Frequency change function and acoustic signals.

E. Wesfreid * V. Wickerhauser **

LMPA, ULCO - France. *Washington University, USA.*
eva.wesfreid@lmpa.univ-littoral.fr *victor@math.wustl.edu*

Abstract

The local cosine4 orthonormal bases [1, 4] *are particularly well adapted for analyzing signals with piecewise time behaviour. There are many acoustic signals in music and speech processing that can be considered as a sequence of overlapping elementary structures such as phonemes in speech signals. The Best Basis algorithm* [2] *computes a local spectrum defined over a dyadic segmentation, however, there is no reason for elementary structures to 'begin' and 'end' near dyadic points. We use Fang's algorithm* [3] which *segments the time axis into intervals of arbitrary length; this algorithm constructs a frequency change function whose local maxima denote structure changes. The smooth cosine4 orthonormal basis defined over this segmentation is used to compute a local spectrum associated with elementary structures. We show that this representation compared with the Best Basis coefficients has less reconstruction distortion and better local pattern description.*

1. Introduction

A signal can be decomposed into a linear combination of elementary *waveforms*, called *time-frequency atom*, each waveform being essentially supported by a rectangle in the time-frequency plane. One now has available a large selection of waveforms, the choice of the *time-frequency atoms* is not unique, the decomposition can therefore be adapted to the analyzed signal.

There are many *acoustic* signals in music or speech processing that can be considered as a sequence of overlapping *elementary structures* like *phonemes* in speech signals. One goal of time-frequency analysis is to decompose these *structures* into *elementary waveforms.*

The *cosine4 Best Basis* algorithm of Coifman and Wickerhauser [2] computes a *local spectrum* over a dyacic time segmentation in $O(N \log N)$ operations. There is no reason, however, for *elementary structures* to 'begin' and to 'end' near dyadic points.

In this paper, we use Fang's algorithm [3] to segment the time axis into intervals associated with *elementary structures.* This algorithm is based on the computation of a *frequency change function* whose *local maxima* denote *structure changes.* These local maxima can therefore be considered as segmentation points. Since two adjacents *elementary structures* are overlapping, the computed segmentation points are only approximate. We therefore say that the segmented time intervals contain *near local elementary structures.*

We use a local *cosine4* orthonormal basis defined over this time segmentation to compute a (piecewise constant) spectra *near local elementary structures* in $O(N^2)$ operations. We show that this representation has less reconstruction distortion. This means that the approximation error is less with coefficients *near local elementary structures* than with the Best Basis spectrum.

2. Block and smooth cosine4 transform

The *block cosine4 spectrum* of a signal S over a segmented time axis,

$$0 = a_0 < a_1 < \ldots < a_s = N,$$

*The authors wish to thank Robert Ryan for helpful discussions and suggestions.
**Research supported in part by AFOSR, NSF, and the Southwestern Bell Telephone Company.

is the set of coefficients

$$D_j = \{d_{j,k} : 0 \le k < \ell_j\} \tag{1}$$

in the decomposition

$$S(t) = \sum_{\substack{j\in Z \\ 0\le k<N}} d_{j,k}\phi_{j,k}(t),$$

where

$$d_{j,k} = \langle S, \chi_{I_j}\phi_{j,k}\rangle$$

is the *block dct4* transform. The function $\phi_{j,k}$ defined as

$$\phi_{j,k} = \frac{\sqrt{2}}{\sqrt{|\ell_j|}} \cos \frac{\pi}{|\ell_j|}(k+\frac{1}{2})(t-a_j),$$

is the *cosine4* function, and $\chi_{I_j}(t)$ is the indicator function of I_j.

We are going to describe the *smooth cosine4* transform algorithm that computes the *smooth local spectrum* of a sampled signal $\{f(t)\}_{t\in Z}$ where

$$\mathbb{Z} = \bigcup_{j\in Z} I_j$$

$I_j = [a_j, a_{j+1})\bigcap \mathbb{Z}$, such that $a_j - \frac{1}{2}$ is an integer, $\inf_{j\in\mathbb{Z}}(a_{j+1}-a_j) > 0$, $\lim_{j\to\pm\infty} a_j = \pm\infty$.

