INTERNATIONAL STANDARD ISO/IEC 14496-3:2001/Amd.2:2004
TECHNICAL CORRIGENDUM 1

Published 2005-08-15

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

Information technology — Coding of audio-visual objects —

Part 3:
Audio

AMENDMENT 2: Parametric coding for high-quality audio

TECHNICAL CORRIGENDUM 1

Technologies de l'information — Codage des objets audiovisuels —

Partie 3: Codage audio

AMENDEMENT 2: Codage paramétrique pour le codage audio de haute qualité

RECTIFICATIF TECHNIQUE 1

Technical Corrigendum 1 to ISO/IEC 14496-3:2001/Amd.2:2004 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

ICS 35.040 **Ref. No. ISO/IEC 14496-3:2001/Amd.2:2004/Cor.1:2005(E)**

Published in Switzerland

In ISO/IEC 14496-3:2001/Amd.2:2004, subpart 1, update Table 1.1. Replacing SSC by SSC(Transient, Sinusoid, Noise) and add a PS tool, as illustrated in the table below.

AOT/Tools	SSC(Transient, Sinusoid, Noise)	Parametric stereo
SSC	X	X

In ISO/IEC 14496-3:2001/Amd.2:2004, subclause 8.5.1, replace Table 8.10 with (modified parts are marked in gray):

Table 8.10 - Decoder level

decoder_ level	Level of complexity	max_nrof_sinusoids	max_nrof_den	#bits for s_nrof_ continuations	#bits for n_nrof_lsf
00	Reserved	Na	Na	Na	Na
01	Medium	60	24	6	4
10	Reserved	Na	Na	Na	Na
11	Reserved	Na	Na	Na	Na

In ISO/IEC 14496-3:2001/Amd.2:2004, subclause 8.5.2, replace (modified parts are marked in gray):

refresh_sinusoids_next_frame – One bit providing an additional frame look ahead for the ADPCM decoding of sinusoidal parameters. If this bit is set to %1, the next frame is a refresh frame. In that case the bit refresh_sinusoids shall be set to %1 in the next frame.

with:
refresh_sinusoids_next_frame – One bit providing an additional frame look ahead for the ADPCM decoding of sinusoidal parameters. If this bit is set to %1, the next frame is a refresh frame. In that case the bit refresh_sinusoids shall be set to %1 in the next frame. If this bit is set to %0, the next frame is not a refresh frame.

In ISO/IEC 14496-3:2001/Amd.2:2004, subclause 8.5.2, replace (modified parts are marked in gray):

t_phi[sf][ch][i] – For a transient of the Meixner type in sub-frame sf of channel ch, these bits represent the phase of the i-th sinusoid under the transient envelope. The decoded value is converted into a phase value in radians in the range [-π, π> and is specified for the start of the transient.

$$tp_q[i] = 2 \cdot tp_e \cdot \text{t\_phi[sf][ch][i]},$$

where tp_e represents the absolute phase error ($tp_e = \frac{\pi}{32}$) and tp_q represents the dequantized absolute phase (in radians). The allowed range for t_phi is [-16, 15]; the representation level +16 is represented by –16 (because +π==-π).

with:
t_phi[sf][ch][i] – For a transient of the Meixner type in sub-frame sf of channel ch, these bits represent the phase of the i-th sinusoid under the transient envelope. The decoded value is converted into a phase value in radians in the range [-π, π> and is specified for the start of the transient.

$$tp_q[i] = 2 \cdot tp_e \cdot \text{t\_phi[sf][ch][i]},$$

where tp_e represents the absolute phase error ($tp_e = \frac{\pi}{32}$) and tp_q represents the dequantized absolute phase (in radians). The allowed range for t_phi is [-16, 15].

In ISO/IEC 14496-3:2001/Amd.2:2004, subclause 8.5.2, replace (modified parts are marked in gray):

If no ICC data is sent in the bit-stream, all ICC parameters are reset to 1 (i.e. index=0). The default quantization grid for ICC is provided in Table 8.B.19.

with:
If no ICC data is sent in the bit-stream, all ICC parameters are reset to 1 (i.e. index=0). The default quantization grid for ICC is provided in Table 8.23

In ISO/IEC 14496-3:2001/Amd.2:2004, subclause 8.5.2, replace (modified parts are marked in gray):

iid_dt[e] – This flag describes for envelope index n, whether the IID parameters are coded differentially over time (iid_dt==%1) or over frequency (iid_dt==%0). In the case iid_mode is different from the previous envelope (e-1), iid_dt[e] shall have the value 0% forcing frequency differential coding.

with:
iid_dt[e] – This flag describes for envelope index e, whether the IID parameters are coded differentially over time (iid_dt==%1) or over frequency (iid_dt==%0). In the case iid_mode is different from the previous envelope (e-1), iid_dt[e] shall have the value 0% forcing frequency differential coding.

