CRITICALLY SAMPLED PR FILTERBANKS OF NONUNIFORM RESOLUTION BASED ON BLOCK RECURSIVE FAMLET TRANSFORM

Unto K. Laine
Laboratory of Acoustics and Audio Signal Processing
Helsinki University of Technology
P.O. Box 3000, FIN-02015 Espoo, Finland
Tel. +358 9 4512492, FAX: +358 9 460224, E-mail: Unto.Laine@hut.fi

ABSTRACT

A new block recursive algorithm is introduced for effective FAMlet transform implementation. When the Fourier transform is combined with the algorithm a nonuniform resolution filterbank is created. The algorithm allows to approximate frequency resolutions of any type, the ERB-rate scale included. The signals can be vector based critically down sampled which allows a perfect reconstruction.

1. INTRODUCTION

In our earlier study [1] a nonuniform resolution filterbank approximating the auditory Bark-scale was created based on the combined FAMlet and Fourier transforms. The FAMlet transform produces frequency warped signal, i.e., a new representation of the original signal in which the frequency components are shifted according to a frequency warping function $\nu(f)$. Finally, the FAMlet transformed, frequency warped signal is Fourier transformed to produce a spectrum of the original signal on a new frequency scale: ν-scale. If the warping function corresponds to the Hz-to-Bark mapping the produced spectrum is auditory, Bark-scale spectrum. When both transforms are carried out sample by sample, the filterbank produces N outputs for each sample. Because the filterbank is of nonuniform type, there is no simple means to do the critical down sampling.

Up to now the only effective implementation method for the FAMlet transform has been the use of the chain of first order allpass filters, i.e., the Laguerre network. Because the phase characteristic of the allpass sections is controlled with one free parameter only, the network can produce only one type of frequency warpings and filterbank resolutions. When, e.g., a logarithmic resolution is needed, there has not been any computationally efficient tool to implement the FAMlet transform. The only way has been the use of relatively long FIR filters for the FAMlet transformation.

One method to increase the design freedom is to use higher order allpass sections or Kautz-type structures [2]. The use of higher order allpass sections increases the complexity of the system and also the treatment of the folding problems is quite expensive. Second and third order allpass sections seem to increase the computational complexity faster than the design freedom needed to implement complicated frequency warpings.

Dispersive delays realised by allpass filters are the most essential part in the generation of FAMlets. In order to avoid the folding, these allpass filters should have average group delay of one unit sample even if the phase has a more complicated structure than that of the first order allpass filter. The following new algorithm is based on the idea of "generalized" chain of allpass filters described by *state space equations*. The state vector of N-elements represents directly the instantaneous values of the N FAMlets.

The state space representation of the FAMlets leads to an interesting observation that if the FAMlet basis (of N functions) is splitted, e.g., in nonoverlapping NxN blocks, every block after the first one can approximately (or in the discrete Laguerre case even exactly!) be produced by multiplying the previous block by a state transition matrix **A**. Thus, the functions are entirely produced from the first block by multiplying repeatedly with **A**. This *block recursion* can be directly applied to effective production of FAMlet functions and transforms of almost any type.

The fact that the FAMlets are *block recursively redundant* is not very surprising when we remember that they are essentially produced by allpass operations which include recursion. Due to the recursive warping the functions are not orthogonal in the first NxN block like the functions in the Fourier base, but need infinitely many recursive NxN blocks to be represented precisely. Moreover, in the case of nonwarped discrete Fourier base the next blocks are identical to the first one (if the basis is extended over the first block). Thus the state transition matrix is simply an identity matrix. The warping means that the state transition matrix is no more an identity. The next question is how this matrix can be optimised to produce the best possible estimate for the FAMlets and the warpings in question.

In this paper we introduce a new fast algorithm which solves the two problems mentioned above: firstly, the critical sampling can be treated by *block based algorithm* which operates on *vectorized data* and secondly, *any kind of warping* can be approximated by a proper *state transition matrix* **A**. Further, we define *critical vectorization*, which means that the input data N-vector is transformed into spectral data N-vector. Since the linear operators involved are of full rank, the resulting filterbank has a *perfect reconstruction* (PR) feature.

