

Source: TSG-SA WG4

Title: 3GPP TS 26.405 version 2.0.0 "Enhanced aacPlus General Audio Codec;
Encoder Specification; Parametric Stereo part" (Release 6)

Agenda Item: 7.4.3

Presentation of Specification to TSG SA Plenary

Presentation to: TSG SA Meeting #25

Document for presentation: TS 26.405 "Enhanced aacPlus General Audio Codec; Encoder
Specification; Parametric Stereo part", Version 2.0.0 (Release 6)

Presented for: Discussion / Decision

Abstract of document:

This Telecommunication Standard (TS) describes the Parametric Stereo encoder part of the Enhanced aacPlus general audio codec.

Changes since last presentation:

None.

Outstanding Issues:

At SA#24 the way forward for the selection of audio codecs was formulated in [TD SP-040481](#). Following the guidance contained therein, SA4 agreed to forward this document to TSG SA#25, as one of the specifications for the Enhanced aacPlus audio codec.

Contentious Issues:

None.

Comment(s):

None.

3GPP TS 26.405 V2.0.0 (2004-09)

Technical Specification

**3rd Generation Partnership Project;
Technical Specification Group Services and System Aspects;
General Audio Codec audio processing functions;
Enhanced aacPlus general audio codec;
Encoder Specification Parametric Stereo part
(Release 6)**



The present document has been developed within the 3rd Generation Partnership Project (3GPP™) and may be further elaborated for the purposes of 3GPP.

The present document has not been subject to any approval process by the 3GPP Organizational Partners and shall not be implemented. This Specification is provided for future development work within 3GPP only. The Organizational Partners accept no liability for any use of this Specification. Specifications and reports for implementation of the 3GPP™ system should be obtained via the 3GPP Organizational Partners' Publications Offices.

Keywords

CODEC, SBR, Enhanced aacPlus, General audio
coder

3GPP

Postal address

3GPP support office address

650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet

<http://www.3gpp.org>

Copyright Notification

No part may be reproduced except as authorized by written permission.
The copyright and the foregoing restriction extend to reproduction in all media.

© 2004, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TTA, TTC).
All rights reserved.

Contents

Foreword	4
1 Scope	5
2 Normative references	5
3 Definitions, symbols and abbreviations	5
3.1 Definitions	5
3.2 Symbols	5
3.3 Abbreviations	6
4 Outline description	6
5 Parametric stereo encoder	6
5.1 System overview	6
5.2 Analysis filterbank	7
5.2.1 QMF analysis filterbank	7
5.2.2 Low frequency filtering	7
5.3 Configurations	10
5.4 Stereo parameter extraction	10
5.4.1 Parameter estimation	10
5.4.2 Quantization of IID and ICC parameters	12
5.5 Writing to bitstream	12
5.6 Downmixing to mono	14
5.7 Synthesis filterbank	14
Annex A (informative): Change history.....	17

Foreword

The present document describes the detailed mapping of the general audio service employing the aacPlus general audio codec within the 3GPP system.

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of this TS, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

- x the first digit:
 - 1 presented to TSG for information;
 - 2 presented to TSG for approval;
 - 3 Indicates TSG approved document under change control.
- y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.
- z the third digit is incremented when editorial only changes have been incorporated in the specification;

1 Scope

This Telecommunication Standard (TS) describes the Parametric Stereo encoder part of the Enhanced aacPlus general audio codec [4].

2 Normative references

This TS incorporates by dated and undated reference, provisions from other publications. These normative references are cited in the appropriate places in the text and the publications are listed hereafter. For dated references, subsequent amendments to or revisions of any of these publications apply to this TS only when incorporated in it by amendment or revision. For undated references, the latest edition of the publication referred to applies.

