AN INTRODUCTION TO THE DISCRETE ORTHOGONAL WAVELET TRANSFORM

David Martin Simpson[1]

ABSTRACT - Wavelet transforms have recently gained considerable attention as a possible alternative (with distinct advantages) to more traditional data transformations (eg. Fourier, Cosine), especially when applied to transient data. In the present work the basic concepts of wavelet transforms will be described and attention is then focused on one of the families of wavelets, the discrete, orthogonal of finite support. The theoretical background will be developed for the direct and inverse transformations as well as the selection of the wavelets. The aim is to explain these in engineering terms, avoiding the more complicated (and for practical use often unnecessary) mathematical techniques generally employed in the description of the methods.

KEYWORDS: Signal Processing, Orthogonal Transforms, Wavelets

INTRODUCTION

In the processing and analysis of signals, data transformations have become some of the best established tools. Two main reasons for their use may be given: firstly, certain signal processing operations are more readily and more efficiently performed in the transform domain (eg. correlation or linear filtering through the Fourier Transform), and secondly the 'important' information may be compressed into fewer terms. The latter may be used in data compression for signal transmission and storage, or may form the first stage in data classification or pattern analysis (eg. the use of the power-spectrum to simplify the analysis of electroencephalographic (EEG) signals).

Through the transformations, the signals are expressed as a linear combination of the 'basis functions'. For the Fourier Series, these are sine and cosine terms of harmonically related frequencies or for the Karhunen-Loeve transform (Principal Components Analysis) the eigenvectors of the autocovariance matrix. This may be viewed as a form of modeling of the signals and may be useful, even if the model is not 'correct' in terms of the physical process generating the signal. In this way, the Discrete Fourier Transform (DFT) may be useful in the analysis of transients, which are evidently not caused by infinite oscillations which happen to combine in such a way as to give the observed function at that instant in time. However, a model that

[1] Currently Visiting Professor at the Biomedical Engineering Program, COPPE/UFRJ, C.P. 68510, 21945-970 Rio de Janeiro - RJ, Brazil.

// Trabalho recebido em 14/04/92 e aceito em 21/08/92

is more realistic may be expected to be more effective in signal analysis. Thus, for transients, a model based on functions concentrated in time (or even finite) is probably more suitable. The wavelets are examples of such functions. Their use in the processing and analysis of a wide range of signals has begun to be explored - including for biomedical signals (eg. Sendhadji et al., 1990; Nagae et al., 1991).

Through the wavelet transform, the signal is modeled as a weighted sum of the 'wavelets', which are translated and scaled (or dilated) in time (Strang, 1989). There exists an infinity of possible wavelet bases, which may be grouped into families (Rioul and Vertterli, 1991). The ones considered here are discrete (i.e. the wavelets are fully defined by a sequence of values and implementation of the transform for discrete signals is thus not only, as with other wavelets, an approximation which is necessary for computation), orthogonal (the wavelets, translated and dilated, are orthogonal to each other) and are of finite duration (finite support, $h(i)=0$, $i<0$, $i\geq M$), as developed by Daubechies (1988). This set of functions still includes a wide variety of wavelet shapes, ranging from the relatively smooth to the very irregular (depending on a choice of parameter in its definition - see for example fig. 5, 6 and 7) and may be classified according to the number of samples involved in the discrete wavelet (wavelets of two, four and six samples will be considered).

In this paper, the conditions the wavelets have to satisfy will be developed, as well as the algorithms for the forward and inverse transforms. It is aimed at using simple mathematics throughout, sufficient to understand and use the transforms but avoiding the more elaborate mathematical terminology and techniques commonly used in its derivation. Only real-valued, one dimensional, deterministic signals of finite energy will be considered.

THEORETICAL BACKGROUND

The wavelet transform of the function $f(x)$, in its general form, is given as

$$(Uf)(a,b) = \langle f,\Psi_{a,b}\rangle = |a|^{-1/2} \int_{-\infty}^{\infty} f(x)\Psi(\frac{x-b}{a})dx \qquad (1)$$

where $\langle .,. \rangle$ represents the inner product and $\Psi_{a,b}(x) = \Psi(\frac{x-b}{a})$ with the parameters a and b representing *dilation* and *translation* respectively. Since $\Psi(x)$ is selected to have a band-pass spectrum, the transformed data is effectively a band-pass filtered version of the signal (with b as the new time-axis), whose bandwidth is varied with a. By sampling the continuous two-dimensional function defined in (1), a discrete form is derived:

$$(Tf)(m,n) = c_{mn} = \langle f,\Psi_{mn}\rangle, \qquad (2)$$

with

$$\Psi_{mn}(x) = a_0^{-m/2}\,\Psi(a_0^{-m}x - nb_0) = \\ = a_0^{-m/2}\,\Psi(a_0^{-m}(x - n\,a_0^{m}\,b_0)) \qquad (3)$$

where x is still a continuous variable, a_0 and b_0 are constants and m and n integers, giving dilation and translation (corresponding to a and b in (1)). The factor $a_0^{-m/2}$ guarantees that the energy of $\Psi_{mn}(x)$ remains independent of dilation level m.

The signal is thus transformed into a series of band-pass filtered versions whose bandwidth decreases with increasing dilation of Ψ (increasing m) and hence may be sampled at a progressively lower rate. The dependence of sampling rate on dilation is evident from (3) and consistent with sampling theory.

The question immediately arises, if there exists an inverse transform, i.e. whether it is possible to reconstruct the continuous f(x) from the transform samples (Tf)(mn). For a restricted set of Ψ_{mn} (the conditions defining a 'frame') this may be carried out through an iterative process and under even tighter conditions (orthogonal Ψ_{mn}), directly. Furthermore, in the interest of minimizing the number of terms necessary to define the signal, the transform coefficients should also be independent: it has been found that with the use of frames, high redundancy of information may arise (Daubechies, 1988). Orthogonality of $\Psi(x)$ guarantees linear independence between transform components (i.e none can be calculated as a linear combination of the others). The similarity with the Fourier transform may be noted, where the sine and cosine bases are also all mutually orthogonal.

ORTHOGONAL WAVELETS

Following the discussion above, only orthogonal wavelets will be considered from now on and furthermore, let $a_0 = 2$ and $b_0 = 1$. To begin with, the functions and wavelets are continuous time, the step to the discrete time signals will follow later.

Now let the function f(x) be a weighted sum of translated versions of some low-pass function $\phi(x)$ at dilation level m (the role of the band-pass function $\psi(x)$ will become evident further along). In accordance with (3), this dilated version of $\phi(x)$ will be denoted as ϕ_{mn}. Thus,

$$f(x) = \Sigma_n\, c_{mn}\, \phi_{mn}(x), \qquad (4)$$

where $\phi(x)$ is now known as the scaling function. It is easy to confirm that indeed $c_{mn} = \langle f, \phi_{mn}\rangle$, as given in (2): with (4),

$$\langle(\Sigma_k\, c_{mk}\, \phi_{mk}), \phi_{mn}\rangle = \Sigma_k\, c_{mk}\langle\phi_{mk}, \phi_{mn}\rangle = c_{mn} \qquad (5)$$

if ϕ_{mn} and ϕ_{mk} are orthonormal, i.e.

$$<\phi_{mn}, \phi_{mk}> = \delta(n-k), \qquad (6)$$

where
$\delta(n) = 1, \; n=0,$
$= 0, \; n\neq 0.$

It then also follows that

$$<f,f> = \Sigma_n \, c_{mn}^2. \qquad (7)$$

Note the similarity to the Fourier Series in the forward and inverse transformation and in energy preservation (equivalent to Parseval's theorem).

