

Diagrams

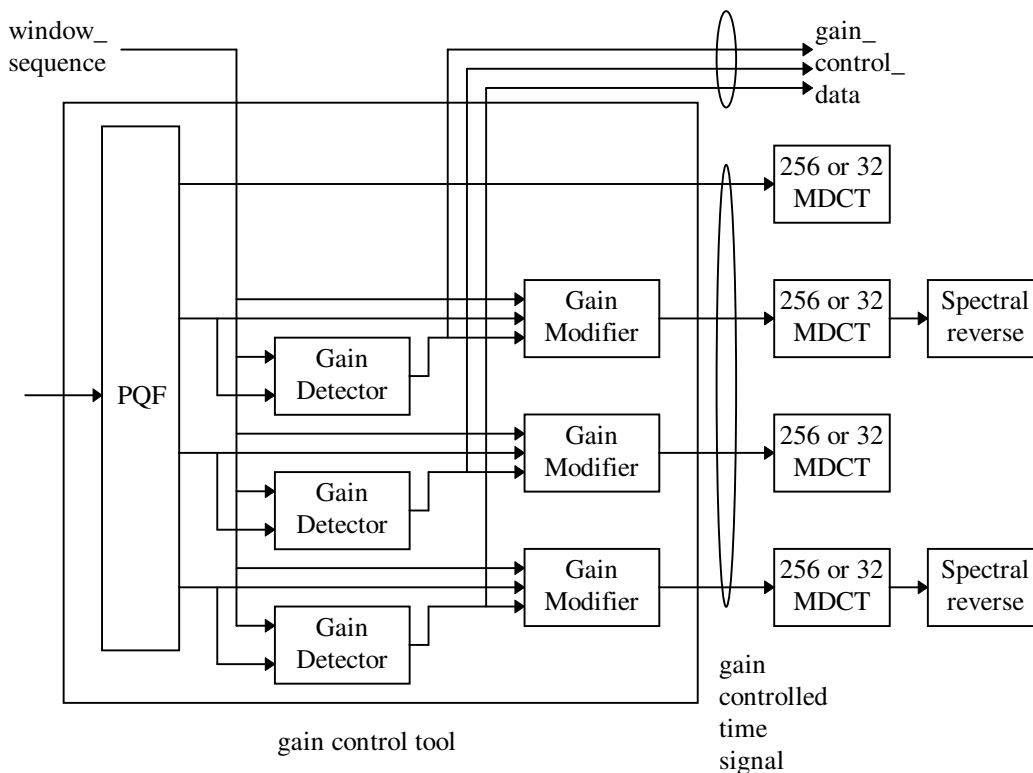


Figure B.2.2.1 -- Block diagram of gain control tool for encoder

Filterbank and Block Switching

A fundamental component in the audio coding process is the conversion of the time domain signals into a time-frequency representation. This conversion is done by a forward modified discrete cosine transform (MDCT).

Encoding process

In the encoder the filterbank takes the appropriate block of time samples, modulates them by an appropriate window function, and performs the MDCT. Each block of input samples is overlapped by 50% with the immediately preceding block and the following block. The transform input block length N can be set to either 2048 or 256 samples. Since the window function has a significant effect on the filterbank frequency response, the filterbank has been designed to allow a change in window shape to best adapt to input signal conditions. The shape of the window is varied simultaneously in the encoder and decoder to allow the filterbank to efficiently separate spectral components of the input for a wider variety of input signals.

Windowing and block switching

The adaptation of the time-frequency resolution of the filterbank to the characteristics of the input signal is done by shifting between transforms whose input lengths are either 2048 or 256 samples. By enabling the block switching tool, the following transitions are meaningful:

from ONLY_LONG_SEQUENCE to	{ ONLY_LONG_SEQUENCE LONG_START_SEQUENCE
from LONG_START_SEQUENCE to	{ EIGHT_SHORT_SEQUENCE LONG_STOP_SEQUENCE
from LONG_STOP_SEQUENCE to	{ ONLY_LONG_SEQUENCE LONG_START_SEQUENCE
from EIGHT_SHORT_SEQUENCE to	{ EIGHT_SHORT_SEQUENCE LONG_STOP_SEQUENCE

Window shape decisions are made by the encoder on a frame-by-frame-basis. The window selected is applicable to the second half of the window function only, since the first half is constrained to use the appropriate window shape from the preceding frame. Figure B.2.3.1 shows the sequence of blocks for the transition (D-E-F) to and from a frame employing the sine function window. The window shape selector generally produces window shape run-lengths greater than that shown in the figure.

