BLOCK-RECURSIVE, MULTIRATE FILTERBANKS WITH ARBITRARY TIME-FREQUENCY PLANE TILING

Unto K. Laine

Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
PO. Box 3000, FIN-02015 Espoo, FINLAND

ABSTRACT

A new method to realize arbitrary time-frequency plane tilings together with critical sampling in block-recursive filterbanks is presented. The method leads to pole-zero approximation of the target channel transfer functions. Perfect reconstruction within the limits of the approximation error can be achieved.

1. INTRODUCTION

Variable filter design was one of the earliest applications of the frequency warping technique [2][14]. The frequency responses of filters, e.g., the cut-off frequency of a lowpass filter was varied by replacing the unit delays of FIRs with allpass sections. Oppenheim et all. [13] introduced frequency warping based nonuniform resolution FFT. Later Strube [15] applied the technique to linear prediction (WLP), which works approximately on the auditory Bark scale. WLP is now more widely used in speech and audio coding [7][4], and it has consistently increased its popularity in the field of perceptual audio signal processing [6].

Laine et al. [8][9] introduced auditory filterbanks which are based on frequency warping and realized by the FAMlet transform. Later a block recursive algorithm was published were only two matrix operations are needed for each down-sampled filterbank output [10]. The block recursive filterbank is a novel structure where the efficient block recursion is based on the short-term cross-correlation between the channel responses. It has been applied to auditory speech analysis, as a front end of a speech recognizer, and automatic speech segmentation [11]. In order to reach full synchrony between the channels, every channel in the bank was sampled at the same rate. This means that the high frequency channels are undersampled and the low frequency ones are correspondingly oversampled. To avoid the loss of information at the highest frequencies the whole bank had to be slightly oversampled. However, a better method has to be found before the filterbank can be applied to maximally decimated subband coding. The new design introduced below solves the problem.

In conventional multirate filterbanks and wavelet transforms the critical sampling and perfect reconstruction is achieved in two special cases only: uniform resolution and octave filterbanks [1]. A novel warped wavelet method has tried to solve the problem of *arbitrary time-frequency plane tiling in critical sampling context* [3]. Unfortunately, the method is computationally so expensive that it hardly can be used in real-time applications.

The present work provides a new method to solve the problem of arbitrary time-frequency plane tiling in critically sampled filterbanks. The method is based on the use of channel grouping and the use of enhanced block-recursive algorithm with optimized coefficients. Now each group can be sampled down by the same ratio. In principle, any type of frequency warping (time-frequency tiling) can be realized.

The new method is tested using an experimental filterbank consisting of 14 channels distributed in a frequency band of 0.05 - 11 kHz according to the auditory ERB-rate scale [12]. The channels are organized in four groups. The sampling in the groups occurs at every 6th, 12th, 24th, and 48th sample (the high frequency group mentioned first). The bank is critically sampled so that every block of 48 input samples creates a block of 48 filter output samples allocated in time and frequency according to the down-sampling rates.

This paper is organized as follows. Section 2 describes the method of block-recursive filterbanks. In Section 3 the method is applied to a simple filterbank. Finally, we give an example how the designed filterbank works in speech analysis.

2. FREQUENCY WARPED BLOCK-RECURSIVE FILTERBANKS

The earlier frequency warped block-recursive filterbank design was based on the use of frequency warped complex exponentials called FAM functions and their time domain representatives, FAMlets [10].

The general goal of the design is to create a maximally decimated filterbank which produces a nonuniform resolution spectrum of the input signal $s(n)$ on a new frequency scale (ν-scale). Typically the ν-scale is an auditory frequency scale.

The ν-scale spectrum $S(b)$ is produced by applying the Fourier transform to the frequency warped signal $s(a)$. The signal $s(a)$ is produced from the spectrum $S(m)$ of the input signal $s(n)$ by FAM transform or by applying the FAMlet transform directly to the signal (1).

$$S(b) = \mathbf{F}\, s_\nu(a) = \mathbf{F}\,\Phi_\nu\, S(m) = \mathbf{F}\,\Psi_\nu\, s(n)$$
$$a, b, m, n \in \mathbf{Z}$$
$$\mathbf{F} \equiv \textit{Fourier transform matrix} \qquad (1)$$
$$\Phi_\nu \equiv \textit{FAM transform matrix}$$
$$\Psi_\nu \equiv \textit{FAMlet transform matrix}$$

When the Fourier transform is combined with the FAM transform a set of warped, discrete (periodic in frequency) sinc functions is created. These functions form the basic building

block for frequency warped filterbanks. The new design in Section 3 also applies these functions.