Now let V_0 denote the space of functions, whose basis is $\phi(x)$ at dilation m=0 ($f(x) = \Sigma_n \, c_{0n} \, \phi_{0n}(x)$). Furthermore, if $f(x) \in V_0$, let the slower, lower frequency (more dilated) signal $f(x/2) \in V_1$, where V_1 forms a subspace of V_0. Continuing this process, a set of such subspaces can be constructed with signals of lower and lower bandwidth, corresponding to the successively more dilated signals. This may be illustrated as a system of concentric circles (fig. 1): the subspace with higher values of m lie closer to the centre and represent a more restricted set of signals. A signal in V_1 will also be in V_0, but not necessarily vice versa.

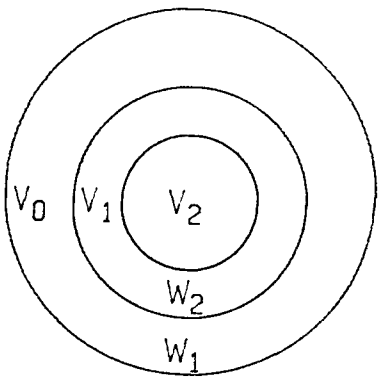

Fig. 1. Function spaces for orthogonal Wavelets. The spaces V_i are represented by the areas of the circles, and W_i by the rings between them.

As an example of such a system of spaces, V_0 may define the set of functions which are piecewise-constant (in time) with step-widths of unity. V_1 would then correspond to the functions

with step-width of 2 units, V_2 with 4 units, etc. This is illustrated in fig. 2 by a sine-wave of varying frequency, approximated (in spaces V_1, V_2, V_3 and V_4) by linear segments whose width doubles with every step of m. The function at m=2 is contained in V_1 but not vice versa. For this set of spaces, $\phi(x) = 1$, $0 \leq x < 1$, and 0 otherwise. This leads to the simplest of orthogonal wavelets with finite support and is equivalent to the well known Haar transform (Mallat, 1989).

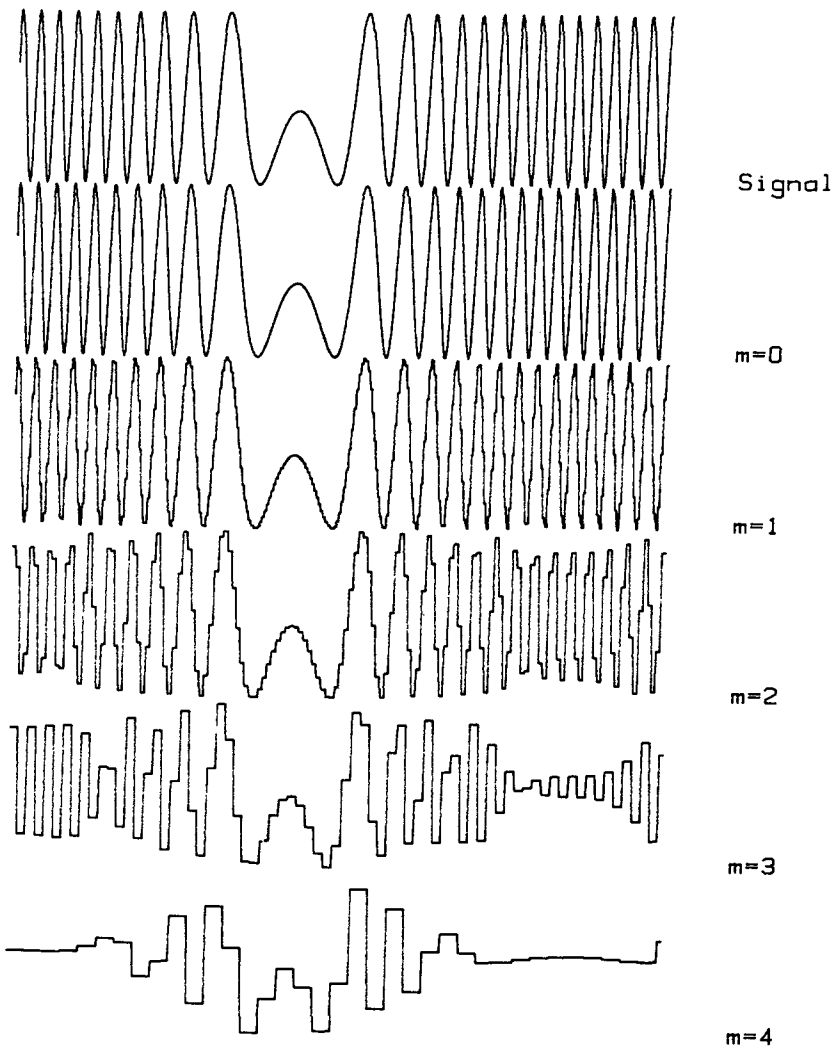

Fig. 2. The original signal and wavelet reconstructions (levels m=0 to 4) using the Haar basis.

A signal which is an element of V_0 but not V_m, can therefore only be approximated by (4). It is interesting to note - and very convenient - that this approximation is optimal in the mean-square sense, as can easily be demonstrated: let the mean-square-error be

$$e = \int_{-\infty}^{\infty} [(f(x) - \Sigma_n\ c_{mn}\ \phi_{mn}(x)]^2\ dx,$$

and minimizing with respect to a c_{mk},

$$\partial e/\partial c_{mk} = 0 = \int_{-\infty}^{\infty} [(f(x) - \Sigma_n\ c_{mn}\ \phi_{mn}(x)]\phi_{mk}(x)dx = \langle f,\phi_{mk}\rangle - c_{mk}$$

since the ϕ_{mn} are orthonormal. Thus, for the minimum mean-square error reconstruction, the c_{mk} are those given by (2).

A high value of m thus leads to a coarse approximation to the function and with a decrease in m, successively better approximations to the signal are achieved, revealing more signal detail until, at m = 0, this signal (if it is in V_0) is reconstructed perfectly. This successive approximation approach is therefore known as Multiresolution Analysis (Mallat, 1989).

The reconstruction of f(x) at dilation level m-1 (to be denoted by f_{m-1}) is given as

$f_{m-1} = \Sigma_n\ c_{(m-1)n}\ \phi_{(m-1)n}$

but may also be given in terms of the lower resolution f_m and a difference term

$$f_{m-1} = \Sigma_n\ c_{mn}\ \phi_{mn} + \Sigma_k\ d_{mk}\ \psi_{mk}\ . \qquad (8)$$

It is the function $\psi(x)$ which is known as the primary wavelet, and ψ_{mk} is obtained from it in accordance with (3).

By requiring orthonormality between ϕ_{mn} and ψ_{mk} and also ψ_{mn} and ψ_{mk} (in addition to that of the ϕ_{mn} already discussed):

$$\langle\phi_{mn},\psi_{mk}\rangle = 0 \qquad (9a)$$
$$\langle\psi_{mn},\psi_{mk}\rangle = \delta(n-k) \qquad (9b)$$
$$\langle\phi_{mn},\phi_{mk}\rangle = \delta(n-k) \qquad (9c)$$

the calculation of the parameters c_{mn} and d_{mn} becomes simply

$$c_{mn} = \langle f_{m-1},\phi_{mn}\rangle = \langle f_m,\phi_{mn}\rangle = \langle f,\phi_{mn}\rangle, \qquad (10a)$$
$$d_{mn} = \langle f_{m-1},\psi_{mn}\rangle = \langle f,\psi_{mn}\rangle. \qquad (10b)$$

This follows by orthogonality from (4) and (8), and the last equality for c_{mn} corresponds to the original definition in (2); d_{mn} may be defined in a similar way, considering that f_{m-1} is the

best mean-square approximation of $f(x)$ in V_{m-1}.