The 2048 time-domain values $x'_{i,n}$ to be windowed are the last 1024 values of the previous `window_sequence` concatenated with 1024 values of the current block. The formula below shows this fact:

$$x'_{i,n} = \begin{cases} x_{(i-1),(n+1024)}, & \text{for } 0 \leq n < 1024 \\ x_{i,n}, & \text{for } 1024 \leq n < 2048 \end{cases}$$

Where i is the block index and n is the sample index within a block. Once the window shape is selected, the **window_shape** syntax element is initialized. Together with the chosen **window_sequence** all information needed for windowing exist.

With the window halves described below all **window_sequences** can be assembled.

For **window_shape** == 1, the window coefficients are given by the Kaiser - Bessel derived (KBD) window as follows:

$$W_{KBD_LEFT, N}(n) = \sqrt{\frac{\sum_{p=0}^n [W'(p, \alpha)]}{\sum_{p=0}^{N/2} [W'(p, \alpha)]}} \quad \text{for } 0 \leq n < \frac{N}{2}$$

$$W_{KBD_RIGHT, N}(n) = \sqrt{\frac{\sum_{p=0}^{N-n} [W'(p, \alpha)]}{\sum_{p=0}^{N/2} [W'(p, \alpha)]}} \quad \text{for } \frac{N}{2} \leq n < N$$

where:

W' , Kaiser - Bessel kernel window function, see also [5], is defined as follows:

$$W'(n) = \frac{I_0 \left[\pi \alpha \sqrt{1.0 - \left(\frac{n - N/4}{N/4} \right)^2} \right]}{I_0[\pi \alpha]} \quad \text{for } 0 \leq n \leq \frac{N}{2}$$

$$I_0[x] = \sum_{k=0}^{\infty} \left[\frac{\left(\frac{x}{2} \right)^k}{k!} \right]^2$$

$$\alpha = \text{kernel window alpha factor, } \alpha = \begin{cases} 4 & \text{for } N = 2048 \\ 6 & \text{for } N = 256 \end{cases}$$

Otherwise, for **window_shape** == 0, a sine window is employed as follows:

$$W_{\text{SIN_LEFT},N}(n) = \sin\left(\frac{\pi}{N}\left(n + \frac{1}{2}\right)\right) \quad \text{for } 0 \leq n < \frac{N}{2}$$

$$W_{\text{SIN_RIGHT},N}(n) = \sin\left(\frac{\pi}{N}\left(n + \frac{1}{2}\right)\right) \quad \text{for } \frac{N}{2} \leq n < N$$

The window length N can be 2048 or 256 for the KBD and the sine window. How to obtain the possible window sequences is explained in the parts a)-d) of this clause. All four window_sequences explained below have a total length of 2048 samples.

For all kinds of window_sequences the window_shape of the left half of the first transform window is determined by the window shape of the previous block. The following formula expresses this fact:

$$W_{\text{LEFT},N}(n) = \begin{cases} W_{\text{KBD_LEFT},N}(n), & \text{if } \text{window_shape_previous_block} == 1 \\ W_{\text{SIN_LEFT},N}(n), & \text{if } \text{window_shape_previous_block} == 0 \end{cases}$$

where:

window_shape_previous_block: **window_shape** of the block (i-1).

For the first block to be encoded the **window_shape** of the left and right half of the window are identical.

a) ONLY_LONG_SEQUENCE:

The **window_sequence** == ONLY_LONG_SEQUENCE is equal to one LONG_WINDOW (see Table 8.3) with a total window length of 2048.

For **window_shape**==1 the ONLY_LONG_SEQUENCE is given as follows:

$$W(n) = \begin{cases} W_{\text{LEFT},2048}(n), & \text{for } 0 \leq n < 1024 \\ W_{\text{KBD_RIGHT},2048}(n), & \text{for } 1024 \leq n < 2048 \end{cases}$$

If **window_shape**==0 the ONLY_LONG_SEQUENCE can be described as follows:

$$W(n) = \begin{cases} W_{\text{LEFT},2048}(n), & \text{for } 0 \leq n < 1024 \\ W_{\text{SIN_RIGHT},2048}(n), & \text{for } 1024 \leq n < 2048 \end{cases}$$

The time domain values after windowing ($z_{i,n}$) can be expressed as:

$$z_{i,n} = w(n) \cdot x'_{i,n};$$

b) LONG_START_SEQUENCE:

The LONG_START_SEQUENCE is needed to obtain a correct overlap and add for a block transition from a ONLY_LONG_SEQUENCE to a EIGHT_SHORT_SEQUENCE.

