IMPROVED ENTROPIC GAIN FOR SPEECH SIGNALS ANALYSIS/SYNTHESIS BASED ON AN ADAPTIVE TIME-FREQUENCY SEGMENTATION SCHEME

G. Gonon[1], *S. Montrésor*[2], *M. Baudry*[1]

[1] Laboratoire d'Informatique,
[2] Laboratoire d'Acoustique UMR CNRS 6613
Université du Maine, 72085 Le Mans Cédex 9, France

ABSTRACT

In the search for adaptive representation of speech signals, the Wavelet Packet Decomposition (WPD) has been proved to be a efficient tool because of its frequency adaptation skills through the best basis search algorithm. The entropic minimization of this algorithm is bounded by two artifacts : the dyadic structure of the decomposition and the lack of temporal segmentation. We propose here a low cost extended tree in the WPD which improves the best basis search by reducing the entropy of the base and which is still compatible with the classical WPD. The decomposition also allows perfect reconstruction. The entropic test is updated to take into account the new basis. The preliminary use of a temporal segmentation, based on the Local Entropic Criterion highly improves the entropic gain of the global analysis. The results are shown on experimental speech signals comparing the gain of our scheme versus a usual WPD.

1. INTRODUCTION

The search for signal-adaptive transforms is of particular interest especially in areas as speech and audio coding, because the better the adaptation, the higher the compression rate.

The two main trends in coding consist in adapting the expansion either temporally [1] through the time-varying MLT or frequencially through the Wavelet Packet Decomposition (WPD) [2] and the frequency-varying MLT [3]. Methods giving a joint time-frequency segmentation in a rate-distortion sense have also been proposed [4] with a perceptual extension for audio coding [5].

The results strongly depend on the coding strategy regarding the signal properties. Many different coding schemes exist because of the variety of signals involved. Transform coding is very efficient for audio signals whereas CELP coders seems better for speech signals. Anyhow, whatever the strategy, a better adaptation of the transform to the signal is always needed.

The aim of this paper is twofold : to show that adapted temporal segmentation improves the entropic gain and to propose an extension of the dyadic WPD Best Basis Search Algorithm (BBSA). To achieve this, we propose to apply a detector before the encoding is done in order to decide which strategy to use. The scheme is shown figure 1.

The detector we use is the Local Entropic Criterion (LEC), presented in [6], which gives informations on the signal behavior. It gives an automatic temporal segmentation of the signal and allows to decide whether the signal is speech or noise. Each segment will then be decomposed using the WPD BBSA.

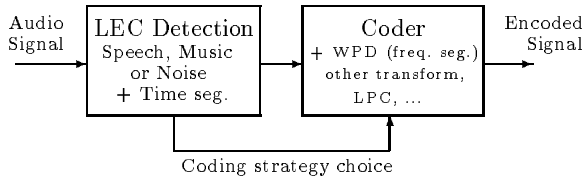

Figure 1: Adaptive coding scheme based on the LEC detector and WPD analysis.

The BBSA expands the signal on an optimal wavelet basis according to a cost function, such as entropy or energy. Thus it finds a partition of the frequency axis adapted to the signal in the sense that single tones are isolated in narrow bands, and that grouped wide bands are noisylike or empty.

A problem subsists due to the dyadic structure of the decomposition. Artificial segmentation can be induced by the WPD because the entropic test does not operate on adjacent bands not coming from the same father in the decomposition.

We propose in the first part of this paper to extend the dyadic basis family by constructing the wavelet packet corresponding to the father of nodes (2,1) and (2,2), and then to include it in the entropic test of the BBSA, also called "Split and Merge" algorithm. The proposed decomposition is noted Sigma-WPD (SWPD) and the new fathers constructed are indexed $(d, 2\,b+\frac{1}{2})$, where d is the depth of the decomposition and b the packet number, $b \in [0, 2^d - 1]$.

Section 2 explains the design of the filter equivalent to the new band, the computation of the SWPD coefficients and the way to take into account the new packets $(d, 2\,b+\frac{1}{2})$ in a "split-merge" bottom-up BBSA. Section 3 presents the Local Entropic Criterion applied to automatic temporal segmentation. Section 4 compared the entropic gains obtained with the different methods on speech signals. The improvement brought by temporal segmentation on a WPD coder is also shown. Finally, section 5 proposes further improve-

ments and applications of our method.