By extension of (8),

$$f_{m-1} = \Sigma_n c_{mn}\, \phi_{mn} + \Sigma_k\, d_{mk}\, \psi_{mk} + \Sigma_k\, d_{(m-1)k}\, \psi_{(m-1)k} + \\ + \ldots + \Sigma_k\, d_{(m-i+1)k}\, \psi_{(m-i+1)k}, \qquad i = 1,2 \ldots m. \tag{11}$$

The signal f_0 is therefore given through a low-resolution term f_m (as determined by c_{mn} and ϕ_{mn}) and m difference terms (given by the d_{jk} and ψ_{jk}) to achieve progressively higher resolution. In terms of the functional spaces illustrated in fig. 1, the rings between V_j and V_{j-1} are the spaces W_j, whose basis is $\psi(x)$ at dilation j. Starting from the restricted space V_m, the successive addition of the W_i builds up the space V_0.

Returning to the earlier example, V_m is the set of piecewise constant functions whose step-width is 2^m. As shown above,

$$\phi(x) = 1, \quad 0 \le x < 1,$$
$$= 0, \quad \text{otherwise,}$$

and, as will be clear later,

$$\psi(x) = 1, \quad 0 \le x < 1/2$$
$$= -1, \quad 1/2 \le x < 1$$
$$= 0, \quad \text{otherwise.}$$

The similarity with the well known Haar transform is now evident. The function f_m is therefore the approximation of $f(x)$ with a step-width of 2^m and each of the remaining terms in (11) adds detail by reducing the step-width by one half. At each stage the best (in the mean-square sense) approximation to the function $f(x)$ with piecewise constant segments is obtained. If the wavelet transform is used in data-transmission, first a broad outline of the signal would be obtained through the c_{mn} which would be sent first, and then progressively the resolution is improved as further detail is added with successive levels of d_{mn} (see fig. 2). This scheme has the great advantage of immediately giving a rough outline of the signal, and transmission may be interrupted when the desired resolution has been achieved. In other schemes, complete transmission of (compressed) data is required to give intelligible results.

Another possible interpretation of (11) is as a set of filters, such that for example, c_{mn} defines the signal in the bandwidth up to 1 Hz, d_{mn} in 1-2 Hz, $d_{(m-1)n}$ in 2-4 Hz etc., thus progressively increasing the bandwidth of the signal. It must be pointed out that this comparison is inaccurate in the sense that the filters involved in the wavelet transform are of course not ideal low or band-pass, but each term does give additional high-frequency detail, and (ψ being time-scaled by a factor of two) has twice the bandwidth of the preceding one. Indeed, ideal band-pass filters would not lead to finite-length wavelets, as desired.

If the function $f(x) \in V_m$, then $f(x) = f_{m-1} = f_m$ and $d_{mn} = 0$, as follows from (8). Therefore

$$d_{mn} = 0 = \langle f_{m-1}, \psi_{mn} \rangle = \langle f_m, \psi_{mn} \rangle = \\ = \Sigma_k \, c_{(m+1)k} \langle \phi_{(m+1)k}, \psi_{mn} \rangle + \Sigma_k \, d_{(m+1)k} \langle \psi_{(m+1)k}, \psi_{mn} \rangle \quad (12)$$

and hence $\langle \phi_{(m+1)k}, \psi_{mn} \rangle = 0$ and $\langle \psi_{(m+1)k}, \psi_{mn} \rangle = 0$, and by extension, through (11),

$$\langle \phi_{(m+j)k}, \psi_{mn} \rangle = 0, \quad j \geq 0 \quad (13a)$$
$$\langle \psi_{(m+j)k}, \psi_{mn} \rangle = 0, \quad j \neq 0 \quad (13b)$$

confirming the orthogonality of the spaces W_i, illustrated in fig. 1. It can readily be confirmed that the above holds true for the case of the Haar wavelet.

Considering that $\phi(x) \in V_0 \subset V_{-1}$ (as follows from the definition) evidently $\phi(x)$ can be transformed (as any other function) and given, at transform level m = -1 as

$$\phi(x) = \sqrt{2} \, \Sigma_k \, c_k \, \phi(2x-k), \quad (14a)$$

(see (3) and (4)), where

$$c_k = \sqrt{2} \langle \phi(x), \phi(2x-k) \rangle. \quad (14b)$$

Similarly, $\psi(x)$ may be transformed (since $\psi(x) \in W_0 \subset V_{-1}$) and in order to satisfy the orthogonality conditions given above, it may be given in terms of the same coefficients c_k:

$$\psi(x) = \Sigma_k \, \sqrt{2} \, (-1)^k \, c_{1-k} \, \phi(2x-k). \quad (15)$$

This guarantees that $\langle \phi(x), \psi(x-n) \rangle = 0$, since

$\langle \phi(x), \psi(x-n) \rangle = \langle \Sigma_k \, c_k \phi(2x-k), \, \Sigma_j \, (-1)^j \, c_{1-j} \, \phi(2x-2n-j) \rangle$

and due to the orthogonality of $\phi(x)$, all terms are zero except if $k = 2n+j$, such that

$$\langle \phi(x), \psi(x-n) \rangle = 2 \, \Sigma_k \, c_k \, (-1)^{-2n+k} \, c_{1-(k-2n)} = \\ = 2 \, \Sigma_{(k \text{ even})} \, [(-1)^{2n-k} \, c_k \, c_{2n-k+1} + \\ + (-1)^{2n-(2n-k+1)} \, c_{2n-k+1} \, c_{2n-(2n-k+1)+1}] \\ = 2 \, \Sigma_{(k \text{ even})} \, (c_k \, c_{2n-k+1} - c_{2n-k+1} \, c_k) = 0.$$

It may be pointed out that the above sum may be taken to the limits of $k=\pm\infty$, (thus avoiding 'edge-effects'), though there are only a finite number of non-zero c_j.

Since $\psi(x/2) = \sqrt{2} \, \Sigma_k \, (-1)^k \, c_{1-k} \, \phi(x-k)$ and $\langle \phi(x+k), \psi(x) \rangle = 0$, it follows from the above that also $\langle \psi(x), \psi(x/2) \rangle = 0$ and clearly $\langle \psi(2x-n), \psi(x) \rangle = 0$, conforming to the orthogonality conditions in (13). Now, $\phi(x) = \sqrt{2} \, \Sigma_k \, c_k \, \phi(2x-k) = \sqrt{2} \, \Sigma_k \, c_k \, (\sqrt{2} \, \Sigma_n \, c_n \, \phi(4x-2k-n))$ and consequently $\langle \phi(x), \psi(2x-n) \rangle = 0$ and $\langle \phi(x), \psi(4x-n) \rangle = 0$. By extending this procedure, it may be shown that the remaining

conditions in (13) are also satisfied. It should be pointed out however, that (15) is not the only possible solution to guarantee orthogonality.

For the case of the Haar function,

$\phi(x) = 1, \quad 0 \le x < 1,$
$= 0, \text{ otherwise,}$

clearly gives $c_0 = c_1 = 1/\sqrt{2}$ (see (14)) and from (15),

$\psi(x) = 1, \quad 0 \le x < 1/2$
$= -1, \quad 1/2 \le x < 1,$
$= 0, \quad \text{otherwise.}$

In the discussion so far, the conditions for suitable scaling functions ($\phi(x)$) and primary wavelets ($\psi(x)$) have been given. The Haar transform is the simplest example of such a system, but the question remains whether others may be found, or indeed, if they exist. Following the above paragraph, it may however be stated, that if a reasonably well behaved $\phi(x)$ is discovered (which luckily is possible - see below), then the corresponding $\psi(x)$ also exists. In the following, the coefficients c_k will become the main focus of attention, since it is they which are used in the discrete forward and inverse transforms (and for this reason the transform is called discrete) and furthermore, $\phi(x)$ and $\psi(x)$ may be determined from them. This will become clear by considering the transform algorithm - which will be developed now.

