# MPEG-4 Low Delay Audio Coding Based on the AAC Codec

Eric Allamanche     Ralf Geiger     Jürgen Herre

Thomas Sporer

Fraunhofer Institut für Integrierte Schaltungen (IIS)
Am Weichselgarten 3
91058 Erlangen, Germany
{alm,ggr,hrr,spo}@iis.fhg.de

### Abstract

Perceptual audio coding is known to deliver high sound quality even at low bit rates for a broad range of audio signals. However, the total delay of the encoder/decoder chain is usually considerably higher than acceptable for two-way communication applications, such as teleconferencing.

This paper discusses the primary sources of algorithmic delay in a perceptual audio codec and describes an MPEG-2 AAC-derived codec which was optimized for very low delay and accepted as the baseline of development for low-delay coding in MPEG-4 version 2 audio.

## 1 Introduction

During the last decade, perceptual audio coding has become one of the most exciting areas of research at the intersection of classical signal processing concepts and more recent knowledge about human perception of sound (psychoacoustics). Additional interest in compressed audio has been stimulated by the advent of multimedia and internet technology. A family of related coding standards have been established e.g. under ISO/MPEG [1, 2, 3, 4] or are currently in progress [5].

Traditionally, however, perceptual audio coding and low-delay coding for communication purposes have been separate worlds. On one hand, perceptual audio codecs provide excellent subjective audio quality for a broad range of signals including speech at bit rates down to 16 kbps. The delay of such a coder/decoder chain usually exceeds 200 ms at very low data rates and in this way is not acceptable for interactive two-way communication, such as telephony or teleconferencing. On the other hand, speech coders (e.g. based on CELP), such as those recommended by ITU-T, meet the delay requirements for these applications, but do not perform very well for non-speech signals. This paper introduces a novel coding scheme which is designed to combine the advantages of perceptual audio coding with the low delay required for two-way communication. The codec is closely derived from the ISO/MPEG-2/4 Advanced Audio Coding (AAC) [3, 4] scheme. It currently forms the baseline of development within version 2 of the MPEG-4 audio standard [5] for the work item 'Low Delay General Audio Coding' meeting the requirement for a maximum algorithmic delay of 20 ms [6].

The paper will first provide a general overview over the structure of perceptual audio coding schemes and then analyze the sources of coding delay inherent in the encoding/decoding chain of such schemes. Subsequently, the design principles of the low-delay AAC coder are introduced. Finally, first results on the performance of the coder will be given and areas of applications identified.

## 2 Perceptual Audio Coding Schemes

Figure 1 shows the basic block diagram of a perceptual audio coding scheme.

- An *analysis filter bank* is used to decompose the input signal in its spectral components. Examples of filter banks used are polyphase filter banks [7] and filter banks based on the modified discrete cosine transform (MDCT) [8].

- A block, called *estimation of masked threshold*, is used to analyze the input signal from a perceptive point of view. The masked threshold depends on the spectral and temporal structure of the input signal [9]. The masking effect of a tonal signal is smaller than the masking effect caused by noise-like signals. For stereo signals the inter-channel dependencies must be taken into account, too [10, 11].

- The audio signal is quantized in a block called *quantization and coding*. On one hand, the quantization must be sufficiently coarse in order not

to exceed the target bit rate. On the other hand, the quantization error should be shaped to be below the estimated masked threshold if possible.

- The quantized values together with side information is *multiplexed* into one *bitstream*.

There are several parameters of coding schemes which need careful adaptation. The most prominent is the number of samples coded together in one frame. Note that the frame length is independent from the length of the impulse response of the filter bank (resp. the length of the analysis window of the MDCT). For each frame of audio data the transmitted data contains the quantized samples and some common side information. The overhead for this side information becomes negligible if the number of samples in a frame is sufficiently large. Using long frames for coding also offers the opportunity to use filter banks with a good frequency separation which is of benefit if the frequency structure of a signal is constant over time. In the cause of rapidly changing input signals (transients) long frames are unfavorable because the temporal spread of quantizations will lead to so-called "pre-echos". For such signals, the size of a frame should therefore correspond closely to the temporal resolution of the human ear. This can be achieved by using rather short frames or by changing the frame length depending on the input signal [8].

Long frames are also necessary to discriminate between tonal and noise-like signals.

