The use of multirate filter banks for coding of high quality digital audio

6th European Signal Processing Conference (EUSIPCO), Amsterdam, June 1992, Vol.1, pages 211-214

Th.Sporer *
Kh.Brandenburg *
B.Edler †

Abstract

In audio coding, the emphasis is on the perceived quality rather than on the statistics of the source signal as in speech coding. The quantization noise has to fit to the masking threshold derived from a psychoacustical model. It can most easily be described in the frequency domain. This is the reason why all major perceptual audio coding systems are frequency domain coders. Historically audio coders are either called subband coders or transform coders depending on the implementation of the filter bank used for analysis and synthesis of the signal. This paper describes filterbanks used for high quality audio coding.

Perceptual coding of digital audio:

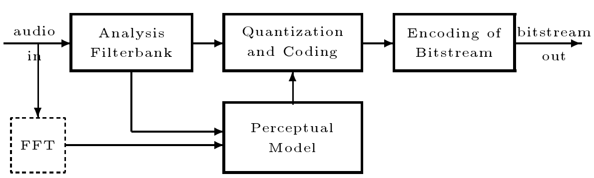

Figure 1: Block diagram of a perceptual coder

The basic structure of a perceptual coder is shown in figure 1 (see [2]):

A multirate filter bank is used to decompose the input signal into subsampled spectral components.

In the frequency domain an estimate of the actual (time dependent) masking threshold using rules known from psychoacoustics is calculated. This is done using the filterbank output or optionally using an additional FFT.

*Universiät Erlangen, Lehrstuhl f. Technische Elektronik, Cauerstr. 9, W-8520 Erlangen, Germany, Tel. +49 9131 85 7201, FAX +49 9131 302951

†Universität Hannover, Lehrstuhl f. Theoretische Nachrichtentechnik und Informationsverarbeitung, Appelstr. 9A, W-3000 Hannover 1, Germany, Tel. +49 511 762 5329, FAX +49 511 762 5333

The spectral components are quantized and coded with the aim of keeping the noise, which is introduced by quantizing, below the masking threshold. Depending on the algorithm, this step is done in very different ways, from simple block companding to analysis-by-synthesis systems using additional noiseless compression.

A multiplexer is used to assemble the bitstream, which typically consists of the quantized and coded spectral coefficients and some side information.

Filter banks:

The following paragraph decribe filter banks which have been used in audio coding.

A typical QMF-tree filter bank for audio coding uses from 4 up to 24 bands [3]. Problems with the non perfect reconstruction can be solved by introducing higher filter length or the use of wave digital filters instead. The computational complexity is high.

Polyphase filter banks are equally spaced filter banks which combine the filter design flexibility of generalized QMF banks with low computational complexity. A typical implementation uses 32 bands [7].

Discrete Fourier Transform and Discrete Cosine Transform with several kinds of windows are used with 128 to 512 bands [2]. Fast algorithms offer a low computational complexity. Blocking effects can be avoided by overlapping of succeeding blocks. They do not offer critical sampling, i.e. the number of time/frequency components is greater than the number of time samples represented by one block length.

Modified Discrete Cosine Transform (MDCT, using time domain aliasing cancellation as proposed by Princen/Bradley [1]) combines critical sampling with the good frequency resolution provided by a sine window and the computational effeciency of a fast FFT like algorithm. Typically 128 to 512 equally spaced bands are used. The MDCT offers also the possibility to change the block length of the transform dynamically. For very dynamic input signals a short block length keeps the quantization error local in time, for quasi static signals a long block length provides a good frequency resolution [4, 6].

Hybrid filter banks offer nonequal frequency resolution. Thus the analysis framework fits better to the ear's filtering performance [5, 7]. The combination of a polyphase filter bank with a MDCT adds the possibility of adaptive block length selection. Typically up to 576 lines are used.

More details on MDCT

Be $x(k)$ the samples in the time domain. $x_t(k),\ k = 0..n-1$ are the samples used to calculate the frequency domain samples $X_t(k),\ k = 0..\frac{n}{2}-1$ of the block number t.