DECOMPOSITION AND RECONSTRUCTION ALGORITHM

At dilation level $m = 0$, $c_{0n} = \langle f, \phi_{0n} \rangle$ and $f_0 = \Sigma_n \, c_{0n}\phi_{0n}$; at the next level, $m = 1$, $c_{1k} = \langle f, \phi_{1k} \rangle = \langle f_0, \phi_{1k} \rangle = \Sigma_n c_{0n} \langle \phi_{0n}, \phi_{1k} \rangle$. The latter inner product may be expanded as

$$\langle \phi_{0n}, \phi_{1k} \rangle = 2^{-1/2} \int_{-\infty}^{\infty} \phi(x-n)\phi(x/2-k) \, dx =$$

$$= 2^{1/2} \int_{-\infty}^{\infty} \phi(2x-(n-2k))\phi(x) \, dx$$

$$= h(n-2k) \qquad (16)$$

where the $h(n-2k)$ are the coefficients c_{n-2k} of (14). It then follows that

$$c_{1k} = \Sigma_n \, c_{0n} \, h(n-2k). \qquad (17)$$

This result is at the basis of the transform algorithm: the transform coefficients at dilation level m=1 may be calculated by discretely transforming those of level m=0, without the need to again turn to the original continuous data. Similarly,

$$d_{1k} = \langle f, \psi_{1k} \rangle = \Sigma_n \, c_{0n} \langle \phi_{0n}, \psi_{1k} \rangle = \Sigma_n \, c_{0n} \, g(n-2k) \qquad (18)$$

where

$$g(n-2k) = \langle \phi_{0n}, \psi_{1k} \rangle = 2^{-1/2} \int_{-\infty}^{\infty} \phi(x-n) \psi(x/2-k) \, dx. \qquad (19)$$

This can readily be extended to the following higher dilation levels where it is clear that the same h(n) and g(n) apply: considering (16),

$$\langle \phi_{mn}, \phi_{(m+1)k} \rangle = 2^{-m/2} . 2^{-(m+1)/2} \int_{-\infty}^{\infty} \phi(2^{-m}x-n) \phi(2^{-(m+1)}x-k) \, dx$$

$$= 2^{-1/2} \int_{-\infty}^{\infty} \phi(x-n) \phi(x/2-k) \, dx =$$

$$= \langle \phi_{0n}, \phi_{1k} \rangle = h(n-2k). \qquad (20)$$

and equivalently

$$\langle \phi_{mn}, \psi_{(m+1)k} \rangle = \langle \phi_{0n}, \psi_{1k} \rangle = g(n-2k). \qquad (21)$$

The algorithm for the discrete wavelet transformation is therefore given by

$$c_{mk} = \Sigma_n \, c_{(m-1)n} \, h(n-2k) \qquad (22a)$$

$$d_{mk} = \Sigma_n \, c_{(m-1)n} \, g(n-2k). \qquad (22b)$$

Thus the continuous data is only used at the lowest dilation level (here called m=0) to obtain c_{0n}. From then on, all transformations are carried out on the transform coefficients of the previous level, the results being identical to those that could be obtained by continuing to apply to continuous transform to the original function. The importance of this result in the practical use of the transform is self-evident.

The reconstruction (inverse transform) algorithm may be found in a similar way: based on (8)
$f_m = \Sigma_k \, c_{(m+1)k} \, \phi_{(m+1)k} + \Sigma_k \, d_{(m+1)k} \, \psi_{(m+1)k}$
and, according to (10), $c_{mn} = \langle f_m, \phi_{mn} \rangle$. Hence

$$c_{mn} = \Sigma_k \, c_{(m+1)k} \langle \phi_{mn}, \phi_{(m+1)k} \rangle + \Sigma_k \, d_{(m+1)k} \langle \phi_{mn}, \psi_{(m+1)k} \rangle =$$
$$= \Sigma_k \, c_{(m+1)k} \, h(n-2k) + \Sigma_k \, d_{(m+1)k} \, g(n-2k) \qquad (23)$$

The coefficients c_{mn} may be thus obtained from the $c_{(m+1)n}$ and $d_{(m+1)n}$ of the next higher dilation level (m+1), by discretely filtering with the functions h(n-2k) and g(n-2k) respectively. It should be noted that in the inverse transform, the summation is with respect to k, whereas in the forward transform it was over n.

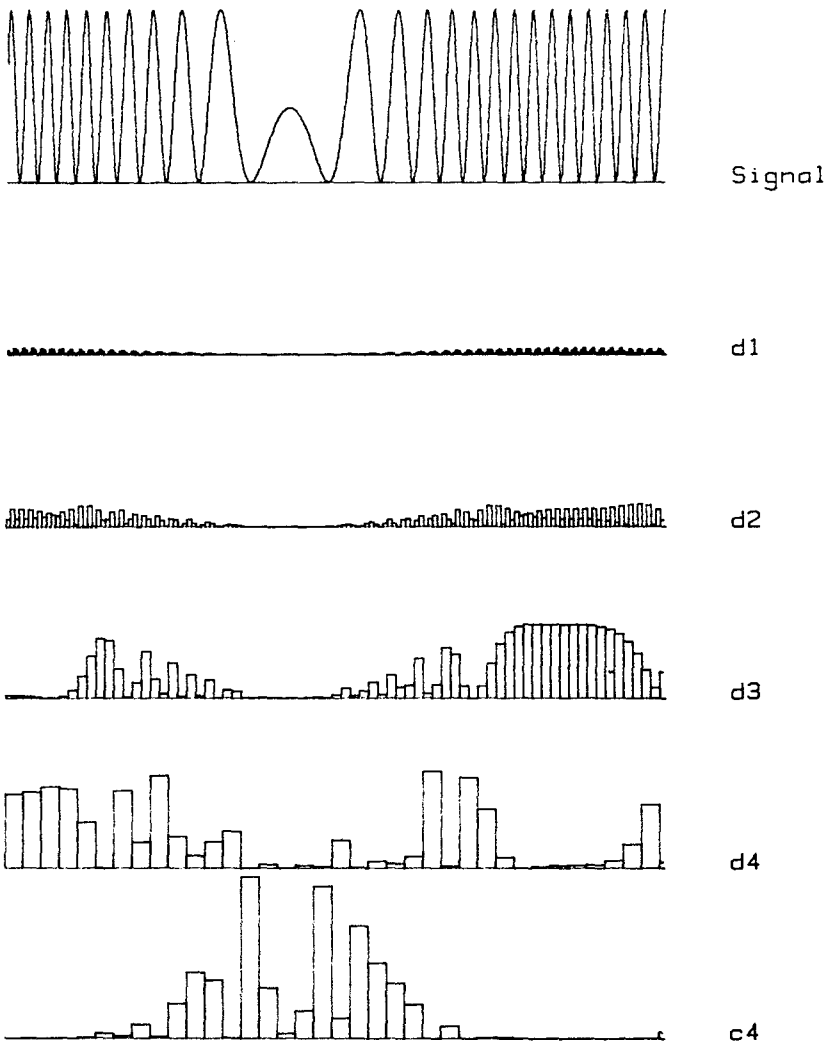

Fig. 3. The coefficients c and d (squared) corresponding to fig 2, with dilation levels m=1 to 4. To improve visualization, the height of the bars is divided by 2^m. The increase in the width of the bars for the heigher levels of m indicates the corresponding reduction in sampling rate.