2. FAM AND FAMLET CLASSES AND FREQUENCY WARPING

The FAM class of orthogonal functions $\phi_\nu(f,a)$ is defined by (1), where $a \in \mathbf{Z}$ is associated to *the order* of the function. FAM stands for Frequency-Amplitude Modulated complex exponentials [3].

$$\phi_\nu(f,a) = \sqrt{d\nu / df}\; e^{j2\pi a\,\nu(f)} \qquad (1)$$

The FAM class consists of sets of orthonormal functions the properties of which are primarily controlled by the function $\nu(f)$, which defines the frequency warping produced by the corresponding FAM or FAMlet transform. The class of FAMlets is defined by:

$$\psi_\nu(t,a) = \mathbf{F}^{-1}\;\phi_\nu(f,a) \qquad (2)$$

In (2) time domain FAMlets are produced by applying the inverse Fourier transform to the frequency domain FAM functions. They also form a class of orthogonal functions [4]. Now, the FAM transform is given by:

$$\mathrm{s}_\nu(a) = \hat{\Phi}_\nu \mathrm{S}(f) = \int_{f_1}^{f_2} \mathrm{S}(f)\;\phi_\nu^*(f,a)\,df \qquad (3)$$

The range of orthogonality is $[f_1, f_2]$. Correspondingly we can define the FAMlet transform.

$$\begin{aligned}\mathrm{s}_\nu(a) &= \hat{\Psi}_\nu \mathrm{s}(t) = \int_{-\infty}^{0} \mathrm{s}(t)\;\psi_\nu(-t,a)\,dt \\ &= \mathrm{s}(t) * \psi_\nu(t,a)\big|_{t=0}\end{aligned} \qquad (4)$$

According to (3) and (4) the FAMlet transform of the signal $s(t)$ is equal to the FAM transform of its spectrum $S(f)$. The new signal $s_\nu(a)$ represented in the a-domain (transform domain) is the *frequency warped* version of the original signal $s(t)$. When the continuous variables t and f are discretized to n and k, the integral transforms are changed to matrix operators.

$$\begin{aligned}&\hat{\Phi}_\nu \rightarrow \Phi_\nu \quad \mathrm{s}(t) \rightarrow \mathrm{s}(n) \quad t,f \in R \\ &\hat{\Psi}_\nu \rightarrow \Psi_\nu \quad \mathrm{S}(f) \rightarrow \mathrm{S}(k) \quad n,k \in Z\end{aligned} \qquad (5)$$

The FAMlets are in principle infinitely long in time (IIR-type), however 99% of their energy is typically given in a limited time window. The corresponding transform matrix Ψ_ν is then not rectangular.

3. NONUNIFORM RESOLUTION FIR TYPE FILTERBANK

The nonuniform resolution spectrum (spectrum on the new, warped ν-scale) is produced by applying the Fourier transform for the frequency warped signal $s(a)$. This procedure is formally given by (6), where $\mathbf{F}$ denotes the Fourier transform.

$$\begin{aligned}S(\nu) &= \mathbf{F}\; s_\nu(a) = \mathbf{F}\;\Phi_\nu S(k) \\ &= \mathbf{F}\;\Psi_\nu\; s(n) \quad \nu,k,a \in \mathbf{Z}\end{aligned} \qquad (6)$$

According to (6) the nonuniform resolution spectrum $S(\nu)$ of the signal $s(n)$ can be also produced by combining the Fourier and the FAMlet transform. If $\nu(f)$ follows the Hz-to-Bark mapping this combined transform defines the impulse responses of the corresponding FIR type *orthogonal auditory filterbank* $\mathbf{F}_\nu$ having Bark-resolution [1]. The FIR coefficients of the channels form the rows of the $\mathbf{F}_\nu$ matrix (7).

$$\mathbf{F}_\nu = \mathbf{F}\;\Psi_\nu \qquad (7)$$

4. BLOCK RECURSIVE FAMLET TRANSFORM

Let $\psi_\nu(n,a)$ denote a member of a set of N orthonormal FAMlets designed for frequency warping $\nu(f)$. Now n denotes the time index and a the order of the function. The block recursive algorithm needs a method to estimate new values for the FAMlets from their present values. In most cases *exact* new values can not be analytically solved. Our method is based on a predictor matrix $\mathbf{P}$ that gives *estimates* for the FAMlet functions based on their earlier values.

