# LOW DELAY FILTERBANKS FOR ENHANCED LOW DELAY AUDIO CODING

*Markus Schnell*, Ralf Geiger*, Markus Schmidt*, Markus Multrus**
*Michael Mellar*, Jürgen Herre*, Gerald Schuller***

*Fraunhofer IIS, Erlangen, Germany, **Fraunhofer IDMT, Ilmenau, Germany

## ABSTRACT

Low delay perceptual audio coding has recently gained wide acceptance for high quality communication. While common schemes are based on the well-known Modified Discrete Cosine Transform (MDCT) filterbank, this paper describes novel coding algorithms that, for the first time, make use of dedicated low delay filterbanks, thus achieving improved coding efficiency while maintaining or even reducing the low codec delay. The MPEG-4 Enhanced Low Delay AAC (AAC-ELD) coder currently under development within ISO/MPEG combines a traditional perceptual audio coding scheme with spectral band replication (SBR), both running in a delay-optimized fashion by using low delay filterbanks.

## 1. INTRODUCTION

Recent years have seen the evolution of traditional communication into a system of most diverse channels, no longer restricted to mere speech transmission. Growing requirements regarding compression efficiency and quality call for continuing enhancements of existing technologies. Quality criteria of modern communication include full audio bandwidth, excellent subjective quality and low latency. Potential comunication scenarios are multi-user Voice-over-IP connections, video-conferencing and wireless intercom systems, all requiring an efficient low delay perceptual audio coder.

A promising technique for such coders is the low delay filterbank framework, presented in [1]. However, current state-of-the-art low delay perceptual audio codecs, such as MPEG-4 Low Delay Advanced Audio Coding (AAC-LD) [2], utilize traditional orthogonal filterbanks (e.g. MDCT). This paper shows how dedicated low delay filterbanks can be applied to enhance a combination of AAC-LD and Spectral Band Replication (SBR) as known from MPEG-4 High Efficiency AAC [2]. The resulting codec shows a higher coding efficiency while maintaining the low delay required for real-time two-way communication. To this end, both the MDCT filterbank in the AAC-LD core codec and the QMF filterbank in the SBR stage are replaced by low delay filterbanks.

## 2. STATE OF THE ART

### 2.1. Low Delay Filterbanks

The purpose of low delay filterbanks is to reduce their reconstruction delay independently of the filter length while still maintaining the perfect reconstruction property. This cannot be achieved with traditional filterbanks, like the TDAC filterbanks [3] such as the MDCT. These so-called para-unitary or orthogonal filterbanks employ symmetric window functions and thus have a system delay identical to the window length minus one.

Some of the first low delay filterbanks were described in [4, 5] in the context of a generalized system delay, i.e. the system delay is no longer connected to the filter length. [4] described a direct design method via a numerical optimization. This approach did not guarantee perfect reconstruction and offered no simple way to obtain a fast implementation. [5] describes an optimization method for cosine modulated filter banks. While this leads to a considerably more efficient implementation, perfect reconstruction still was not a feature.

The design method used in this paper was first described in [6, 7], and later in [1, 8] combining the desired properties. The resulting filterbanks have the same cosine modulation function as the traditional MDCT, but can have longer window functions which can be non-symmetric, with a generalized or low reconstruction delay.

#### 2.1.1. Mathematical Description

Although the design method allows extensions of the MDCT in both directions, only an extension of $E$ blocks to the past is applied here, where each block comprises $M$ samples.

**Analysis:** The frequency coefficient $X$ of band $k$ and block $i$ inside an $M$-channel filterbank is defined as

$$X_{i,k} = -2 \sum_{n=-E\cdot M}^{2M-1} p_A(n) \cdot x(n) \cos[\frac{\pi}{M}(n + \frac{1}{2} - \frac{M}{2})(k + \frac{1}{2})] \tag{1}$$

for $0 \leq k < M$, where $n$ is a sample index and $p_A$ is an analysis window function.

**Synthesis:** The demodulated vector $z$ is defined as

$$z_{i,n} = -\frac{1}{M} \sum_{k=0}^{M-1} p_S(n) \cdot X_{i,k} \cos[\frac{\pi}{M}(n+\frac{1}{2}-\frac{M}{2})(k+\frac{1}{2})] \tag{2}$$

for $0 \leq n < M(2+E)$ and $p_S$ is a synthesis window compatible with $p_A$.