- [1] ISO/IEC 14496-3:2001/AMD1:2003, Bandwidth Extension.
 - [2] ISO/IEC 14496-3:2001/Amd.1:2003/DCOR1.
 - [3] ISO/IEC 14496-3:2001/ Amd.2:2004, Parametric Coding for High Quality Audio.
 - [4] 3GPP TS 26.401 : Enhanced aacPlus general audio codec; General Description
-

3 Definitions, symbols and abbreviations

3.1 Definitions

For the purposes of this TS, the following definitions apply:

- hybrid QMF:** a QMF filterbank combined with additional filters to achieve higher frequency resolution for the lower QMF bands
- stereo band:** a group of consecutive hybrid QMF subbands used for coding one stereo parameter

3.2 Symbols

For the purposes of this TS, the following symbols apply:

- $l_k(n)$ Subsample in hybrid QMF matrix: left channel, band k , subsample n .
- $r_k(n)$ Subsample in hybrid QMF matrix: right channel, band k , subsample n .

3.3 Abbreviations

For the purposes of this TS, the following abbreviations apply.

SBR	Spectral Band Replication
AAC	Advanced Audio Coding
aacPlus	Combination of MPEG-4 AAC and MPEG-4 Bandwidth extension (SBR)
Enhanced aacPlus	Combination of MPEG-4 AAC, MPEG-4 Bandwidth extension (SBR) and MPEG-4 Parametric Stereo
QMF	Quadrature Mirror Filter
MPEG	Moving Picture Expert Group
IID	Inter Intensity Difference, (stereo parameter)
ICC	Inter Channel Coherence, (stereo parameter)

4 Outline description

This TS is structured as follows:

Section 5.2 describes the hybrid QMF filterbank and its integration in the Parametric Stereo system.

Section 5.3 describes the hybrid QMF filterbank and its integration in the Parametric Stereo system.

Section 5.4 describes the parameter estimation algorithms and quantization.

Section 5.5 describes how to convey the estimated parameters in the bitstream.

Section 5.6 and section 5.7 describes preparation of the signal that should feed the aacPlus mono encoder after the Parametric Stereo encoding.

5 Parametric stereo encoder

5.1 System overview

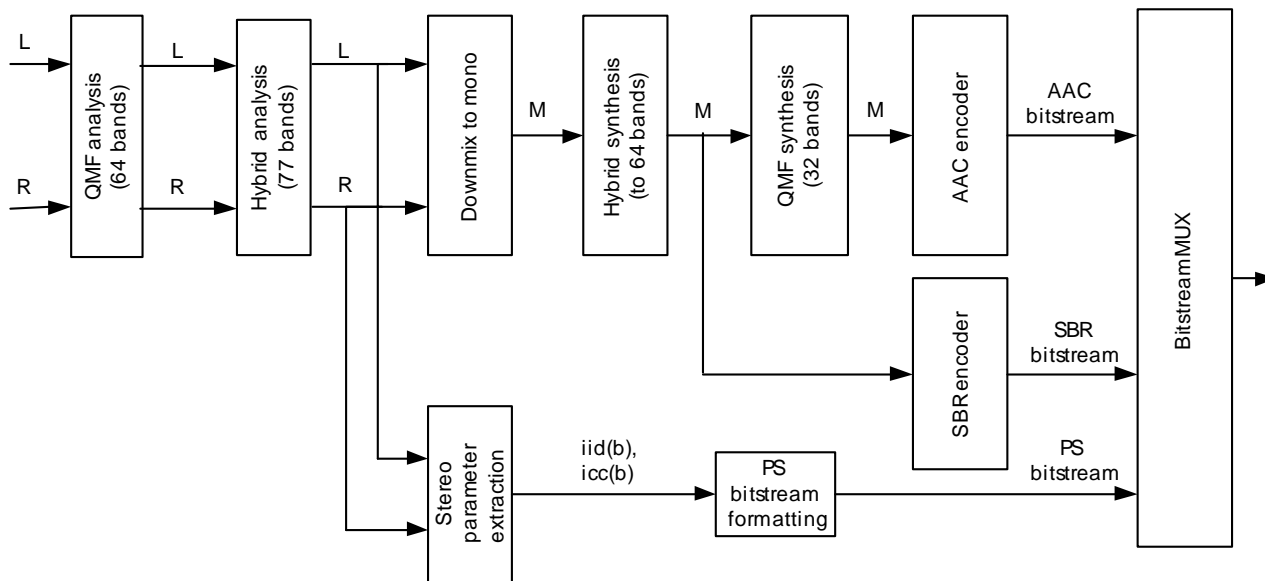


Figure 1 - Encoder overview

The interface between the parametric stereo encoder tool and the aacPlus encoder is depicted in Figure 1. In the figure L and R denotes the left and right channel respectively, while M denotes the down-mixed mono signal which the aacPlus encoder operates on.