The equation of the MDCT is [1]:

$$X_t(m) = \sum_{k=0}^{n-1} f(k)x_t(k)\cos(\frac{\pi}{2n}(2k+1+\frac{n}{2})(2m+1))$$

$$\text{for}\, m = 0..\frac{n}{2}-1$$

The equation of the inverse MDCT is [1]:

$$y_t(p) = f(p)\frac{n}{4}\sum_{m=0}^{\frac{n}{2}-1} X_t(m)\cos(\frac{\pi}{2n}(2p+1+\frac{n}{2})(2m+1))$$

$$\text{for}\, p = 0..n-1$$

Cancellation of time domain alias terms is done by an overlap add operation:

$$\tilde{x}_t(q) = y_{t-1}(q+\frac{n}{2}) + y_t(q) \quad \text{for}\, q = 0..\tfrac{n}{2}-1$$

To cancel alias terms the shape of each window must keep the following conditions: The shapes of the windows in succeeding blocks must fit to each other only in the overlapping part. It is possible to split each long block into shorter blocks [4]. Overlapping these shorter blocks must result in the same window-shape as used by the overlapping part of a long block (figure 2).

$$f_{t-1}(\frac{n}{2}+k)^2 + f_t(k)^2 = 1 \qquad for\ k = 0..\frac{n}{2}-1$$

There must be a symmetry in each half of a window:

$$f_t(k)^2 + f_t(\frac{n}{2}-k)^2 = 1 \qquad for\ k = 0..\frac{n}{2}-1$$

$$f_t(k+\frac{n}{2})^2 + f_t(n-1-k)^2 = 1 \qquad for\ k = 0..\frac{n}{2}-1$$

Several fast algorithms are known [9]. The highly optimized algorithm described below has been used in the real-time implementation of high quality coders [6] for several years. It has the following advantages:

- Low numbers of operations (additions, multiplications, storage operations)
- Minimum size of storage needed (in-place algorithm)
- High robustness against rounding errors
- Simple implementation on general purpose DSPs

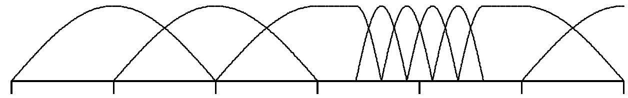

Figure 2: Typical sucession of window shapes

The algorithm uses only $n+4$ words of storage. n is the size of a window in the time domain. Both MDCT and IMDCT use the same kernel.

<u>MDCT:</u>

The windowed time domain samples are x_k with $k = 0..n-1$.

$$u_k = \begin{cases} -x_{k+\frac{3n}{4}} & \text{for } k = 0..\frac{n}{4}-1 \\ x_{k-\frac{n}{4}} & \text{for } k = \frac{n}{4}..n-1 \end{cases}$$

The rest of the MDCT is done by the kernel. The results are in X_k with $k = 0..\frac{n}{2}-1$.

<u>IMDCT:</u>

The spectral coefficients are in Y_k with $k = 0..\frac{n}{2}-1$.

$$u_k = \begin{cases} Y_k & \text{for } k = 0..\frac{n}{2}-1 \\ -Y_{n-k-1} & \text{for } k = \frac{n}{2}..n-1 \end{cases}$$

The next steps are done by the kernel.

$$y_k = \frac{n}{4}\begin{cases} X_{k+\frac{n}{4}} & \text{for } k = 0..\frac{n}{4}-1 \\ -X_{\frac{3n}{4}-k-1} & \text{for } k = \frac{n}{4}..\frac{3n}{4}-1 \\ -X_{k-\frac{3n}{4}} & \text{for } k = \frac{3n}{4}..n-1 \end{cases}$$

The results are in y_k with $k = 0..n-1$.