**Overlap Add:** The reconstructed signal $\hat{x}$ can be obtained by

$$\hat{x}_{i,n} = \sum_{j=-(E+1)}^{0} z_{i+j,n-j\cdot M} \tag{3}$$

978-1-4244-1619-6/07/$25.00  

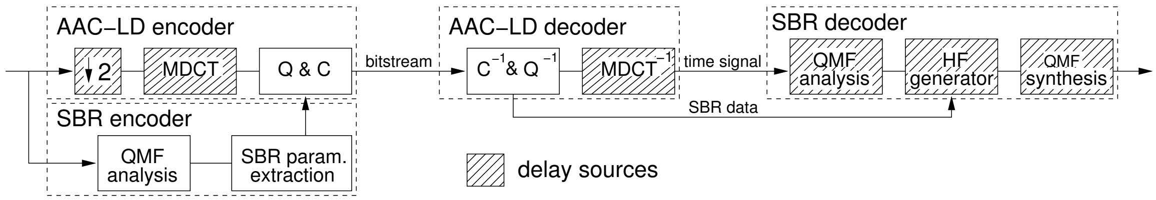


Figure 1: Sources of delay in the encoder/decoder process of AAC-LD in combination with SBR

### 2.2. MPEG-4 Low Delay AAC and High-Efficiency AAC

MPEG-4 Low Delay AAC (AAC-LD) is the low delay variant of the Advanced Audio Coding (AAC) codec standardized within MPEG-4 [2]. It shares the general structure of AAC Low Complexity (AAC-LC) [2][9], and is designed as a filterbank based audio coder (see Figure 1). A time domain audio signal is transformed into a spectral representation using an MDCT. The spectral components are quantized according to the requirements of a psychoacoustic model before they are entropy coded and multiplexed into a bitstream ('Q & C'). For a more detailed description see [10].

AAC-LD achieves an algorithmic delay of down to 20 ms by utilizing a reduced transform size and omitting the block switching mechanism with its associated look-ahead delay [10][11].

A further descendant of AAC-LC is High-Efficiency AAC (HE-AAC) [2]. It additionally utilizes the Spectral Band Replication (SBR) technique, which reconstructs higher frequency components with the help of the low frequency base band signal and a very compact parametric description of the high frequency band [12]. The low frequency base band of the signal is coded by a conventional core coder. The high frequency band is derived by using a Quadrature Mirror Filterbank (QMF).

Both AAC-LD and HE-AAC serve as a basis for the new Enhanced Low Delay AAC (AAC-ELD) which will be described in the next section.

## 3. ENHANCED LOW DELAY AAC

The Enhanced Low Delay AAC (AAC-ELD) aims at combining the low delay feature of AAC-LD with the high coding efficiency of HE-AAC by utilizing SBR in conjunction with AAC-LD. The SBR decoder acts as a postprocessor which is applied after the core decoder including a complete analysis and synthesis filterbank, thus adding further decoding delay, as illustrated in Figure 1. In addition, also the delay of the core coder is doubled by operating it at half the original sampling rate ('dual-rate'). In the following, such delay sources are examined and their associated delay is minimized.

### 3.1. Modified SBR Framing and HF-Generator

In order to avoid delay caused by different framing of the core coder and the SBR module, the SBR framing is adapted to fit the frame length of 480 or 512 samples of AAC-LD. Furthermore, the variable time grid of the HF-generator, which implies 384 samples of delay, is restricted regarding the spreading of SBR data over adjacent AAC-LD frames. Thus, the only remaining source of delay in the SBR module is the QMF filterbank.

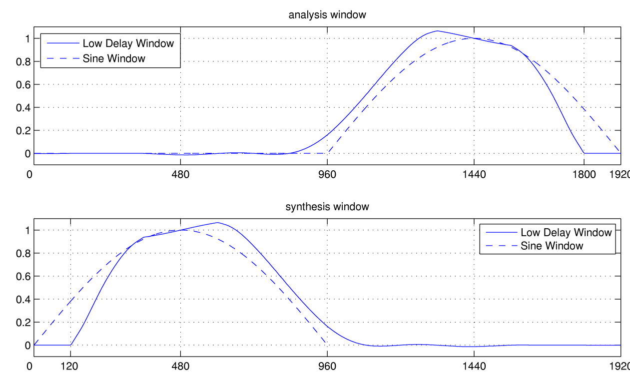


Figure 2: Impulse response of low delay windows

### 3.2. Low Delay Filterbank in AAC-LD Core

For AAC-ELD, a substantial delay reduction is achieved by replacing the MDCT/IMDCT by a low delay filterbank.

