WELL ADAPTED NON DYADIC LOCAL SPECTRUM FOR SOME ACOUSTIC SIGNALS

*E. Wesfreid, V. Wickerhauser *and R. Bouguerra.*

LMA, Université du Littoral. Calais - France.
Department of Mathematics, Washington University in St. Louis - USA.
Hopital Laennec, Paris, France.

ABSTRACT

We show that the Fang's segmentation algorithm[1] of *nearly constant instantaneous frequency* is *well adapted* to some *noisy vocal command signals* and that the orthonormal trigonometric basis of $l^2(Z)$ [2] defined over this segmentation offers an *optimal, non dyadic* time-frequency tiling. We use this basis in speech processing to compute a local spectrum *nearly phonemes* and in biomedical applications to *timing velopharyngeal closure* for swallowing sound.

1. INTRODUCTION

Given a time axis segmentation into intervals of arbitrary length, a *smooth local spectrum* can be computed using an orthonormal trigonometric basis of $l^2(Z)$[3][4][5] defined over this segmentation.
On one hand, the *Best Basis*[2] of Coifman and Wickerhauser offers a local spectrum defined over dyadic segments which can be obtained in $O(NlogN)$ operations.
On the other hand, the orthonormal basis defined over a Fang's segmented signal provides a *well adapted* spectrum defined over non dyadic segments which can be obtained in $O(N^2)$ operations.
We use Fang's segmentation in speech processing to compute local spectrum *nearly phonemes* and in biomedical applications to *timing velopharyngeal closure.*

2. SMOOTH LOCAL TRIGONOMETRIC TRANSFORM[6]

Let us consider

- 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 following *orthonormal* window over $I_j = [a_j, a_{j+1}]$

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

  with $\ell_j = (a_{j+1}-a_j)$, $t \in Z+1/2$, $j \in Z$ and $0 < \eta < \ell_j/2$ (η_j is the adjacent window overlap).

- $b_j(t) = \frac{r(t-a_j)}{\eta}$.
- A sampled signal $\{f(j)\}_{0 \le j < N}$.

The *folding* operator defined in a neighbour of a_j:

$$U_j f(t) = \begin{cases} b_j(t)f(t) + b_j(2a_j-t)f(2a_j-t) & \text{if } a_j < t < a_j+\eta \\ b_j(2a_j-t)f(t) - b_j(t)f(2a_j-t) & \text{if } a_j-\eta < t < a_j \end{cases}$$

and its adjoint, the *unfolding* operator:

$$U_j^* f(t) = \begin{cases} b_j(t)f(t) - b_j(2a_j-t)f(2a_j-t) & \text{if } a_j < t < a_j+\eta \\ b_j(2a_j-t)f(t) + b_j(t)f(2a_j-t) & \text{if } a_j-\eta < t < a_j \end{cases}$$

- The *folded* function over $I_j = [a_j, a_{j+1}]$

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

  where $\chi_{I_j}(t)$ is equal to 1 when $t \in I_j$ and null elsewhere.

The given *orthogonal* window (1) can also be written as follows

$$\omega_j(t) = U_j^* U_{j+1}^* \chi_{I_j}. \qquad (2)$$

Given a time axis segmentation into intervals

$$I_j = [a_j, a_{j+1}]$$

of arbitrary length, the associated *orthonormal trigonometric basis* of $l^2(Z)$[3][4]

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

Research supported in part by AFOSR, NSF, and the Southwestern Bell Telephone Company

consists of *orthonormal* windows $w_j(t)$ modulated by trigonometric functions

$$g_{j,k}(t) = \frac{\sqrt{2}}{\sqrt{|\ell_j|}} \cos \frac{\pi}{|\ell_j|}(k+\frac{1}{2})(t-a_j). \qquad (3)$$

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

$$C_j = \{c_{j,k} : 0 \leq k < \ell_j\} \qquad (4)$$

where

$$c_{j,k} =< f, \Psi_{j,k} >=< f, \omega_j g_{j,k} >$$

and

$$c_{j,k} =< f, U_j^* U_{j+1}^* \chi_{I_J} g_{j,k} >=< F_{a_j,a_{j+1}}, g_{j,k} > . \quad (5)$$

Therefore the *dct4* transform of the folded signal $F_{a_j,a_{j+1}}$ is the *smooth dct4* transform of f over I_j.