The earlier block-recursive design was based on the block-wise approximation of the FAMlet transform. However, the method can be generalized and applied directly to almost any type of a set of finite energy impulse responses. The earlier design resulted in the following algorithm [10]:

$$S_t(b) = \mathbf{T}\,S_{t-1}(b) + \mathbf{U}\,\mathbf{s}_t,\ t \in \{1,2,3,\ldots\},\ S_0 = \phi, \qquad (2)$$

where t is the time index of the input signal block $\mathbf{s}_t$, **U** is a (spectral) state control matrix and **T** a (spectral) state transition matrix. From the design point of view **U** realizes an FIR part containing the first samples of the channel impulse responses and **T** produces an IIR type approximation of the rest of the channel impulse responses. Thus the block recursive model forms a rational transfer function approximation of the actual target filterbank. The optimal design of the recursive part is based on the short-term cross-correlation between the channel impulse responses.

In the following (2) is enhanced by extending the FIR part to two separate blocks. The new block-recursive filterbank is based on the use of the following equations:

$$\begin{aligned} &S_t(b) = \mathbf{A}\,\mathbf{s}_t + \mathbf{z}_t \quad t \in \{1,2,3,\ldots\} \\ &\mathbf{z}_t = \mathbf{P}\,\mathbf{z}_{t-1} + \mathbf{B}\,\mathbf{s}_{t-1},\ \mathbf{z}_0 = \phi,\quad \mathbf{s}_0 = \phi \end{aligned} \qquad (3)$$

where the matrix **A** contains the first samples of the channel impulse responses, **B** contains the next samples of the impulse responses, and **P** is so called block predictor matrix which recursively produces the rest of the impulse responses (IIR part).

The design example of the next Section demonstrates that by grouping the channels in a proper way and then applying the block-recursive algorithm to each of the groups, maximal decimation and perfect reconstruction can be realized in this nonuniform resolution (ERB-rate) filterbank. The channel responses are considerably improved by carefully optimizing the matrices **B** and **P** in each channel group.

3. EXAMPLE OF THE NEW DESIGN

3.1 The Target Filterbank

The target filterbank is made of frequency warped discrete time sinc functions defined by (4).

$$H(b,\omega) = \frac{1 - e^{-jB\pi\nu(11.025\omega/\pi)/\nu 0}}{e^{-jb2\pi/B} - e^{-j\pi\nu(11.025\omega/\pi)/\nu 0}}, \qquad (4)$$

$$where \quad \nu(f) = sign(f)\log(1 + 4.37|f|).$$

The function $\nu(f)$ defines the ERB-rate scale [12] based frequency warping (f in kHz). The term $\nu 0 = \nu(11.025)$ is used for ν-scale normalization, the parameter B defines the total number of channels (bands), and ω is the normalized frequency $-\pi < \omega \le \pi$.

The variable b defines the actual channel $-B \le b \le B$. The target filterbank was designed with $K = 16$, however, the lowest ("DC" like channel) and the highest one have little use in speech analysis and were left out. The target filterbank can be easily orthonormalized by using a proper weighting function. The channel frequency responses without weighting (nonorthogonal case) and without windowing (without sidelobe attenuation) is depicted in Figure 1.

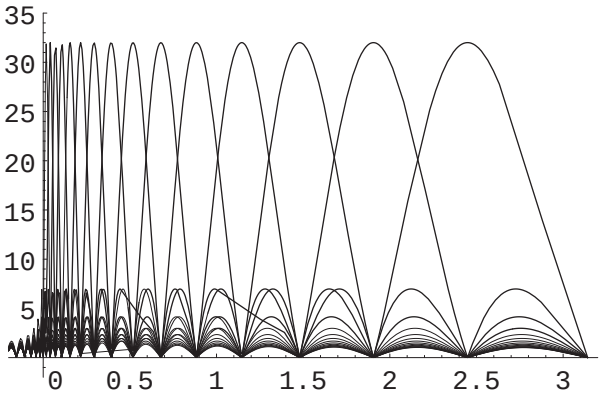

Figure 1. Frequency responses of the ERB-rate target filterbank (y: amplitude, x: normalized frequency ω).

3.2 Properties of the Target Filterbank

One interesting question is how the frequency warping affects the basic time-frequency properties of the target filterbank. When the frequency responses of (4) are inverse Fourier transformed a set of analytic (complex valued) impulse responses is created. This type of bank provides a high time resolution because the channel magnitude information is available at every sample.