## 2.1 Examples of Perceptual Audio Coding Schemes

### 2.1.1 MPEG-1 Layer-3 ("MP3")

- Filter bank
  MPEG-1 Layer-3 uses a hybrid filter bank consisting of a 32 band polyphase filter bank and a modified discrete cosine transform (MDCT) in each of these bands (see Figure 2). The windows of succeeding MDCTs overlap by 50 % of the window length. The MDCT provides the possibility of switching between different lengths and shapes of the analysis window. Figure 3 shows the window types used in Layer-3. The overlapping parts of succeeding windows must fit to each other. This limits the possibility of switching between different types. Figure 4 shows a typical sequence of windows. In the case of the window type "SHORT" each long transform is split into three short transforms. This window type reduces the "smear-out" of the quantization error over time in the cause of attacks. The "LONG" window is best used for

3

quasi stationary signals. Due to the fact that a "START" window must be inserted between "LONG" and "SHORT" a look-ahead for the analysis of the input signal is necessary.

- Quantization and coding
  Layer-3 uses a non-linear quantizer and different Huffman code tables. Quantization and coding is done in two nested loops: The inner loop (rate control loop) checks the amount of bits necessary and increases the quantizer step size for all frequency components. The outer loop (distortion control loop) compares the quantization error with the estimated masked threshold and decreases the quantization step size for frequency ranges where the distortion is above the threshold.

- Bit reservoir
  The necessary data rate to encode audio signals in a perceptually lossless way depends on the input signal. It is useful to vary the data rate over time to track this behavior. In Layer-3 this is done via a bit reservoir: If a frame is easy to code then the spare bits are put into the bit reservoir. If a frame needs more than the average amount of bits this extra bit allocation is taken from the bit reservoir. The size of the bit reservoir depends on bit rate and sampling rate of the input signal. The maximum deviation from the average number of bits in a frame is 4096, the maximum number of bits in one frame is 7680.

- Multiplexing
  Each frame of 1152 samples consists of two sub-frames, the so-called granules. In principle the two granules of a frame are coded independently. In the cause of quasi-stationary signals the second granule can reuse part of the side information of the first granule.

### 2.1.2  MPEG-2 Advanced Audio Coding (AAC)

The basic structure of AAC is similar to the basic structure of Layer-3. Only the main differences, which are important in the context of low delay audio coding, are discussed here. A more comprehensive introduction into AAC can be found in [12] .

- Filter bank
  The frame length of AAC is 1024 samples. A switched MDCT filterbank is used which allows to chose between a time/frequency resolution of 1024 lines or 8 sets of 128 lines each. Figure 5 shows the window types used in AAC and Figure 6 shows a typical sequence of windows.

4

- Temporal Noise Shaping (TNS)
  A tool called TNS [13] uses a predictor along the frequency axis to shape the quantization error in the time domain. This moves the distortion below the peaks in the time signal without influencing the frequency structure of the error. Due to some properties of the MDCT (see [8]) TNS works best if the overlap between succeeding windows is small.

- Bit reservoir
  The maximum number of bit per channel is 6144 bit. The size of the bit reservoir of AAC is 6144 bit minus the average number of bits per frame.

- Changes of lossless coding
  No subdivision of frames into granules is done for "LONG", "START" and "STOP" blocks. Succeeding windows in a "SHORT" block can share part of the side information. The encoding of scale factors is improved. An optional prediction scheme reduces the amount of bits necessary to encode quasi-stationary signals.

### 2.1.3  MPEG-4 Advanced Audio Coding (AAC)

MPEG-2 AAC was used as the basis of MPEG-4 general audio coding. Several tools were added to improve the quality of AAC at low bit rates:

- Long Term Predictor (LTP)
  A new predictor uses the decoded signal of preceeding frames as an estimate of the signal in the current frame. If this estimate is sufficiently good, the lag and the gain are transmitted together with a vector signaling in which scale factors the predictor is active. The LTP reduces the amount of bits necessary to encode signals with quasi-stationary tonal components.

- Perceptual Noise Substitution (PNS)
  For frequency ranges where all lines contain noise-like components only the average power of this noise is transmitted instead of each individual line [14].

## 3  Delay in Perceptual Audio Coding

One of the main objectives of an audio coding scheme is to provide the best trade-off between quality and bit rate, or to achieve transparent audio quality

at the smallest possible bit rate. In general, this goal can only be achieved at the expense of a certain encoding/decoding delay.

For a generic audio coder, the overall delay can be viewed as the sum of the delay contributions related to the following codec parameters:

- Frame length
- Filter bank delay
- Look-ahead time for block switching
- Use of bit reservoir

This section will give an overview of the background of the different delay contributions occurring in a perceptual audio codec. All calculations will be based on the so-called "algorithmic delay" which describes the theoretical minimum delay allowed by an algorithm assuming negigible delay contributions due to speed of calculation, bitstream transmission or other implementation or application specific circumstances. As an example, the algorithmic delay of an MPEG-2 AAC codec at low data rates will be calculated in order to give some impression about the order of magnitude of delay for a modern high–performance perceptual codec.