3. FANG'S SEGMENTATION ALGORITHM[1]

A segmentation of a sampled signal is a strictly increasing sequence of integers which are the initial indices of each segment. The Fang's segmentation is obtained computing the *local maxima* of a *frequency change function*, which is the average of an *instantaneous frequency change function*.

3.1. Instantaneous frequency change function

This function is the difference between the *flatness* of the spectrum over $[j-\ell, j+\ell]$ with $(\ell > 0)$ and the *flatness* of the combined spectra over $[j-\ell, j]$ and $[j, j+\ell]$. This *flatness* can be measured with the following *cost function* of $x = (x_0, x_1, \ldots, x_n)$

$$\lambda(x) = \sum_{k=0}^{n-1} |x_k| \qquad (6)$$

or

$$\lambda(x) = \sum_{k=0}^{n-1} |x_k|^2 log(|x_k|^2). \qquad (7)$$

Let A_j, B_j and C_j denote the *smooth dct4* transform over $[j-\ell, j[$, $[j, j+\ell[$ and $[j-\ell, j+\ell[$, then

$$IFJ(j) = \lambda(C_j) - (\lambda(A_j) + \lambda(B_j)) \qquad (8)$$

with $j \in \{\eta+\ell, \ldots, N-\eta-\ell\}$ is the *Instantaneous Frequency Change* function. This function oscillates, even when the signal is periodic as it is shown in Fig.1.

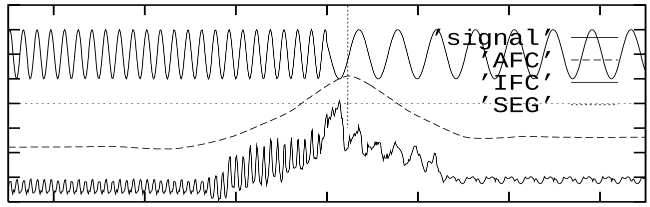

Figure 1: IFC and AFC frequency change functions

3.2. Segmentation algorithm

This algorithm consists of the following four steps:

1. Compute $IFC(j)$ for $j \in]\ell+\eta, N-\ell-\eta[= I$ as follows:

   Let consider $IFC(j) = 0 \ \forall j \in I$ and compute C_j, the *dct4* transform of $F_{j-\ell,j+\ell}$ and B_j, the *dct4* transform of $F_{j,j+\ell}$, then

   $$IFC(j) = IFC(j) + \lambda(C_j) - \lambda(B_j)$$

   and

   $$IFC(j+\ell) = IFC(j+\ell) - \lambda(B_j)$$

   because $A_{j+\ell} = B_j$.

2. Filter $IFC(j)_{j\in I}$ to obtain an *averaged frequency change function* $AFG(j)_{j\in I}$ as follows:

   If H and G denote a *biorthogonal lowpass filter* and its dual then

   $$AFC = GH^d(IFC)$$

   where $H^d = HHH\ldots H$.

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

4. Squeltch the local maxima above some threshold.

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.

In particular, we use this algorithm with $\eta = 16$, $\ell = 256$ and $d = 9$ to obtain a *nearly phoneme* segmentation of *noisy vocal signals* recorded in fly.

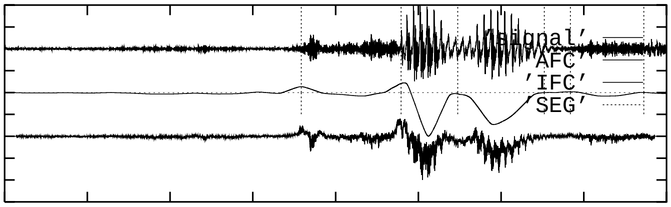

Figure 2: vocal signal segmentation

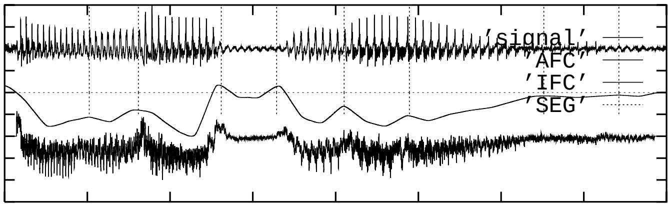

Figure 3: vocal signal segmentation

Fig2 and Fig3 show the signal at the top, the IFC function at the bottom, the AFC function at the middle and the segmentation, with vertical lines at the AFC local maxima:

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

This is a *non dyadic* segmentation *nearly phonemes.*

4. OPTIMAL SPECTRUM

In a previous paper[7] the speech signal segmentation was realized using an orthonormal trigonometric *Best Basis* followed by a *split and mergin*[8] algorithm.
Fig.4 and Fig5 show a local spectrum, in absolute value, computed using the orthonormal trigonometric bases defined over the *Fang's segmented* signal. The coefficients of this spectrum are defined in (4) and (5) over each interval $[a_j, a_{j+1}]$.

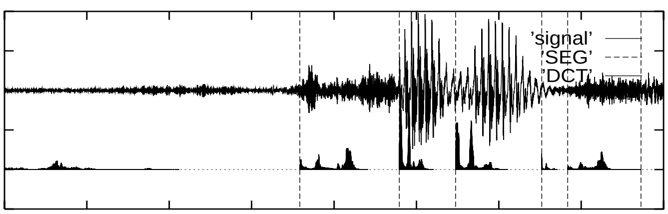

Figure 4: well adapted, non dyadic local spectrum

The *Fang's segmentation* algorithm was also applied in biomedicine to *timing velopharyngeal closure* for swallowing sound. The described algorithm was used with $\eta = 16$, $\ell = 128$ and $d = 7$, the local spectrum was computed over each interval $[a_j, a_{j+1}]$ of this segmentation

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

using (4) and (5). We searched a time interval $[T_0, T_1]$ with $T_0 = a_s$ and $T_1 = a_p$ such that the absolute value

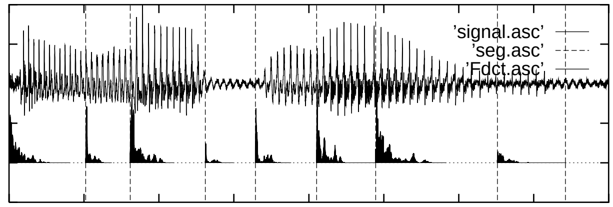

Figure 5: well adapted, non dyadic local spectrum

of the local spectrum over $[T_0, T_1]$ has their maxima greater than some preallocated threshold (Fig6).

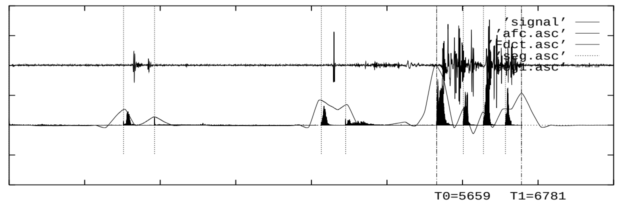

Figure 6: timing velopharyngeal closure

References

[1] X. Fang, *Automatic Phoneme Segmentation of Continuous Speech Signals*, IEEE Transactions on Speech and Audio Processing, (1994).

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

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

[4] H. Malvar, *Lapped transforms for efficient transform/subband coding*, IEEE Trans. Acoustics, Speech and Signal Processing, **38**, pp. 969-978.(1990)

[5] H. Malvar, *Signal Processing with Lapped transforms*, ARTECH HOUSE, Boston, London (1992).

[6] V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, A. K. Peters. Wellestey Massachusetts, (1994).

[7] E. Wesfreid and V. Wickerhauser. *Adapted trigonometric transform and speech processing*, IEEE Trans. Acoustic Speech Processing. Dec. 1993.

[8] Y. Meyer, *Wavelts Algorithms and Applications*, Siam,1993.(Translated and Revised by Robert D. Ryan).