The bandwidths of the channels were chosen to be clearly broader than in the human auditory system. The aim was to further improve the time resolution to study rapid spectral variations in speech (e.g., pitch-synchronous effects).

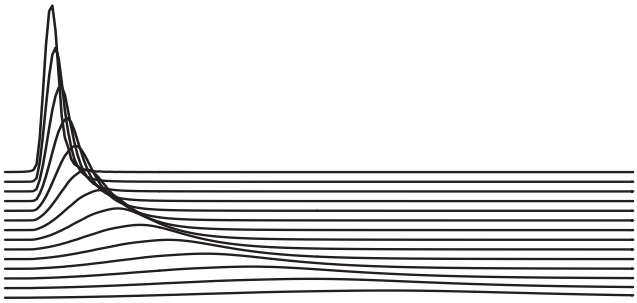

Figure 2. Hilbert envelopes of the analytic impulse responses of the target filterbank (z: channel number, y: amplitude, x: time).

The Hilbert envelopes of the channel impulse responses of Figure 2 give some idea of the varying time resolution of the filterbank. The envelopes follow approximately the shape of the gamma function (rapid growth in the amplitude followed by slowly decreasing tail). The high-frequency channels have naturally narrowest time envelopes and the time resolution decreases gradually towards the low-frequency end. The amplitude maximum has increasing latency towards the low frequencies. This approximates the traveling wave phenomenon in the cochlea.

The instantaneous frequencies of the responses show a chirp-like behavior [11]. This type of gammachirp responses were

proven to be superior to the conventional gamma-tone filter [5].

When the -3 dB points of the time envelopes and the frequency magnitude curves are numerically estimated, the classical Gabor time-frequency resolution measure of the filters can be calculated. Figure 3 collects the data. Each channel is depicted by a dot on this time-frequency resolution plane. The solid curve represents the case df •dt = 0.5 which is the optimum in Gabor sense.

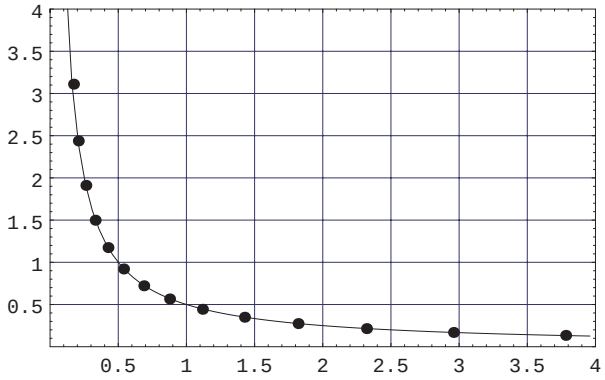

Figure 3. Time-frequency selectivity of the target filterbank (dots). Theoretical Gabor limit (line) (y: bandwidth in kHz, x: duration in ms).

The results show that *the frequency warping does not affect the time-frequency optimality of the filterbank*. On the average the products of the time and frequency resolutions of the channels are 0.5. This is true at least when ERB-rate warping is used.

3.3 The new Block-Recursive Filterbank

The sampling frequency for the target filterbank was chosen to be 22.05 kHz. The fourteen channels where grouped into four groups with 4, 4, 3 and 3 channels listed from the low frequency channels up. Correspondingly, the down sampling ratios were chosen to be: 48, 24, 12, and 6. These numbers equal to the block sizes used in the groups. Thus 48 input samples produces 4+2•4+4•3+8•3=48 channel output samples. In other words, the block-recursive filterbank is maximally decimated. Note that even though the filterbank is not exactly an octave bank the block-recursive structure allows (in this case) down sampling in the steps of octaves.

The matrix **A** of the algorithm (3) was chosen directly from the channel impulse responses of the target system. An attempt to optimize this further was made without any noticeable improvements in the approximation of the frequency response.

The block predictor **P** can be solved based on a set of normal equations. A closer derivation of the equations is given in [11]. The final formula for **P** is given in (5).

$$\mathbf{P} = \left[\left(\mathbf{G}\mathbf{G}^T \right)^{-1} \mathbf{G}\mathbf{G}_d^T \right]^T, \qquad (5)$$

where the rows of the matrix **G** contain the windowed impulse responses of the corresponding targets and the rows of the matrix $\mathbf{G}_d$ contain the windowed impulse responses of the corresponding targets taken from the d-th sample forward. The parameter d defines the block size used in the prediction. Formally **P** is a ratio of two cross-correlation matrices (one matrix multiplied by the inverse of another).