The parametric stereo coding tool is able to capture the stereo image into a limited number of parameters, requiring only a small overhead of a few kbit/s. Together with a controlled monaural downmix of the stereo input signal, the parametric stereo coding tool is able to regenerate the stereo signal at the decoder side.

The encoder operates as a non-modifying analyzer prior to the aacPlus encoder, though it shares the same QMF analysis filterbank. The decoder operates as a post process to aacPlus using the Parametric Stereo data conveyed by the bitstream to synthesize the stereo properties of the output signal. Part from the parametric stereo tool, the aacPlus runs in mono mode not affected by Parametric Stereo.

The bitstream syntax and decoder description of the parametric stereo tool in combination with aacPlus is defined in [3]. This system includes only the baseline level defined in that standard.

5.2 Analysis filterbank

5.2.1 QMF analysis filterbank

This filterbank is identical to the 64 complex QMF analysis filterbank as defined in ISO/IEC 14496-3/AMD1:2003, sub clause 4.B.18.2 [1], [2]. However, in the equation for matrix $\mathbf{M}(k,n)$ and in Figure 4.B.20, the term “ $(2*n+1)$ ” has to be substituted by “ $(2*n-1)$ ”. The input to the filterbank are blocks of 64 samples of the monaural synthesized signal M. For each block the filterbank outputs one slot of 64 QMF samples.

5.2.2 Low frequency filtering

The lower QMF subbands are further split in order to obtain a higher frequency resolution enabling a proper stereo analysis and synthesis for the lower frequencies. To achieve those, in total 77 frequency bands, a hybrid filterbank configurations have been defined. The filter used for this sub subband filtering, Q_p^q is defined according to:

$$G_q^p = g^p(n) \exp\left(j \frac{2\pi}{Q^p} \left(q + \frac{1}{2}\right) (n-6)\right)$$

where g^p represents the prototype filters in QMF subband p. Q^p represents the number of sub-subbands in QMF subband p, q the sub-subband index in QMF channel p and n the time index. The prototype filters are all of length 13 and have a delay of 6 QMF samples. The prototype filters are listed in Table 1.

Table 1- Prototype filter coefficients for the filters that split the lower QMF subbands

n	$g^0(n), Q^0 = 8$	$g^{1,2}(n), Q^{1,2} = 4$
0	0.00746082949812	-0.00305151927305
1	0.02270420949825	-0.00794862316203
2	0.04546865930473	0
3	0.07266113929591	0.04318924038756
4	0.09885108575264	0.12542448210445
5	0.11793710567217	0.21227807049160
6	0.125	0.25000000000000
7	0.11793710567217	0.21227807049160
8	0.09885108575264	0.12542448210445
9	0.07266113929591	0.04318924038756
10	0.04546865930473	0
11	0.02270420949825	-0.00794862316203
12	0.00746082949812	-0.00305151927305

Figure 2 and Figure 3 illustrate the hybrid analysis and synthesis filterbank for the 77 frequency bands configuration.

