEG Surround	16	15, 31, 47, 63
HE-AAC 960/1920	30	29, 59
- with downsampled MPEG Surround	60	59
- with upsampled MPEG Surround	15	14, 29, 44, 59
BSAC	16	15, 31, 47, 63
- with downsampled MPEG Surround	32	31, 63
- with upsampled MPEG Surround	8	7, 15, 23, 31, 39, 47, 55, 63, 71
BSAC with SBR	32	31, 63
- with downsampled MPEG Surround	64	63
- with upsampled MPEG Surround	16	15, 31, 47, 63
AAC LD 512	8	7, 15, 23, 31, 39, 47, 55, 63, 71
- with downsampled MPEG Surround	16	15, 31, 47, 63
AAC ELD 512	8	7, 15, 23, 31, 39, 47, 55, 63, 71
- with downsampled MPEG Surround	16	15, 31, 47, 63
AAC ELD with SBR 512/1024	16	15, 31, 47, 63
- with downsampled MPEG Surround	32	31, 63
- with upsampled MPEG Surround	8	7, 15, 23, 31, 39, 47, 55, 63, 71
MPEG1/2 Layer II	18	17, 35, 53, 71
- with downsampled MPEG Surround	36	35, 71
MPEG1/2 Layer III	18	17, 35, 53, 71
- with downsampled MPEG Surround	36	35, 71

If a Baseline MPEG Surround Profile decoder is used to decode MPEG Surround data conveyed as buried data in a PCM downmix signal (as defined in 7.3), the temporal relationship between downmix data and spatial data must be such that the sacTimeAlign parameter (defined in 7.2.5) either has the value 0 or has the value -N*(bsFrameLength+1), where N is the number of QMF bands used in MPEG Surround, i.e. N=64 for normal operation, N=32 for downsampled operation, or N=128 for upsampled operation (see 6.3.3).

In the first paragraph of 6.1.2.3.2, replace:

```
case DIFF_TIME:
    if ( (pg > 0) || (mixedTimePairXXX[pi][setIdx]) ) {
```

with:

```
case DIFF_TIME:
    if ( (pg > 0) || !(mixedTimePairXXX[pi][setIdx]) ) {
```

In the first paragraph of 6.10.3.2, replace:

parameter time slot l

with:

parameter set l

In 6.1.13, replace:

The allowed values for *bsResidualSamplingFrequencyIndex* or *bsArbitraryDownmixResidualSamplingFrequencyIndex* depend on *bsFrameLength*, *bsSamplingFrequencyIndex* and *bsResidualFramesPerSpatialFrame* or *bsArbitraryDownmixResidualFramesPerSpatialFrame*, respectively, as shown in Table 88.

Table 88 — Allowed combinations of *bsSamplingFrequencyIndex* and *bsResidualSamplingFrequencyIndex* or *bsArbitraryDownmixResidualSamplingFrequencyIndex*

(*bsFrameLength***+1)/ (***bsResidualFramesPerSpatialFrame***+1) or (***bsFrameLength***+1)/ (***bsArbitraryDownmixResidualFramesPerSpatialFrame***+1)**	**Allowed combinations of {***bsSamplingFrequencyIndex***, *bsResidualSamplingFrequencyIndex***} or {***bsSamplingFrequencyIndex***, *bsArbitraryDownmixResidualSamplingFrequencyIndex***}**
15	{0x0, 0x0}, {0x1, 0x1}, {0x2, 0x2}, {0x3, 0x3}, {0x4, 0x4}, {0x5, 0x5}, {0x6, 0x6}, {0x7, 0x7}, {0x8, 0x8}, {0x9, 0x9}, {0xa, 0xa}, and {0xb, 0xb}
16	{0x0, 0x0}, {0x1, 0x1}, {0x2, 0x2}, {0x3, 0x3}, {0x4, 0x4}, {0x5, 0x5}, {0x6, 0x6}, {0x7, 0x7}, {0x8, 0x8}, {0x9, 0x9}, {0xa, 0xa}, and {0xb, 0xb}
18	{0x0, 0x2}, {0x1, 0x2}, {0x2, 0x3}, {0x3, 0x5}, {0x4, 0x5}, {0x5, 0x6}, {0x6, 0x8}, {0x7, 0x8}, {0x8, 0x9}, {0x9, 0xb}, and {0xa, 0xb}
24	{0x0, 0x3}, {0x1, 0x3}, {0x2, 0x5}, {0x3, 0x5}, {0x4, 0x5}, {0x5, 0x8}, {0x6, 0x9}, {0x7, 0x9}, and {0x8, 0xb}
30	{0x0, 0x3}, {0x1, 0x3}, {0x2, 0x5}, {0x3, 0x7}, {0x4, 0x7}, {0x5, 0x8}, {0x6, 0x9}, {0x7, 0x9}, and {0x8, 0xb}
32	{0x0, 0x3}, {0x1, 0x4}, {0x2, 0x5}, {0x3, 0x6}, {0x4, 0x7}, {0x5, 0x8}, {0x6, 0x9}, {0x7, 0xa}, {0x8, 0xb}, {0x9, 0xb}, {0xa, 0xb}, and {0xb, 0xb}

