ON NONUNIFORM FILTER BANKS FOR SUBBAND SPEECH CODING AND THEIR EFFICIENT IMPLEMENTATION

*Jürgen Schnitzler and Michael Kappelan**

RWTH Aachen, University of Technology
Institute of Communication Systems and Data Processing (IND), D-52056 Aachen, Germany
Tel: +49 241 80-6982; fax: +49 241 8888-186
E-mail: `Juergen.Schnitzler@ind.rwth-aachen.de`

ABSTRACT

We address design and implementation aspects of filter banks for the use in speech compression schemes. Particularly, we focus upon wideband (7 kHz) coding algorithms that use a few, nonuniformly spaced subbands and aim at system requirements of high speech quality at a low bit rate and moderate delay and complexity. We discuss requirements to the filter bank structure and filter design, and describe very efficient implementations for fractional rate conversion and modulated filter bank structures that are also useful for general applications.

1 INTRODUCTION

Subband coding (SBC) generally exploits the non-flat spectrum of the input signal to achieve a coding gain. If the signal spectrum is divided into only a few bands, however, the subband signals are not optimally decorrelated: they can be subject to further subordinate, e.g. time-domain compression algorithms instead. The wideband (7 kHz) speech and audio standard CCITT G.722 [1] uses a QMF filter bank to split the input spectrum into two subbands signals (0-4, 4-7 kHz) which are both coded by an ADPCM configuration, resulting to a bit rate between 48 and 64 kbit/s. In [2] we have proposed a 16 kbit/s split-band scheme that includes an efficient ACELP (Algebraic Code-Excited Linear Prediction) coding scheme tailored for the lower 0-6 kHz band, whereas a rather simple quantizer suffices for the upper (6-7 kHz) band. Typical applications aim at high quality voice communication systems such as ISDN, video-conferencing or multimedia systems where low to medium bit rate (1-2 bits per sample) and medium delay (20-40 ms) properties are required.

Fig. 1 shows the overall structure of a general subband codec. The analysis filter bank section splits the input signal (sampling frequency $f_{s,x}$) into N channels, with each channel j ($j = 1 \ldots N$) containing the $f_{j-1} \ldots f_j$ band of the spectrum of the input signal $x(n)$. An individual codec is allocated in each channel to quantize the critically subsampled subband signal $y_j(m)$ with sampling frequencies $f_{s,j}$.

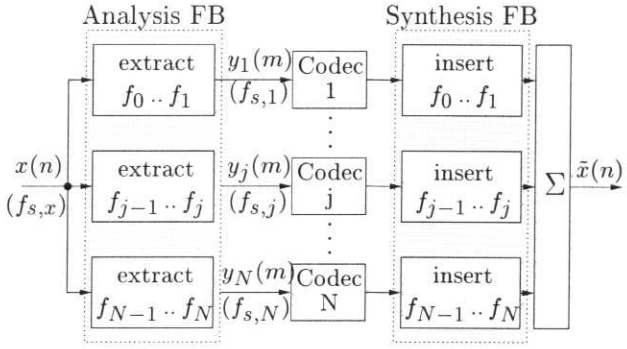

Figure 1: General structure of a subband codec.

In the scope of our contribution, we will refer to an example of a wideband scheme ($f_{s,0} = 16$ kHz) with $N = 2$, comprising a lower band at 0-5 kHz and an upper band at 5-7 kHz. Such a structure has also been used in [3]. A brief overview will be given in section 5.

2 FRACTIONAL SAMPLING RATE CONVERSION

Fig. 2 depicts the basic structure for a fractional decimation or interpolation unit, consisting of an expansion by L, FIR (lowpass or bandpass) filtering and decimation by M. This structure defines a key element to realize nonuniform filter banks.

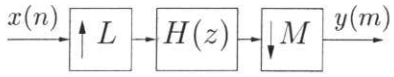

Figure 2: Basic fractional ratechange circuit.