2. THE SIGMA-WPD AND EXTENDED BBSA

2.1. Basis Construction

In order to simplify the notations, we work at depth 2 of the decomposition. The construction remains the same for any other depth.

Using the analogy between the WPD and Filter Banks, the WPD can be seen as an iterative two ways Quadrature Mirror Filter (QMF) bank. The problem is then to design the filter equivalent to the father of packets (2,1) and (2,2), which are normally coming from two different fathers in the WPD. The new filter, if divided in two with the QMF bank, must also be coherent with the WPD and so lead to the coefficients of packets (2,1) and (2,2). We are in fact searching another way to obtain the WPD coefficients of depth 2 in a different iterative way than with the WPD. The construction method for this filter bank is explained in [7] and only the main steps are summed up here.

The desired filter bank must split the frequency axis into 3 bands with frequency ranges $[0, \frac{F_s}{8}]$, $[\frac{F_s}{8}, \frac{3F_s}{8}]$ and $[\frac{3F_s}{8}, \frac{F_s}{2}]$, respectively corresponding to packets $(2,0)$, $(1,\frac{1}{2})$ and $(2,3)$.

It must also satisfy the perfect reconstruction properties of the QMF bank.

Letting h_n be the coefficients of the lowpass QMF $h(n)$ and $g_n = (-1)^n h_{-n+1}$ be the highpass QMF, $h(n)$ satisfying the QMF and perfect reconstruction properties (see [2]).

The method consists in upsampling by a factor 2 the Impulse Response (IR) of h_n and g_n. This operation insures that the QMF properties and perfect reconstruction are left unchanged for the upsampled filters. Let $h_k^{\uparrow 2}$ and $g_k^{\uparrow 2}$ denote the IR of these filters.

$h_k^{\uparrow 2}$ ranges over bands $[0; \frac{F_s}{8}] \cup [\frac{3F_s}{8}; \frac{F_s}{2}]$ (stopband behavior) and $g_k^{\uparrow 2}$ over $[\frac{F_s}{8}; \frac{3F_s}{8}]$ (bandpass behavior).

Filtering $h_k^{\uparrow 2}$ and $g_k^{\uparrow 2}$ by h and g gives 4 filters equivalent to the depth 2 filters of the WPD. Downsampling each subband by a factor 4 leads to the WPD coefficients at depth 2. As expected the SWPD is just an alternative way to obtain directly the WPD coefficients at depth 2. In the SWPD $g_k^{\uparrow 2}$ is the father of $(2,1)$ and $(2,2)$ we are looking for.

In the end, the involved filters, if indexed with their relative packet, are

$$\begin{aligned} h_{(2,0)}(n) &= \sum_k g_k^{\uparrow 2} h_{n-k}, \\ h_{(1,\frac{1}{2})}(n) &= h_k^{\uparrow 2}, \\ h_{(2,3)}(n) &= \sum_k g_k^{\uparrow 2} g_{n-k}. \end{aligned} \quad (1)$$

2.2. The critically sampled case

In a critical sampling sense, the coefficients of packet $(1,\frac{1}{2})$ of the SWPD cannot be obtained directly with a downsampling operation as for the WPD coefficients. Downsampling by a factor 2 the subband $(1,\frac{1}{2})$ would lead to unrecoverable aliasing.

The filter bank defined by (1) is an even type of modulated filter bank [8]. In this case, the central band has to be modulated before being downsampled. We choose the Single-Side Band modulation (SSB) presented in [8]. The SSB leads to real coefficients so that there is no expansion of the global number of coefficients in the decomposition. The principle of the SSB modulation is given in figure 2. The SWPD coefficients of band $(1,\frac{1}{2})$ result of the SSB modulation followed by a downsampling of a factor 2.

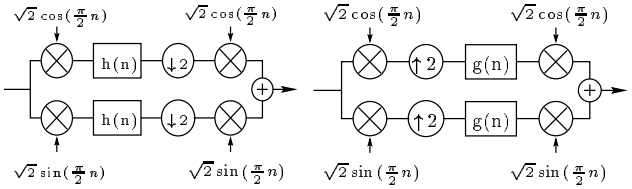

Figure 2: Single Side Band Modulation and downsampling scheme.