with:

The allowed values for *bsResidualSamplingFrequencyIndex* or *bsArbitraryDownmixResidualSamplingFrequencyIndex* are shown in Table 88.

Table 88 — Allowed values of *bsResidualSamplingFrequencyIndex* or *bsArbitraryDownmixResidualSamplingFrequencyIndex*

Parameter	**Allowed values**
bsResidualSamplingFrequencyIndex	0x3, 0x4, 0x5, 0x6, 0x7, 0x8, 0x9, 0xa, 0xb
bsArbitraryDownmixResidualSamplingFrequencyIndex	0x3, 0x4, 0x5, 0x6, 0x7, 0x8, 0x9, 0xa, 0xb

In 6.3.3, replace:

$$\exp\left(i\frac{pi}{256}(k+0.5)(2n-2)\right),\quad 0\le k<128, 0\le n<256$$

with:

$$0.5\cdot\exp\left(i\frac{pi}{256}(k+0.5)(2n-2)\right),\quad 0\le k<128, 0\le n<256$$

and replace:

using a 1280 sample version of the window function c[i] where the additional intermediate samples are obtained by linear interpolation of neighboring samples of the original 640 sample window function specified in 14496-3 subclause 4.A.6.2Table 4.A.87.

with:

replacing the window function $c[i]$ by a 1280 sample version $c_{128}[i]$, which is obtained from the original 640 sample window function specified in ISO/IEC 14496-3:2009, Table 4.A.87 according to:

$$c_{128}[i]=\begin{cases} c[i/2] & \text{if } i \text{ even} \\ \left(c[(i-1)/2]-c[(i+1)/2]\right)/2 & \text{if } i\in\{255,511,767,1023\} \\ c[(i-1)/2]/2 & \text{if } i=1279 \\ \left(c[(i-1)/2]+c[(i+1)/2]\right)/2 & \text{else} \end{cases}$$

and replace:

$$\exp\left(i\frac{pi}{256}(k+0.5)(2n-510)\right),\quad 0\le k<128, 0\le n<256$$

with:

$$\frac{1}{64}\exp\left(i\frac{pi}{256}(k+0.5)(2n-510)\right),\quad 0\le k<128, 0\le n<256$$

In the first paragraph of 6.4.3.2.1, replace:

configuration 6 rows and 3 columns, according to:

with:

configuration 6 rows and 5 columns, according to:

In the first paragraph of 6.4.4.2.1, replace:

configuration 8 rows and 3 columns, according to:

with:

configuration 8 rows and 5 columns, according to:

In 6.9.2.5.1, replace:

$$L_{\text{qmf}} = \begin{cases} \max\left(64, ceil\left(\dfrac{1024}{N_{\text{qmf,LONG}}}\right)\right) & \text{for long windows} \\ \max\left(64, ceil\left(\dfrac{128}{N_{\text{qmf,SHORT}}}\right)\right) & \text{for short windows.} \end{cases}$$

with:

$$L_{\text{qmf}} = \begin{cases} \min\left(64, ceil\left(\dfrac{1024}{N_{\text{qmf,LONG}}}\right)\right) & \text{for long windows} \\ \min\left(64, ceil\left(\dfrac{128}{N_{\text{qmf,SHORT}}}\right)\right) & \text{for short windows.} \end{cases}$$

In 6.10.2.2, replace:

$$\mathbf{M}_r(k,n) = 2 \cdot \cos\left(\frac{\pi \cdot (k+0.5) \cdot (2 \cdot n - 192)}{128}\right) \quad , \begin{cases} 0 \le k < 64 \\ 0 \le n < 128 \end{cases}$$

with:

$$\mathbf{M}_r(k,n) = \cos\left(\frac{\pi \cdot (k+0.5) \cdot (2 \cdot n - 192)}{128}\right) \quad , \begin{cases} 0 \le k < 64 \\ 0 \le n < 128 \end{cases}$$