$$\psi_\nu(n+d) = \mathbf{P}\;\psi_\nu(n) \quad n, d \in \mathrm{Z} \qquad (8)$$

In (8) $\psi_\nu(n)$ denotes a N-vector taken from the FAMlet basis in the time instant n and d is a positive integer denoting the time span over which the values are predicted. It is easy to see that this "vector predictor" can be solved by solving N "scalar predictors" of the conventional form

$$\psi_\nu(n+d,i) = \sum_{j=1}^{N} p_{ij}\;\psi_\nu(n,j) \qquad (9)$$

$$i = 1,\ 2,\ \ldots\ ,\ N$$

where the value of the i-th FAMlet at the time instant $n+d$ is estimated from the values of all N FAMlets at the time instant n. The predictor coefficients p_{ij} can be solved in the conventional way based on the autocorrelation function.

$$p_{ij} = \mathbf{R}^{-1}\mathbf{r} \qquad (10)$$

where matrix $\mathbf{R}$ is the autocorrelation matrix and vector $\mathbf{r}$ the corresponding correlation vector. The correlation is now computed between FAMlets of different order and the matrix $\mathbf{R}$ is in fact a cross correlation matrix. Moreover, because the FAMlets are orthonormal this matrix is an identity matrix ($\mathbf{R} = \mathbf{I}$). So the predictor coefficients are easily produced by correlating different FAMlets where one of them is shifted in time d units.

In the orthonormal case the predictor matrix $\mathbf{P}$ is simply a matrix containing correlation coefficients

$$\mathbf{P} = \{\mathbf{r}_d(\psi_i, \psi_j)\}, \quad i,j \in \mathrm{Z} \qquad (11)$$

where $\mathbf{r}_d$ is correlation between FAMlet of order i shifted in time d units and FAMlet of order j. Let $d = N$ and the matrix $\Psi_{\nu,0}$ consist of the first N samples of the N

FAMlets (a N x N matrix). The next N samples of the FAMlets can now be predicted by

$$\Psi_{\nu,\mathbf{m}+1} = \mathbf{P}_N \, \Psi_{\nu,\mathbf{m}} \qquad m \in Z^+ \qquad (12)$$

where m denotes the block number. Finally, the *block recursive FAMlet transform* is defined by:

$$\mathbf{x}_{m+1} = \mathbf{P}_{\mathrm{N}} \, \mathbf{x}_m + \Psi_{\nu,0} \, \mathbf{s}_m \qquad m \in Z^+ \qquad (13)$$

where $\mathbf{s}_{\mathrm{m}}$ denotes the m-th N-vector formed from the incoming discrete-time signal $s(n)$, and $\mathbf{x}_{\mathrm{m}+1}$ is the frequency warped form of $\mathbf{s}_{\mathrm{m}}$.

The block recursive FAMlet transfom (13) is easy to interpret. At the beginning the state vector is zeroed. The first transform is made with the *base block* $\Psi_{\mathrm{v},0}$. Because the next FAMlet block is recursively generated from the first one, so is the transform vector. The history of the transform is collected recursively, whereas the new information is added trough the base matrix analysis.

The transform (13) has also a state space interpretation. The warping is done applying a generalized version of an allpass chain. Matrix **P** corresponds the state transition matrix (**A**) and the matrix $\Psi_{\mathrm{v},0}$ the state control matrix (**B**). From this point of view it is easy to understand that the absolute values of the eigenvalues of **P** must be less than one in order to quarantee the stability.

Note also that this recursion can be used to generate the FAMlet basis. When s_0 is replaced by an identity matrix and the following inputs with zero filled matrices the state vector is changed to matrix $\Psi_{\mathrm{v,i}}$, where the index i runs from zero upwards. When all the generated blocks are joined the FAMlet basis is created.

Almost any type of frequency warping can be realised by the block recursive FAMlet transform. The accuracy of the warping depends on the shape of the $\nu(f)$ function and on the ability of the predictor to estimate the values of the actual FAMlets.