**Frame length**

For block-based processing, a certain amount of time has to pass to collect the samples belonging to one block. The delay caused by this collection process ("framing delay") increases linearly with the frame length. This delay (in samples) is denoted as $N_{framing}$.

$$N_{framing} = frame\_size \tag{1}$$

**Filter bank delay**

In order to exploit the spectral masking properties of the human auditory system, perceptual audio coding schemes employ an analysis/synthesis filter bank pair. While numerous types of filter banks have been used for audio coding, the Modified Discrete Cosine Transform (MDCT) [8] has been used extensively for modern audio codecs, like MPEG-2 AAC and MPEG-4, and has shown its merits for compression at very low bit rates. Due to the overlap-add characteristics of the MDCT with an overlap of 50% between subsequent windows, this filter bank causes an additional delay identical to the framing delay.

$$N_{filter\_bank\_MDCT} = frame\_size \tag{2}$$

6

### Look-ahead delay for block switching decision

As mentioned above, most modern audio codecs that are designed to operate at very low bit rates use an MDCT filter bank with a high spectral resolution. In particular, high efficiency can be reached with long frames (about 20 ms and more) for stationary signals. In the case of transient signals, like percussive sounds or "attacks", this would lead to the so-called pre-echo phenomenom which is well known in audio coding. This can be avoided by using dynamic block switching [8]. The main idea is to dynamically switch between different filter bank analysis/synthesis window sizes, and thus reduce the noise spreading in time over one block. Due to restrictions in the permissible sequence of window types, no "instantaneous" switching between long and short windows is possible with this block switching strategy but an intermediate transition window type ("start block") has to be inserted in between long and short windows. Therefore the detection of the optimum window type requires a look-ahead and thus requires a further delay in the encoder. In general, an encoder using block switching incurs an additional delay of

$$N_{look\text{-}ahead} = frame\_size \cdot \frac{numShortWindows + 1}{2 \cdot numShortWindows} \tag{3}$$

where $numShortWindows$ is the number of short windows which fit in a frame (e.g. 8 for MPEG-2 AAC).

### Use of bit reservoir

Since not all segments of an audio signal are equally demanding to code, the number of bits needed to code a specific frame will vary. To end up with a constant bit rate, the bit reservoir mechanism has proven to be useful. Since the use of the bit reservoir is equivalent to a local variation in bit rate, the size of the input buffer of the decoder must be adapted to the maximum local bit rate (i.e. the maximum number of bits which can be allocated for a single frame per channel). The decoder has to wait at least until this input buffer is read before audio output can be started. Thus, increasing the size of the bit reservoir will also increase the overall codec delay. In fact, the overall delay of the audio coder may be dominated entirely by the size of the bit reservoir.

The delay expressed in terms of samples caused by the bit reservoir is

$$N\_bitres = \frac{bitres\_size}{bitrate} \cdot F_s \tag{4}$$

where $bitres\_size$ is the bit reservoir size expressed in bits and $F_s$ is the sampling rate in Hz.

7

**Overall delay**

From the discussion above, the overall delay can be calculated as follows:

$$t_{delay} = \frac{N_{framing} + N_{filter\_bank} + N_{look\text{-}ahead} + N_{bitres}}{F_s} \qquad (5)$$

with:

$F_s$:     the coder sampling rate (in Hz)

$N_{framing}$:     the framing delay
(MPEG-2 AAC: 1024 samples)

$N_{filter\_bank}$:     the delay caused by the filter bank
(MPEG-2 AAC: 1024 samples)

$N_{look\text{-}ahead}$:     look-ahead delay for block switching
(MPEG-2 AAC: $1024 \cdot \frac{9}{16}$ samples)

$N_{bitres}$:     delay due to the use of the bit reservoir
(MPEG-2 AAC: maximal $6144 \cdot \frac{F_s}{bitrate} - 1024$ samples)

Note that the overall delay scales inversely with the sampling frequency.

For a standard MPEG-2 AAC coder running at 24 kbps and a sampling rate of 24 kHz[1], the resulting overall codec delay is about 109.3 ms without the use of the bit reservoir. Assuming the nominal size of the input buffer as indicated in the MPEG-2 AAC standard (6144 bit/channel), a maximum additional delay of 213.3 ms is incurred, leading to a total delay of 322.6 ms.

# 4 The MPEG-4 Low Delay Audio Coder

The overall approach taken for the design of the low delay codec was to rely as much as possible on the proven architecture of MPEG-2/4 AAC [3, 4] and to achieve the desired low delay functionality with a minimum number of changes. In particular, the low delay codec is derived from the so-called MPEG-4 General Audio object type, i.e. a codec consisting of the standard MPEG-2 AAC codec plus the PNS (Perceptual Noise Substitution) [14] and the LTP (Long Term Prediction) tools [4]. Furthermore, stereo and low sampling rate modes are supported [15, 16, 17].