<u>The kernel:</u>

STEP 1

for $k = 0..\frac{n}{4}-1$:

$$\begin{aligned} v_{n-4k-1} &= (u_{4k} - u_{n-4k-1})A_{2k} - \\ &\quad -(u_{4k+2} - u_{n-4k-3})A_{2k+1} \\ v_{n-4k-3} &= (u_{4k} - u_{n-4k-1})A_{2k+1} + \\ &\quad +(u_{4k+2} - u_{n-4k-3})A_{2k} \end{aligned}$$

STEP 2

for $k = 0..\frac{n}{8}-1$:

$$w_{\frac{n}{2}+3+4k} = v_{\frac{n}{2}+3+4k} + v_{4k+3}$$

$$
\begin{aligned}
w_{\frac{n}{2}+1+4k} &= v_{\frac{n}{2}+1+4k} + v_{4k+1} \\
w_{4k+3} &= (v_{\frac{n}{2}+3+4k} - v_{4k+3})A_{\frac{n}{2}-4-4k} - \\
&\quad -(v_{\frac{n}{2}+1+4k} - v_{4k+1})A_{\frac{n}{2}-3-4k} \\
w_{4k+1} &= (v_{\frac{n}{2}+1+4k} - v_{4k+1})A_{\frac{n}{2}-4-4k} + \\
&\quad +(v_{\frac{n}{2}+3+4k} - v_{4k+3})A_{\frac{n}{2}-3-4k}
\end{aligned}
$$

STEP 3

ld is the logarithm with base 2
for $l = 0..ld(n) - 4$ $k_0 := \frac{n}{2^{l+2}}$ $k_1 := 2^{l+3}$

for $r = 0..\frac{n}{2^{l+4}} - 1$ and $s = 0..2^{l+1} - 1$

$$
\begin{aligned}
\hat{u}_{n-1-k_0 2s-4r} &= w_{n-1-k_0 2s-4r} + w_{n-1-k_0(2s+1)-4r} \\
\hat{u}_{n-3-k_0 2s-4r} &= w_{n-3-k_0 2s-4r} + w_{n-3-k_0(2s+1)-4r}
\end{aligned}
$$

$$
\begin{aligned}
\hat{u}_{n-1-k_0(2s+1)-4r} = {}& \\
&(w_{n-1-k_0 2s-4r} - w_{n-1-k_0(2s+1)-4r})A_{rk_1} - \\
&-(w_{n-3-k_0 2s-4r} - w_{n-3-k_0(2s+1)-4r})A_{rk_1+1} \\
\hat{u}_{n-3-k_0(2s+1)-4r} = {}& \\
&(w_{n-3-k_0 2s-4r} - w_{n-3-k_0(2s+1)-4r})A_{rk_1} + \\
&+(w_{n-1-k_0 2s-4r} - w_{n-1-k_0(2s+1)-4r})A_{rk_1+1}
\end{aligned}
$$

STEP 4

for $i = 1..\frac{n}{8} - 2$
$j =$BITREVERSE(i)
IF$(i < j)$THEN

$$
\begin{aligned}
\hat{v}_{8j+1} &= \hat{u}_{8i+1} & \hat{v}_{8i+1} &= \hat{u}_{8j+1} \\
\hat{v}_{8j+3} &= \hat{u}_{8i+3} & \hat{v}_{8i+3} &= \hat{u}_{8j+3} \\
\hat{v}_{8j+5} &= \hat{u}_{8i+5} & \hat{v}_{8i+5} &= \hat{u}_{8j+5} \\
\hat{v}_{8j+7} &= \hat{u}_{8i+7} & \hat{v}_{8i+7} &= \hat{u}_{8j+7}
\end{aligned}
$$

STEP 5

$$
\hat{w}_k = \hat{v}_{2k+1} \qquad \text{for } k = 0..\frac{n}{2} - 1
$$

STEP 6

for $k = 0..\frac{n}{8} - 1$

$$
\begin{aligned}
\tilde{u}_{n-1-2k} &= \hat{w}_{4k} & \tilde{u}_{n-2-2k} &= \hat{w}_{4k+1} \\
\tilde{u}_{\frac{3n}{4}-1-2k} &= \hat{w}_{4k+2} & \tilde{u}_{\frac{3n}{4}-2-2k} &= \hat{w}_{4k+3}
\end{aligned}
$$

It is possible to combine step 5 and 6 for some n.