5. NONUNIFORM RESOLUTION, BLOCK RECURSIVE, CRITICALLY SAMPLED, PR FILTERBANK

According to (6) the ν-resolution spectrum $S(\nu)$ of the incoming signal $s(n)$ can be produced by Fourier transforming the frequency warped signal $s(a)$.

$$S_{m+1}(\nu) = \mathbf{F} \, \mathbf{x}_{\mathrm{m}+1} = \mathbf{F} \, \mathbf{P}_{\mathrm{N}} \, \mathbf{x}_{\mathrm{m}} + \mathbf{F} \, \Psi_0 \, \mathbf{s}_{\mathrm{m}} \qquad (14)$$

We may develop (14) further and get the final form for the corresponding block recursive filterbank.

$$S_{m+1}(\nu) = \mathbf{F} \, \mathbf{P}_{\mathrm{N}} \, \mathbf{F}^{-1} S_m(\nu) + \mathbf{F} \, \Psi_0 \, \mathbf{s}_{\mathrm{m}}$$

$$S_{m+1}(\nu) = \mathbf{T} \, S_m(\nu) + \mathbf{U} \, \mathbf{s}_{\mathrm{m}}, \qquad (15 \text{ a b})$$

where $\mathbf{T} = \mathbf{F} \, \mathbf{P}_{\mathrm{N}} \, \mathbf{F}^{-1}$ and $\mathbf{U} = \mathbf{F} \, \Psi_0$.

In the case the matrix **U** has an inverse the original signal $s(n)$ can be fully reconstructed from its ν-resolution spectrum $S(\nu)$.

$$\mathbf{s}_m = \mathbf{U}^{-1}\left[S_{m+1}(\nu) - \mathbf{T} S_m(\nu)\right] \qquad (16)$$

6. SIMULATIONS

A simple 17-channel block recursive filterbank was simulated with Mathematica™ 3.0. The signal bandwidth was 22.05 kHz. A Gaussian window was used between the FAMlet and Fourier transforms in order to attenuate the sidelobes. The bank produces approximately 1.5 Bark resolution. Fifteen of the channel outputs are complex valued. The bank gives magnitude estimates after each input vector of 32 samples. The synthetic test signals was chosen the coming speech analysis in mind.

The first test signal (1024 samples) consists of five positive half waves of sinusoids which are not synchronized with the sampling rate so that each pulse differs slightly from the others (Fig. 1 upper part). The lower part of the Fig. 1 shows the Bark-scaled time-frequency distribution of the pulses. The high frequency channels react first and are able to detect the rapid changes at the beginning and at the end of each pulse. Because these points differ in each pulse the spectral pictures of them differs too. There is increasing group delay towards the low frequency end. However, in the middle of the pulse there is a pure sinusoidal part which is detected eventhough its duration is only half cycle.

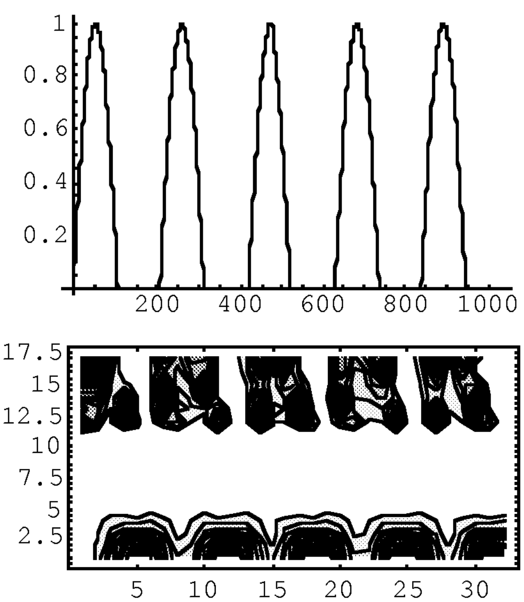

Fig. 1. Upper frame: sinusoidal test pulses.
Lower frame: auditory spectrogram (x: vector number, y: channel number)

In the next test a middle frequency sinusoidal signal is amplitude modulated with an low frequency rectangular wave. Also here the bursts are not equal. The spectrogram of Fig. 2 shows small variation too. The amplitude modulation is also here clearly detected. The reaction time of the bank depends naturally on the frequency of the carrier.

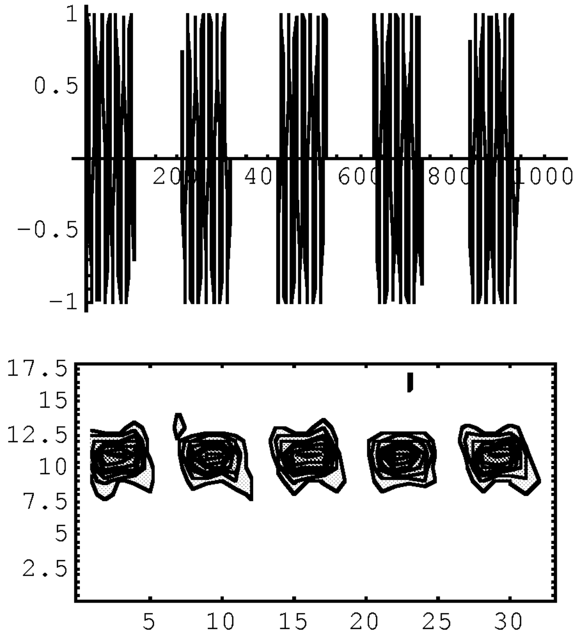

Fig. 2. Test with sinusoidal bursts.