The results of forward and inverse transformation are illustrated in fig. 2, based on the example of the Haar function.

As already stated, fig. 2 shows the signal and the reconstructed versions at different levels of m. Fig. 3. shows the coefficients c_{4n} used in the reconstruction at level m=4 (f_4), and the d_{mn}, $m \leq 4$, successively incorporated to obtain the remaining higher resolution signals.

Fig. 4 illustrates the transform álgorithm. Here, starting with the N samples of c_{0n}, N/2 samples of c_{1k} are obtained, as well as N/2 of d_{1k}. The reduction in the number of samples in c and d is the result of the translation of h(n-2k) and g(n-2k) by 2k. This is the practical result of the sampling frequency being halved when the dilation is doubled, as was indicated earlier. Fig. 3 also illustrates this point. It should be underlined, that the sampling interval at level m+1 is twice that at level m, or equivalently, the sample c_{mn} is at the same point on the time-axis as $c_{(m+1)n/2}$. At the next stage of the transform, the N/2 samples of c_{1k} result in N/4 c_{2k} and N/4 d_{2k}. This process may be continued until finally only one value each of c and d is obtained.

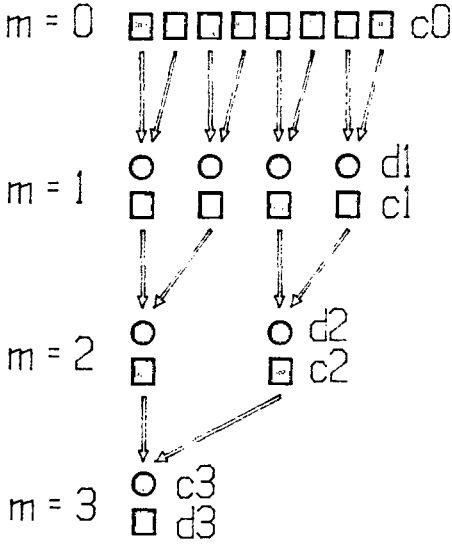

Fig. 4. Schematic Diagram of the Wavelet transform.

For the reconstruction of the signal (inverse transform), the c's and d's at dilation level m are combined to find the $c_{(m-1)k}$, and then, including the $d_{(m-1)k}$, the $c_{(m-2)k}$ are obtained. This process may be continued, until finally arriving at the c_{0k}. Thus the reconstruction algorithm requires the c's and d's of the ultimate dilation level (m), together with the d's of all the lower levels m-i, $0 < i < m$. The samples of the signal at m=0 are then identical to those of the original signal

(This was confirmed for fig. 2). The lower resolution versions (m>0) in fig. 2 were reconstructed by neglecting successively more of the d's: thus at m=1, the d_{1n} were set to zero and at m=2, $d_{1n} = d_{2n} = 0$, etc.

It may easily be verified by considering fig. 4, that the number of coefficients necessary for complete signal reconstruction is equal to N, the number of samples in c_{0n}. This holds true, whether or not the transformation is carried out up to the ultimate dilation level possible, i.e when there is only one coefficient each of c and d (at m=3 in fig. 4). Each of the transform coefficients (c_{mk} and d_{mk}) is formed through a weighted sum of the samples c_{0k} and this may therefore be viewed as a set of simultaneous linear equations. Furthermore, they are linearly independent since the wavelets are orthogonal (and therefore may not be obtained as a linear sum of each other). In the inverse transform the equations are solved and evidently N equations (i.e. N transform coefficients) are required in order to solve for the N samples c_{0k}. More than N coefficients would lead to overdetermination and redundancy of information (or worse, contradictory information), less would not allow a solution. It should however be pointed out that when the wavelets involve more than two samples (or the signal involves an odd number of samples), the signal segment has to be zero-padded and hence the number of transform coefficients is slightly larger than N.

In order to reconstruct the continuous signal f(x), the continuous form of the transform (2) would be applied to the c_{0k}. The signal thus reconstructed is accurate if f(x) forms part of V_0, else it will be f_0, the best mean-square approximation through $\phi(x)$ at dilation m=0. However, the level m=0 is an arbitrary choice, and c_{-1k} could be calculated by continuing to apply the discrete inverse transform with the assumption that $f(x) \in V_0$, thus setting $d_{0k}=0$. In such a way discrete versions of the signal with higher and higher resolution (in terms of sample spacing, but without additional information content) can be reconstructed. These discrete functions converge to the continuous version f_0, as the corresponding ϕ_{mn} are reduced in time-scale, becoming impulses in the limit.

The Haar function may be used to illustrate the point: the discrete versions (c_{mn}, m<0) place more samples on each step, but the basic step-width remains constant. The continuous signal may be obtained at any stage by interpolation with the rectangular pulse (boxcar) function, whose size decreases with a progressive reduction in dilation level. The equivalent operation for the discrete Fourier transform is obtained when the spectrum is zero-padded. This does not alter the signal spectrum (zero power above the Nyquist frequency is part of the basic requirement for aliasing not to occur) or increase the information content, but does increase the number of samples obtained on inverse transformation. Low-pass filtering either the original sampled signal, or that obtained after spectral zero-padding, would lead

to exactly the same continuous function.

The question arises, whether the c_{0n} can be considered samples of the continuous signal f(x). The answer is clearly no, not even if the signal forms part of the space V_0 and therefore is perfectly defined by the c_{0n}. This is evident from the reconstruction formula, since $\phi(x)$ does not have zeros at the location of the remaining samples, such that generally $f(n) = \Sigma_k c_{0k}\phi(n-k) \neq c_{0n}$. In contrast, when dealing with the samples of band-limted signals, the equivalent to $\phi(x)$ is the sinc function $\sin(x)/x$ which does pass through zero at the location of adjacent samples.

In most applications of signal processing, only sampled signals, and not the original continuous versions, are available. These samples might however be treated as c_{0n} and form the basis for the application of the discrete wavelet transform. This is permissible, since the independence of the c_{0n} leads to no inherent restrictions on their value (other than finite energy, which in practice is of little concern). In the reconstruction of the continuous signals f(x) from these 'samples' it would of course not be justified to interpolate based on $\phi(x)$. Rather, this should be carried out through the usual low-pass filter (or the equivalent in the time-domain, the sinc-function).

It is interesting to observe that the signals in V_0 are not band-limited, since the $\phi(x)$ of finite support have infinite bandwidth. In spite of this the signals are defined through a set of coefficients c_{0n} at a finite sampling rate, when according to the sampling theorem, the sampling rate would have to be infinite. The explanation is of course that we are now not dealing with samples of the signal.

CONDITIONS FOR THE DISCRETE WAVELET FUNCTIONS

Above the wavelet transform was developed based on the continuous signals. It was shown that once the transition to the discrete version was made, all further transformations may be carried out in the discrete domain. It was then seen that in practice, the discrete transformation may start with the sampled signal, such that all computational work involves only h(n) and g(n) and no longer $\phi(x)$ and $\psi(x)$. The conditions for the admissibility of $\phi(x)$ have been given and it was seen that when such a $\phi(x)$ is chosen, $\psi(x)$ follows ((14b) and (15)) as do h(n) (16) and g(n) (19). The choice of wavelet may in fact commence with the selection of h(n) (and g(n) by implication) - according to conditions similar to those for the continuous functions.

