Analysis/synthesis of sound signals using a discrete wavelet transform (DWT)

Christian Dorize, Olivier Rioul, Antoine Chaigne

ANALYSIS SYNTHESIS OF SOUND SIGNAL USING A DISCRETE WAVELET TRANSFORM (DWT)

C. Dorize*, O. Rioul, A. Chaigne

Lab. Acoustique, Département Signal, Ecole Nationale Supérieure des Télécommunications, 46 Rue Barrault, 75634 Paris cedex 13, France

Abstract

The DWT can be viewed as a frequency analysis of a time dependent signal carried out locally in time, but where high frequency components are studied with a sharper time resolution than low frequency components. The sampled signal is expanded into an orthonormal wavelet basis. The lack of redundancy leads to a fast and reversible algorithm (FWT), and several examples are given highlighting its efficiency: detection of low amplitude discontinuities in industrial acoustics and speech signals, and also extraction of frequency modulation.

Introduction

Wavelets $\psi_{a,b}(t)$ are analysis functions obtained from a single function $\psi(t)$ called "analyzing wavelet", and are located both in time and frequency [1]. Their aim is to reveal information contained in a signal s(t) at a given time and around a fixed frequency. In fact, "b" is a time location parameter (we translate $\psi(t)$ to point "b"), and "a" a scaling parameter to focus on a given range of oscillations (we contract or dilate $\psi(t)$ to the "a" scale):

$$\psi_{a,b}(t) = a^{-1/2}\ \psi\left(\frac{t-b}{a}\right);\quad a \in \mathbb{R}^{*+},\ b \in \mathbb{R} \tag{1}$$

The resulting scalar products of the signal s(t) and the wavelets $\psi_{a,b}(t)$ are called wavelet coefficients $C_{a,b}$:

$$C_{a,b} = \langle s,\ \psi_{a,b} \rangle_{L^2(\mathbb{R})} \tag{2}$$

They represent elementary information quantities present in a signal and are located in both time and scale. The set of wavelet coefficients can be used to form a time-frequency representation of the analyzed signal.

Let us name two fundamental properties of this transformation, i.e. energy conservation and reversibility:

$$\int |\, s(t)\,|^2\, dt = \int\int a^{-2}\, |\, C_{a,b}\,|^2\ da\, db \tag{3}$$

$$s(t) = \int\int a^{-2}\, C_{a,b}\, \psi_{a,b}(t)\ da\, db \tag{4}$$

Nevertheless, to preserve these two properties by calculating only a limited number of coefficients in the time-scale semi-plane (for future computer aided implementation), one has to reconsider the wavelets' theory, developed in the continuous case.

For this matter, the multi-resolution analysis concept [2] helps foresee the construction of orthonormal bases of wavelets, where the signal is entirely represented by a network of independent wavelet coefficients $C_{j,k}$ [4]:

$$C_{j,k} = \langle s, \psi_{j,k} \rangle_{l^2(Z)};\quad j \in \mathbb{N}^*,\ k \in Z \tag{5}$$

A wavelet positioning compatible with multi-resolution analysis is the following:

$$\psi_{j,k}(n) = 2^{j/2}\ \psi(\,2^{j}n\ -\ k) \tag{6}$$

where "j" is a so-called "voice" index used for discretization of the scale axis, and "k" a parameter adjusting the time position of the wavelet. We should note that such a discretization leads to a separation of the information into octaves [5].

* *Present address: Brüel & Kjær, Nærum Hovedgade 18, DK-2850 Nærum, Denmark*

I. Daubechies has recently discovered finite and causal sequences $\psi(n)$ causing the $\psi_{j,k}$ to form orthonormal wavelet bases [3]. The $\psi_{j,k}(n)$ wavelets are defined from the sequence $\psi(n)$ by successive dilations (D^2 operator) and interpolations (convolutions with a sequence $\varphi(n)$), [4] & [5]:

$$\psi_{j,k} = (D_\varphi^2)^j \left[\psi(n-k) \right] \tag{7}$$

where $\varphi(n)$ is an interpolation sequence describing a multi-resolution analysis of the signal [2]. When "j" is increased towards infinity, the discrete wavelet $\psi_{j,0}(n)$ tends to become a continuous one $\psi(x)$. Such a function as well as its Fourier transform are shown in Fig. 1 where we can see the simultaneous time and frequency location.