The following modifications were performed on the standard algorithm to achieve low delay operation:

---

[1]For MPEG-2 AAC the optimum quality at a bit rate of 24 kbps is achieved at a sampling rate of 24 kHz.

## Frame length and filter bank delay

The frame length has been reduced to 512 or 480 samples. The length of the analysis window has been reduced to 1024 or 960 time domain samples corresponding to 512 and 480 spectral values, respectively. This leads to a framing delay of 512 or 480 samples. As described above, the MDCT analysis/synthesis filter bank processing causes a further delay of the same size.

## Block switching

Due to the considerable contribution of the look-ahead time to the overall delay, no block switching is used. The temporal spread of quantization noise ("pre-echo") is handled by the Temporal Noise Shaping (TNS) [13] module.

## Window shape

Besides the standard sine window shape, the low delay codec uses a new window shape which exhibits a lower overlap between subsequent frames (see Figure 8). Selection of this window shape allows the TNS module to provide even better protection against pre-echo effects by minimizing the temporal aliasing effect which is inherent in the MDCT's Time Domain Aliasing Cancellation (TDAC) concept [18]. A typical sequence of windows is shown in Figure 9. Figure 10 illustrates the improvement by using the low overlap window shape for coding of transient signal parts. Note that this dynamic adaptation of the window shape does not imply any additional delay.

## Bit reservoir

Use of the bit reservoir is minimized in order to reach the desired target delay. As one extreme case, no bit reservoir is used at all.

## Overall delay

For the low delay codec the overall delay can be calculated as followed:

$$t_{delay} = \frac{N_{framing} + N_{filter\_bank} + N_{reduced\_bitres}}{F_s} \tag{6}$$

or without the bit reservoir:

$$t_{delay} = \frac{N_{framing} + N_{filter\_bank}}{F_s} \tag{7}$$

with:

$F_s$:      the coder sampling rate (in Hz)

$N_{framing}$:      the framing delay
(512 or 480 samples)

$N_{filter\_bank}$:      the delay caused by the filter bank
(512 or 480 samples)

$N_{reduced\_bitres}$:      size of the reduced bit reservoir, expressed in samples

So for a window length of 960 samples, a sampling frequency of 48 kHz and without use of a bit reservoir, the overall algoritmic delay of the low delay codec is 20 ms which is commensurate with widely used speech codecs.

Note that it would be difficult to achieve similar low algorithmic delay values by using MPEG-1 Layer-3, mainly due to two reasons: Firstly, the composite hybrid filterbank used in Layer-3 exhibits a higher filterbank related delay than a plain MDCT with the same spectral resolution. Furthermore, since no TNS tool is available with Layer-3, it is necessary to resort to block switching techniques (and accept the associated look-ahead delay) in order to avoid pre-echo problems for transient signals.

# 5 Results

In order to assess the sound quality delivered by the low delay codec, a listening test was carried out according to the usual guidelines provided by the MPEG-4 core experiment methodology. Specifically, this test should answer the question how much penalty in sound quality arises from restricting the delay of a codec. To this end, the performance of a low delay codec (960 points sine window, no bit reservoir, running at 32 kbps at $F_s$ =48kHz, overall delay 20 ms) was compared to a standard MPEG-2 AAC Main Profile codec running at 24 kbps at $F_s$=24kHz [16].

- The low delay codec was compared to the MPEG-2 AAC codec using the comparison test methodology. The order of presentation was OAB-OAB, where O was the original signal, A was coded with one of the two codecs and B was coded with the other codec.

- To compensate for positional effects, each test sequence was presented a second time with reversed order of the coded signals. So the signal A in the first comparison was presented as signal B in the second comparison.

- The seven grade comparison scale was used (see Table 1). The listeners were asked to give integer grades.

- Nine experienced listeners participated in the test.

- The playback was done on Stax Lambda Pro and Stax Lambda Nova headphones.

- All 12 items of the MPEG standard test set were used (see Table 2).

Figure 7 and Table 3 show the results of the test. Coder A corresponds to the low delay coder while coder B is the MPEG-2 AAC coder. As can be seen from these test results, the performance of the low delay codec is roughly comparable to that of the unconstrained MPEG-2 AAC for most of the test items. For tonal items with a densely populated spectrum (e.g. Harpsichord or Plucked Strings) it is visible that the low-delay coder does not achieve as much coding gain as the unconstrained coder. Note that the unconstrained coder can make use of a more than 4.2 times finer frequency resolution.

