A FLEXIBLE TILING OF THE TIME AXIS FOR ADAPTIVE WAVELET PACKET DECOMPOSITIONS

Antonio S. Pena, Nuria González Prelcic and Carlos Serantes

E.T.S.E.Telecomunicación
Universidad de Vigo
36200 VIGO (Spain)
nuria@tsc.uvigo.es

ABSTRACT

A segmentation procedure of time sequences based on a time-frequency analysis is presented in this paper. The use of both a wavelet packet transform and the original time signal provides a set of spectral and time parameters that allows the algorithm to locate some proper break points to split the input frame into a discrete number of smaller segments. Some examples showing the performance of the method are also presented. An application to wavelet-based audio coding is briefly discussed.

1. INTRODUCTION

One of the main problems of the analysis of 1-D time signals such as speech or audio is their non-stationary nature. The adaptive Wavelet Packet Transform, implemented by means of time-varying orthogonal filter banks, has been recently proposed as an efficient analysis tool when dealing with this kind of sequences [1, 2].

When using any analysis structure based on filter banks, some critical points have to be considered: first, the choice of the prototype filters of the analysis tree; second, the selection of the structure of the tree or, in other words, the choice of the basis used to decompose the signal; and finally the determination of the optimal length of the segment to analyze, that is, how often the filters and the basis have to be updated because of the appearance of changes in the features of the signal. The consideration of the influences of the filter characteristics and the different basis selection algorithms have been studied in the literature [3, 4, 5] but the last problem has not been successfully addressed so far.

In this work a wavelet packet based subframer has been developed. It makes use of the temporal and spectral information provided by both the wavelet packet coefficients and the time samples to decide how to split a fixed length frame into shorter segments or subframes. In this way, both sharp variations in the temporal envelope and changes in the spectral content of the signal are isolated.

2. SEGMENTATION PROCEDURE

Let us consider a signal frame of length L. In order to perform an adaptive analysis we desire to split this signal into subframes of different lengths. The final result will be a set of frame segments or subframes which do not present abrupt changes in their features.

To achieve this goal it is necessary to locate spectral and time changes in the original frame, to decide later if these variations are strong enough to lead to a splitting of the original segment. In the following subsections two procedures are proposed to detect spectral and time changes.

2.1. Spectral clustering

We will note as *time slots* the segments obtained when splitting a frame into NTS parts of equal length, where the NTS parameter is chosen according to the sampling rate.

For every time slot a **time-frequency pair** is computed in the following way:

Firstly, the entire frame is transformed using a one-stage wavelet packet decomposition. In order to achieve a good time resolution. short filters have been selected as the basic transformation cell,

After this analysis stage, we will have a sequence of temporally ordered coefficients for the high and low frequency bands, out of which the ones corresponding to every time slot can be extracted. We will note $l_i(k)$ and $h_i(k)$ the sequences of low and high frequency coefficients, respectively, of the time slot i. Figure 1 illustrates the computation of these signals for $L = 256$ samples and $NTS = 4$.

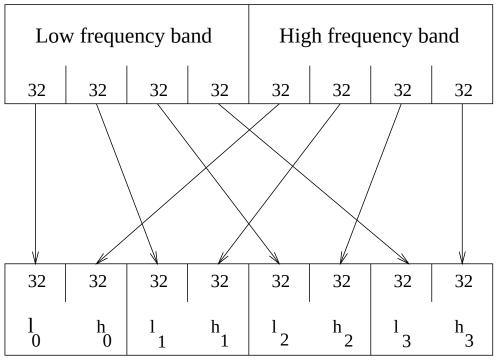

Figure 1: Composition of the sequences l_i and h_j.