The quality of the result can be controlled by varying (optimizing) the windows used for the impulse responses before the computation of the cross-correlations. Best results were obtained using windows having the shape of rised cosine ($0 \le x \le \pi$) to power α, where the coefficient α must be optimized for each channel.

In all optimizations the quality of channel frequency responses was monitored by computing the log-magnitude differences at frequency points where the side-lobe maxima of the target are located (uniform distribution in the ERB-rate scale). If the magnitude of the target was larger than that of the model that point was not taken with in the average error (i.e., the model was allowed to be better than the target).

Finally, the matrix **B** was optimized based on the over-all quality of the frequency responses of the channels. The initial values of the elements are picked up from the target impulse responses just after the samples taken in the matrix **A**. The further optimization was based on a method where a random number matrix was generated and added to **B.** If the average error between the model and the target at every channel in the group was decreased the random number matrix was stored. When fifteen such matrices were found they were averaged and added to **B** by using an optimized scaling.

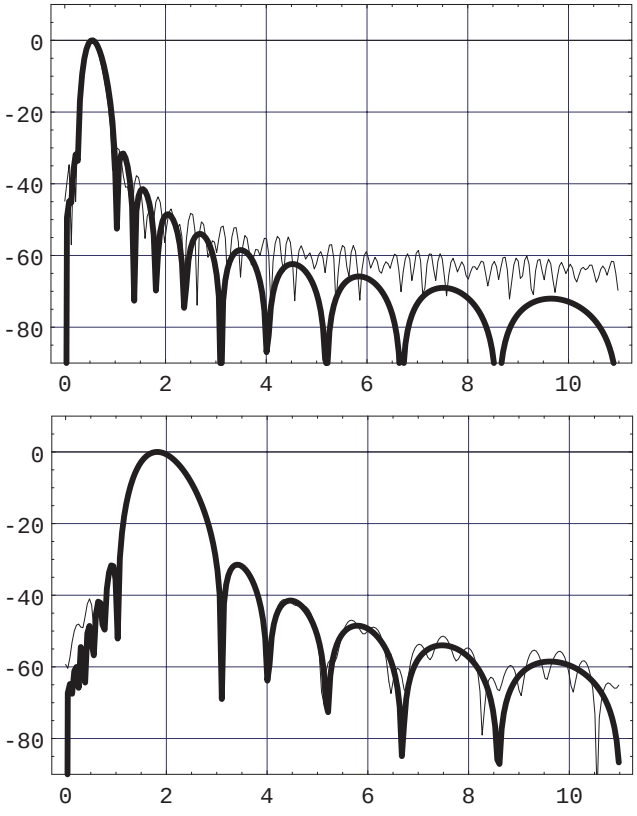

Figure 4. Channel 4 (upper frame) and 8 (lower frame) target magnitudes (thick lines) and their block recursive approximations (thin lines). (y: dB, x: kHz).

The idea behind this procedure is that by allowing some error in the second FIR part of the responses the recursive (IIR) continuation may fit better to the nonrecursive (FIR) part thus minimizing the average error in the frequency response.

The impulse responses of the block-recursive filterbank can easily be produced by allowing matrix form data to be stored to the vector-valued delay elements (3) and by feeding in an identity matrix followed by a chain of zero matrices. Finally, the impulse response matrix is created by joining the generated output matrices.

Figure 4 depicts two typical magnitude responses of the created filterbank compared to the corresponding target responses. In low frequency channels (upper frame) the block-recursive algorithm is not able to generate the rather long tail of the impulse response accurately. The introduced approximation error is seen in the high frequency range.

The situation at middle and high frequency channels is better due to the shorter impulse responses. The FIR matrices include large part of the impulse response and the recursive part is short. The optimization of the matrix **B** is of great importance in improving the magnitude responses of these channels.

3.4 Simulations and Results

When the impulse responses of the target system are windowed so that the introduced magnitude error corresponds to that of the block-recursive (BR) approximation, the average length of the impulse responses is 90 taps. The average computational load of the block-recursive realization is about 49 taps per channel which is only 54% of the FIR realization. The difference will be even larger when narrower filters with longer impulse responses are used.