The new low delay filterbank, called LD-MDCT, with $M = 480$, reduces the filterbank delay from 959 samples $(2M - 1)$ to 719 $(2M - \frac{M}{2} - 1)$ samples, due to the reduced overlap of 120 samples towards the future. At the same time the impulse response is extented to the past by 960 samples ($E = 2$). Figure 2 shows the new analysis and synthesis window functions and, for comparison, the traditional sine window. Note that the analysis window is simply a time-reversed replica of the synthesis window, i.e. $p_A(n) = p_S(4M - 1 - n)$.

In the analysis window, the part that accesses future input values (and thus would cause delay) is reduced by 120 samples. Correspondingly, in the synthesis window the overlap with future output samples, which is needed in order to complete the overlap-add operation, is reduced by another 120 samples, resulting in an overall delay reduction of 240 samples. The extended overlap does not result in any additional delay, as it only involves adding values from the past.

A comparison of the traditional AAC-LD windows and the LD-MDCT window in [13] showed that a frequency response similar to that of the AAC-LD sine window is achieved for the LD-MDCT. As an additional option, the AAC-LD offers to switch to a low-overlap window [10] in order to eliminate pre-echo artifacts for transients. Together with Temporal Noise Shaping (TNS), the reduced overlap enables to avoid pre-echo artifacts better, i.e. spreading of the quantization noise before the signal's attack. The new LD-MDCT window offers the same property, but provides a better frequency response. Thus the low delay window replaces both traditional AAC-LD windows, and a dynamic window shape adaption is not necessary any more.

In [13] the result of a listening test is presented showing that replacing the MDCT by the LD-MDCT keeps the audio quality at the same level. Furthermore, this test shows that adding SBR to AAC-LD results in a significant improvement of the codec. The next section presents additional enhancements over [13] .

### 3.3. Using a Complex Low Delay Filterbank for SBR

#### *3.3.1. Description*

In order to achieve a further reduction in delay, the QMF inside the SBR decoder is replaced by a complex low delay filterbank (CLDFB). This replacement is done in a compatible way by keeping the number of bands (64), the length of the impulse response (640) and by using a similar complex modulation. Figure 3 shows a comparison of the CLDFB prototype filter with that of the original SBR QMF prototype. Furthermore, it illustrates that the delay of modulated filterbanks can be determined by analyzing the overlap delay introduced by the prototype filter in addition to the framing delay of the modulation core (i.e. a $\mathrm{DCT}_{\mathrm{IV}}$ of length $M$). In the analysis and the synthesis only the overlap to the right resp. left side of the modulation core adds delay. It can be observed that the synthesis prototype for the SBR QMF introduces an overlap delay of 288 samples while the CLDFB synthesis prototype introduces an overlap delay of only 32 samples. The same delay is introduced at the analysis stage as well.

Similar to the LD-MDCT, the CLDFB is based on the principle decribed in Section 2.1. The low delay extension towards past samples is $E = 8$. However, the delay reduction is achieved by a shift of the modulation core only. In contrast to the LD-MDCT, the impulse response of the MDCT core is not truncated.

The resulting overlap delay for $M = 64$ is 64 samples. The framing delay of 63 samples is already covered by the AAC-LD core coder in a complete coding scheme described in Figure 1.

For the standard SBR mode, the complex version is obtained by simply adding a sine modulation to the given cosine one (see Section 2.1.1). However, the SBR module can also run in the low-power mode by using only the real valued part of the CLDFB.

#### *3.3.2. Filterbank Evaluation*

In order to show that this delay-optimized, non-symmetric filterbank approach does provide additional value compared to a classical filterbank with a symmetric prototype, the asymmetric prototype is compared with symmetric prototypes with the same delay in the following. For the CLDFB with 64 bands the overall delay is 127 samples (framing+overlap). A modulated filterbank with a symmetric prototype and the same delay would therefore have a prototype of length 128. For these filterbanks with 50% overlap (such as the MDCT), sine or Kaiser-Bessel-Derived (KBD) windows generally provide a good choice of prototypes. In Figure 4 the frequency response of the CLDFB prototype is compared to the frequency response of alternative symmetric prototypes with the same delay (sine window, KBD windows with $\alpha = 4$ and $\alpha = 6$). This comparison clearly shows that a much better frequency response can be achieved with the CLDFB having an unsymmetric prototype.