Since the filtering according to Fig. 2 is performed in the interpolated domain, however, the basic structure is inefficient, particularly for sampling ratios demanding large L. Our solution adopts a polyphase approach similar to [4] or [5], but chooses a modified representation for the resulting, very efficient structure. First, the filter described by $H(z)$ is subject to a polyphase decomposition with respect to the interpolation by L:

* Now with Siemens AG München,
E-mail: Michael.Kappelan@pn.siemens.de

$$H(z) = \sum_{k=0}^{L-1} H_k(z^L) \cdot z^{-k} \quad .$$

This can be exploited by moving the expander into the polyphase branches: Fig. 3 a) shows the redrawing of the k-th polyphase branch, whose output is fed to the overall output summation. Provided L and M being relatively

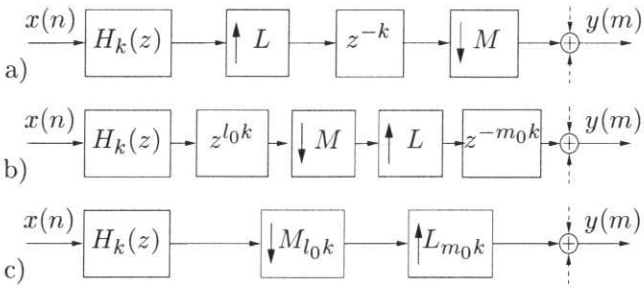

Figure 3: Redrawing of k-th polyphase branch.

prime, Euclid's theoreme $-1 = l_0 \cdot L - m_0 \cdot M$ grants the existence of two integers l_0, m_0 that allow to decompose the delay z^{-k} and to interchange the expander and the decimator, see Fig. 3 b). In our representation, however, we show that the combination of the non-causal delay $z^{l_0 k}$ and the subsequent sampling at time instants $0 + \mu M$ (said M_0), $\mu \in \mathbb{Z}$, is equivalent to a sampling at instants $l_0 k + \mu M$ (denoted as $M_{l_0 k}$). Therefore, instead of introducing a non-causal delay that has to be compensated later on [4][5], this representation uses a shifted sampling. Furthermore, expanding a signal by a factor of L (said L_0) and then delaying by $m_0 k$ is equivalent to generating a zero sequence at an L-times higher sampling rate and then inserting the data stream at instants $\lambda \cdot L + m_0 k$ (denoted as $L_{m_0 k}$), $\lambda \in \mathbb{Z}$. The redrawing in Fig. 3 c) illustrates the result of this consideration. Now it is evident that only every M-th output sample of the polyphase subfilter has to be computed, i.e. a further polyphase decomposition of the branch k allows the subband filter to actually operate at the lowest sampling rate $\frac{f_{s,x}}{M} = \frac{f_{s,y}}{L}$. Please note that this shifted sampling concept does not produce any additional system delay apart from the filter delay.

In the examples of Fig. 4 and 5, the rate conversion modules for the extraction and interpolation of 0-5 kHz band of a wideband speech codec are depicted. The transfer functions $H'_k(z)$, $k = 0 \ldots 7$, in Fig. 5 denote the polyphase filters of the prototype filter ($H(z)$) with respect to the interpolation by 8. In section 5, an overview of this coding algorithm will be presented.

3 MODULATED FILTER BANK

The solution of the previous section may directly apply to extract a baseband (lowpass) signal. To transfer an arbitrary subband to a critically sampled baseband signal, a generic solution could be achieved by cascading