Figure 5 demonstrates the use of the BR filterbank in speech analysis. Five periods of Finnish /æ/ are analyzed. Strong energy peaks are seen at the glottal closures and secondary excitations with quite strong high frequency energy are seen at the glottal openings. At the low frequency area the open periods with flow pulses produce "bubbles". The high frequency selectivity due to the higher sampling rate is clearly improved when compared to the earlier design [10].

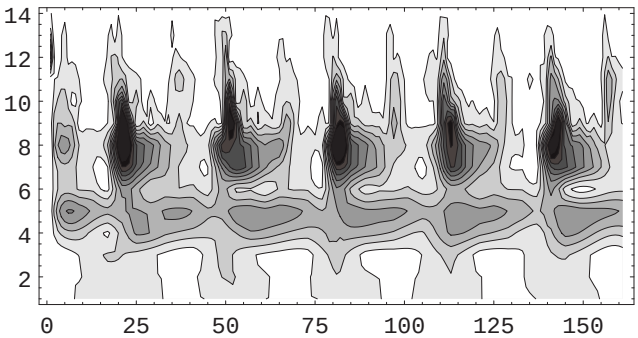

Figure 5. Finnish /æ/ analyzed by the block recursive ERB-rate filterbank (y: channel number, x: time in blocks of 6 samples).

The simulations showed that the perfect reconstruction is mainly limited by the approximation error of the BR filterbank. The measured noise level was about 60 dB below the signal. Thus the filterbank may find applications in speech and audio coding, too.

4. ACKNOWLEDGMENT

This study was partially financed by the Academy of Finland.

5. REFERENCES

[1] Akansu A. N., Haddad R. A., *Multiresolution Signal Decomposition*, Academic Press Inc., Boston, 1992.

[2] Constantinedes A. G., "Spectral transformations for digital filters, *Proc. IEE*, 117, pp. 1585-1590, Aug., 1970*).

[3] Evangelista G. and Cavaliere S., "Discrete frequency warped wavelets: theory and applications," *IEEE Tr. on Signal Processing,* vol. 46, 4, pp. 874-885., 1998.

[4] Härmä A., Laine U. K., and Karjalainen M., "WLPAC - a perceptual audio codec in a nutshell," *AES 102nd Conv. preprint 4420*, 1997.

[5] Irino, T., "A Gammachirp function as an optimal auditory filter with Mellin transform," Paper AE3.2 in *Proc. of ICASSP-96*, Atlanta, 1996.

[6] Karjalainen M., Härmä A., Laine U. K., and Huopaniemi J., "Warped filters and their audio applications," *Proc. 1997 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'97),* New Paltz, New York, USA, 1997.

[7] Koishida K., Tokuda K., Kobayashi T., and Imai S., "CELP coding system based on mel-generalized cepstral analysis," *Proc. of ICSLP'96*, vol. 1, 1996.

[8] Laine U. K., "Speech analysis using complex orthogonal auditory transform (COAT)," *Proc. Int. Conf. on Spoken Language Processing (ICSLP'92),* pp. 69-72, Banff, Alberta, Canada, 1992.

[9] Laine U. K. and Härmä A., "On the design of Bark-FAMlet filterbanks," *Proc. Nordic Acoustical Meeting* (NAM'96), pp. 277-284, Helsinki, Finland, June 12-14, 1996.

[10] Laine U. K., "Critically sampled PR filterbanks of nonuniform resolution based on block recursive FAMlet transform," *Proc of European Conf. on Speech Communication and Technology (EUROSPEECH'97),* vol. 2, pp. 697-700, Rhodes, Greece, 1997.

[11] Laine U. K., "Speech analysis by using a novel block recursive algorithm for auditory spectrograms (BRASS)", *Lingvistica Uralica*, XXXIV 3, pp. 213-219, 1998

[12] Moore B. C J., Peters R. W., and Glasberg B. R., "Auditory filter shapes at low center frequencies," *JASA*, 88, pp. 132-140, 1990.

[13] Oppenheim A. V., Johnson D. H., Steiglitz K., "Computation of spectra with unequal resolution using the fast Fourier transform, *Proc. of the IEEE*, 59, pp. 299-301, Feb. 1971.

[14] Schlüssler W., Winkelnkemper W., "Variable digital filters, *Arch. Elek. Übertr*., 24, pp. 524-525, 1970*)..

[15] Strube H. W., "Linear prediction on a warped frequency scale," J. Acoust. Soc. Am., 64 (4), pp. 1071-1076, Oct. 1980.

_____

*) Reprinted also in: Rabiner L.R., Rader C. M. (Eds.), Digital Signal Processing, IEEE Press, New York, 1972.