#### *3.3.3. Listening Test*

For assessing the quality of AAC-ELD when replacing the QMF in the SBR module by the CLDFB, a MUSHRA test [14] was carried

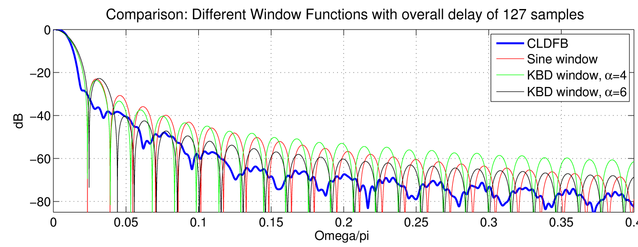


Figure 4: Frequency responses of different filterbank prototypes with same filterbank delay of 127 samples

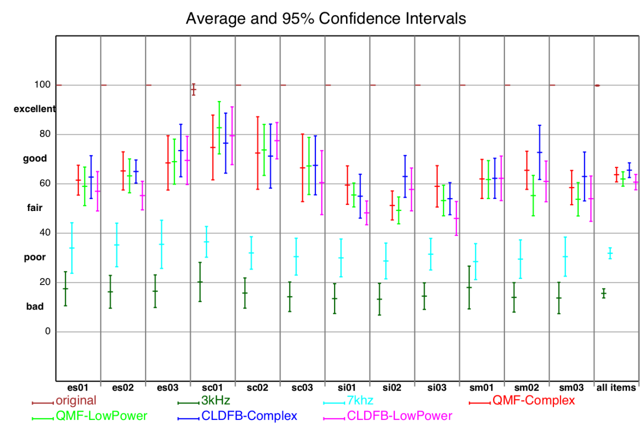


Figure 5: Results of MUSHRA test (8 listeners) CLDFB vs. QMF

out using the standard MPEG test set. The systems under test were:

- Original (hidden reference)
- 3.5 and 7 kHz anchors
- QMF-Complex: ELD with SBR (complex QMF)
- QMF-LowPower: ELD with low power SBR (real QMF)
- CLDFB-Complex: ELD with SBR (CLDFB)
- CLDFB-LowPower: ELD with low power SBR (CLDFB, real part)

All ELD versions were coded at a bitrate of 40kbit/s with a sampling rate of 48/24 kHz. The test was taken by 8 experienced listeners. Figure 5 shows the results of this listening test. It can be concluded that the CLDFB keeps the audio quality of AAC-ELD at the same level and does not introduce any degradation, neither for the complex SBR mode nor for the low power SBR mode. Thus, the delay-optimized CLDFB does not introduce any burden on the audio quality. For the transient items one can even observe some slight (but not statistically significant) improvement, especially for si02 (castanets) and sm02 (glockenspiel).

## 4. DELAY SUMMARY

Table 1 provides an overview of the delay with the different modification stages assuming a frame length of 480 samples and a sampling rate of 48 kHz.

It can be seen that the combination of the described delay reduction methods indeed results in a delay saving of 29 ms, i.e. an overall algorithmic delay of 31 ms rather than 60 ms for the straight forward combination of AAC-LD and SBR.

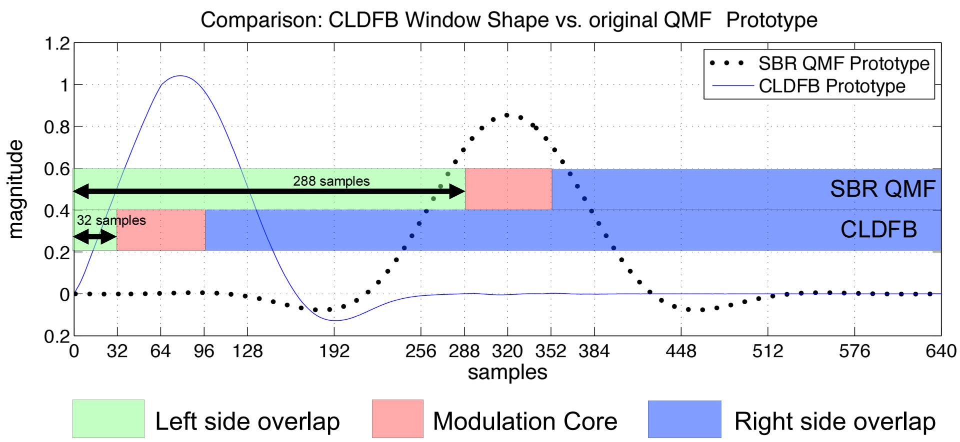


Figure 3: Impulse response of synthesis prototypes, CLDFB vs. SBR QMF

| Codec | Delay Source | delay [samples] | delay [ms] |
|---|---|---|---|
| AAC-LD + SBR | MDCT/IMDCT/ dual-rate | $960 \cdot 2$ | 40 |
| | QMF | 577 | 12 |
| | SBR-Overlap | 384 | 8 |
| | | 2881 | 60 |
| AAC-ELD | LD-MDCT/LD-IMDCT/ dual-rate | $720 \cdot 2$ | 30 |
| | CLDFB | 64 | 1 |
| | SBR-Overlap | 0 | - |
| | | 1504 | 31 |