# 6   Applications

Due to their large system delay "state of the art" audio coding schemes are not applicable for two way communication. Telephone and video conferencing applications today use either speech coding schemes, which can only provide speech quality and usually fail when stressed with more complex audio signals like music, or use higher data rates.

The proposed low delay audio coding scheme based on AAC (LD-AAC) can now bridge the gap between speech coding schemes and high quality audio coding schemes. Two way communication with LD-AAC is possible on usual analog telephone lines and via ISDN connections.

Usual telephone lines provide a maximum data rate of about 28.8 kbps (V.34, [19]). The audio quality of LD-AAC at that bit rate is similar to AAC at 20 kbps. The bandwidth of the audio signal is about 7 kHz, the perceived quality is far above the usual "telephone quality level".

ISDN lines provide a data rate of 64 kbps. Using LD-AAC the audio bandwidth can be up to 15 kHz. The quality is expected to fulfill the ITU-R requirements for commentary channels [20]. The low system delay of LD-AAC open new possibilities for broadcasting: Live interviews via (ISDN-) lines with far better audio quality can enrich the programme.

# 7 Conclusions

This paper presents a modified MPEG-2/4 AAC coder which fulfills an algorithmic delay requirement of 20 ms and thus enables applications which require full-duplex communication. Compared to known CELP coders, the codec is capable of coding both music and speech signals with good quality. Unlike speech coders, however, the achieved coding quality scales up nicely with bit rate.

A listening test demonstrated that the low delay codec running at a bit rate of 32 kbps achieved a quality close to the one of a Main Profile AAC codec running at a bit rate of 24 kbps for many items. The computational complexity is significantly lower than AAC Main Profile.

The new codec was accepted as the baseline of development within version 2 of MPEG-4 audio.

# References

[1] ISO/IEC JTC1/SC29/WG11 Moving Pictures Expert Group. Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s. International Standard 11172-3, ISO/IEC, 1993.

[2] ISO/IEC JTC1/SC29/WG11 Moving Pictures Expert Group. Generic Coding of Moving Pictures and Associated Audio: Audio. International Standard 13818-3, ISO/IEC, 1994.

[3] ISO/IEC JTC1/SC29/WG11 Moving Pictures Expert Group. Generic Coding of Moving Pictures and Associated Audio: Advanced Audio Coding. International Standard 13818-7, ISO/IEC, 1997.

[4] ISO/IEC JTC1/SC29/WG11 Moving Pictures Expert Group. Coding of Audio-Visual Objects: Audio. International Standard 14496-3, ISO/IEC, 1999.

[5] ISO/IEC JTC1/SC29/WG11 Moving Pictures Expert Group. Coding of Audio-Visual Objects: Audio. Working Draft 14496-3 Amd 1, ISO/IEC, 1999.

[6] ISO/IEC JTC1/SC29/WG11 Moving Pictures Expert Group. MPEG-4 Requirements, version 10. Document N2562, ISO/IEC, Roma, December 1998.

[7] Kh. Brandenburg and G. Stoll. The ISO/MPEG-audio codec: A generic standard for coding of high quality digital audio. In *92nd AES-Convention*, Vienna, 1992. preprint 3336.

[8] Th. Sporer, K. Brandenburg, and B. Edler. The use of multirate filter banks for coding of high quality digital audio. In *6th European Signal Processing Conference (EUSIPCO)*, volume 1, pages 211–214, Amsterdam, June 1992. Elsevier.

[9] E. Zwicker and H. Fastl. *Psychoacoustics - Facts and Models*. Springer-Verlag, Berlin, 1990.

[10] Jens Blauert. *Räumliches Hören*. Hirzel-Verlag, Stuttgart, 1974. (in german).

[11] Jens Blauert. *Räumliches Hören - Nachschrift: Neue Ergebnisse und Trends seit 1972*. Hirzel-Verlag, Stuttgart, 1985. (in german).

[12] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, M. Dietz H. Fuchs, J. Herre, G. Davidson, and Y. Oikawa. ISO/IEC MPEG-2 Advanced Audio Coding. In *101st AES-Convention*, Los Angeles, 1996. preprint 4382.

[13] Jürgen Herre and James D. Johnston. Enhancing the Performance of Perceptual Audio Coders by Using Temporal Noise Shaping (TNS). In *101st AES-Convention*, Los Angeles, 1996. preprint 4384.

[14] Jürgen Herre and Donald Schulz. Extending the MPEG–4 AAC Codec by Perceptual Noise Substitution. In *104th AES-Convention*, Amsterdam, 1998. preprint 4720.