In ISO/IEC 14496-3:2001/Amd.2:2004, subclause 8.6.2.4, replace (modified parts are marked in gray):

8.6.2.4 Synthesis of sinusoids for segments without a transient
with:
8.6.2.4 Synthesis of sinusoids for segments with a transient

In ISO/IEC 14496-3:2001/Amd.2:2004, subclause 8.6.4.3, replace (modified parts are marked in gray):

In order to compensate for the overall delay of the hybrid analysis filterbank, the first 10 sets (6 from delay and 4 from QMF filter) of hybrid subbands are flushed and therefore not taken into account for processing.
with:
In order to compensate for the overall delay of the hybrid analysis filterbank, the first 10 sets (6 from delay and 4 from QMF filter) of hybrid subbands are flushed and therefore not taken into account for processing. Note that in Figure 8.24 this delay has already been accounted for.

In ISO/IEC 14496-3:2001/Amd.2:2004, subclause 8.6.4.4, Figure 8.24, replace caption (modified parts are marked in gray):

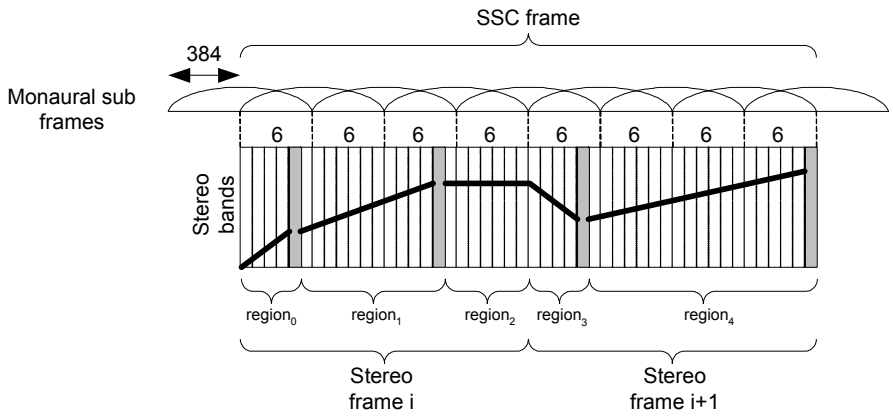

Figure 8.24 - One SSC frame comprises two stereo frames of data. The solid line illustrates the interpolation between stereo parameters for slots that have not been assigned stereo parameters to

with:

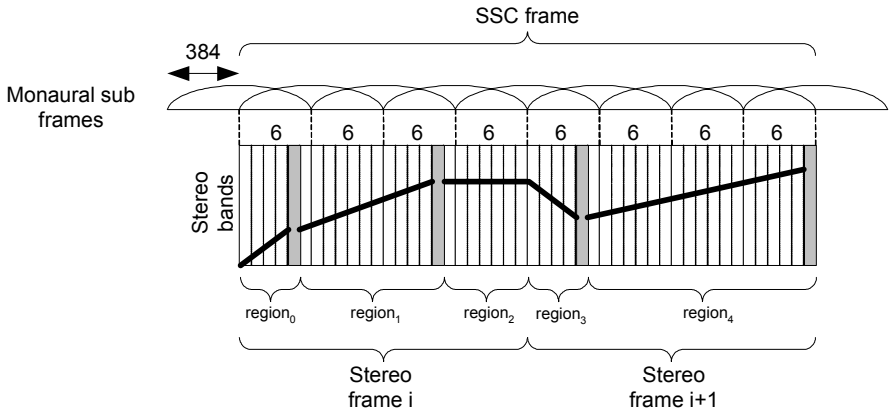

Figure 8.24 - One SSC frame comprises two stereo frames of data. The solid line illustrates the interpolation between stereo parameters for slots that have not been assigned stereo parameters to. Note that the delay introduced by the hybrid QMF analysis filterbank has been compensated for in this figure.