Table 1: Delay values for modifications on AAC-ELD

## 5. CONCLUSIONS

This paper investigates the utilization of low delay filterbanks for low delay audio coding schemes. For the Enhanced Low Delay AAC codec, which is based on a combination of MPEG-4 AAC-LD and SBR, two variants of low delay filterbanks are used. By replacing both the MDCT in the AAC core and the QMF in the SBR module by low delay filterbanks, the overall delay is significantly reduced, while maintaining the high audio quality.

The ongoing ISO/IEC MPEG standardization process of AAC-ELD is expected to be finalized by the end of 2007.

## 6. REFERENCES

[1] G. Schuller, M. J. T. Smith, “New Framework for Modulated Perfect Reconstruction Filter Banks”, *IEEE Transactions on Signal Processing*, vol. 44, no. 8, pp. 1941–1954, Aug. 1996

[2] ISO/IEC 14496-3:2005, “Coding of Audio-Visual Objects, Part 3: Audio”, 2005

[3] J. P. Princen, A. W. Johnson, A. B. Bradley, “Subband/Transform Coding Using Filter Bank Designs Based on Time Domain Aliasing Cancellation”, *IEEE ICASSP*, pp. 2161–2164, 1987

[4] K. Nayebi, T. P. Barnwell, M. J. T. Smith, “Design of Low Delay FIR Analysis-Synthesis Filter Bank Systems”, *Proc. Conf. on Info. Sci. and Sys. (CISS)*, Mar. 1991

[5] T. Q. Nguyen, “A Class of Generalized Cosine-Modulated Filter Bank”, *International Symposium on Circuits and Systems (ISCAS)*, San Diego, USA, pp. 943–946, 1992

[6] G. Schuller, M. J. T. Smith, “A General Formulation for Modulated Perfect Reconstruction Filter Banks with Variable System Delay”, *NJIT 94 Symposium on Appl. of Subbands and Wavelets*, Mar. 1994

[7] G. Schuller, M. J. T. Smith, “Efficient Low Delay Filter Banks”, *DSP Workshop*, Yosemite, CA, USA, Oct. 1994

[8] G. Schuller, T. Karp, “Modulated Filter Banks with Arbitrary System Delay: Efficient Implementations and the Time-Varying Case”, *IEEE Transactions on Signal Processing*, vol. 48, no. 3, Mar. 2000

[9] ISO/IEC International Standard 13818-7, “Information Technology - Generic Coding of Moving Pictures and Associated Audio, Part 7: Advanced Audio Coding”, 1997

[10] E. Allamanche, R. Geiger, J. Herre, T. Sporer, “MPEG-4 Low Delay Audio Coding Based on the AAC Codec”, *106th AES Convention*, Munich, Germany, preprint 4929, May 1999

[11] J. Hilpert, M. Gayer, M. Lutzky, T. Hirt, S. Geyersberger, J. Hoepfl, R. Weidner, “Real-Time Implementation of the MPEG-4 Low Delay Advanced Audio Coding Algorithm (AAC-LD) on Motorola DSP56300”, *108th AES Convention*, Paris, France, Feb. 2000

[12] M. Dietz, L. Liljeryd, K. Kjörling, O. Kunz, “Spectral Band Replication, a Novel Approach in Audio Coding”, *112th AES Convention*, Munich, Germany, preprint 5553, Apr. 2002

[13] M. Schnell, R. Geiger, M. Schmidt, M. Jander, M. Multrus, G. Schuller, J. Herre, “Enhanced MPEG-4 Low Delay AAC - Low Bitrate High Quality Communication”, *122nd AES Convention*, Vienna, Austria, preprint 6998, May 2007

[14] International Telecommunication Union, *Method for the subjective assessment of intermediate sound quality (MUSHRA)*, Geneva, Switzerland, ITU-R, Recommendation BS. 1543-1, 2001