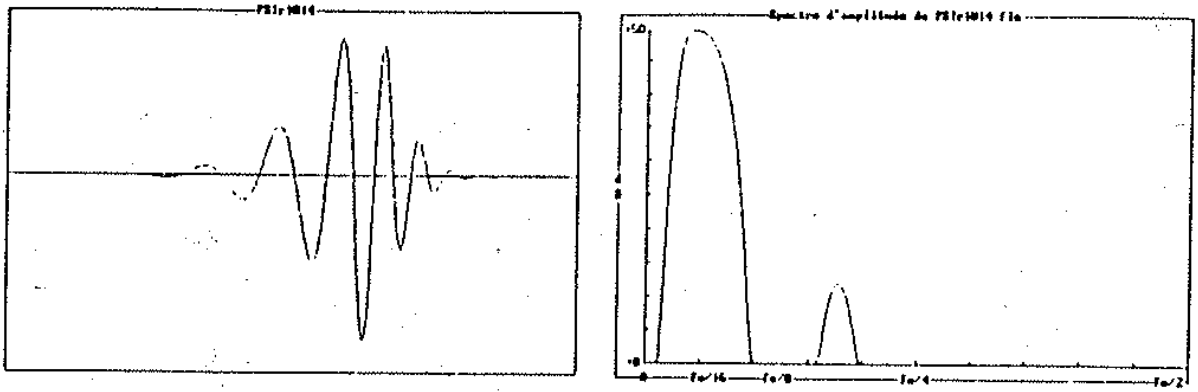

Fig. 1 On the left, the continuous wavelet version corresponding to a sequence $\psi(n)$ found by I. Daubechies. Note the oscillations and the time localisation. On the right, the Fourier transform of $\psi(t)$ which corresponds to a band-pass filter (octave band)

In order to decompose the signal into wavelet coefficients, one has to compute all the scalar products $<s, \psi_{j,k}>$. Considering the adjoining operator $D_\varphi^{1/2}$ (resp. $D_\psi^{1/2}$) of D_φ^2 (resp. D_ψ^2), where $D^{1/2}$ is a decimation operator, and noting their duality property for the scalar product, we can obtain an algorithm which reduces the numerical complexity of this transformation [4].

Whatever the number of voices "J" chosen for the decomposition ($J \geqslant 1$), the signal is entirely defined by the set of wavelet coefficients $C_{j,k}$ ($1 \leqslant j \leqslant J$) and by its approximation at the worst resolution s_J (see Fig. 2). The $C_{j,k}$ coefficients contain oscillations of frequencies varying between $F_s/2^{j+1}$ and $F_s/2^j$, where F_s is the sampling frequency of the signal s(n). The s_J approximation contains the slowest varying information, not collected by the wavelet coefficients, the frequency of which lies within the range $[0; F_s/2^{J+1}]$.