Secondly, the quadratic norm of these sequences has to be computed, defining

$$\begin{aligned} L_i &= ||l_i(k)||_2^2, \\ H_i &= ||h_i(k)||_2^2. \end{aligned} \tag{1}$$

for $i = 1, ..., NTS - 1$. And the normalized *time-frequency pair* P_i is defined as

$$P_i = \left(\frac{L_i}{C}, \frac{H_i}{C} \right) \tag{2}$$

where C is a normalization factor

$$C = max\{L_0, ..., L_{NTS-1}, H_0, ..., H_{NTS-1}\}. \tag{3}$$

The NTS time-frequency pairs can be represented in a normalized high-low frequency plane, as shown in Figure 2. The so obtained *constellation* provides the following information:

- An estimation of the energy for each time slot; this may be used to locate changes in the envelope of the original frame, specially attacks.
- A temporal evolution of the signal energy distribution between high and low frequency bands.

Once the time-frequency pairs have been obtained, a **spectral clustering** stage decides how the NTS time slots are merged into different subframes.

This step is performed from the previously built constellation. When there exists a change in a specific time slot, a movement of the associated pair away from the previous one takes place in the constellation. The magnitude of the change can be measured as a function of the Euclidean distance between the two considered pairs

$$d_{i,j} = ||(P_i - P_j)||_2 \tag{4}$$

In order to decide if the change is significant enough to assign the two time slots to different subframes, a threshold Th is established. If the distance between pairs does not surpass this threshold, the corresponding time slots are merged. Obviously, the value of Th heavily depends on the final application of the subframer. Figure 2 illustrates the evolution of the clustering algorithm for an example constellation with $NTS = 3$. Note that the distances between not adjacent pairs also have to be considered.

2.2. Enhancement of time resolution

As the time resolution of the segmentation performed by the spectral clustering may not be enough to locate all the time changes, a refinement is performed by means of a **time clustering** procedure. With this aim, a new segmentation of the original frame into what we call *time bins* must be done. The number of time bins, NTB, is fixed to a higher value than NTS, with NTS being a multiple of NTB. The philosophy of the time clustering is similar to the spectral one; the only difference is that the distances are computed from the **localized energy** per time bin, instead from the time-frequency pairs

$$e_j = \frac{NTB}{L} \sum_{k=0}^{NTB-1} x_j[k]^2. \tag{5}$$

where x_j denotes the samples of the j-th bin, $j = 0, ..., NTB - 1$.

This is a simple and precise measure that let us locate variations on the amplitude of the time envelope that may not lead to a significant spectral variation in the previous clustering.

2.3. Description of the algorithm

The block diagram of the complete algorithm is shown in Figure 3. For each input frame, a spectral clustering and a time clustering are performed. The results of these operations go through the decision stage, that performs the final subframing by simply considering the segmentation provided by the spectral clustering and adding the hypothetic new break points introduced by the time clustering.

For a general analysis scheme, the minimum subframe size is then L/NTB. However, for an application as signal coding, in order to avoid an excessive increase of side information, the minimum subframe size is set to L/NTS. In this situation, the role of the refinement stage will be to add break points in a subframe returned by the spectral clustering of size equal or larger than $2L/NTS$.