We consider the following functions and sets over $\mathbb{Z}$:

- the *raising function*

$$r(t) = \begin{cases} 0 & t \in]-\infty, -1[\\ \sin[\frac{\pi}{4}(1+\sin\frac{\pi}{2}t)] & t\in[-1,1] \\ 1 & t\in[1,\infty[\end{cases}$$

- the *smooth orthogonal* window associated with $I_j = [a_j, a_{j+1})\bigcap\mathbb{Z}$

$$\omega_j(t) = r(\frac{t-a_j}{\eta})\, r(\frac{a_{j+1}-t}{\eta}) \tag{2}$$

  where η is the adjacent window overlap, $0 < \eta < \ell_j/2$ and $\ell_j = (a_{j+1}-a_j)$ is the number of points belonging to $[a_j, a_{j+1})\bigcap\mathbb{Z}$.

- $b_j(t) = r(\frac{(t-a_j)}{\eta})$.

- $O_j^+ =]a_j, a_j+\eta[, \quad O_j^- =]a_j-\eta, a_j[, \quad O_j = O_j^- \cup O_j^+$.

We use the *folding* operator [8]

$$U_jf(t) = \begin{cases} b_j(t)f(t)+b_j(2a_j-t)f(2a_j-t) & if\ t\in O_j^+, \\ b_j(2a_j-t)f(t)-b_j(t)f(2a_j-t) & if\ t\in O_j^-. \end{cases}$$

and its adjoint, the *unfolding* operator [8]

$$U_j^*f(t) = \begin{cases} b_j(t)f(t)-b_j(2a_j-t)f(2a_j-t) & if\ t\in O_j^+, \\ b_j(2a_j-t)f(t)+b_j(t)f(2a_j-t) & if\ t\in O_j^-. \end{cases}$$

that verify $U_jU_j^* = U_j^*U_j = id$ to compute the *folded* function

$$F_{a_j,a_{j+1}} = \chi_{I_j}U_jU_{j+1}f.$$

The *smooth orthogonal* window (2) is equal to the rectangular window χ_{I_j} *unfolded* at a_j and at a_{j+1}:

$$\omega_j = U_j^*U_{j+1}^*\chi_{I_j}. \tag{3}$$

The associated *orthonormal cosine4 basis* of $\ell^2(\mathbb{Z})$

$$\{\Psi_{j,k}\}_{j\in\mathbb{Z},0\le k<\ell_j},$$

where

$$\Psi_{j,k}(t) = w_j(t) g_{j,k}(t) \tag{4}$$

consists of *smooth orthogonal* windows w_j modulated by *cosine4* functions.

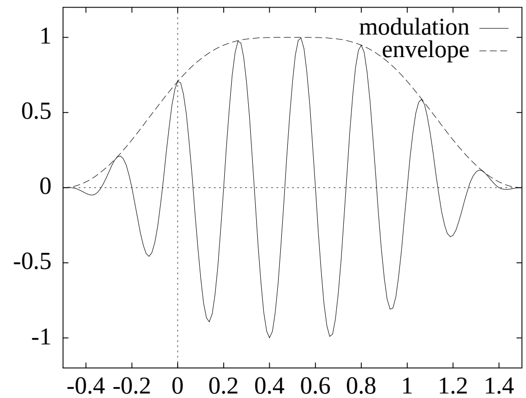

Figure 1: Smooth cosine4 basis function

The *smooth spectrum* of f over $I = [a_j, a_{j+1}]$ is the set of coefficients

$$C_j = \{c_{j,k} : 0 \le k < \ell_j\}$$

of the signal decomposition:

$$f(t) = \sum_{\substack{j\in\mathbb{Z}\\ k\in N}} c_{j,k}\Psi_{j,k}(t),$$

where

$$c_{j,k} = \langle f, \Psi_{j,k}\rangle = \langle f, \omega_j g_{j,k}\rangle \tag{5}$$

is the *smooth cosine4 transform*.
Since

$$c_{j,k} = \langle f, U_j^* U_{j+1}^* \chi_{I_J} g_{j,k}\rangle = \langle F_{a_j,a_{j+1}}, g_{j,k}\rangle,$$

the *smooth cosine4* transform $c_{j,k} = \langle f, \Psi_{j,k}\rangle$ is equal to the *block cosine4* transform of the folded signal

$$c_{j,k} = \langle F_{a_j,a_{j+1}}, g_{j,k}\rangle. \tag{6}$$

3. Fang's segmentation algorithm

Fang's segmentation algorithm computes the *local maxima* of a *frequency change function*. This function is the average of an *instantaneous frequency change function* that oscillates even when the signal has constant frequencies.