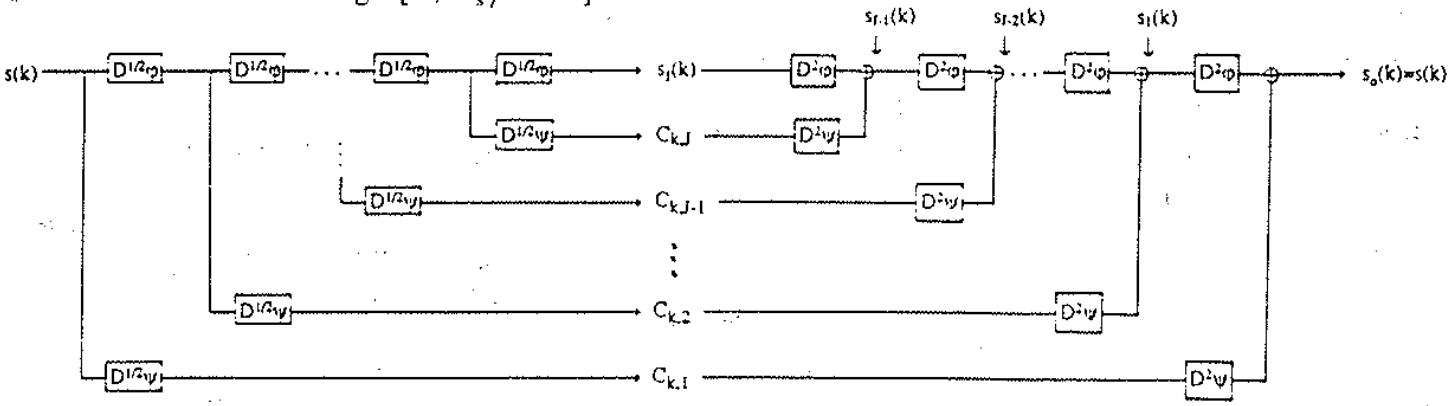

Fig. 2 The FWT algorithm splits the signal into J voices, which contain independent wavelet coefficients $C_{j,k}$. The low frequency information not collected in these coefficients are contained in the approximation $s_J(k)$. The reversible algorithm is able to perform an exact reconstruction of the signal

However, this decomposition differs from a conventional filter bank [5]. First because of the double orthogonality of the wavelet coefficients which leads to a lack of information redundancy and to a number of analysis points equal to the number of samples in the original signal; second, because this orthogonal decomposition allows a perfect recovery of the signal due to a similar reconstruction algorithm; finally, the decomposition algorithm presents a low numerical complexity, which is comparable to that of an FFT [4]. Hence, the name Fast Wavelet Transform (FWT).

Our aim is now to present some examples showing the potential applications of this transformation.

Detection of plosives

The reason for highlighting plosives in a speech signal is in order to split the message into syllables, and represents one step in the general process of speech recognition. However, these plosives might be masked by vowels of much higher energy. It is for example the case with the "**b**" of "tambour" in the sentence "dès que le tambour bat", of which the time representation is given in Fig.3.a. In order to extract such a plosive, we can calculate the prediction error a LPC (Linear Prediction Coding) has led to. The prediction error increases when it is difficult for the filter to adapt to the signal, which is particularly the case within the non-stationary periods associated with plosives.

An alternative method would be to observe the wavelet coefficients corresponding to the most rapidly varying oscillations present in the signal. At this scale, because of the high time resolution provided by the transformation, one should expect an easy detection of phenomena highly concentrated in time. This is shown in Fig.3.c, where the "**b**" in "tambour" is very distinct from the neighbouring vowels, while it is still masked with the prediction error (Fig.3.b).