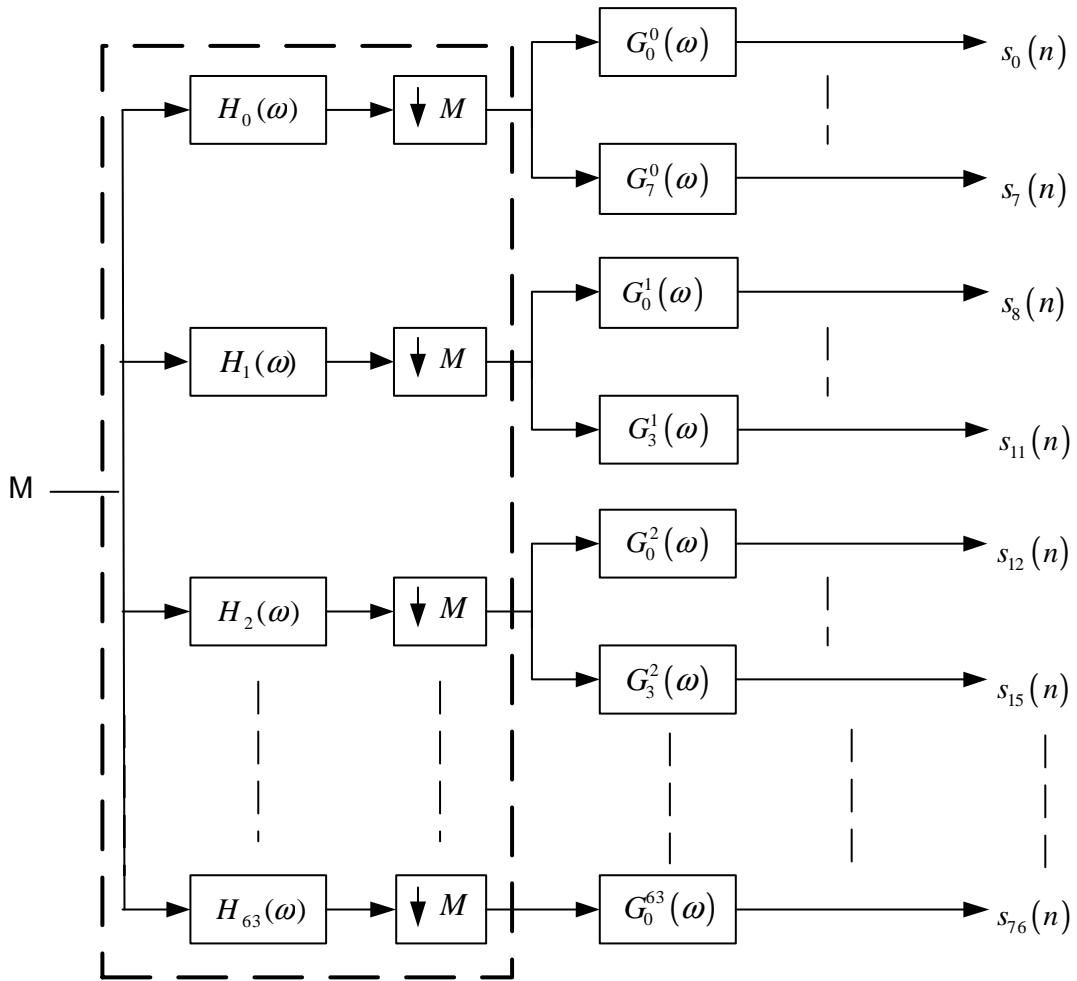


Figure 2 - Hybrid QMF analysis filterbank providing 77 output bands. The three lower subbands of the 64 QMF (see dashed box) are further split to provide for increased resolution for the lower frequencies