2.3. best basis search adaptation

In order to include the band $(1,\frac{1}{2})$ in the BBSA, we need to update the test proposed in [2]. We also propose a way to represent the best tree of the basis. Both stays compatible with the dyadic case so that they can be added on any "split and merge" algorithm.

2.3.1. New Entropic Test

The entropic test is based on the minimization of the entropy of the basis. In the dyadic case, each father's entropy is compared to its sons' entropy sum and the minimum fixes the basis. Now at each depth, starting from depth 2, we are dealing with 3 fathers for 4 sons.

Let $E_{(d,b)}$ denote the entropy or any other cost function of packet (d,b). We define the following test which allows to retain band $(1,\frac{1}{2})$ as a final node of the basis :

1. Dyadic test

$$\begin{aligned} E_{(1,0)} &\leftarrow \min(E_{(1,0)}, (E_{(2,0)} + E_{(2,1)})) \\ E_{(1,1)} &\leftarrow \min(E_{(1,1)}, (E_{(2,2)} + E_{(2,2)})) \end{aligned} \quad (2)$$

2. S-dyadic test : if

$$E_{(1,\frac{1}{2})} < E_{(2,1)} + E_{(2,2)} \text{ then test 3.} \quad (3)$$

3. Compatibility

$$\text{If } (E_{(1,\frac{1}{2})} + E_{(2,0)} + E_{(2,3)}) < E_{(1,0)} + E_{(1,1)}$$

$$\text{Then } E_{(0,0)} \leftarrow E_{(1,\frac{1}{2})} + E_{(2,0)} + E_{(2,3)} \quad (4)$$

$$\text{Else } E_{(0,0)} \leftarrow \min(E_{(0,0)}, (E_{(1,0)} + E_{(1,1)})) \quad (5)$$

In case (4) the packet $(1,\frac{1}{2})$ is marked as a final basis node else we are in the dyadic case.

2.3.2. S-Dyadic tree representation

When a s-dyadic packet is a final node in the best basis, we propose to insert it into the best tree representation. Figure 3 shows the adopted representation and its corresponding tiling in the time-frequency plane.

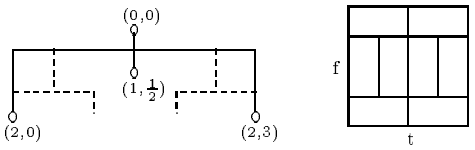

Figure 3: s-dyadic tree representation (dashed lines stands for removed dyadic part) and corresponding tiling of the TF plane.

2.3.3. Complexity

In D-depth WPD, $2^{D+1}-1$ subbands are computed leading to 2^D-1 entropic tests. The SWPD adds 2^D-1 s-dyadic fathers used in $2^{D-1}-1$ more entropic test. The whole complexity raises from 2^{D+1} to $3*2^D$.

3. THE LOCAL ENTROPIC CRITERION

The LEC is defined in [6] as a relative difference of entropy for a sliding length N window and the entropy of its two sub-windows of length $\frac{N}{2}$. Shannon's entropy is used to compute the CEL, as it can be considered as an indicator of the signal spectrum energy concentration. Letting $W_N^{nk} = e^{-j2\pi kn/N}$ denote the Discrete Fourier Transform Basis of length N, $s = \{s_n\}_{0\leq n<N}$ be the current windowed discrete signal and its halves $s_1 = \{s_n\}_{0\leq n<\frac{N}{2}-1}$ and $s_2 = \{s_n\}_{\frac{N}{2}\leq n<N}$, it is defined by :

$$E(s, W_N^{nk}) = -\sum_k \frac{|C_k|^2}{\|s\|^2} \ln(\frac{|C_k|^2}{\|s\|^2}), \text{ where } C_k = \sum_{n=0}^{N-1} s_n W_N^{nk}. \quad (6)$$

We assume $s_1 = \{s_n\}_{0\leq n<\frac{N}{2}-1}$ and $s_2 = \{s_n\}_{\frac{N}{2}\leq n<N}$ denote the left and right halves of the signal, and $E = E(s, W_N^{nk})$, $E_1 = E(s_1, W_{N/2}^{nk})$, $E_2 = E(s_2, W_{N/2}^{nk})$. The LEC function is then defined by :

$$LEC_s(\frac{N}{2}) = \frac{E-(E_1+E_2)}{|E+E_1+E_2|}. \quad (7)$$

Time dependency is obtained by sliding the window. The function LEC varies with respect to spectrum energy concentration of analyzed signal. Positive values traduce dispersion of the signal's spectral energy while negative ones are linked to its concentration.