In 6.10.2.5, replace:

$$\mathbf{M}_r(k,n) = 2 \cdot \cos\left(\frac{\pi \cdot (k+0.5) \cdot (2 \cdot n - 384)}{256}\right) \quad , \begin{cases} 0 \le k < 128 \\ 0 \le n < 256 \end{cases}$$

with:

$$\mathbf{M}_r(k,n) = 0.5 \cdot \cos\left(\frac{\pi \cdot (k+0.5) \cdot (2 \cdot n - 384)}{256}\right) \quad , \begin{cases} 0 \le k < 128 \\ 0 \le n < 256 \end{cases}$$

and replace:

using a 1280 sample version of the window function c[i] where the additional intermediate samples are obtained by linear interpolation of neighboring samples of the original 640 sample window function specified in 14496-3 subclause 4.A.6.2Table 4.A.87.

with:

replacing the window function $c[i]$ by the 1280 sample version $c_{128}[i]$, which is defined in 6.3.3.

In 6.11.4.2.2.2, replace:

where $w_{\mathrm{L}}^{l,m} = \frac{\left(\sigma_{\mathrm{L}}^{l,m}\right)^2}{\left(\sigma_{\mathrm{L}}^{l,m}\right)^2 + \left(\sigma_{\mathrm{Ls}}^{l,m}\right)^2}$, $w_{\mathrm{Ls}}^{l,m} = \frac{\left(\sigma_{\mathrm{Ls}}^{l,m}\right)^2}{\left(\sigma_{\mathrm{L}}^{l,m}\right)^2 + \left(\sigma_{\mathrm{Ls}}^{l,m}\right)^2}$, $w_{\mathrm{R}}^{l,m} = \frac{\left(\sigma_{\mathrm{R}}^{l,m}\right)^2}{\left(\sigma_{\mathrm{R}}^{l,m}\right)^2 + \left(\sigma_{\mathrm{Rs}}^{l,m}\right)^2}$, $w_{\mathrm{Rs}}^{l,m} = \frac{\left(\sigma_{\mathrm{Rs}}^{l,m}\right)^2}{\left(\sigma_{\mathrm{R}}^{l,m}\right)^2 + \left(\sigma_{\mathrm{Rs}}^{l,m}\right)^2}$ and where g_c is the down mix gain for the centre channel.

with:

where $w_{\mathrm{L}}^{l,m} = \frac{\left(\sigma_{\mathrm{L}}^{l,m}\right)^2 \left(P_{\mathrm{R,L}}^{m}\right)^2}{\left(\sigma_{\mathrm{L}}^{l,m}\right)^2 \left(P_{\mathrm{R,L}}^{m}\right)^2 + \left(\sigma_{\mathrm{Ls}}^{l,m}\right)^2 \left(P_{\mathrm{R,Ls}}^{m}\right)^2}$, $w_{\mathrm{Ls}}^{l,m} = \frac{\left(\sigma_{\mathrm{Ls}}^{l,m}\right)^2 \left(P_{\mathrm{R,Ls}}^{m}\right)^2}{\left(\sigma_{\mathrm{L}}^{l,m}\right)^2 \left(P_{\mathrm{R,L}}^{m}\right)^2 + \left(\sigma_{\mathrm{Ls}}^{l,m}\right)^2 \left(P_{\mathrm{R,Ls}}^{m}\right)^2}$,

$w_{\mathrm{R}}^{l,m} = \frac{\left(\sigma_{\mathrm{R}}^{l,m}\right)^2 \left(P_{\mathrm{L,R}}^{m}\right)^2}{\left(\sigma_{\mathrm{R}}^{l,m}\right)^2 \left(P_{\mathrm{L,R}}^{m}\right)^2 + \left(\sigma_{\mathrm{Rs}}^{l,m}\right)^2 \left(P_{\mathrm{L,Rs}}^{m}\right)^2}$, $w_{\mathrm{Rs}}^{l,m} = \frac{\left(\sigma_{\mathrm{Rs}}^{l,m}\right)^2 \left(P_{\mathrm{L,Rs}}^{m}\right)^2}{\left(\sigma_{\mathrm{R}}^{l,m}\right)^2 \left(P_{\mathrm{L,R}}^{m}\right)^2 + \left(\sigma_{\mathrm{Rs}}^{l,m}\right)^2 \left(P_{\mathrm{L,Rs}}^{m}\right)^2}$.