If **window_shape**==1 the LONG_START_SEQUENCE is given as follows:

$$W(n) = \begin{cases} W_{\text{LEFT},2048}(n), & \text{for } 0 \leq n < 1024 \\ 1.0, & \text{for } 1024 \leq n < 1472 \\ W_{\text{KBD_RIGHT},256}(n + 128 - 1472), & \text{for } 1472 \leq n < 1600 \\ 0.0, & \text{for } 1600 \leq n < 2048 \end{cases}$$

If **window_shape**==0 the LONG_START_SEQUENCE looks like:

$$w(n) = \begin{cases} W_{LEFT,2048}(n), & \text{for } 0 \leq n < 1024 \\ 1.0, & \text{for } 1024 \leq n < 1472 \\ W_{SIN_RIGHT,256}(n+128-1472), & \text{for } 1472 \leq n < 1600 \\ 0.0, & \text{for } 1600 \leq n < 2048 \end{cases}$$

The windowed domain values can be calculated with the formula explained in a).

c) EIGHT_SHORT

The **window_sequence** == EIGHT_SHORT comprises eight overlapped SHORT_WINDOWs (see Table 8.3) with a window length of 256 samples each (2048 in total). Each of the eight short windows are windowed separately first. The short window number is indexed with the variable $j = 0, \dots, 7$.

The **window_shape** of the previous block influences the first of the eight short windows ($W_0(n)$) only.

If **window_shape**==1 the window functions can be given as follows:

$$W_0(n) = \begin{cases} W_{LEFT,256}(n), & \text{for } 0 \leq n < 128 \\ W_{KBD_RIGHT,256}(n), & \text{for } 128 \leq n < 256 \end{cases}$$

$$W_{1-7}(n) = \begin{cases} W_{KBD_LEFT,256}(n), & \text{for } 0 \leq n < 128 \\ W_{KBD_RIGHT,256}(n), & \text{for } 128 \leq n < 256 \end{cases}$$

Otherwise, if **window_shape**==0, the window functions can be described as:

$$W_0(n) = \begin{cases} W_{LEFT,256}(n), & \text{for } 0 \leq n < 128 \\ W_{SIN_RIGHT,256}(n), & \text{for } 128 \leq n < 256 \end{cases}$$

$$W_{1-7}(n) = \begin{cases} W_{SIN_LEFT,256}(n), & \text{for } 0 \leq n < 128 \\ W_{SIN_RIGHT,256}(n), & \text{for } 128 \leq n < 256 \end{cases}$$

An EIGHT_SHORT sequence $x'_{i,n}$ is subdivided into eight overlapping segments first and then multiplied with the appropriate window function:

$$z_{i,n} = \begin{cases} x'_{i,n+448} \cdot W_0(n), & \text{for } 0 \leq n < 256 \\ x'_{i,n+576} \cdot W_1(n-256), & \text{for } 256 \leq n < 512 \\ x'_{i,n+704} \cdot W_2(n-512), & \text{for } 512 \leq n < 768 \\ x'_{i,n+832} \cdot W_3(n-768), & \text{for } 768 \leq n < 1024 \\ x'_{i,n+960} \cdot W_4(n-1024), & \text{for } 1024 \leq n < 1280 \\ x'_{i,n+1088} \cdot W_5(n-1280), & \text{for } 1280 \leq n < 1536 \\ x'_{i,n+1216} \cdot W_6(n-1536), & \text{for } 1536 \leq n < 1792 \\ x'_{i,n+1344} \cdot W_7(n-1792), & \text{for } 1792 \leq n < 2048 \end{cases}$$

d) LONG_STOP_SEQUENCE

This window_sequence is needed to switch from a EIGHT_SHORT_SEQUENCE back to a ONLY_LONG_SEQUENCE.

If **window_shape**==1 the LONG_STOP_SEQUENCE is given as follows:

$$W(n) = \begin{cases} 0.0, & \text{for } 0 \leq n < 448 \\ W_{LEFT,256}(n - 448), & \text{for } 448 \leq n < 576 \\ 1.0, & \text{for } 576 \leq n < 1024 \\ W_{KBD_RIGHT,2048}(n), & \text{for } 1024 \leq n < 2048 \end{cases}$$

If **window_shape**==0 the LONG_START_SEQUENCE is determined by:

$$W(n) = \begin{cases} 0.0, & \text{for } 0 \leq n < 448 \\ W_{LEFT,256}(n - 448), & \text{for } 448 \leq n < 576 \\ 1.0, & \text{for } 576 \leq n < 1024 \\ W_{SIN_RIGHT,2048}(n), & \text{for } 1024 \leq n < 2048 \end{cases}$$

The windowed time domain values can be calculated with the formula explained in a).