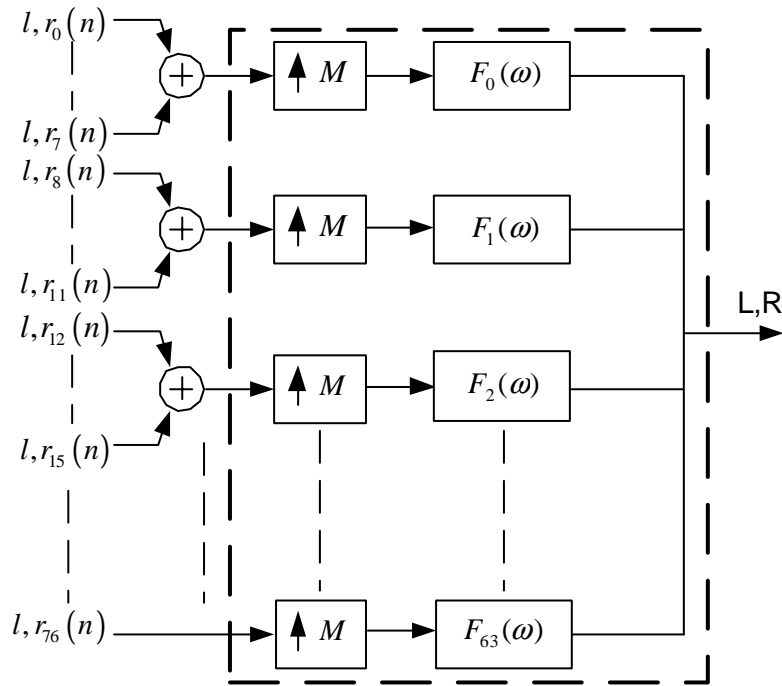


Figure 3 - Hybrid QMF synthesis filterbank using 77 input bands. The coefficients offering higher resolution for the lower QMF subbands are simply added prior to the synthesis with the 64 subbands QMF (see dashed box)

In order to time align all the samples originating from the hybrid filterbank, the remaining QMF subbands that have not been filtered are delay compensated. This delay amounts to 6 QMF subband samples. This means $G_0^k(z) = z^{-6}$ for $k=3\dots 63$. 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.

The resultant of this operation is a slot of hybrid subband samples consisting of a LF (low frequency) sub QMF subband portion and HF (high frequency) QMF subband portion.

5.3 Configurations

The parametric stereo encoder uses two different configurations depending on desired frequency resolution. The configuration parameter, *num_stereo_bands* determines what frequency resolution should be used for the stereo parameters. For all bitrates below 21000 bit/s, *num_stereo_bands* is set to 10 otherwise *num_stereo_bands* is set to 20.

5.4 Stereo parameter extraction

5.4.1 Parameter estimation

In order to estimate the stereo parameters the signals M, L and R are analyzed using the hybrid filterbank as in Figure 2 for providing the 77 frequency bands addressed by the index, $0 \leq k < 77$. This results in the (sub-)subband domain signals, $m(k, n)$, $l(k, n)$ and $r(k, n)$.

To estimate the parameters for the current frame the following is calculated:

$$e_l(b) = \sum_{k=k_l}^{k_h} \sum_{n=0}^{L-1} \left| l \left(k, -\frac{L}{2} + 1 + n \right) \right|^2 + \epsilon$$

$$e_r(b) = \sum_{k=k_l}^{k_h} \sum_{n=0}^{L-1} \left| r\left(k, -\frac{L}{2} + 1 + n\right) \right|^2 + \varepsilon$$

$$e_R(b) = \sum_{k=k_l}^{k_h} \sum_{n=0}^{L-1} l\left(k, -\frac{L}{2} + 1 + n\right) r^*\left(k, -\frac{L}{2} + 1 + n\right) + \varepsilon$$

where $e_l(b)$, $e_r(b)$ and $e_R(b)$ are the left channel excitation, the right channel excitation and the non-normalized cross-channel excitation between left and right channel for stereo bin b respectively, L the segment length, ε a very small value preventing division by zero ($\varepsilon = 10^{-10}$). The summation over k is shown in Table 2 for 20 bands case. For the 10 bands case additional formulas below are used.

Table 2 – Summation range in 71 sub subbands in case of 20 bands

Parameter index b	Sub subband index	QMF channel
0	0	0
1	1	0
2	2	0
3	3	0
4	10	1
5	11	1
6	12	2
7	13	2
8	16	3
9	17	4
10	18	5
11	19	6
12	20	7
13	21	8
14	22-23	9-10
15	24-26	11-13
16	27-30	14-17
17	31-35	18-22
18	36-47	23-34
19	48-76	35-63

For the 10 band case, summation will be use the same table as the 20 band case but with the additional summation:

$$e_{l,10}(b) = \frac{e_l(2b) + e_l(2b+1)}{2}$$

$$e_{r,10}(b) = \frac{e_r(2b) + e_r(2b+1)}{2}$$

$$e_{R,10}(b) = \frac{e_R(2b) + e_R(2b+1)}{2}$$

Where $e_{l,10}$ and $e_{r,10}$ and $e_{R,10}$ should replace e_l and e_r and e_R in the following formulas for the 10 band case.