Clearly for any reasonable transform $\Sigma\, h^2(n) < \infty$ and hence $\Sigma\, g^2(n) < \infty$. Through forward and then inverse transformation, the orthogonality conditions may be derived:

$c_{mj} = \Sigma_k\ h(j-2k)\ \Sigma_n\ h(n-2k)\ c_{mn} + \Sigma_k\ g(j-2k)\ \Sigma_n\ g(n-2k)\ c_{mn} =$
$= \Sigma_n\ [\Sigma_k\ h(j-2k)\ h(n-2k) + \Sigma_k\ g(j-2k)\ g(n-2k)]\ c_{mn}$
and therefore

$$\Sigma_k\ h(j-2k)\ h(n-2k) + \Sigma_k\ g(j-2k)\ g(n-2k) = \delta(j-n) \qquad (24)$$

By inverse and then forward transformation
$c_{mj} = \Sigma_n\ h(n-2j)\ \Sigma_k\ h(n-2k)\ c_{mk} + \Sigma_n\ h(n-2j)\ \Sigma_k\ g(n-2k)\ d_{rk}$
it follows that

$$\Sigma_n\ h(n-2k)h(n-2j) = \delta(j-k) \qquad (25a)$$
$$\Sigma_n\ g(n-2k)h(n-2j) = 0, \text{ all } j,k \qquad (25b)$$

underlining the equivalence between the discrete and continuous versions of the wavelet basis. Energy preservation

$$\Sigma_n\ c^2_{mn} = \Sigma_n\ c^2_{(m+1)n} + \Sigma_n\ d^2_{(m+1)n} \qquad (26)$$

may thus be confirmed.

Consider the boxcar function $f(x)=a$, $-N \le x < N$, 0 otherwise, then near x=0,

$$c_{mn} = a \int_{-\infty}^{\infty} \phi_{mn} dx \quad \text{and} \quad c_{(m-1)n} = a \int_{-\infty}^{\infty} \phi_{(m-1)n} dx = 1/\sqrt{2}\ c_{mn},$$

as follows from (3) and $a_0 = 2$. Considering the discrete version of the wavelet transform, $c_{mk} = c_{(m-1)n}\ \Sigma_n h(n-2k)$. Thus in the region near the centre of f(x) (and hence generally),

$$\Sigma_n h(n) = \sqrt{2}. \qquad (27)$$

From energy preservation (26), it is clear that $\Sigma_n\ c^2_{(m-1)n} = \Sigma_n\ c^2_{mn} + \Sigma_n\ d^2_{mn}$. Now taking N to its limits such that all c_{mn} are equal and considering that $c_{mn} = \sqrt{2}\ c_{(m-1)n}$ and that the sampling rate at dilation level m-1 is twice as high as at level m, evidently $d_{mn} = 0$ and consequently

$$\Sigma_n\ g(n) = 0. \qquad (28)$$

According to (15), (16) and (19),

$$g(n) = (-1)^n\ h(1-n) \qquad (29)$$

which guarantees the orthonormality of g(n) and its orthogonality to h(n). It should be stressed again that this is not a unique solution (eg. time shifted versions will also work). It may easily be verified that under these conditions (24) is also satisfied. As a result of (28) and (29), $\Sigma_n\ h(2n) = \Sigma_n\ h(2n+1) = 1/\sqrt{2}$, which furthermore guarantees that a smooth $c_{(m-1)n}$ leads to

a smooth c_{mn}.

The derivation of the discrete orthogonal wavelet transform is thus complete. A Pascal implementation of the forward and inverse transform is given in the Appendix.

EXAMPLES OF DISCRETE ORTHOGONAL WAVELETS OF FINITE SUPPORT

Having given the necessary and sufficient conditions on h(n) and g(n) for the discrete orthogonal wavelet transform, examples of these functions can be given. These will be restricted to those of finite support and in particular those of length 2 samples (the Haar function already cited), and the families of length 4 an 6 samples respectively. There does not appear to exist any general theory on how to obtain these discrete functions, but it may readily be verified that the ones given below satisfy the necessary conditions.

For the Haar function:

$$h(0) = 1/\sqrt{2}, \quad h(1) = 1/\sqrt{2} \tag{30a}$$
$$g(0) = 1/\sqrt{2} \quad g(1) = -1/\sqrt{2}. \tag{30b}$$

The admissibility is confirmed as $\Sigma_n h(n) = \sqrt{2}$, $\Sigma_n h^2(n) = 1$ and $\Sigma_n h(n)h(n-2k) = 0$, $k \neq 0$. These values are in accordance with (16) and (19), and the continuous version of the Haar transform given above.

For the wavelets of length 4 samples,

$$h(0) = 1/\sqrt{2}\ \nu(\nu-1)/(\nu^2+1) \tag{31a}$$
$$h(1) = 1/\sqrt{2}\ (1-\nu)/(\nu^2+1) \tag{31b}$$
$$h(2) = 1/\sqrt{2}\ (\nu+1)/(\nu^2+1) \tag{31c}$$
$$h(3) = 1/\sqrt{2}\ \nu(\nu+1)/(\nu^2+1) \tag{31d}$$

all other samples being zero. There exists thus a family of wavelets, and their shape (i.e that of $\phi(x)$) varies quite radically with the choice of the parameter ν - as shown in fig. 5 and 6. It is easily confirmed that the sum and sum-square values of the samples satisfy the conditions given above. The samples g(n) may be obtained as

$$g(n) = (-1)^n\, h(3-n) \tag{32}$$

which guarantees the orthogonality of h(n) and g(n) and conveniently locates both functions in the same position on the time-axis.

The discrete orthogonal wavelet of 6 samples will be defined next. In this case, there are 6 unknowns (the h(n)), 4 restrictions (the mean and mean-square values, and two for orthogonality, see below). The function is therefore defined

through the choice of two parameters (α and β):

$$
\begin{aligned}
h(0) &= (1/4\sqrt{2})[(1+\cos\alpha+\sin\alpha)(1-\cos\beta-\sin\beta)+2.\sin\beta.\cos\alpha)] \\
h(1) &= (1/4\sqrt{2})[(1-\cos\alpha+\sin\alpha)(1+\cos\beta-\sin\beta)-2.\sin\beta.\cos\alpha)] \\
h(2) &= (1/2\sqrt{2})[1+\cos(\alpha-\beta)+\sin(\alpha-\beta)] \\
h(3) &= (1/2\sqrt{2})[1+\cos(\alpha-\beta)-\sin(\alpha-\beta)] \\
h(4) &= 1/\sqrt{2} - h(0) - h(2) \\
h(5) &= 1/\sqrt{2} - h(1) - h(3)
\end{aligned} \tag{33}
$$

and

$$
g(n) = (-1)^n h(5-n), \tag{34}
$$

again locating h(n) and g(n) in the same positions on the time-axis. The orthogonality condition on h(n) now leads to two equations:

$h(0)h(2)+h(1)h(3)+h(2)h(4)+h(3)h(5) = 0$

and

$h(0)h(4) + h(1)h(5) = 0.$

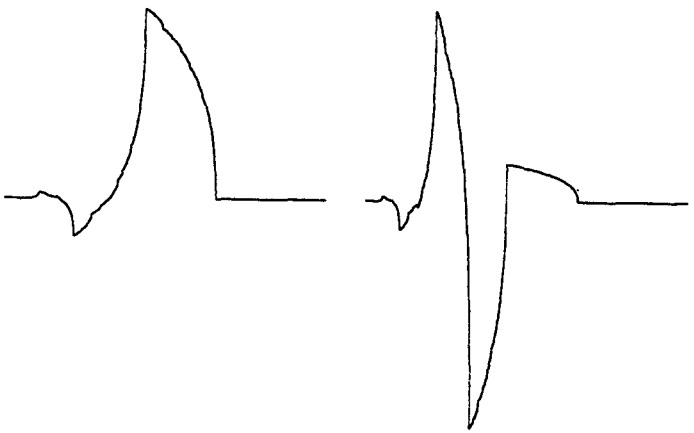

Fig. 5. $\phi(x)$ and $\psi(x)$ for the wavelet with 4 samples and ν=0.67.