In ISO/IEC 14496-3:2001/Amd.2:2004, subclause 8.6.4.5.1, remove:

F_s is the output sampling rate.

Replace equation:

$$NR\_ALLPASS\_BANDS = \begin{cases} 53 & ,F_s < 32kHz, 10 \text{ or } 20 \text{ stereo bands} \\ 73 & ,F_s < 32kHz, 34 \text{ stereo bands} \\ 30 & ,F_s \geq 32kHz, 10 \text{ or } 20 \text{ stereo bands} \\ 50 & ,F_s \geq 32kHz, 34 \text{ stereo bands} \end{cases}$$

with:

$$NR\_ALLPASS\_BANDS = \begin{cases} 30 & ,10 \text{ or } 20 \text{ stereo bands} \\ 50 & ,34 \text{ stereo bands} \end{cases}$$

Replace equation:

$$SHORT\_DELAY\_BAND = \begin{cases} 71 & ,F_s < 32kHz, 10 \text{ or } 20 \text{ stereo bands} \\ 91 & ,F_s < 32kHz, 34 \text{ stereo bands} \\ 42 & ,F_s \geq 32kHz, 10 \text{ or } 20 \text{ stereo bands} \\ 62 & ,F_s \geq 32kHz, 34 \text{ stereo bands} \end{cases}$$

with:

$$SHORT\_DELAY\_BAND = \begin{cases} 42 & ,10 \text{ or } 20 \text{ stereo bands} \\ 62 & ,34 \text{ stereo bands} \end{cases}$$

Replace equation:

$$a_{Smooth} = \begin{cases} 0.6 & ,F_s < 32kHz \\ 0.25 & ,F_s \geq 32kHz \end{cases}$$

with:

$$a_{Smooth} = 0.25$$

In ISO/IEC 14496-3:2001/Amd.2:2004, subclause 8.6.4.5.3, replace equation:

$$\mathbf{G}_{TransientRatio}(i,n) = \begin{cases} \dfrac{\mathbf{P}(i,n)}{\gamma \cdot \mathbf{P}_{SmoothPeakDecayDiffNrg}(i,n)} & , \gamma \cdot \mathbf{P}_{SmoothPeakDecayDiffNrg}(i,n) > \mathbf{P}(i,n) \\ 1 & , otherwise \end{cases}$$

with:

$$\mathbf{G}_{TransientRatio}(i,n) = \begin{cases} \dfrac{\mathbf{P}_{SmoothNrg}(i,n)}{\gamma \cdot \mathbf{P}_{SmoothPeakDecayDiffNrg}(i,n)} & , \gamma \cdot \mathbf{P}_{SmoothPeakDecayDiffNrg}(i,n) > \mathbf{P}_{SmoothNrg}(i,n) \\ 1 & , otherwise \end{cases}$$

In ISO/IEC 14496-3:2001/Amd.2:2004, subclause 8.6.4.5.4, replace Table 8.34 with:

Table 8.34 - Filter coefficient vector and delay length vector $\mathbf{d}(m)$.

m	$\mathbf{a}(m)$	$\mathbf{d}(m)$
0	0.65143905753106	3
1	0.56471812200776	4
2	0.48954165955695	5

Remove equation:

Delay length vector, $\mathbf{d} = \begin{cases} \mathbf{d}_{24kHz} & , F_s < 32kHz \\ \mathbf{d}_{48kHz} & , F_s \geq 32kHz \end{cases}$.

In ISO/IEC 14496-3:2001/Amd.2:2004, subclause 8.6.4.5.4, replace Table 8.35 with (modified parts are marked in gray):

Table 8.35 - Delay length vector $\mathbf{f}_{center\_20}$.

k	$\mathbf{f}_{center\_20}(k)$	k	$\mathbf{f}_{center\_20}(k)$
0	-3/8	5	7/8
1	-1/8	6	5/4
2	1/8	7	7/4
3	3/8	8	9/4
4	5/8	9	11/4

Replace Table 8.38 with:

Table 8.36 - Peak Decay Factor α.

α	0.76592833836465

Remove equation:

Peak decay factor, $\alpha = \begin{cases} \alpha_{Decay\,24kHz} & , F_s < 32kHz \\ \alpha_{Decay\,48kHz} & , F_s \geq 32kHz \end{cases}$.