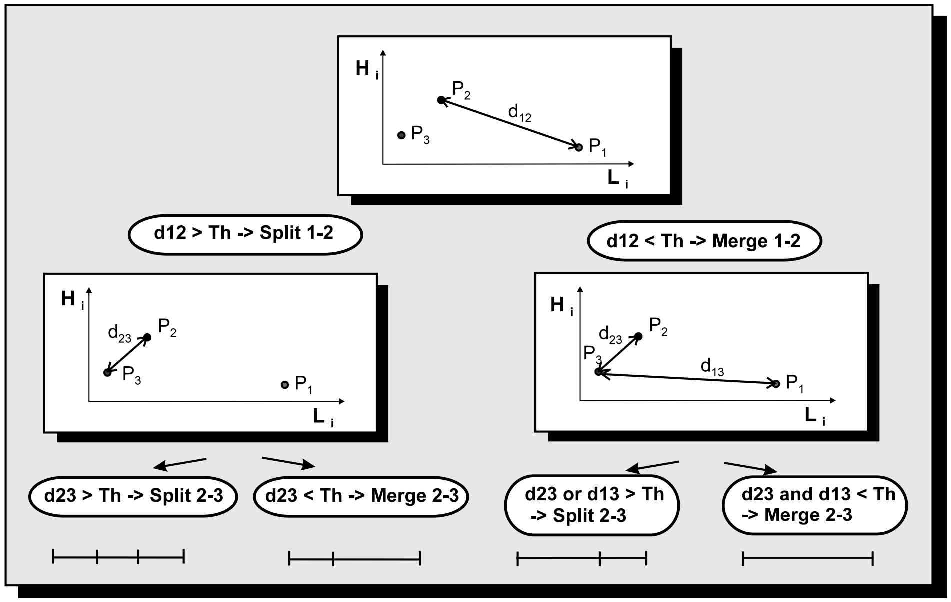

Figure 2: Evolution of the merging process by spectral clustering.

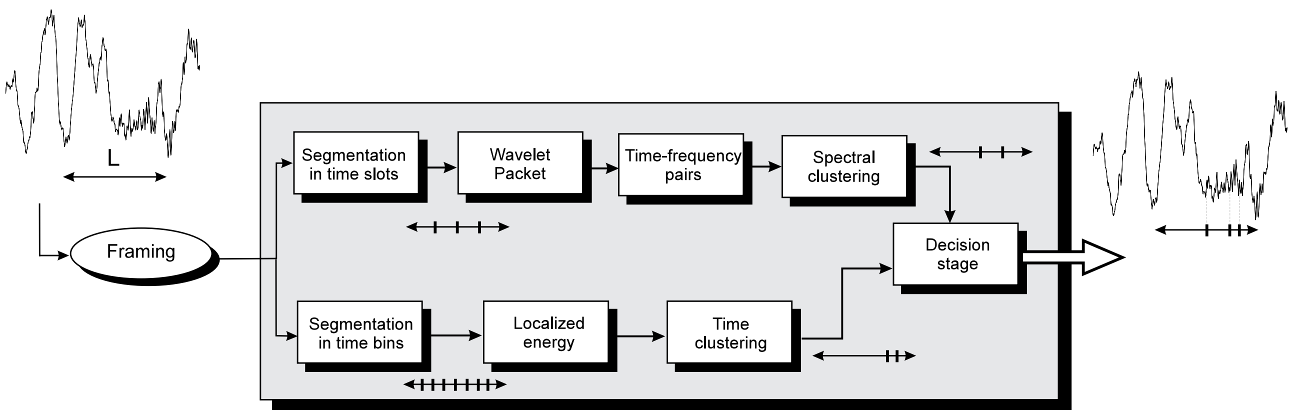

Figure 3: Block diagram of the subframing algorithm.

3. PERFORMANCE

The segmentation of an audio signal such as an attack is not usually a difficult task because the sharp variation of the time envelope can be easily detected. However, when just a change in the spectral distribution of the signal energy appears, the problem is not trivial. To illustrate the usefulness of the algorithm, we will analyze the 1024 samples audio frame with a slow change in the time envelope that appears in Figure 4(a). NTS has been fixed to 4. The time-frequency pairs associated to each slot of 256 samples appear in Figure 4(b). The circles in Figure 4(a) shows the points returned by the clustering algorithm, that decides to consider 3 subframes of length 256, 256 and 512.

It is important to remark the speed and low cost of the algorithm, an interesting feature if it has to be included in a preprocessing stage, previous to a coding procedure for example.

4. APPLICATION TO AUDIO CODING

Several high quality audio coding schemes based on nonuniform time-varying filter banks have been proposed in the recent literature [6, 7, 8]. This kind of systems makes use of a psychoacoustic model to adapt the frequency resolution of the filter bank to properly shape the quantization noise per band.