In 7.2.1, replace:

The spatial frame length is preferred to be an integer multiple of the frame length of the underlying downmix coder. Asynchronous framing of spatial data and the downmix data (i.e., different frame lengths) is possible. However, in this case, additional buffering of the spatial data in the decoder might be needed.
In general spatial data is conveyed in such a manner that it is available to the MPEG Surround decoder in time when it is required to process the decoded downmix signals, assuming the most efficient connection of downmix decoder to the MPEG Surround decoder. This is a direct connection of HE-AAC and MPEG Surround in the QMF domain in case of MPEG Surround using normal operation (as opposed to upsampled or downsampled operation as defined in subclause 6.3.3), and a connection in the PCM time domain in all other cases. When HE-AAC and MPEG Surround are connected in the time domain even though the most efficient connection would have been in the QMF domain, the spatial parameters have to be delayed accordingly in order to maintain the time alignment between spatial data and downmix data. Information about this delay is given in subclause 6.4.1.
In the case that the spatial data is embedded in the downmix data stream (see subclause 7.2.2, 7.2.3, and 7.2.4), the temporal relationship between spatial frames and downmix frames is indicated by the value of sacTimeAlign (see subclause 7.2.5).
In the case that the downmix data and the spatial data are conveyed in separate streams, the temporal relationship between spatial frames and downmix frames is indicated by the time stamps of the corresponding streams. If the transport layer does not provide time stamps (as e.g. in case of LATM), the transport layer needs to define the temporal relationship between the data of these both streams by other means.

with:

The spatial frame length is preferred to be an integer multiple of the frame length of the underlying downmix coder. Asynchronous framing of spatial data and the downmix data (i.e. different frame lengths) is possible. However, in this case, additional buffering of the spatial data in the decoder might be needed.
In general, spatial data is preferably conveyed in such a manner that it is available to the MPEG Surround decoder in time when it is required to process the decoded downmix signals, assuming the most efficient connection of downmix decoder to the MPEG Surround decoder. This is a direct connection of HE-AAC and MPEG Surround in the QMF domain in the case that both use the same number of OMF bands (see 4.4), and a connection in the PCM time domain in all other cases. When HE-AAC and MPEG Surround are connected in the time domain even though the most efficient connection would have been in the QMF domain, the spatial parameters have to be delayed accordingly in order to maintain the time alignment between spatial data and downmix data. Information about this delay is given in 4.5.
In the case that the spatial data is embedded in the downmix data stream (see 7.2.2, 7.2.3, and 7.2.4), the temporal relationship between spatial frames and downmix frames is indicated by the value of sacTimeAlign (see 7.2.5). If sacTimeAlign has the value 0, this indicates that the spatial data is conveyed in the preferred manner outlined above.
In the case that the downmix data and the spatial data are conveyed in separate streams, the temporal relationship between spatial frames and downmix frames is indicated by the time stamps of the access units of the corresponding streams. If a downmix coder other than HE-AAC is used, the time stamp of an access unit carrying an SAC frame identifies the first PCM sample of the corresponding time domain downmix signal

frame that is input to the MPEG Surround decoder. If HE-AAC is used as downmix coder, the time stamp of the SAC frame identifies the first PCM sample of the corresponding time domain downmix signal frame at the output of the AAC core decoder.
If the transport layer does not provide time stamps, the temporal relationship between the data of these both streams needs to be defined by other means. In case of LATM (see ISO/IEC 14496-3), the first MPEG Surround access unit and the first downmix coder access unit in an AudioMuxElement() are considered to have the same time stamp.

In 7.2.2, replace:

The transmission of spatial audio data requires a spatial elementary stream that depends on the elementary stream containing the related coded audio downmix data. The actual spatial data is either conveyed in the spatial elementary stream or multiplexed into the downmix data stored in the elementary stream upon which the spatial elementary stream depends. The latter is specified for MPEG-2/4 AAC payloads (see subclause 7.2.3) and for MPEG-1/2 Layer I/II/III payloads (see subclause 7.2.4).
Backwards compatibility with decoders that can decode the coded audio downmix data but not the spatial audio data is achieved in both scenarios.
If the downmix signal is encoded with the combination of a mono AAC downmix coder and SBR bandwidth extension, it is possible that both data for the MPEG-4 Parametric Stereo (PS) tool as well as data for MPEG Surround in a 5-1-5 configuration is present simultaneously in the bitstream conveying the downmix signal. Such a bitstream can be decoded into a 2 channel stereo signal in accordance with the MPEG-4 HE-AAC v2 Profile, whereby the MPEG Surround data is ignored. When such a bitstream is used in combination with an MPEG Surround decoder, the PS data in the bitstream is ignored and the downmix bitstream is decoded in accordance with the MPEG-4 HE-AAC profile. This provides a mono downmix signal in the QMF domain that is used as input to the subsequent MPEG Surround decoding process for a 5-1-5 configuration as described in subclause 6.4.2.
The interface to ISO/IEC 14496-1 is in line with the specification given in ISO/IEC 14496-3 subclause 1.6. An elementary stream carrying spatial audio data is identified by the Audio Object Type “MPEG Surround” (Object Type ID 30). The AudioSpecificConfig() for this object carries the SpatialSpecificConfig() data and a sacPayloadEmbedding flag that indicates whether the SpatialFrame() payload is conveyed as an elementary stream or embedded into the downmix data, as defined in ISO/IEC 14496-3 subclause 1.6.3.14.

with:

The signaling of the availability of spatial audio data is possible either by means of a spatial elementary stream that depends on the elementary stream containing the related coded audio downmix data (as, e.g., indicated by the dependsOn_ES_ID field defined in ISO/IEC 14496-1:2004, 6.5.2) or by means of including the SpatialSpecificConfig() at the end of the AudioSpecificConfig() of the downmix elementary stream in a backward compatible way (as defined in ISO/IEC 14496-3:2009, 1.6). The actual spatial data is either conveyed in the spatial elementary stream or multiplexed into the downmix data stored in the elementary stream upon which the spatial elementary stream, if present, depends. The latter is specified for MPEG-2/4 AAC payloads (see 7.2.3) and for MPEG-1/2 Layer I/II/III payloads (see 7.2.4).
Backwards compatibility with decoders that can decode the coded audio downmix data but not the spatial audio data is achieved in all these scenarios.
If the downmix signal is encoded with the combination of a mono AAC downmix coder and SBR bandwidth extension, it is possible that both data for the MPEG-4 Parametric Stereo (PS) tool as well as data for MPEG Surround in a 5-1-5 configuration is present simultaneously in the bitstream conveying the downmix signal. Such a bitstream can be decoded into a 2 channel stereo signal in accordance with the MPEG-4 HE-AAC v2 Profile, whereby the MPEG Surround data is ignored. When such a bitstream is used in combination with an MPEG Surround decoder, the PS data in the bitstream is ignored and the downmix bitstream is decoded in accordance with the MPEG-4 HE-AAC profile. This provides a mono downmix signal in the QMF domain that is used as input to the subsequent MPEG Surround decoding process for a 5-1-5 configuration as described in 6.4.2.
The interface to ISO/IEC 14496-1 is in line with the specification given in 1.6 of ISO/IEC 14496-3. An elementary stream carrying spatial audio data is identified by the Audio Object Type “MPEG Surround” (Object Type ID 30). The AudioSpecificConfig() for this object carries the SpatialSpecificConfig() data and a sacPayloadEmbedding flag that indicates whether the SpatialFrame() payload is conveyed as an elementary stream or embedded into the downmix data, as defined in of ISO/IEC 14496-3:2009, 1.6.3.17.

In 7.2.3, replace:

Spatial audio data can be conveyed in the AAC extension_payload() mechanism using extension_type EXT_SAC_DATA ("1100"), as defined in ISO/IEC 13818-7 subclause 8.8 and ISO/IEC 14496-3 subclause 4.5.2.9. The extension_payload() for type EXT_SAC_DATA is used to carry a SacDataFrame(), complete or split into several fragments, using the same syntax elements ancType, ancStart, and ancStop as defined in the next subclause.

with:

Spatial audio data can be conveyed in the AAC extension_payload() mechanism using extension_type EXT_SAC_DATA ("1100"), as defined in ISO/IEC 13818-7:2006, 8.8 and ISO/IEC 14496-3:2009, 4.5.2.9. The extension_payload() for type EXT_SAC_DATA comprises the sac_extension_data(), as defined in ISO/IEC 13818-7:2006, 6.3 and ISO/IEC 14496-3:2009, 4.4.2.7, which is used to carry a SacDataFrame(), complete or split into several fragments, using the same syntax elements ancType, ancStart, ancStop, and ancDataSegmentByte as defined in 7.2.4, and where in the semantics of the syntax element ancDataSegmentByte, the term AncDataElement is to be replaced by sac_extension_data.