3.1. Instantaneous frequency change function

This function can be obtained using the signal spectrum computed with either the *block* or the *smooth cosine4* transform. This function is the difference between the *flatness* of the spectrum over an interval

$[n-\ell, n+\ell]$ with fixed $\ell > 0$ and the *flatness* of the combined spectra over $[n-\ell, n]$ and $[n, n+\ell]$. This *flatness* can be measured with one of the following *cost functions*:

$$\lambda(x_0, x_1, \ldots, x_m) = \sum_{k=0}^{m-1} |x_k| \tag{7}$$

or

$$\lambda(x_0, x_1, \ldots, x_m) = -\sum_{k=0}^{m-1} |x_k|^2 \log(|x_k|^2). \tag{8}$$

where $(x_0, x_1, \ldots, x_m)$ is a point of $\mathbb{R}^m$.

Let A_n, B_n, and C_n denote the *cosine4* spectrum over $[n-\ell, n+\ell]$, $[n-\ell, n]$, and $[n, n+\ell]$. Then

$$IFC(n) = \lambda(C_n) - (\lambda(A_n) + \lambda(B_n)) \tag{9}$$

is called the *instantaneous frequency change* function, where $n \in \{\eta+\ell, \ldots, N-\eta-\ell\}$, $0 < \eta < 2\ell$, and $\eta = 0$ if the block cosine4 is used.

This function oscillates even when the signal is periodic as shown in Fig.1. The IFC function is plotted in the bottom and its average, the AFC function, is plotted in the middle. The signal over $[n-\ell, n+\ell]$ changes with n.

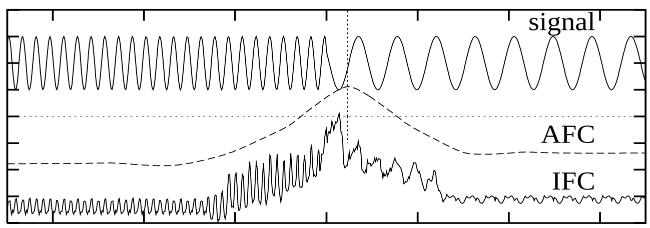

Figure 2: IFC and AFC frequency change functions

3.2. Segmentation algorithm

This algorithm consists of the following five steps:

1. Compute $IFC(n)$ for $n \in]\ell+\eta, N-\ell-\eta[= I$ as follows:
   Consider $IFC(n) = 0 \ \forall n \in I$ and compute C_n, the *dct4* transform of the signal over $[n-\ell, n+\ell]$, and B_n, the *dct4* transform of signal over $[n, n+\ell]$. Then

   $$IFC(n) = IFC(n) + \lambda(C_n) - \lambda(B_n),$$

   and

   $$IFC(n+\ell) = IFC(n+\ell) - \lambda(B_n),$$

   since $A_{n+\ell} = B_n$.

2. Filter $IFC(n)_{n \in I}$ to obtain an *averaged frequency change function* $AFC(n)_{n \in I}$.

3. Find the *local maxima* by detecting zero crossings of the adjacent differences of $AFC(n)_{n \in I}$.

4. Squelch the local maxima above some threshold.

5. **Improvement**
We consider only the local maxima of AFC such that its second derivative is lower than a given negative threshold. This condition eliminates those maxima that are too flat.

There are three parameters to set:
1) the adjacent window overlap η,
2) the window size ℓ,
3) the number d of iterations of the *lowpass filter* H.

Fig.3 shows the first half of a second of a flute signal plotted in the top. It was segmented with Fang's algorithm *near elementary structures* using $\eta = 16$, $\ell = 256$, and $d = 7$. The IFC function is shown in the bottom. Its average, the AFC function, is in the middle. Vertical lines are plotted at segmentation points given by AFC local maxima.