The segmentation is achieved with a fast algorithm where ruptures are localized at local positive maximas of the LEC function computed for an arbitrary sliding step, depending on the precision required.

In this paper, we only use the LEC as a segmentation criterion but further informations can be extracted for audio coding such as decision over the signal type (speech/noise in [6]) or even an pre-allocation rate in the case of a variable rate coder. This last result is mainly due to the fact that better compression factors can be obtained on long stationnary parts of the signal, which are denoted by the minimas of the LEC. Examples of segmentation obained on speech signal are given in next section.

4. EXPERIMENTS

The described segmentation scheme has been tested on various audio and speech signals and gives coherent results. In order to test both improvements of our scheme we compare the entropic gain obtained on the LEC-segmented signal and on the complete sequence with the WPD and the S-WPD. The tested signal is the word "Eurospeech" sampled at 8kHz and recorded in a natural noisy environment, with a SNR close to 9 dB. It is shown in figure 4, with the segmentation resulting from the positive local maximas of the LEC. The WPD and S-WPD analysis have been performed at depth 8 of the decomposition with 10^{th} order "Symmlets" wavelets. The resulting entropic gains for differents

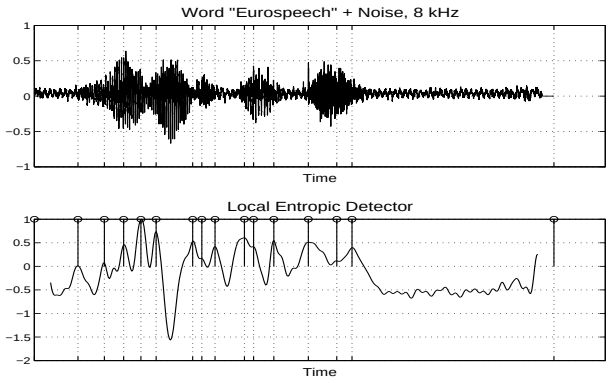

Figure 4: Top : Word "*Eurospeech*", bottom : automatic segmentation provided by local positive maximas of the LEC.

methods are shown in table 1. First part of table 1, allows to measure the gain provided by the S-WPD for the non-segmented signal. The results for each frame resulting form the CEL segmentation shows how dependant the entropic gain is to the signal's entropy. Though the results are strongly signal-dependant, we can say that the average gain provided by our extension of the WPD is about 3% . For the different frames, it ranges from 0% when no central band is retained up to 30% (frame 2). A high gain is generaly obtained when a low depth central node is kept as final.

The most relevant gain results from the comparison between the segmented and non-segmented analysis where the total entropic gain has doubled. This prooves the efficiency of the LEC segmentation, which contributes to get closer to the entropic boundary of the signal.

Assuming that negative minimas of the LEC localize stationnary centers, we can use the LEC to switch between different coding strategies. Such information may be included in a variable rate coder, in order to optimize each

Table 1: Entropic gain (in %) comparison for both methods WPD and SWPD BBSA on segmented and whole signal.

Non Segmented Signal					
Entropy	8.40	BB	5.97	S-BB	5.76
Gain in %	0	BB	29.8	S-BB	32.1
Segmented Signal (Gain in %)					
Frame	BB	S-BB	Frame	BB	S-BB
1	48.7	50.1	9	45.3	51.3
2	41.9	70.8	10	35.3	36.6
3	38.3	38.5	11	40.1	43.2
4	47.8	48.1	12	43.0	43.9
5	40.7	41.6	13	17.1	22.1
6	39.4	40.1	14	19.6	24.3
7	39.3	39.4	15	39.5	41.1
8	36.4	37.7			
Total Entropic Gain in %					
BB	56.9		S-BB	59.6	

frame rate and thus achieve better compression.