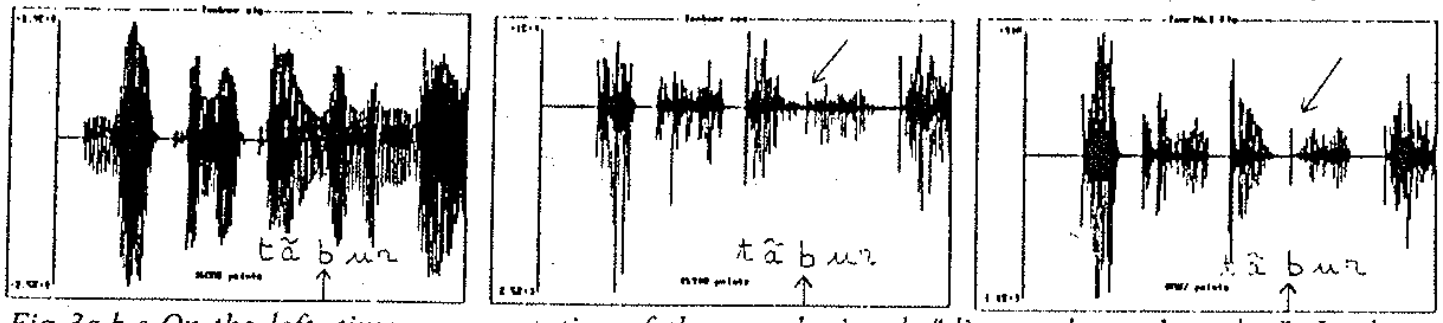

Fig. 3a,b,c On the left, time representation of the speech signal: "dès que le tambour bat". In the middle, prediction error obtained with a 16 coefficient linear prediction filter. On the right, wavelet coefficients corresponding to the fastest oscillations (voice j = 1) of the signal. The burst for the "b" of "tambour" is easily detected by the wavelet coefficients

This example characterizes very well the focusing abilities of the wavelet coefficients on such peaks.

Extraction of a frequency modulation

Revealing the existence of a frequency modulation in a time signal may be perceived as an observation of its slowest varying time oscillations. And the ability to focus on a fixed oscillation range is one feature of the wavelet transform.

During the signal decomposition, one has to concentrate on the lowest resolution s_j approximations (j is a high value), or on the $C_{j,k}$ wavelet coefficients on the voices with a high j index.

Fig. 4.a represents a modulated sinusoïd with a time varying modulation frequency f_m. Here is the general formula, where f_0 is the instantaneous frequency of the signal, f_c its central frequency, and Δf the maximum variation in frequency:

$$f_0 = f_c + \Delta f \sin(2\pi f_m t) \qquad (8)$$

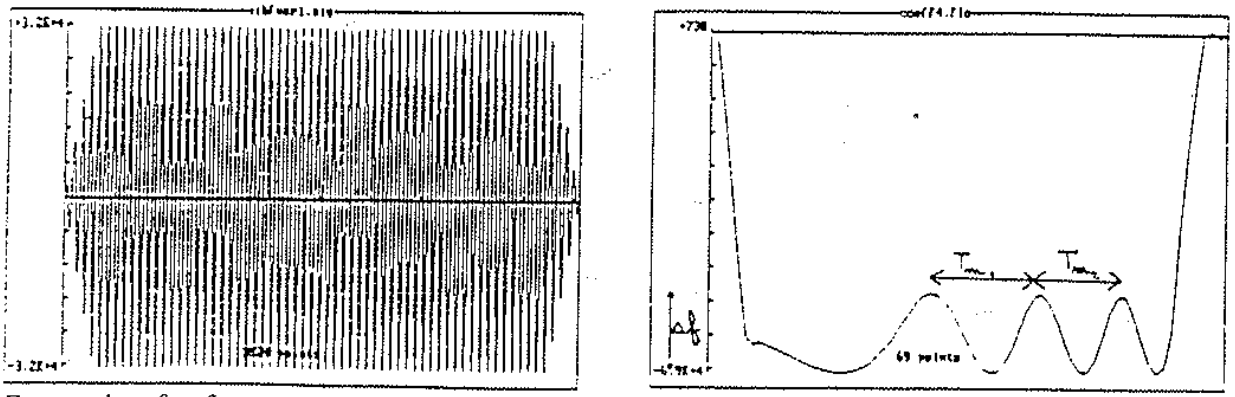

Fig. 4a,b Example of a frequency modulated signal, with a time varying modulation period T_m (a). The evolution of this modulation period appears clearly in the 4th voice wavelet coefficients (b)

Fig. 4.b, representing wavelet coefficients at a scale with a high j index, clearly reveals the time evolution of the modulation period T_m. The variation in frequency appears on the y axis.
This feature of the wavelet decomposition may be used for instance in musical signal analysis, in order to detect the existence of vibrato phenomena.