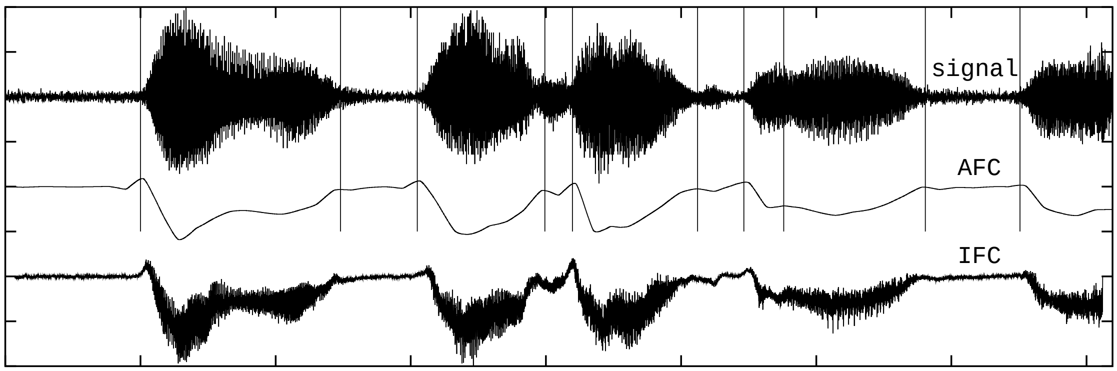

Figure 3: Music signal segmentation

4. Smooth spectrum near local structures

In previous papers, we analyzed speech signals first [9] using the orthonormal *cosine4 Best Basis* algorithm which computes the *smooth local spectrum* over dyadic segments. We then compute the *smooth local spectrum* of speech signals *near phonemes*[10][11] using Fang's segmentation algorithm.

In this paper, we compute *smooth spectrum near elementary structures.* We first *fold* the signal at segmentation points and takes it's restriction over each interval

$$F_{a_j,a_{j+1}} = \chi_{I_j} U_j U_{j+1} f.$$

The *smooth spectrum*

$$c_{j,k} = \langle f, \Psi_{j,k} \rangle$$

over $I_j = [a_j, a_{j+1}]$ is then computed using the *block cosine4* transform

$$c_{j,k} = \langle F_{a_j,a_{j+1}}, g_{j,k} \rangle$$

of the folded signal, where $0 \leq k < l_j$ is the frequency variable. This spectrum is constant over each segment $I_j = [a_j, a_{j+1}]$; we say that this *spectrum* is *near elementary structure.*

Fig.4 shows this *spectrum* in absolute value separated by vertical lines at time segmentation points. Each segment in the bottom of this graph represents the whole frequency interval. The previous segmented signal is shown in the top.

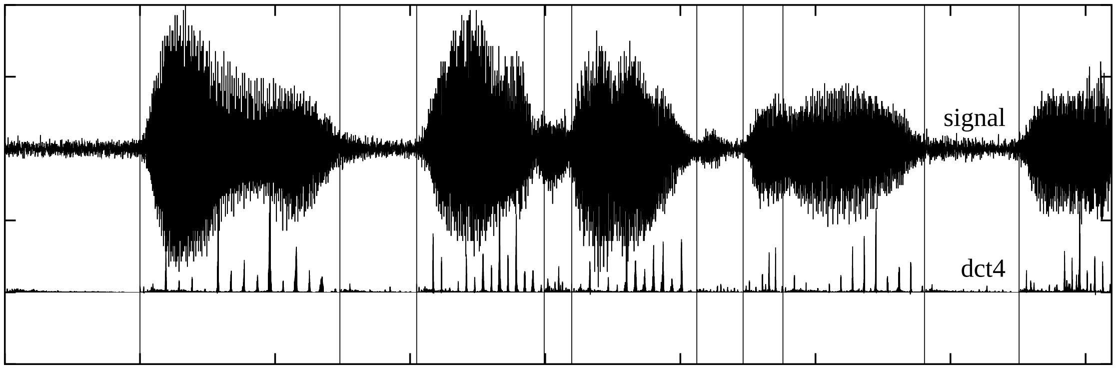

Figure 4: Smooth local spectrum near *elementary structures*

5. Non linear approximation

Each signal was approached using p percent of the smooth local spectrum, choosing the largest energy coefficients.