STEP 7

for $k = 0..\frac{n}{8} - 1$

$$
\begin{aligned}
\tilde{v}_{\frac{n}{2}+2k} &= (\tilde{u}_{\frac{n}{2}+2k} + \tilde{u}_{n-2-2k} + \\
&\quad +C_{2k+1}(\tilde{u}_{\frac{n}{2}+2k} - \tilde{u}_{n-2-2k}) + \\
&\quad +C_{2k}(\tilde{u}_{\frac{n}{2}+2k+1} + \tilde{u}_{n-2-2k+1}))/2 \\
\tilde{v}_{n-2-2k} &= (\tilde{u}_{\frac{n}{2}+2k} + \tilde{u}_{n-2-2k} - \\
&\quad -C_{2k+1}(\tilde{u}_{\frac{n}{2}+2k} - \tilde{u}_{n-2-2k}) - \\
&\quad -C_{2k}(\tilde{u}_{\frac{n}{2}+2k+1} + \tilde{u}_{n-2-2k+1}))/2 \\
\tilde{v}_{\frac{n}{2}+1+2k} &= (\tilde{u}_{\frac{n}{2}+1+2k} - \tilde{u}_{n-1-2k} + \\
&\quad +C_{2k+1}(\tilde{u}_{\frac{n}{2}+1+2k} + \tilde{u}_{n-1-2k}) - \\
&\quad -C_{2k}(\tilde{u}_{\frac{n}{2}+2k} - \tilde{u}_{n-2-2k}))/2 \\
\tilde{v}_{n-2k-1} &= (-\tilde{u}_{\frac{n}{2}+1+2k} + \tilde{u}_{n-1-2k} + \\
&\quad +C_{2k+1}(\tilde{u}_{\frac{n}{2}+1+2k} + \tilde{u}_{n-1-2k}) - \\
&\quad -C_{2k}(\tilde{u}_{\frac{n}{2}+2k} - \tilde{u}_{n-2-2k}))/2
\end{aligned}
$$

STEP 8:

$k = 0..\frac{n}{4} - 1$

$$
\begin{aligned}
X_k &= \tilde{v}_{2k+\frac{n}{2}} B_{2k} + \tilde{v}_{2k+1+\frac{n}{2}} B_{2k+1} \\
X_{\frac{n}{2}-1-k} &= \tilde{v}_{2k+\frac{n}{2}} B_{2k+1} - \tilde{v}_{2k+1+\frac{n}{2}} B_{2k}
\end{aligned}
$$

<u>Twiddle-Factors:</u>

for $k = 0..\frac{n}{4} - 1$

$$
A_{2k} = \cos(\frac{4k\pi}{n}) \qquad A_{2k+1} = -\sin(\frac{4k\pi}{n})
$$

for $k = 0..\frac{n}{4} - 1$

$$
B_{2k} = \cos(\frac{(2k+1)\pi}{2n}) \qquad B_{2k+1} = \sin(\frac{(2k+1)\pi}{2n})
$$

for $k = 0..\frac{n}{8} - 1$

$$
C_{2k} = \cos(\frac{2(2k+1)\pi}{n}) \qquad C_{2k+1} = -\sin(\frac{2(2k+1)\pi}{n})
$$

More details on Hybrid filter banks

In hybrid filter banks frequency domain alias terms introduced in the first filter bank of the cascade limit the overall performance of the filter bank.

Alias reduction techniques [10] can improve the characteristics of the hybrid filter bank.

Cascading of the polyphase filterbank and the MDCT causes aliasing problems, because the frequency responses of the polyphase filters originally were designed for non-cascaded applications. This results in very high stopband attenuations and in transition widths, which avoid crosstalk between more than two neighboring subbands. However, considering the increased frequency resolution of the hybrid filter bank this causes crosstalk between subbands of the hybrid filterbank over a distance of several times their bandwidth. Thus, the overall subband filter frequency responses of the analysis as well as the synthesis filter banks show peaks within their stopbands.