[15] Jürgen Herre, Eric Allamanche, Ralf Geiger, and Thomas Sporer. Proposal for a low delay MPEG-4 audio coder based on AAC. MPEG98/M4139, ISO/IEC JTC1/SC29/WG11, October 1998.

[16] Jürgen Herre, Eric Allamanche, Ralf Geiger, and Thomas Sporer. Information on MPEG-4 low delay audio coding. MPEG98/M4306, ISO/IEC JTC1/SC29/WG11, October 1998.

[17] Jürgen Herre, Eric Allamanche, Ralf Geiger, and Thomas Sporer. Update on MPEG-4 low delay audio coding. MPEG98/M4307, ISO/IEC JTC1/SC29/WG11, October 1998.

[18] J. Princen, A. Johnson, and A. Bradley. Subband/transform coding using filter bank designs based on time domain aliasing cancellation. In *Proceedings of the ICASSP*, pages 2161–2164, New York, 1987. IEEE.

[19] ITU-T. Recommendation V.34 "A modem operating at data signalling rates of up to 33 600 bit/s for use on the general switched telephone network and on leased point-to-point 2-wire telephone-type circuits", February 1998.

[20] ITU-R. Recommendation BS.1115 "Low bit-rate audio coding", November 1993.
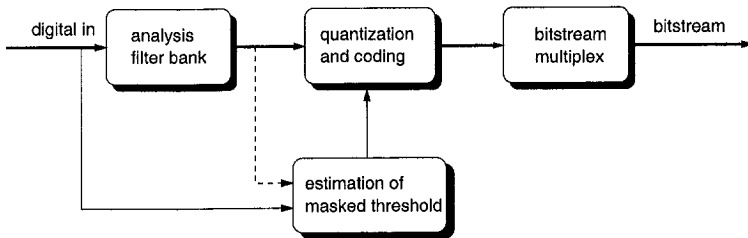
**Figure 1:** Basic black diagram of perceptual audio coding
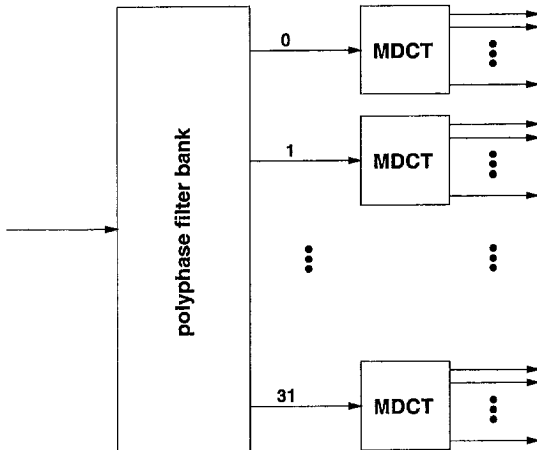


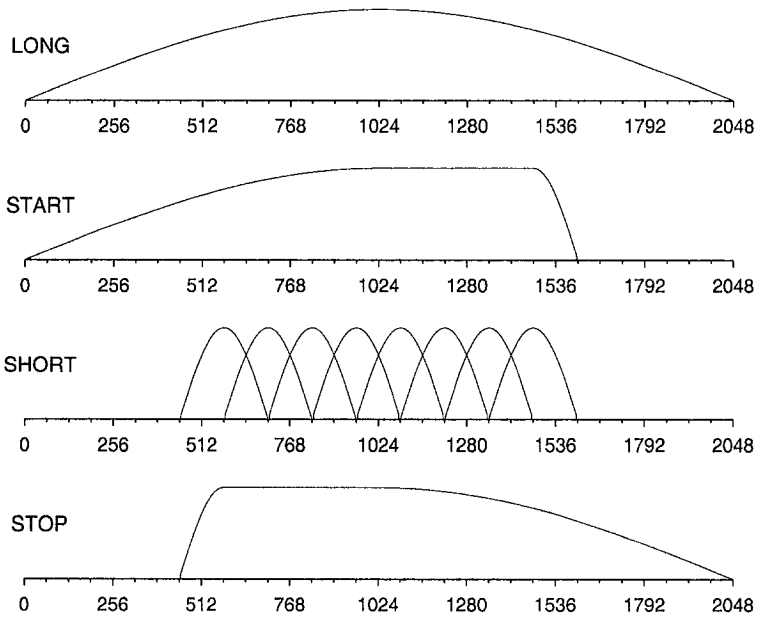**Figure 2:** Hybrid filter bank used in Layer-3