Preliminary results of a temporal-adaptive + WPD coding scheme are shown in 5. For high compression factors (> 50), on the non segmented sentence an energetic allocation in the resulting best basis coefficients, removes first the fricatives because of their spread spectrum, then low energy phonemes. In the case of an adapted segmentation, with the same compresion factor, each phoneme is allocated and the sentence remains intelligible. The quality of the decoded signal is still very poor, mainly because of the simple allocation algorithm and the high compression factor. Our goal was to reach the limit of intelligibility for the non segmented scheme and show that the decoded sentence remains intelligible with the LEC segmentation.

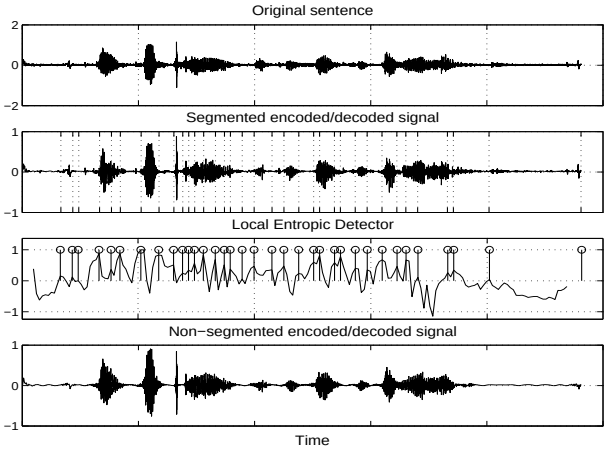

Figure 5: French sentence : "*Un fort crédit est consenti par une banque*", 8kHz sampled. From top to bottom : 1. Original, 2. Signal compressed 60 times (< 1.3kBits) with segmentation, 3. LEC detector and its segmentation, 4. Compressed signal without segmentation

5. CONCLUSION

The proposed time-frequency adaptive scheme improves both the entropic gain of the decomposition and the quality of a WPD coding scheme with energetic allocation. The inclusion of SWPD the best basis search algorithm extends the dyadic bases library of the WPD and allows perfect reconstruction, with a fair complexity raise. The results on speech signals prove that the extended library of bases improve the adaptation to the signal in a entropic diminution sense. Such an adaptive scheme allows to get closer to the minimal signal entropy bound and increases the global compression factor affordable for a given quality.

In future work we will include this decomposition in a wavelet based perceptual audio coder and adapt the test to take into account rate-distortion function and psychoacoustical phenomenas in the best basis search, as proposed in [5]. A finer analysis of the LEC will also be used in order to pre-allocate variable rates and switch between different coding strategies.

6. REFERENCES

[1] Seymour Shlien, "The modulated lapped transform, its time-varying forms, and its applications to audio conding standards," *IEEE Trans. on speech and audio processing*, vol. 5, no. 4, pp. 359–366, July 1997.

[2] R.R. Coifman and M.V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. on Information Theory*, vol. 38, no. 2, pp. 713–718, Mar. 1992.

[3] Purat M. and Noll P., "Audio coding with a dynamic wavelet packet decomposition based on frequency-varying modulated lapped transform," *ICASSP*, May 1996.

[4] Herley C., Kovačević J., Ramchandran K., and Vetterly M., "Tilings of the time-frequency plane: Construction of arbitrary orthogonal bases and fast tiling algorithms," *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3341–3359, Dec. 1993.

[5] Erne M. and Moschytz G.S., "Audio coding based on rate-distortion and perceptual optimization," *Proc. of the SPIE Wavelet Applications Conference, Orlando*, vol. 4056, pp. 235–246, Apr. 2000.

[6] Imad Abdallah, Silvio Montresor, and Marc Baudry, "Speech signal detection using a local entropic criterion," in *5th EUROSPEECH'97*, 1997, vol. 5, pp. 2595–2598.

[7] Gilles Gonon, Silvio Montrésor, and Marc Baudry, "Extension de la recherche de meilleure base pour la décomposition en paquets d'ondelettes. application à l'analyse en sous-bandes de la parole," in *XXIII^{èmes} Journées d'Études sur la Parole - JEP 2000*, June 2000.

[8] Ronald Crochiere and Lawrence Rabiner, *Multirate Digital Signal Processing*, Prentice-Hall Signal processing series, 1983.