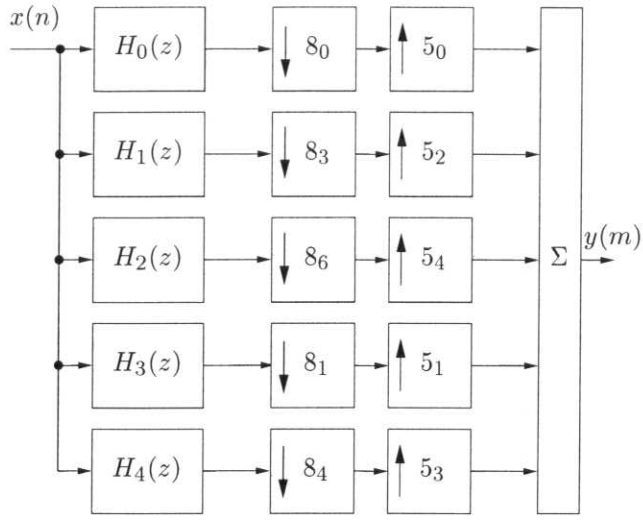

Figure 4: Example: extraction of the 0-5 kHz band

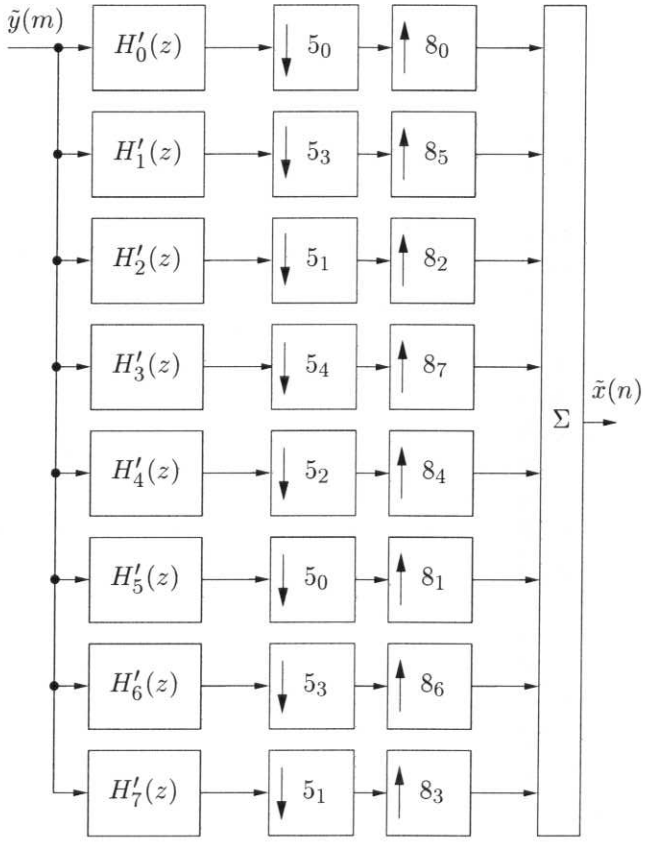

Figure 5: Example: interpolation of the 0-5 kHz band

two stages according to Fig. 2. This option obviously exhibits several drawbacks in terms of complexity and delay, however.

Our proposal overcomes these problems by using the well-known technique of complex modulation. We restrict our description to the example of extracting the 5-7 kHz band from the wideband input signal. It has to be noted, however, that the extent of possible optimizations depends on the relation between the subband edge frequencies and the input sampling rate. For our ex-

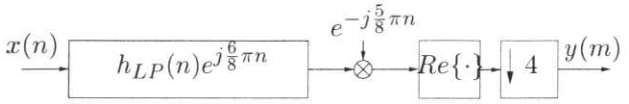

Figure 6: Modulated filter bank using a lowpass filter.

ample, the basic idea is shown in Fig. 6. A real-valued FIR prototype lowpass ($f_c = 1\,\text{kHz}$), designed for the input sampling rate and complex modulated to a center frequency at 6 kHz, is applied. This yields a complex bandpass signal portion at 5-7 kHz which is shifted to the single-sided baseband by means of the second modulation. The critical decimation of the real-valued portion produces the desired subband signal. A very efficient, real-valued implementation can be achieved by carefully inspecting the redundant operations due to the decimation, real-part and sine/cosine operations, and by a polyphase decomposition of the quadrature filters. The resulting analysis network that is shown in Fig. 7 leads to complexity savings of more than 90 % with regard to the basic scheme. The decimated impulse responses of the polyphase branches are computed from the prototype filter response h_{LP} of Fig. 6 as follows ($k = 0 \ldots 7$):

$$h_{ck}(n) = h_{LP}(8 \cdot n + k) \cdot cos\left(\frac{6\pi}{8}k\right) \qquad (1)$$

$$h_{sk}(n) = h_{LP}(8 \cdot n + k) \cdot sin\left(\frac{6\pi}{8}k\right) \qquad (2)$$

The branches marked dashed in Fig. 7 indicate that the associated impulse responses equal to zero; these branches need not to be computed. Concluding, the two branches of this modulated filter bank operate at a sampling rate of 2 kHz. The modulation has been reduced to a sign operation, and all samples computed in the branch filters are interleaved to the out data stream. Similarly, a corresponding synthesis structure can be found.