In ISO/IEC 14496-3:2001/Amd.2:2004, subclause 8.6.4.6.1, replace (modified parts are marked in gray):

The number of stereo bands that is actually used for the processing of the cues depends on the number of available parameters for IID and ICC according to the relation given in Table 8.39.

with:
The number of stereo bands that is actually used for the processing of the cues depends on the number of available parameters for IID and ICC according to the relation given in Table 8.39. In case no IID or no ICC parameters have been transmitted in the current frame (enable_iid==%0 or enable_icc==%0), the number of IID or ICC parameters, respectively, is assumed to be 20 for the purpose of Table 8.39. In case no IID and no ICC parameters have been transmitted in the current frame (enable_iid==%0 and enable_icc==%0), the number of stereo bands in the previous frame is kept unchanged and used also for the processing of the current frame.

In ISO/IEC 14496-3:2001/Amd.2:2004, subclause 8.6.4.6.1, replace (modified parts are marked in gray):

If the number of stereo bands changes from 10,20 in the previous frame to 34 in the current frame, the stereo parameters from the previous frame are mapped to 34 stereo bands according to Table 8.40 prior to further processing of the current. The frequency resolution of the hybrid QMF analysis filterbank (see subclause 8.6.4.3) is changed instantaneously to the 34 stereo band configuration. The state variable of the decorrelation process are mapped from *NR_BANDS*==71 to *NR_BANDS*==91 configuration according to Table 8.42. The state variables for sub subbands in *NR_BANDS*==91 configuration not listed in Table 8.42 are reset to zero.
If the number of stereo bands changes from 34 in the previous frame to 10,20 in the current frame, the stereo parameters from the previous frame are mapped to 20 stereo bands according to Table 8.41 prior to further processing of the current. The frequency resolution of the hybrid QMF analysis filterbank (see subclause 8.6.4.3) is changed instantaneously to the 20 stereo band configuration. The state variable of the decorrelation process are mapped from *NR_BANDS*==91 to *NR_BANDS*==71 configuration according to Table 8.42. The state variables for sub subbands in *NR_BANDS*==91 configuration not listed in Table 8.42 are discarded.

with:
If the number of stereo bands changes from 10,20 in the previous frame to 34 in the current frame, the coefficients $h_{11}(b)$, $h_{12}(b)$, $h_{21}(b)$, and $h_{22}(b)$ at the end of the previous frame are mapped from 20 to 34 stereo bands according to Table 8.40 (by substituting idx_b by $h_{ij}(b)$, where ij is 11, 12, 21, or 22) prior to further processing as defined in subclause 8.6.4.6.3 and the IPD/OPD smoothing state variables are reset, i.e., $opd(b,n_{e-1})=0$, $ipd(b,n_{e-1})=0$, $opd(b,n_e)=0$, and $ipd(b,n_e)=0$. The frequency resolution of the hybrid QMF analysis filterbank (see subclause 8.6.4.3) is changed instantaneously to the 34 stereo band configuration. The state variables of the decorrelation process are reset to zero (see Table 8.42).
If the number of stereo bands changes from 34 in the previous frame to 10,20 in the current frame, the coefficients $h_{11}(b)$, $h_{12}(b)$, $h_{21}(b)$, and $h_{22}(b)$ at the end of the previous frame are mapped from 34 to 20 stereo bands according to Table 8.41 (by substituting idx_b by $h_{ij}(b)$, where ij is 11, 12, 21, or 22) prior to further processing as defined in subclause 8.6.4.6.3 and the IPD/OPD smoothing state variables are reset, i.e., $opd(b,n_{e-1})=0$, $ipd(b,n_{e-1})=0$, $opd(b,n_e)=0$, and $ipd(b,n_e)=0$. The frequency resolution of the hybrid QMF analysis filterbank (see subclause 8.6.4.3) is changed instantaneously to the 20 stereo band configuration. The state variables of the decorrelation process are reset to zero (see Table 8.42).

In ISO/IEC 14496-3:2001/Amd.2:2004, subclause 8.6.4.6.1, replace Table 8.42 – Mapping of state variables of decorrelation process between 10,20 and 34 stereo band configurations with:

Table 8.42 – Changing number of stereo bands.

	current frame	
previous frame	10/20 bands	34 bands
10/20 bands	-	map $h_{ij}(b)$ according to Table 8.40, reset state variables
34 bands	map $h_{ij}(b)$ according to Table 8.41, reset state variables	-

In ISO/IEC 14496-3:2001/Amd.2:2004, after subclause 8.6.4.6.4, insert new subclause 8.6.4.6.5:

8.6.4.6.5 Procedure for incomplete parameter sets

In the case no parameters have been transmitted in the current frame for either IID, ICC, nor IPD/OPD or a combination thereof, the parameters values for the current frame are obtained according to the num_env variable as given in Table 8.45, 8.46 and 8.47.