High coding gain can be achieved if the structure of the filter bank is updated for every frame to match

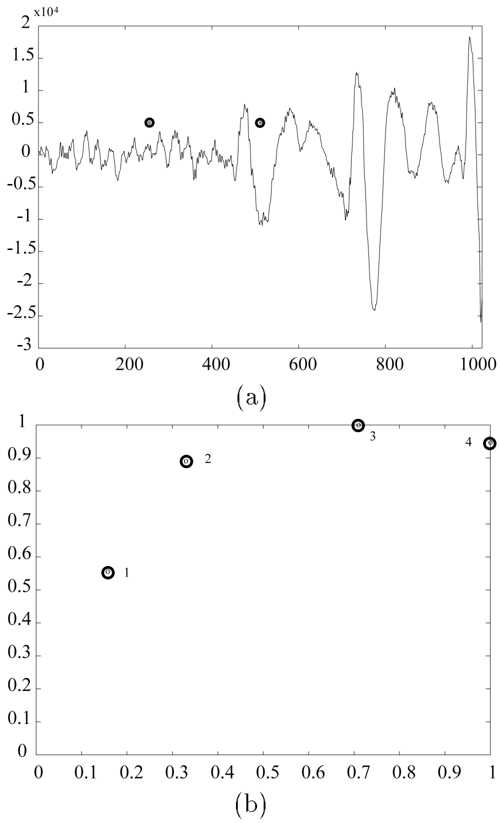

Figure 4: (a) Continuous line: original audio frame; circles: break points detected by the algorithm. (b) Time-frequency pairs.

the frequency decomposition to the masking threshold as described in [5]. At this point the following question arises: how often is it necessary to update this structure?

Some points have to be considered, at least: when the features of the audio signal have changed enough to significantly change the shape of the masking threshold or when the variation of the time envelope is so abrupt that the use of a simultaneous masking threshold does not make any sense, but taking into account the "frame size / side information" trade-off that has to be discussed considering the codec structure itself. Therefore, an algorithm as the one described in Section 2 will be adequate to fix an adaptive subframe size.

In the ARCO audio coder described in [8], the proposed segmentation scheme has been successfully introduced. In addition to the good performance of the scheme, we can enumerate some advantages:

- It is not necessary to add an extra block to compute the wavelet packet transform since it is already available to code the signal.
- If the subframer decides not to split the frame, the coefficients obtained during the computation of the time-frequency pairs can be reused in the coding stage.

5. CONCLUSIONS

A wavelet-packet based subframing procedure has been introduced and the evolution of the split-and-merge process fully described. This algorithm can be used as a preprocessing module for a general time-frequency analysis and its efficiency into an audio coding environment has been also justified.

6. REFERENCES

[1] C. Herley and M. Vetterli. Orthogonal Time-Varying Filter Banks and Wavelet Packets. *IEEE Transactions on Signal Processing*, 42(10):2650–2663, October 1994.

[2] I. Sodagar, K. Nayebi, and T.P. Barnwell. Time-Varying Filter Banks and Wavelets. *IEEE Transactions on Signal Processing*, 42(11):2983–2996, November 1994.

[3] R.R. Coifman, I. Meyer, and W.V. Wickerhauser. Signal Processing and Compression with Wave Packets. In *Proceedings of the Conference on Wavelets*, 1989.

[4] R.R. Coifman and M.V. Wickerhauser. Entropy-Based Algorithm for Best Basis Selection. *IEEE Transactions on Information Theory*, 38(2):713–718, March 1992.

[5] N. González Prelcic and A. S. Pena. An Adaptive Tree Search Algorithm with Application to Multiresolution based Perceptive Audio Coding. In *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, pages 117–120, 1996.

[6] D. Sinha and J.D. Johnston. Audio Compression at Low Bit Rates using a Signal Adaptive Switched Filterbank. In *Proceedings of ICASSP*, pages 1053–1056, 1996.

[7] J. Princen and J.D. Johnston. Audio Coding with Signal Adaptive Filterbanks. In *Proceedings of ICASSP*, pages 3071–3074, 1995.

[8] A. S. Pena, C. Serantes, and N. González Prelcic. ARCO (Adaptive Resolution COdec): A Hybrid Approach to Perceptual Audio Coding. In *100th AES Convention*, 1996.