4 DESIGN OF SUBBAND FILTERS

The aspect of practical system requirements has dominated our access to the subband filter design. The use of low bit rate subband codecs implies distortions that prohibit perfect reconstruction approaches. Instead, a good overall performance, even for multiple transcodings, calls for narrow transition bands and optimum passband properties, but only near-optimum reconstruction behaviour. We therefore prefer highly selective linear-phase interpolator filters (Nyquist filters [4]) which are appropriate for the polyphase implementation methods. The linear-phase and interpolator properties are highly preferrable in order to minimize in-band distortions that accumulate especially in case of multiple transcodings.

For the fractional rate conversion algorithm of section 2, the filters have to be designed for the sampling rate of

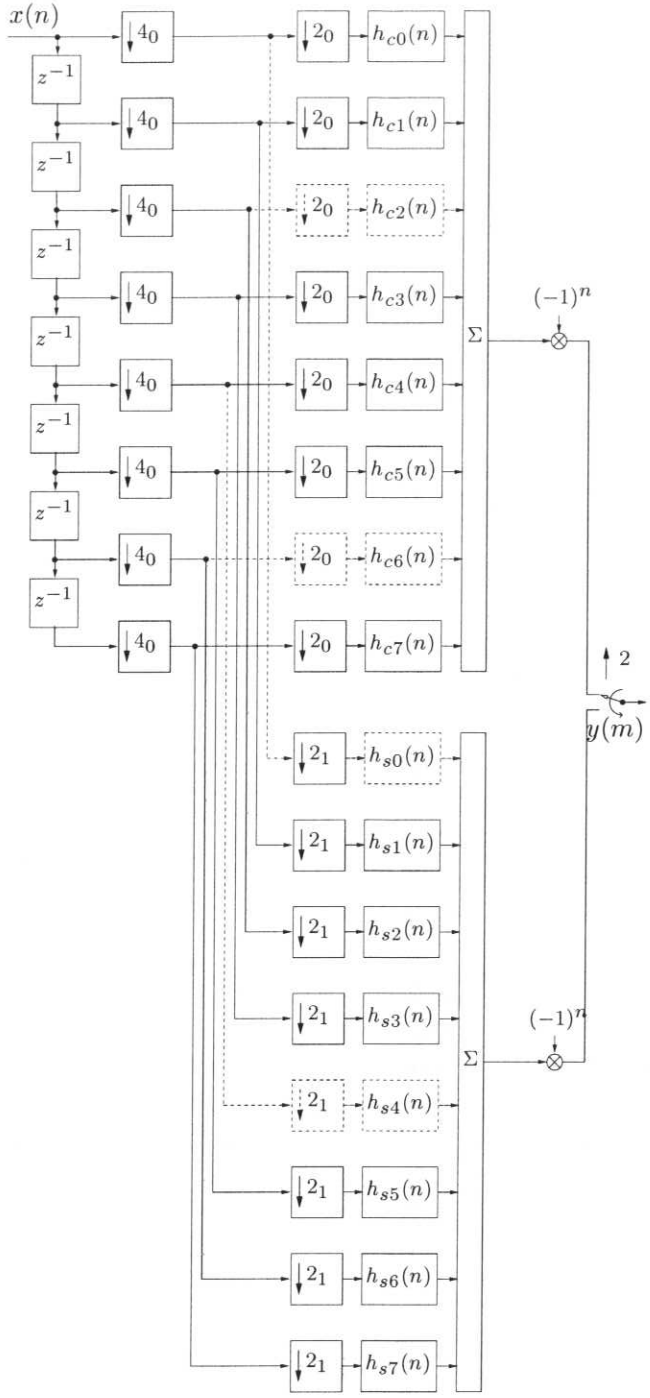

Figure 7: Example: extraction of the 5-7 kHz band.

the signal expanded by L. On the other hand, the modulated filter bank solution of section 3 requires a filter design for the input sampling rate only. The stopband attenuation is chosen to exceed 50 dB, the passband attenuation to about 0.03 dB. The exact edge frequencies are adjusted to achieve a good overall reconstruction performance of the analysis-synthesis filter bank. For a Modified Chebychev approximation using the algorithm of Parks and McClellan, the Nyquist filter property must be enforced by modifying the filter impulse response [6]. This modification degrades the frequency response prop-

erties and restricts the design to certain parameter configurations. Therefore, among different approaches, a straightforward design based on a Modified-Fourier approximation using a Kaiser window (see e.g. [6]) gave the best results and allowed analysis/synthesis filter banks with an overall delay of 5 ... 10 ms, requiring transition bandwidths between 300 and 700 Hz, respectively.