15
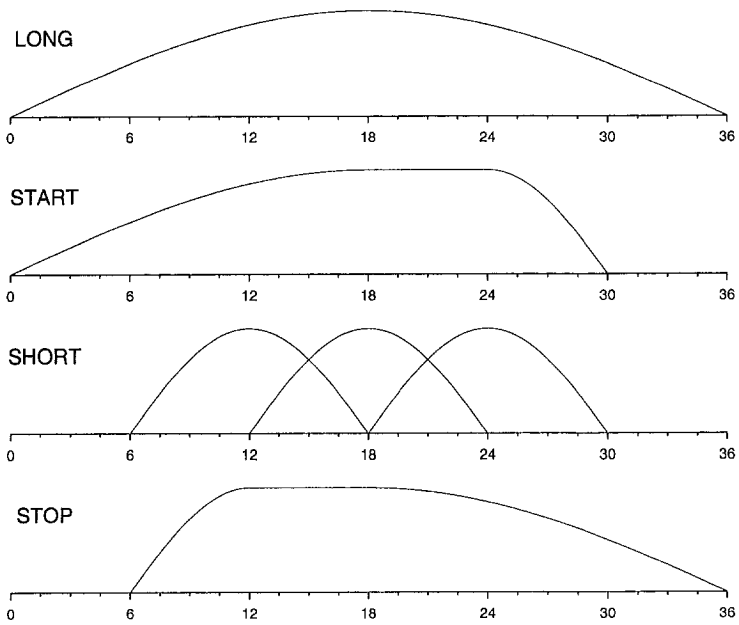
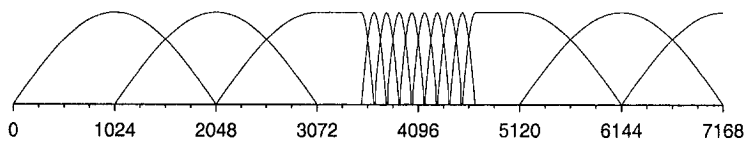**Figure 3:** Window types used in Layer-3
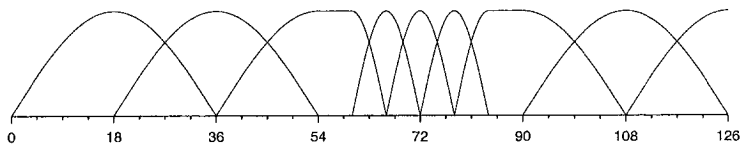


**Figure 4:** Typical sequence of windows as used in Layer-3

16

**Figure 5:** Window types used in AAC



**Figure 6:** Typical sequence of windows as used in AAC

17

| Comparison of A and B | Score |
|---|---|
| A much better than B | +3 |
| A better than B | +2 |
| A slightly better than B | +1 |
| A equal to B | 0 |
| A slightly worse than B | -1 |
| A worse than B | -2 |
| A much worse than B | -3 |

**Table 1:** Seven grade comparison scale

| Test signal | Content |
|---|---|
| sc01 | Trumpet solo & orchestra |
| sc02 | Symphonic orchestra |
| sc03 | Contemporary pop music |
| es01 | English female speaker |
| es02 | German male speaker |
| es03 | English female speaker |
| sm01 | Bagpipes |
| sm02 | Glockenspiel |
| sm03 | Plucked strings |
| si01 | Harpsichord |
| si02 | Castanets |
| si03 | Pitch pipe |

**Table 2:** Standard set of test items

| items | mean | size of 95% conf.int. | upper boundary of 95% conf.int. | lower boundary of 95% conf.int. |
|---|---|---|---|---|
| sc01 | -0,61 | 0,34 | -0,27 | -0,95 |
| sc02 | -0,72 | 0,31 | -0,41 | -1,03 |
| sc03 | -0,78 | 0,38 | -0,40 | -1,16 |
| es01 | -0,06 | 0,50 | 0,44 | -0,55 |
| es02 | -0,67 | 0,41 | -0,26 | -1,07 |
| es03 | -0,28 | 0,44 | 0,16 | -0,72 |
| sm01 | -0,83 | 0,15 | -0,68 | -0,99 |
| sm02 | 1,44 | 0,42 | 1,86 | 1,02 |
| sm03 | -1,06 | 0,36 | -0,70 | -1,41 |
| si01 | -1,94 | 0,10 | -1,84 | -2,05 |
| si02 | -0,17 | 0,63 | 0,47 | -0,80 |
| si03 | -0,17 | 0,46 | 0,30 | -0,63 |
| overall mean: | -0.48611 | | | |