It may be noted that if $\alpha = \beta$, the Haar function is obtained and if $\sin(\alpha-\beta) = (1-\nu^2)/(1+\nu^2)$ the wavelet with 4 samples (31) results, with $h(4) = h(5) = 0$, $h(1) = 1/\sqrt{2} - h(3)$ and $h(2) = 1/\sqrt{2} - h(0)$.

Fig. 7 shows an example of a wavelet of six samples - one that is relatively smooth.

FINDING $\phi(x)$ AND $\psi(x)$ FROM h(n)

Having defined h(i), the continuous functions $\phi(x)$ and $\psi(x)$ may be obtained, as has already been indicated. Let c_{0n} be the transform coefficients at dilation level m = 0 of the function $\phi(x)$. It is clear that due to the orthonormality of $\phi(x)$, and since $\phi(x)$ is in the space V_0 (by definition)

$c_{0n} = 1$, n=0,

$= 0$, otherwise.

It also follows that for $m < 0$, $d_{mn} = 0$. Thus c_{mn}, $m < 0$ can be reconstructed by first applying the reconstruction formula (23) to the impulse at level m=0 ($c_{0n} = \delta(n)$), and then recursively to the result, setting all d_{mn}=0. As discussed above, the c_{mn} give a discrete approximation to the continuous function, which improves as m is decreased and ϕ_{mn} approaches the impulse function. Thus the shape of $\phi(x)$ progressively emerges with repeated application of the reconstruction algorithm. It may be shown that for h(n) of length M samples, the $\phi(x)$ converges to a length of M-1, if it is considered that with each decrease in m, the distance between samples is halved.

Fig. 6. $\phi(x)$ and $\psi(x)$ for the wavelet with 4 samples and ν=1.5.

The function $\psi(x)$ may be obtained in a similar way. However, now $c_{0n} = 0$, and $d_{0n} = \delta(n)$. Thus c_{-1n} is obtained from (23) by setting the first sum to zero. Now $\psi(x) \in W_0$ and thus in V_{-1}, so $d_{mn} = 0$, $m < 0$. Thus $\psi(x)$ can be approximated by recursively applying the reconstruction algorithm to the c_{mn}. The difference between the reconstruction of $\phi(x)$ and $\psi(x)$ therefore lies only in the calculation of c_{-1n}, where in the former case the $c_{0n} = \delta(n)$, $d_{0n} = 0$, and in the latter $d_{mn} = \delta(n)$, $c_{0n} = 0$. From then on the procedures are identical, both only employing the h(n). It may be noted that the magnitude of the $\phi(x)$ and $\psi(x)$ thus approximated decrease with each decreasing level of m (and increase in number of samples), as follows from

energy conservation (26). In order to maintain the amplitudes over successive levels of m, the results should be multiplied by $\sqrt{2}$ at each iteration. It is through this algorithm that fig. 5, 6 and 7 were obtained (iterating 7 times). Fig. 8 shows the reconstruction of a signal using the wavelet shown in fig. 5.

Fig. 7. $\phi(x)$ and $\psi(x)$ for the wavelet of with 6 samples and $\alpha=1.15$, $\beta=0.4$.

FINAL COMMENTS

In the above, conditions for the admissibility of discrete orthogonal wavelets have been derived and some such families given. However, it is found that many of these $\phi(x)$ (and hence $\psi(x)$) do not converge to 'nice' functions (see for example fig. 6), but are very spiky with a noise-like appearance. These would therefore probably not form a convenient basis for signal modeling and data compression through the wavelet transform (see the rather irregular reconstructions for m=3 and 4 in Fig. 8). The so-called regularity conditions were derived by Daubechies (1988) to select the more 'reasonable' functions out of the set of permissible ones. Continuity criteria can be found by considering the Fourier domain. These are beyond the scope of the present work.

It should be pointed out that wavelet transforms have certain disadvantages: one of particular interest in the present application to the analysis of Visual Evoked Potentials (VEPs), is that they do not hold delay (latency) information in a very obvious way. In this respect the DFT has advantages, since delay simply results in a linear phase-trend, whose gradient is proportional to the delay. With the wavelets here presented, delay may lead to a rather complex redistribution of energy

between the different transform levels. It may thus be necessary to pre-process (align) the signals in order not to obtain confusing results.

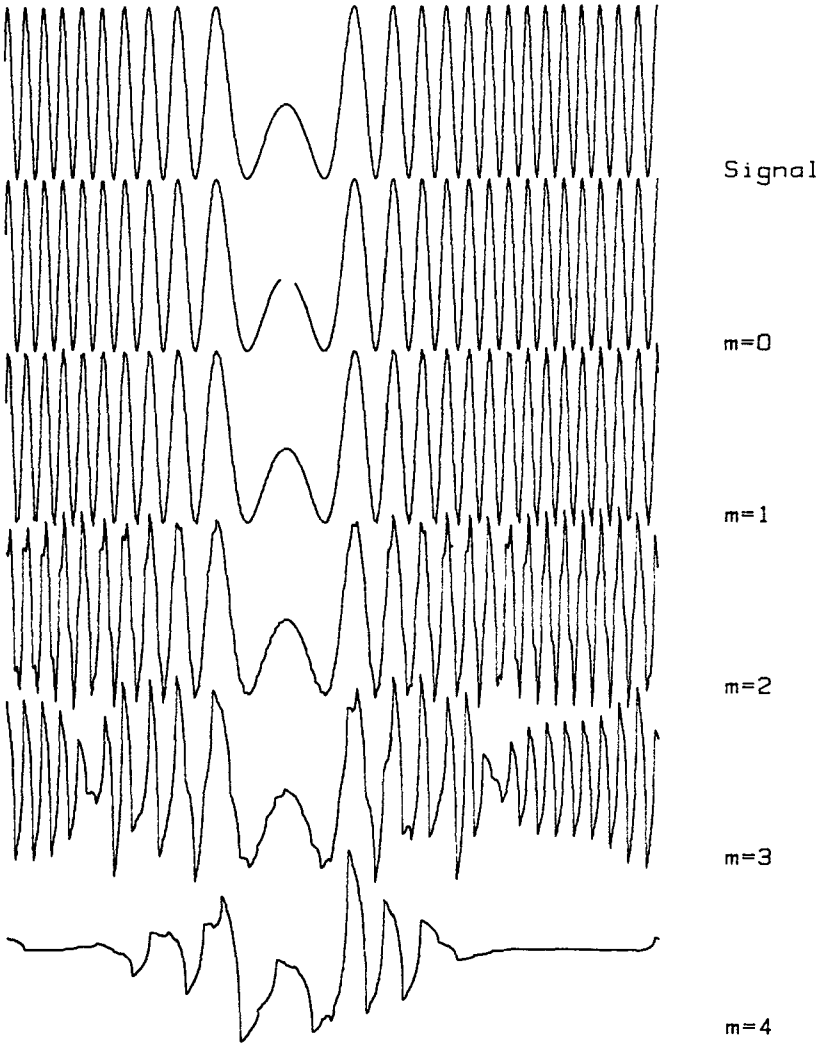

Fig. 8. A signal and its reconstruction using the wavelet of fig. 5.