5 EXAMPLE OF A SPLIT BAND CODEC

The filter bank structures described in the previous sections are used in a coding scheme for wideband (7 kHz) speech and music. A more detailed description can be found in [3]. The codec operates at bit rates of 16, 24 and 32 kbit/s and is designed for high quality ISDN and video-conferencing applications. Depending on the characteristics of the input signal and the bit rate, different codec modes are selected: at 16 and 24 kbit/s, a split-band ACELP (SB-ACELP) algorithm is active for speech signals, whereas for music signals, an Adaptive Transform Coding (ATC) technique is applied. At 32 kbit/s, only the ATC mode is used.

For the SB-ACELP mode, the codec operates on signal frames of 20 ms length. The algorithm is based on a split-band scheme with two unequal subbands using an ACELP codec in the lower subband (0-5 kHz) and a simplified CELP codec in the upper band (5-7 kHz). The filter bank performs the unequal band splitting and critical subsampling of the two subbands. In particular, the structures shown in the Figs. 4, 5 and 7 are used.

The ACELP algorithm in the lower subband contains a short term synthesis filter (LPC) of order 12 and a long-term synthesis filter (LTP) realized as an adaptive codebook. For the fixed excitation, different algebraic codebooks are searched at 16 and 24 kbit/s. A perceptual weighting filter is used during the codebook search, and an adaptive postfilter enhances the perceptual output speech. In the upper band, a simplified CELP algorithm consists of a fixed random codebook and a short-term synthesis filter of order 8.

In formal subjective listening tests [7], the speech quality at 16, 24 and 32 kbit/s was rated comparable to the CCITT G.722 reference codec at 48, 56 and 64 kbit/s, respectively.

6 CONCLUSION

In this paper, design and implementation aspects of filter banks for the use in speech compression schemes have been described. The focus was put on wideband (7 kHz) coding algorithms that use a few, nonuniformly spaced subbands. Very efficient implementations of fractional rate conversion algorithms and modulated filter bank structures have been shown. Finally, an application example of a coding algorithm for wideband speech has been given.

ACKNOWLEDGEMENTS

This work has been supported by the Deutsche Telekom Berkom GmbH. The authors would like to thank especially Mr. J. Stegmann.

REFERENCES

[1] CCITT, "7 kHz Audio Coding within 64kbit/s," in *Recommendation G.722*, vol. Fascile III.4 of *Blue Book*, pp. 269–341, International Telecommunication Union, Melbourne 1988.

[2] J. Paulus und J. Schnitzler, "16 kbit/s Wideband Speech Coding Based on Unequal Subbands" in *Proc. Int. Conf. Acoust., Speech, Signal Processing, ICASSP*, (Atlanta, Georgia, USA), pp. 651–654, 1996.

[3] Deutsche Telekom and France Telecom/CNET, "High-level description for the ITU-T wideband (7kHz) ATCELP speech coding algorithm of Deutsche Telekom, Aachen University of Technology (RWTH) and France Telecom (CNET)" Delayed contribution COM16-130, ITU-T Study Group 16, Q.20, Geneva, January 1998.

[4] P. Vaidyanathan, *Multirate Systems and Filter Banks.* Prentice Hall Signal Processing Series, 1993.

[5] N. Fliege, *Multiraten-Signalverarbeitung.* Teubner Verlag, 1993.

[6] W. Hess, *Digitale Filter.* Teubner Studienbücher, 1988.

[7] Deutsche Telekom and France Telecom/CNET, "Subjective qualification test results for the ITU-T wideband (7kHz) speech coding algorithm of Deutsche Telekom, Aachen University of Technology (RWTH) and France Telecom (CNET)" Delayed contribution COM16-129, ITU-T Study Group 16, Q.20, Geneva, January 1998.