Table 8.45 – Derivation of paramers for IID in the case no parameters are transmitted.

	enable_iid	
	0	1
num_env=0	IID parameters set to default	IID parameters held
num_env>0	IID parameters set to default	n.a.

Table 8.46 – Derivation of paramers for ICC in the case no parameters are transmitted.

	enable_icc	
	0	1
num_env=0	ICC parameters set to default	ICC parameters held
num_env>0	ICC parameters set to default	n.a.

Table 8.47 – Derivation of paramers for IPD/OPD in the case no parameters are transmitted.

	enable_ipdopd	
	0	1
num_env=0	IPD/OPD parameters set to default	IPD/OPD parameters held
num_env>0	IPD/OPD parameters set to default	n.a.

In the case parameters are to be set to default, the parameters at the positions defined by n_e are set to their default value (index=0).
In the case the parameters are to be held, two situations are distinguished. If enable_ipdopd==%1, the four vectors $H_{11}(k,n)$, $H_{12}(k,n)$, $H_{21}(k,n)$ and $H_{22}(k,n)$ for all n=[0,…, *numQMFSlots*-1], are copied from those same four vectors at position n=*numQMFSlots*-1 in the previous ps_data() element. If enable_ipdopd==%0, the four vectors $H_{11}(k,n)$, $H_{12}(k,n)$, $H_{21}(k,n)$ and $H_{22}(k,n)$ for all n=[0,…, *numQMFSlots*-1], are set to the four vectors $h_{11}(k,n)$, $h_{12}(k,n)$, $h_{21}(k,n)$ and $h_{22}(k,n)$ respectively, where n=*numQMFSlots*-1 in the previous ps_data() element.

In ISO/IEC 14496-3:2001/Amd.2:2004, at the end of subclause 8.6.5.1. Start decoding, append:

A conformant decoder that receives PS data shall output the mono signal in the two output channels until a first ps_data() element with enable_ps_header==1 is received and in which for all enabled parameters frequency differential coding is employed and num_env>0, ensuring that the PS data can be decoded correctly.

In ISO/IEC 14496-3:2001/Amd.2:2004, subclause 8.A.1, replace (modified parts are marked in gray):

The usage of this parametric stereo extension to HE-AAC is signalled implicitly in the bitstream. Hence, if an sbr_extension() with bs_extension_id==EXTENSION_ID_PS is found in the SBR part of the bitstream, a decoder supporting the combination of SBR and PS shall operate the PS tool to generate a stereo output signal. If no ps_data() element is available in the SBR part of a monaural HE-AAC bitstream, the normal monaural signal is generated by the SBR tool.

with:
The usage of this parametric stereo extension to HE-AAC is signalled implicitly in the bitstream. Hence, if an sbr_extension() with bs_extension_id==EXTENSION_ID_PS is found in the SBR part of the bitstream, a decoder supporting the combination of SBR and PS shall operate the PS tool to generate a stereo output signal. If no ps_data() element is available in the SBR part of a monaural HE-AAC bitstream, the normal monaural signal is generated by the SBR tool and mapped to a stereo output signal in which the left and right channel both contain the same monaural signal.

*In ISO/IEC 14496-3:2001/Amd.2:2004, subclause 8.A.3, after "Furthermore, numQMFSlots=numTimeSlots*RATE.", insert the following new paragraph:*

In order to allow an efficient implementation of the PS tool, a partial reset of the decorrelator state variables is performed for each stereo frame for all QMF subbands above the highest QMF subband generated by the SBR tool by forcing the states

$$d_k(n) = 0$$
$$s_k(n) = 0$$

where $n < n_e$, $k_{\max} \le k < NR\_BANDS$, n_e is the first sample in the current stereo frame and

$$k_{\max} = k_x + M + \begin{cases} 7 & ,10 \text{ or } 20 \text{ stereo bands} \\ 27 & ,34 \text{ stereo bands} \end{cases}$$

where k_x and M are defined in subclause 4.6.18.3.2.2.