Fig. 3 showed the transform of a relatively simples signal (a sine-wave of varying frequency) and, as might have been expected, the energy tends to move to higher dilation levels as the frequency decreases. However, the behaviour in this respect is somewhat erratic, as the transform coefficients depend also on

the alignment of the wavelets with the oscillations of the signal (see comments above). It may thus be stated that these wavelets are rather sensitive to signal 'phase' and are therefore not a simple alternative to power-spectral estimation. This may however also be considered an advantage, since power-spectra can be criticized for their lack of responsiveness to signal phase and therefore signal morphology.

To what extend these wavelets do have a useful role in the processing of biomedical signals will only become clear as we gain more experience in their use on 'real data'. Some results in this respect will be presented in a companion paper to follow. It seems doubtful that wavelets will ever replace the Fourier Transform but they do suggest some interesting new possibilities.

APPENDIX

Pascal listing of the forward and inverse transform algorithm.

```
const
      max_buffer_length = 580 ;{no. of samples in data block}
      max_transform_level = 7 ;{max. m}
      max_wavelet_size = 6 ;{max. size of wavelet}

type
     buffer_t = array [ 0 .. max_buffer_length - 1 ] of real ;
     coefficient_store = array [ 0 .. max_transform_level] of
       buffer_t ;
     coefficient_type = array [ 0 .. max_wavelet_size - 1 ] of
       real ;

procedure reconstruction_filter
  ( var h : coefficient_type ;
  wavelet_size : integer ;
  var c : buffer_t );
{
Perform the inverse transform of coefficients 'c', using the
discrete wavelet 'h' of size 'wavelet_size'.
The same routine may be used with 'g' and 'd'.
}
var n, k : integer ;
    temp : buffer_t ;
    start_offset : integer :
begin
  start_offset := - (wavelet_size div 2 - 1) ;
  for n := 0 to max_buffer_length - 1   do
    begin
      temp [n] := 0 ;
      k := n div 2 ;
      while ( k >= start_offset) and ( (n-2*k) < wavelet_size )
      do
        begin
          temp[n] := temp[n] + h[n-2*k]*c[k-start_offset] ;
          k := k - 1 ;
        end ;
```

```
    end ;
  c := temp ;
end ;

procedure decomposition_filter
  ( var h : coefficient_type ;
  wavelet_size : integer ;
  var c : buffer_t );
{
Perform the forward transform of coefficients 'c', using the
discrete wavelet 'h' of size 'wavelet_size'.
The same routine may be used with 'g' to obtain 'd'.
}
var n, k : integer ;
    temp : buffer_t ;
    start_offset : integer ;
    r : real ;
begin
  start_offset := - (wavelet_size div 2 - 1) ;
  for k := 0 to max_buffer_length - 1  do
    temp [k] := 0 ;
  for k := start_offset to max_buffer_length div 2 - 1 do
    begin
      r := 0 ;
      for n := k*2 to (k*2 + wavelet_size - 1 ) do
        begin
          if (n >= 0) and (n < max_buffer_length) then
            r := r + h[n-2*k]*c[n];
        end ;
      temp [k-start_offset] := r ;
    end ;
  c := temp ;
end ;

procedure block_wavelet_transform
  ( var signal : buffer_t ;
  var c_m : buffer_t ;
  var ds : coefficient_store ;
  var h, g : coefficient_type ;
  wavelet_size, max_m : integer );
{
Transform the signal 'signal' to 'max_m' levels, using the 'h'
and 'g' of size 'wavelet_size', and store the results in 'c_m' (c
at level max_m) and 'ds' (the d's up to level max_m)
}
var m : integer ;
begin
  c_m := signal ;
  for m := 1 to max_m do
    begin
      ds[m] := c_m ;
      decomposition_filter ( g, wavelet_size , ds[m]);
      decomposition_filter ( h, wavelet_size , c_m );
    end ;
end ;
```

```
procedure block_inverse_wavelet_transform
  ( var c_m : buffer_t ; var ds : coefficient_store ;
  var h, g : coefficient_type ;
  wavelet_size, max_m : integer ;
  reconstruction_levels : integer ;
  var signal : buffer_t );
{
Inverse transform the results of 'block_wavelet_transform', using
only the highest 'reconstruction_levels' of d's., i.e. if
reconstruction_levels = max_m, it performs complete
reconstruction, if reconstruction_levels = 0, it only uses c_m
and none of the ds. However, the reconstruction is always
performed to the number of samples in the original data.
}
var m, i, levels_reconstructed : integer ;
    temp : buffer_t ;
begin
  signal := c_m;
  levels_reconstructed := 0 ;
  for m:= max_m downto 1 do
    begin
      reconstruction_filter ( h, wavelet_size , signal );
      if levels_reconstructed < reconstruction_levels then
        begin
          temp := ds[m] ;
          reconstruction_filter ( g, wavelet_size , temp );
          for i := 0 to max_buffer_length - 1 do
            signal[i] := signal[i] + temp[i] ;
        end ;
      levels_reconstructed := levels_reconstructed + 1 ;
    end ;
end ;
```

ACKNOWLEDGEMENT

The author would like to thank the agencies CNPq and FAPERJ for their financial support in this work and to the anonymous reviewers for the improvements they suggested.

REFERENCES

DAUBECHIES, I. (1988), "Orthonormal Bases of Compactly Supported Wavelets", Communications on Pure and Applied Mathematics, Vol. XLI, 909-996.

DAUBECHIES, I. (1989) "Orthonormal Bases of Wavelets with Finite Support - Connection with Discrete Filters", in: J.M. Combes, A. Grossman, PH. Tchamitchian (Eds), Wavelets: Time-Frequency Methods and Phase Space, Springer Verlag, Berlin, pp.38-67.

MALLAT, S.G., (1989) "A Theory Multiresolution Signal Decomposition: The Wavelet Representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 7, pp.674-693.

NAGAE, N., ITOH, S., UTSUNAMIYA, T. (1991), "Identification of Letter Cognitive Process in Evoked EEG Using Discriminant Analysis of Orthonormal Wavelet Transform of EEG", Digest of the World Congress on Medical Physics and Biomedical Engineering, July 7-12, 1991, Kyoto, Japan, Medical and Biological Engineering & Computing, Vol. 29, Suppl., p.385.

RIOUL, O., VETTERLI, M. (1991), "Wavelets and Signal Processing", IEEE Signal Processing Magazine Vol. 8, No. 4, (October), pp.14-38.

SENDHADJI, L., BELLANGER, J.J., CARRAULT, G., COATRIEUX, J.L. (1990), "Wavelet analysis of E.C.G. Signals", Proc. 12th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Philadelphia, Nov. 1-4, Vol. 12, No. 2, pp.0811-0812.

STRANG, G. (1989), "Wavelets and Dilation Equations: A Brief Introduction", SIAM Review, Vol. 31, No. 4, pp.614-627.

UMA INTRODUÇÃO À TRANSFORMADA ONDELETTE, DISCRETA E ORTOGONAL

por

David Martin Simpson

RESUMO -- A transformada ondelete ('wavelet') tem recebido consideravel atenção como possível alternativa (com certas vantagens) a transformadas mais tradicionais, tais como a de Fourier ou de Cosseno, especialmente quando aplicada a sinais transientes. No presente trabalho, os conceitos básicos da transformada ondelete são descritos e uma das famílias de ondeletes é então focalizada: a discreta, ortogonal de suporte finito. A base teórica é desenvolvida para transformação direta e inversa bem como para seleção de ondeletes. O trabalho visa explicar a técnica em termos acessíveis para engenheiros, evitando os métodos matemáticos mais complicados (e frequentemente desnecessários) geralmente usados na sua descrição.