MDCT

The spectral coefficient, $X_{i,k}$, are defined as follows:

$$X_{i,k} = 2 \cdot \sum_{n=0}^{N-1} z_{i,n} \cos\left(\frac{2\pi}{N}(n + n_0)\left(k + \frac{1}{2}\right)\right) \text{ for } 0 \leq k < N / 2.$$

where:

$z_{i,n}$ = windowed input sequence

n = sample index

k = spectral coefficient index

i = block index

N = window length of the one transform window based on the window_sequence value

$n_0 = (N / 2 + 1) / 2$

The analysis window length N of one transform window of the mdct is a function of the syntax element **window_sequence** and is defined as follows:

$$N = \begin{cases} 2048, & \text{if ONLY_LONG_SEQUENCE (0x0)} \\ 2048, & \text{if LONG_START_SEQUENCE (0x1)} \\ 256, & \text{if EIGHT_SHORT_SEQUENCE (0x2) (8 times)} \\ 2048, & \text{if LONG_STOP_SEQUENCE (0x3)} \end{cases}$$

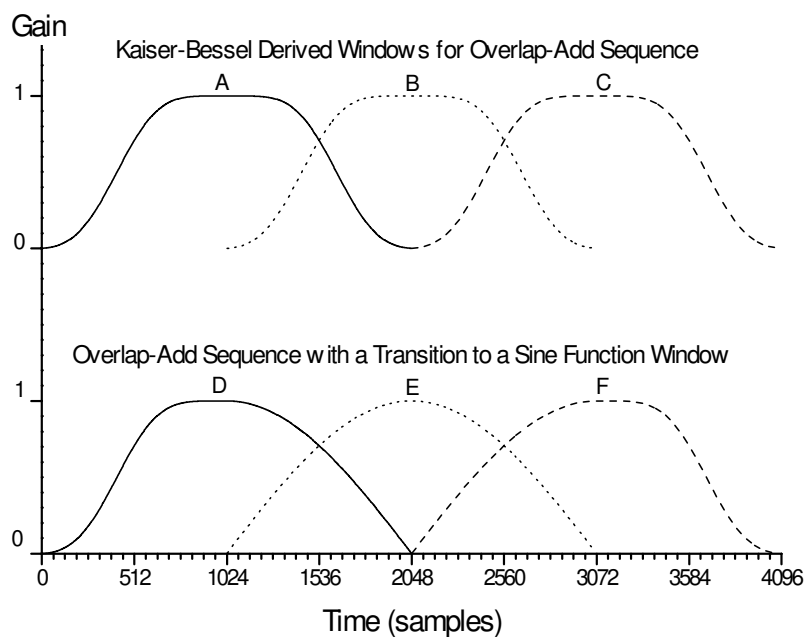
Diagrams

Figure B.2.3.1 -- Example of the Window Shape Adaptation Process.

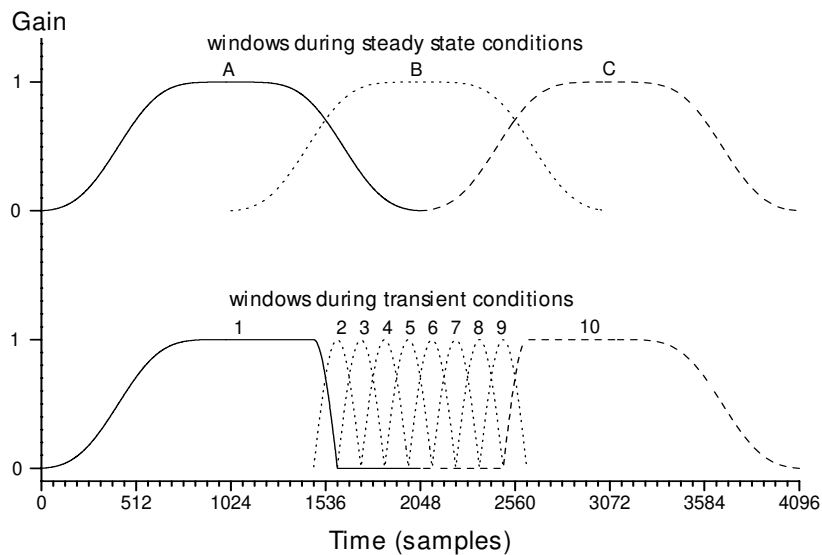


Figure B.2.3.2 – Example of Block Switching During Transient Signal Conditions