The IID, denoted as iid(b) and the ICC, denoted as icc(b), for each stereo band b are calculated as:

$$iid(b) = 10 \log_{10} \left(\frac{e_l(b)}{e_r(b)} \right)$$

$$icc(b) = \begin{cases} \sqrt{1 - \min \left(\frac{\operatorname{Re}(e_R(b))}{\sqrt{e_l(b)e_r(b)}}, 1 \right)} & , b < 5 \text{ and } num_stereo_bands = 10 \text{ or} \\ & , b < 11 \text{ and } num_stereo_bands = 20 \\ \sqrt{1 - \min \left(\frac{|e_R(b)|}{\sqrt{e_l(b)e_r(b)}}, 1 \right)} & , otherwise \end{cases}$$

5.4.2 Quantization of IID and ICC parameters

The obtained values for IID and ICC are quantized to the nearest lower value given in Table 3 and Table 4, respectively, then coded into Huffman words according to Huffman tables found in [4] section 8.B.1 Huffman tables ps_data.

Table 3 - Quantization grid for iid

Index	-7	-6	-5	-4	-3	-2	-1	0
IID [dB]	-21.5	-16	-12	-8.5	-5.5	-3	-1	0
Index	1	2	3	4	5	6	7	
IID [dB]	1	3	5.5	8.5	12	16	21.5	

Table 1 – Quantization grid for icc

index	0	1	2	3	4	5	6	7
icc	0	0.088900	0.229800	0.364250	0.504500	0.635100	0.799600f	0.945650

5.5 Writing to bitstream

The parametric stereo bitstream should be placed in the SBR extension field according to syntax in Table 5. The EXTENSION_ID_PS should be used as extension identifier. Note that due to SBR extension is design, the total size, in bytes, of the ps_data element has to be added to the SBR extension element before writing the actual ps_data.

Table 5 – Syntax of sbr_extension()

Syntax	No. of bits	Mnemonic
<pre>sbr_extension(bs_extension_id, num_bits_left) { switch (bs_extension_id) { case EXTENSION_ID_PS: num_bits_left -= ps_data(); break; default: bs_fill_bits; num_bits_left = 0; break; } }</pre>		Note 1 num_bits_left bslbf
Note 1: ps_data() returns the number of bits read.		

Table 6 – Values of the bs_extension_id field

Symbol	Value	Purpose
EXTENSION_ID_PS	2	Parametric Stereo Coding
	all other values	reserved

The syntax of the ps_data is defined in [4] Table 8.1. The bitstream elements in ps_data should be assigned values according to the following list.

enable_ps_header – Is set to 1 for all frames containing an SBR header. Is set to 0 for all other cases.

enable_iid – Is set to 0 if all values in the iid vector after quantization are zero. Is set to 1 for all other cases.

enable_icc – Is set to 0 if all values in the icc vector after quantization are zero. Is set to 1 for all other cases.

iid_mode – Is set to 0 if 10 stereo bands resolution is chosen. Is set to 1 for all other cases.

icc_mode – Is set to 0 if 10 stereo bands resolution is chosen. Is set to 1 for all other cases.

enable_ext – Is set to 0.

frame_class – Is set to 0.

num_env_idx – Is set to 0 if all values in the iid and icc vectors after quantization are unchanged since previous frame. Is set to 1 for all other cases.

iid_dt – Is set to 1 if differential coding of iid parameters over time gives a lower total Huffman code length than differential coding of iid parameters with respect to the previous parameter position. Is set to 0 for all other cases.

icc_dt – Is set to 1 if differential coding of icc parameters over time gives a lower total Huffman code length than differential coding of iid parameters with respect to the previous parameter position. Is set to 0 for all other cases.

iid_par_dt[n] – In case of differential coding of IID parameters over time (iid_dt==1), iid_par_dt[n] describes the IID difference with respect to the previous parameter position. If no previous parameter position is available, iid_par_dt[n] represents the IID difference with respect to the decoded value 0 (i.e. index=0).