**Table 3:** Comparison of LD-AAC at 32 kbps and AAC at 24 kbps
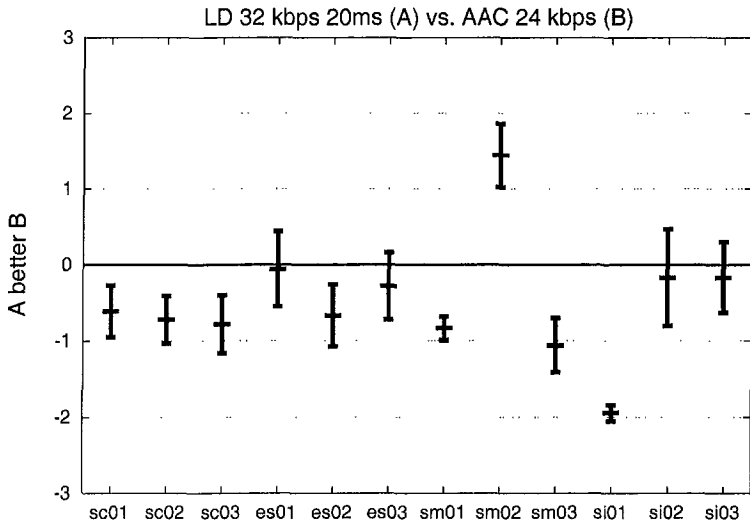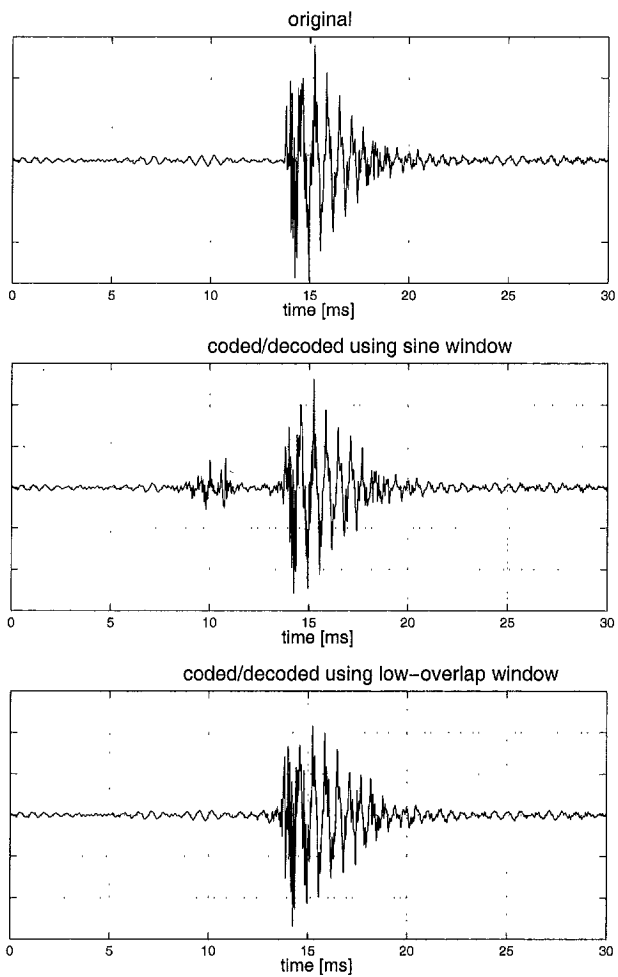


**Figure 7:** Comparison of LD-AAC at 32 kbps and AAC at 24 kbps
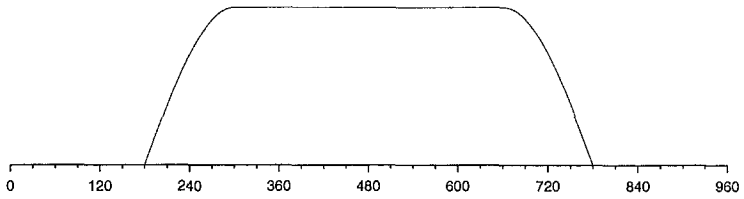
**Figure 8:** Low overlap window
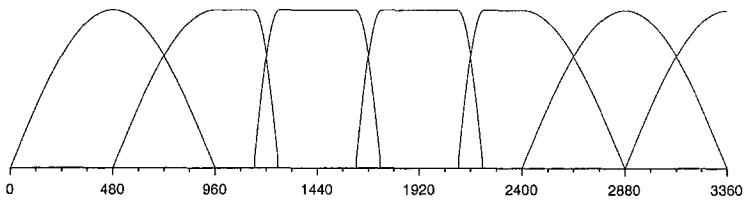


**Figure 9:** Typical sequence of windows with window shape adaptation

20

**Figure 10:** Example: Reduction of temporal aliasing by using a low overlap window sequence like shown in Figure 9.
Item: si02 (castanets), bit rate: 64 kbps, sampling rate: 48 kHz, system delay: 20 ms (no bit reservoir used).