In 7.2.5, replace:

sacTimeAlign Identifies the PCM sample in the output frame of the downmix decoder that corresponds to the beginning of the present SAC frame. The position of the first sample of the output frame is represented as 0. The present SAC frame is the first SAC frame that is completed (i.e., ancStop==1) in the present downmix decoder frame.

with:

sacTimeAlign Identifies the PCM sample in the time domain output frame of the downmix decoder that corresponds to the beginning of the present SAC frame (i.e. the first sample of the time domain input signal that is consumed by the MPEG Surround decoding process for the present SAC frame). The position of the first sample of the output frame is represented as 0. The present SAC frame is the first SAC frame that is completed (i.e. ancStop==1) in the present downmix decoder frame. If HE-AAC is used as downmix coder, the time domain output frame of the AAC core decoder (delay-free upsampled by a factor of two in case of normal operation of MPEG Surround with 64 QMF bands) is considered here.

In 7.3.3, replace Table 114 with the following

Table 114 — bsBDType

bsBDType	Type of data
0	MPEG Surround frame, i.e. SacDataFrame(0)
1	MPEG Surround header+frame, i.e. SacDataFrame(1)
2 ... 3	reserved
4 ... 5	reserved for use by ISO/IEC 23003-2
6	user specific
7	padding

At the end of F.6.1, add the following new paragraph and table:

Suitable values for *bsResidualSamplingFrequencyIndex* or *bsArbitraryDownmixResidualSamplingFrequencyIndex* can depend on *bsFrameLength*, *bsSamplingFrequencyIndex* and *bsResidualFramesPerSpatialFrame* or *bsArbitraryDownmixResidualFramesPerSpatialFrame*, respectively, as shown in Table F.0A.

Table F.0A — Suitable combinations of *bsSamplingFrequencyIndex* and *bsResidualSamplingFrequencyIndex* or *bsArbitraryDownmixResidualSamplingFrequencyIndex*

(*bsFrameLength***+1)/ (***bsResidualFramesPerSpatialFrame***+1) or (***bsFrameLength***+1)/ (***bsArbitraryDownmixResidualFramesPerSpatialFrame***+1)**	**Allowed combinations of {***bsSamplingFrequencyIndex***, *bsResidualSamplingFrequencyIndex***} or {***bsSamplingFrequencyIndex***, *bsArbitraryDownmixResidualSamplingFrequencyIndex***}**
15	{0x0, 0x3}, {0x1, 0x3}, {0x2, 0x5}, {0x3, 0x3}, {0x4, 0x3}, {0x5, 0x5}, {0x6, 0x3}, {0x7, 0x3}, {0x8, 0x5}, {0x9, 0x6}, {0xa, 0x6} and {0xb, 0x8}
16	{0x0, 0x3}, {0x1, 0x4}, {0x2, 0x5}, {0x3, 0x3}, {0x4, 0x4}, {0x5, 0x5}, {0x6, 0x3}, {0x7, 0x4}, {0x8, 0x5}, {0x9, 0x6}, {0xa, 0x7} and {0xb, 0x8}
18	{0x0, 0x4}, {0x1, 0x4}, {0x2, 0x5}, {0x3, 0x4}, {0x4, 0x4}, {0x5, 0x5}, {0x6, 0x4}, {0x7, 0x4}, {0x8, 0x5}, {0x9, 0x7}, {0xa, 0x7} and {0xb, 0x8}
24	{0x0, 0x5}, {0x1, 0x5}, {0x2, 0x7}, {0x3, 0x5}, {0x4, 0x5}, {0x5, 0x7}, {0x6, 0x5}, {0x7, 0x5}, {0x8, 0x7}, {0x9, 0x8}, {0xa, 0x8} and {0xb, 0xa}
30	{0x0, 0x6}, {0x1, 0x6}, {0x2, 0x8}, {0x3, 0x6}, {0x4, 0x6}, {0x5, 0x8}, {0x6, 0x6}, {0x7, 0x6}, {0x8, 0x8}, {0x9, 0x9}, {0xa, 0x9} and {0xb, 0xb}
32	{0x0, 0x6}, {0x1, 0x7}, {0x2, 0x8}, {0x3, 0x6}, {0x4, 0x7}, {0x5, 0x8}, {0x6, 0x6}, {0x7, 0x7}, {0x8, 0x8}, {0x9, 0x9}, {0xa, 0xa} and {0xb, 0xb}