iid_par_df[n] – In case of differential coding of IID parameters over frequency (iid_dt==0), iid_par_df[n] describes the Huffman encoded IID difference with respect to the parameter (n-1).

iid_par_df[0] – represents the IID difference with respect to the decoded value 0 (i.e. index=0).

icc_par_dt[n] – In case of differential coding of ICC parameters over time ($\text{icc_dt}=1$), $\text{icc_par_dt}[n]$ describes the difference with respect to the previous parameter position. If no previous parameter position is available, $\text{icc_par_dt}[n]$ represents the ICC difference with respect to the decoded value 1 (i.e. $\text{index}=0$).

icc_par_df[n] – In case of differential coding of ICC parameters over frequency ($\text{icc_dt}=0$), $\text{icc_par_df}[n]$ describes the ICC difference with respect to the parameter (n-1). $\text{icc_par_df}[0]$ represents the ICC difference with respect to the decoded value 1 (i.e. $\text{index}=0$).

5.6 Downmixing to mono

A monaural signal is needed for encoding of the HE AAC mono core. A simple stereo-to-mono downmixing scheme is performed after extraction of the stereo parameters. It combines equal parts of left and right channel and scales the resulting mono signal in order to preserve most of the original total power. The process is defined according to:

$$m(k,n) = \frac{l(k,n) + r(k,n)}{2} \cdot \gamma(k,n)$$

where the stereo scale factor, γ is defined by:

$$\gamma(k,n) = \min \left(2, \sqrt{\frac{|l(k,n)|^2 + |r(k,n)|^2}{0.5 \cdot |l(k,n) + r(k,n)|^2}} \right)$$

$$0 \leq k < \text{NUM_OF_BANDS}$$

5.7 Synthesis filterbank

The stereo to mono down-mixed hybrid subband signal $m(k,n)$ is fed into the hybrid synthesis filterbank, that is implemented as an adder of sub QMF samples. This is illustrated in Figure 3.

The synthesis filtering and implicit down-sampling of the 64 subband signals is achieved using a 32-channel QMF bank. The output from the filterbank is real-valued time domain samples. The process is given by the flowchart in Figure 4. The synthesis filtering comprises the following steps, where an array \mathbf{v} consisting of 640 samples is assumed:

- Shift the samples in the array \mathbf{v} by 64 positions. The oldest 64 samples are discarded.
- The array of 32 complex-valued subband samples \mathbf{Z} is separated into the real and imaginary components as $\mathbf{Z} = \mathbf{R} + i\mathbf{I}$. The components are scaled and DCT and DST type IV transformed as

$$\begin{aligned} r(n) &= \frac{1}{64} \sum_{k=0}^{31} R(k) \cos \left[\frac{\pi}{32} \left(k + \frac{1}{2} \right) \left(n + \frac{1}{2} \right) \right] \\ i(n) &= \frac{1}{64} \sum_{k=0}^{31} I(k) \sin \left[\frac{\pi}{32} \left(k + \frac{1}{2} \right) \left(n + \frac{1}{2} \right) \right] \end{aligned}, 0 \leq n < 32$$

- The arrays \mathbf{r} and \mathbf{i} are combined and stored in the positions 0 to 63 in array \mathbf{v} as

$$\begin{aligned} v(n) &= i(n) - r(n) & , 0 \leq n < 32 \\ v(63-n) &= i(n) + r(n) & , 0 \leq n < 32 \end{aligned}$$

- Extract samples from \mathbf{v} according to the flowchart in Figure 4 to create the 320-element array \mathbf{g} .
- Multiply the samples of array \mathbf{g} by every other coefficient of window \mathbf{w} . The window coefficients of \mathbf{c} can be found in Figure 4, and are the same as for the analysis filterbank.

- Calculate 32 new output samples by summation of samples from array \mathbf{w} according to the last step in the flowchart of Figure 4

Every SBR frame produces an output of $numTimeSlots \cdot RATE \cdot 32$ time domain samples. In the flowchart of Figure 4 $\mathbf{X}[k][l]$ corresponds to subband sample l in the QMF subband k , and every new loop produces 32 time domain samples as output.