The reduction of the aliasing problem is based on the fact that every frequency component of the input signal influences two subbands of the hybrid filterbank, one as a signal component and the other one as an aliasing component [10]. Furthermore, the overall filter frequency responses for these two bands show a symmetrical behavior. This implies the possibility to compensate an aliasing component in one band by subtracting the corresponding signal component in the other band multiplied by an appropriate weighting factor and vice versa (see figure 3). However, because the relation of signal and aliasing amplitudes depends on the frequency offset within the subbands, not all aliasing components can be eliminated with this operation, but the weighting factors can be optimized in order to give the best overall subband filter frequency responses.

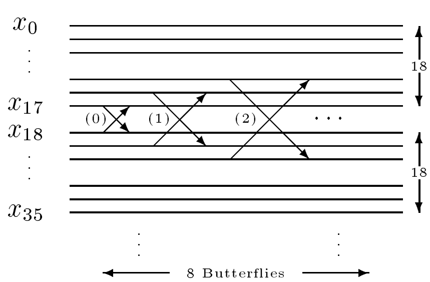

Figure 3: Aliasing reduction operation

Wish list for future filter bank designs:

For future developements of audio coding algorithms we would like filter banks with the following properties:

Perfect reconstruction: Within the resolution of the input signal (today 16 bit, in the future 20 bit and more) the analysis and synthesis filter bank should be perfect reconstructing for every input signal.

Locality in frequency: The quantization error introduced in one subband should be confined to the same subband and should not spread over the whole spectrum.

Implementation: There should be a fast algorithm available to enable cheap implementations.

Delay: For many applications a short system delay is a key property.

Perceptually adapted filter bank: The filter bank should fit the time and frequency resolution of the human ear. That means a good frequency resolution at lower frequencies and a good time resolution at higher frequencies.

Critical sampling: The number of spectral coefficients should be the same as the number of time samples represented by the block. The sum coefficient rate in the frequency domain should be the same as sampling rate in the time domain.

References

[1] J.Princen, A.Johnson, A.Bradley: "Subband / Transform Coding Using Filter Bank Designs Based on Time Domain Aliasing Cancellation", Proc. of the ICASSP 1987, pp. 2161 - 2164

[2] J.D.Johnston, K.Brandenburg: "Wideband Coding — Perceptual Considerations for Speech and Music", in: S.Furui, M.M.Sondhi (ed.): "Advances in Speech Signal Processing", New York 1991

[3] M.A.Krasner: "Digital Encoding of Speech and Audio Signals based on the Perceptual Requirements of the Auditory System", Technical Report 535 of the MIT Lincoln Laboratory, Lexington 1979

[4] B.Edler: "Coding of Audio Signals with Overlapping Block Transform and Adaptive Window Functions" (in German), Frequenz, vol.43(1989), pp. 252 - 256

[5] K.Brandenburg, J.D.Johnston: "Second Generation Perceptual Audio Coding: The Hybrid Coder", 88th. AES-Convention, Montreaux 1990, preprint 2937

[6] K.Brandenburg, J.Herre, J.D.Johnston, Y.Mahieux, E.F.Schroeder: "ASPEC:Adaptive Spectral Perceptual Entropy Coding of High Quality Music Signals", 90th. AES-Convention, Paris 1991, preprint 3011

[7] K.Brandenburg, G.Stoll, Y.F.Dehery, J.D.Johnston, L.v.d.Kerkhof, E.F.Schroeder: "The ISO/MPEG-Audio Codec: A Generic Standard for Coding of High Quality Digital Audio", 90th. AES-Convention, Paris 1991, preprint 3011

[8] J.Herre, E.Eberlein,K.Brandenburg: "A Real Time Perceptual Theshold Simulator", IEEE Workshop on Applications of Signal Processing to Audio and A-coustics, Mohonk 1991

[9] H.S.Malvar: "Extended lapped transforms: fast algorithms and applications", Proc. of IEEE Int. Conf. on ASSP, pp. 1797-1800, Toronto 1991

[10] B.Edler: "Alias Reduction for cascaded Filter Banks", submitted to Electronic Letters