The third test is made with frequency modulated sinusoid. The modulating signal changes up and down approximately logarithmically (modulation frequency 170 Hz) and the change in the auditory spectrogram is nearly linear (triangular). An interesting detail here is that the up going part gives little larger magnitude values than the down going part. A related phenomenon is found in psychoacoustical listening tests [5], [6].

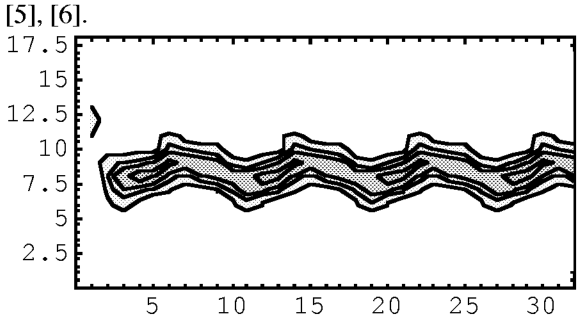

Fig. 3. Auditory spectrogram of frequency modulated sinusoid.

The fourth example is analysis of open Finnish /ae/ vowel produced by a male speaker (MK). The signal consists of five glottal periods and the signal bandwidth in this test is 11.025 kHz. The first formant of the vowel is around 750 Hz, close to the channel number seven (about in the middle). The first formant has clear pitch synchronous amplitude modulation. The amplitude is attenuated strongly during the open glottal period. Also a small frequency modulation can be detected. The second formant fluctuates from period to period and so do the third one which is very close to the second formant. The most interesting founding in this picture is the clear pitch synchronous resonance around 450 Hz (channel four). It appears just before the glottal closure. It must be a tracheal resonance which is seen in the spectrogram only during the open glottal periods.

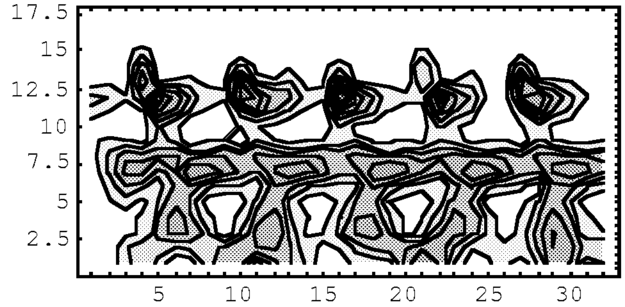

Fig. 4. Auditory spectrogram of Finnish /ae/..

7. ACKNOWLEDGEMENT

This study has been supported by the Academy of Finland. The author is grateful to Prof. Matti Karjalainen for many illuminating discussions during the development of this algorithm. Also the criticism from M.Sc. Aki Härmä has helped to keep the wheels running.

8. REFERENCES

[1] Laine U. K., Härmä A., "*On the design of Bark-FAMlet filterbanks*", Paper D7, Proc. of NAM-96, Helsinki, pp. 277-284.

[2] Kautz W. H., "Transient synthesis in the time domain", *IRE Trans. Circuit Theory*, CT-1, pp. 29-39, 1954.

[3] Laine U. K., Altosaar T., "*An orthogonal set of Frequency and Amplitude Modulated (FAM) functions for variable resolution signal analysis*", Proc. of ICASSP-90, pp. 1615-1618.

[4] Laine, U. K., "*FAMlet to be or not to be a wavelet*", IEEE-SP Int. symp. on time-frequency and time-scale analysis, Victoria, 1992, pp. 335-338.

[5] Kohlrausch A. and Sander A., "Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets", *J. Acoust. Soc. Am.*, Vol. 97, No 3, pp. 1817-1829, March 1995.

[6] Laine U. K. and Huotilainen M.., "*A study of auditory temporal fine resolution by using short Bark-FAMlet clicks*", Paper D6, Proc. of NAM-96, Helsinki, pp. 267-276.