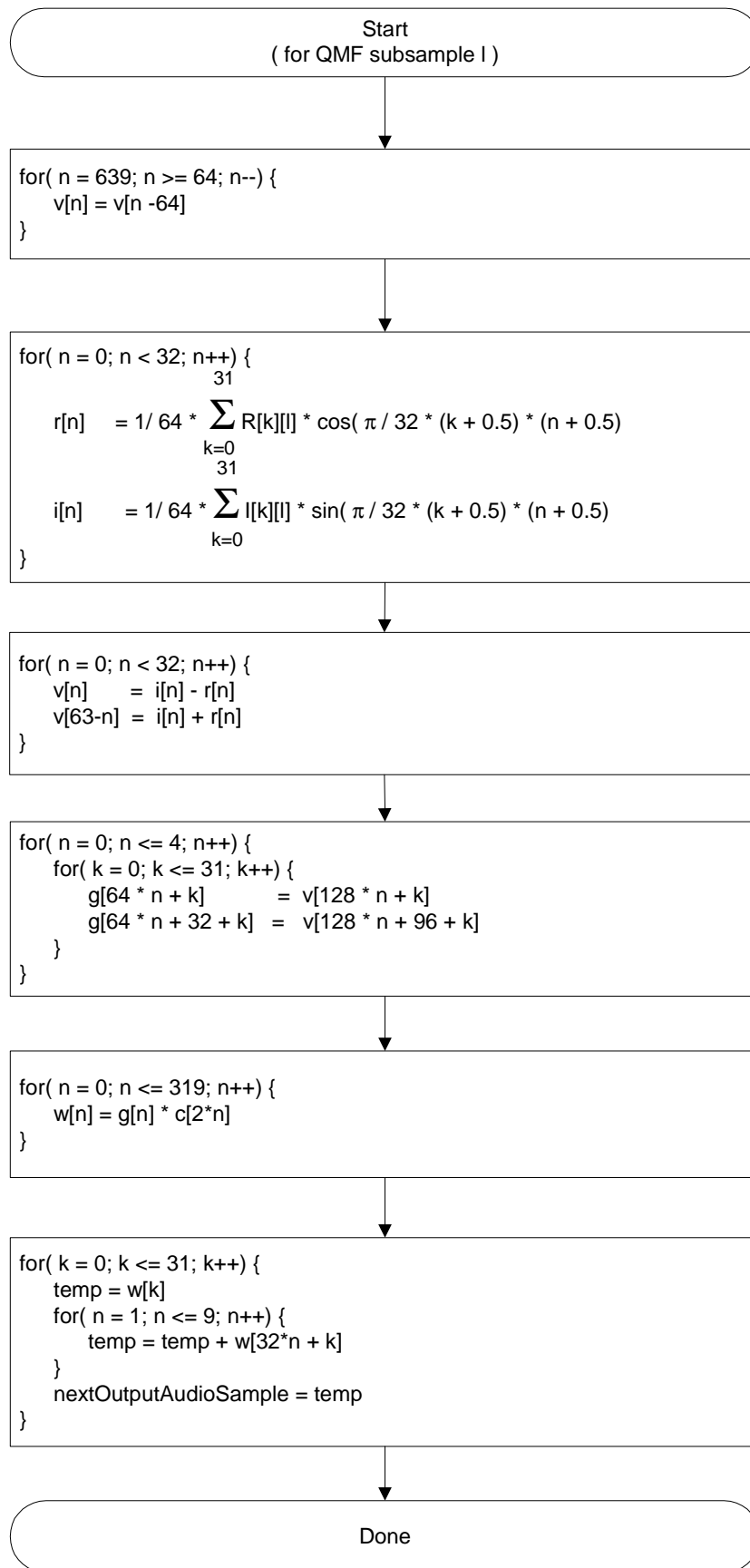


Figure 4 - Flowchart of encoder synthesis QMF bank

Annex A (informative): Change history

Change history							
Date	TSG SA#	TSG Doc.	CR	Rev	Subject/Comment	Old	New
2004-09	25	SP-040637			Presented at SA#25	1.0.0	2.0.0