Let us denote

- k_0 the integer part of $\ell_j * p/100$,
- $(s_{j,k})_k$ the sequence $(|c_{j,k}|)_k$ sorted in decreasing order,
- $T_j = s_{j,k_0}$,

$$\tilde{c}_{j,k} = \begin{cases} c_{j,k} & \text{if} \quad |c_{j,k}| \geq T_j \\ 0 & \text{if} \quad |c_{j,k}| < T_j . \end{cases}$$

Each *folded* signal is approached using p percent of the coefficients:

$\tilde{F}_{a_j,a_{j+1}} = \sum_{k\in N} \tilde{c}_{j,k} g_{j,k}(t)$ therefore $\tilde{f}_p(t) = \sum_{\substack{j\in\mathbb{Z}\\k\in N}} \tilde{c}_{j,k}\Psi_{j,k}(t)$.

The approximation error $error(p) = ||f - \tilde{f}_p||_2$, is less for the local spectrum *near elementary structure* than for the Best Basis representation. Fig.5 shows this performance for p between 0 and 40.

6. Conclusion

The *cosine4* time-frequency representation has better approximation using Fang's segmentation algorithm then the Best Basis dyadic segmentation. It has less reconstruction distortion, however the number of operations is $O(N^2)$ instead of $O(NlogN)$.

References

[1] R.R. Coifman and Y. Meyer, *Remarques sur l'analyse de Fourier à fenêtre*, C. R. Acad. Sci. Paris 312, pp. 259-261, 1991.

[2] R.R. Coifman and M.V. Wickerhauser, *Entropy-based algorithms for best-basis selection*, IEEE Trans. Info. Theory, March, 1992.

[3] X. Fang, *Automatic Phoneme Segmentation of Continuous Speech Signals*, Report, Dept. of Mathematics, Washington University, Saint Louis, USA, 1994.

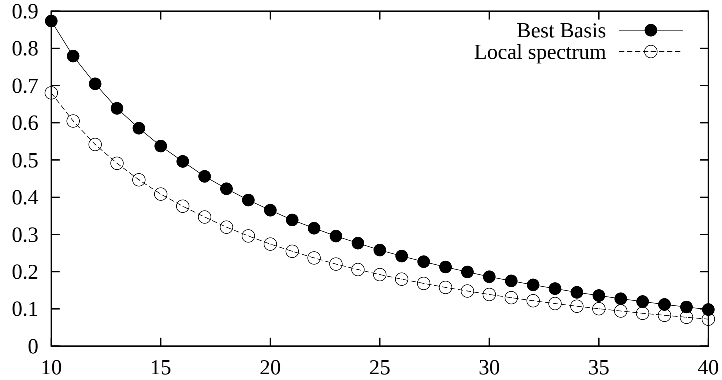

Figure 5: Comparison between Best Basis and local spectrum error approximation

[4] H. Malvar, *Signal Processing with Lapped transforms*, Artech House, Norwood, MA, 1992.

[5] S. Mallat, *A wavelet tour of signal processing*, Academic Press, 1998.

[6] Y. Meyer, *Wavelets: Algorithms and Applications*, Siam, 1993. Translated and Revised by R.D. Ryan.

[7] Y. Meyer, *Ondelettes et Algorithmes Conccurrents*, Hermann, 1992.

[8] V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, A.K. Peters, Wellesley, MA, 1994.

[9] E. Wesfreid and V. Wickerhauser, *Adapted trigonometric transform and speech processing*, IEEE Trans. Acoustic Speech Processing, Dec. 41, 12, 3596-3600, 1993.

[10] E. Wesfreid and V. Wickerhauser, R. Bouguerra, *Well adapted non dyadic local spectrum for some acoustic signals*, International Wavelet Conference, IWC-Tanger98.

[11] E. Wesfreid and V. Wickerhauser, *Vocal command signal segmentation and phonemes classification* CIMAF 99, II Symposium on Artificial Intelligence, 45-50, (1999).

[12] S. Jaffard, Y. Meyer and R. D. Ryan, *Wavelets: Tools for Science and Technology*, SIAM, Philadelphia, PA (to appear in April 2001).