Machine-condition monitoring

Our objective is to show how an analysis by orthogonal discrete wavelets can also reveal faults in rotating machinery. We consider here a signal taken from a gear by means of an accelerometer. Fig. 5 shows an energy representation of the coefficients related to the first 4 voices in a decomposition by orthogonal wavelets. The signal section chosen corresponds to 3 gear rotation periods and has been obtained 3 days before a partial crack appeared on one of the 16 cogs. Note that the corresponding time signal is not shown here for it does not reveal relevant information.

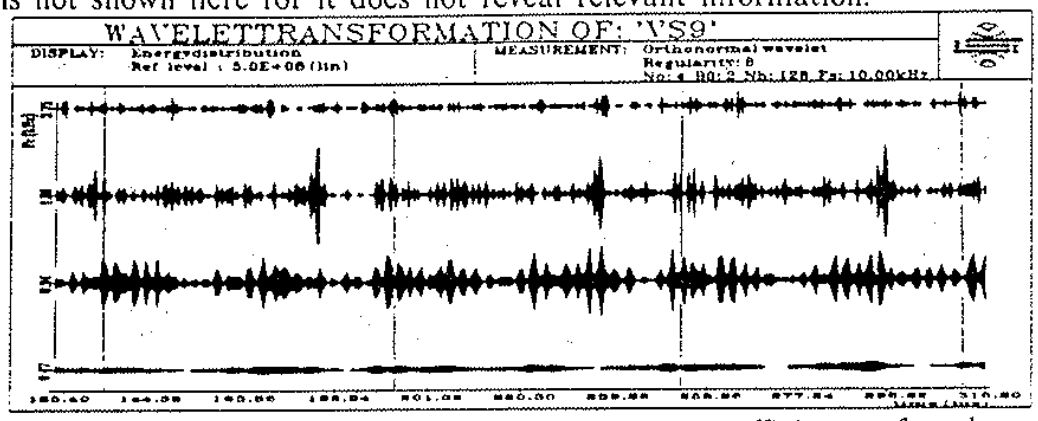

Fig. 5 Energy representation of the four first voices' wavelet coefficients, after decomposition of a rotating machine signal. The periodic peaks appearing on the second voice wavelet coefficients reveal a malfunction on one of the gear's cogs

The presence of periodic discontinuities (periods identical to that of the gear rotation) is clearly highlighted on the index 2 voice in the decomposition, and proves the existence of a fault. The rotation indicator represented by lines in Fig. 5 helps assess that this is the sign of malfunction that has led to the crack. Moreover, the high time resolution in the decomposition helps locate precisely the fault on the gear.

Conclusion

These first experiments highlight the new features associated to the Fast Wavelet Transform algorithm; it can mainly be used for applications requiring a high time resolution without the simultaneous need for an accurate frequency resolution. Note that this reversible algorithm could also be used for sub-band coding applications, which is not the purpose of this paper.

References

[1] R. Kronland-Martinet, J. Morlet, A. Grossmann, *"Analysis of sound patterns through wavelet transform"*, CNRS, Centre de physique théorique, Marseille, CPT-87, 1987

[2] Y. Meyer, *"Les ondelettes"*, Université de Paris-Dauphine, Paris, to be published.

[3] I. Daubechies, *"Orthonormal bases of compactly supported wavelets"*, AT&T Bell Labs, Murray Hill, USA, to be published.

[4] O. Rioul, *"Un algorithme rapide de décomposition discrète de signaux échantillonnés en ondelettes orthogonales"*, IGE, Ecole Nationale Supérieure des Télécommunications, Paris, to be published.

[5] C. Dorize, *"Ondelettes discrètes orthogonales et signaux acoustiques"*, Thèse Ingénieur CNAM, Ecole Nationale Supérieure des Télécommunications, Paris.