From Calculus to Wavelets: A New Mathematical Technique

Wavelet Analysis

Gerald B Folland

A 'wavelet' is a function that exhibits oscillatory behaviour in some interval and then decays rapidly to zero outside this interval. A remarkable discovery of recent years is that the translations and dilations of certain wavelets can be used to form sets of 'basic' functions for expanding general functions into infinite series, and these expansions have many theoretical and practical applications.

G B Folland is with the University of Washington since 1973. He received his Ph.D from Princeton University in 1971. He is the author of eight advanced texts and monographs and a number of research papers.

The expansion of functions into infinite series is one of the most powerful techniques of mathematical analysis. The general idea is to write a function $f(t)$ (defined on the real line or an interval in the real line) as

$$f(t) = \sum c_n \phi_n(t), \tag{1}$$

where the ϕ_n's are a family of 'basic' functions with properties that are significant for the problem at hand and the coefficients c_n are real or complex numbers depending on f. (As we shall see, sometimes it is more convenient to choose an index set for the ϕ's other than the natural numbers $0, 1, 2, \ldots$) For such an expansion to be useful, the coefficients c_n must be computable in terms of f in a reasonably straightforward way, and the series $\sum c_n \phi_n$ must encode important information about the function f in an easily accessible form.

Usually the first type of series expansion that one meets as a mathematics student is *power series*, that is,

$$f(t) = \sum_{n=0}^{\infty} c_n (t-a)^n \tag{2}$$

The coefficients of such a series are given by Taylor's formula,

$$c_n = \frac{f^{(n)}(a)}{n!}.$$

In particular, a function cannot be represented as a power series unless it possesses derivatives of all orders (and even this is not a guarantee of the validity of (2)). However, if f has derivatives up to order N, one can still use Taylor's formula with remainder,

$$f(t) = \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(t-a)^n + R_N(t), \tag{3}$$

in which the remainder R_N vanishes more rapidly than $|t-a|^N$ as $t \to a$. This formula provides a very precise description of $f(t)$ for t near a but is usually not of much use when t is far away from a.

Another very useful type of series is the *trigonometric series* or *Fourier series.* These series come in several closely related forms, of which the simplest for analytical purposes is the one involving the complex exponentials $e^{2\pi int} = \cos 2\pi nt + i \sin 2\pi nt$:

$$f(t) = \sum_{n=-\infty}^{\infty} b_n e^{2\pi int}. \tag{4}$$

It was Fourier's great discovery that such series can be used to represent more or less arbitrary functions on the real line that are *periodic* with period 1 (that is, $f(t+1) = f(t)$ for all t). More precisely, if f is any periodic function — even one with lots of discontinuities and infinite singularities — for which the area under the graph of $|f|^2$ over one period is finite,

$$\int_{-1/2}^{1/2} |f(t)|^2 \, dt < \infty, \tag{5}$$

then f admits an expansion of the form (4). (If f is a rather 'rough' function, there may be some exceptional points where the series does not converge to $f(t)$, but it converges to $f(t)$ 'almost everywhere.') The *Fourier coefficients* b_n are given by a simple integral formula, whose meaning is explained in *Box 1*:

$$b_n = \int_{-1/2}^{1/2} f(t) e^{-2\pi int} \, dt. \tag{6}$$

One of the virtues of the Fourier representation is that it converts differentiation into a simple algebraic operation.

One of the virtues of the Fourier representation (4) is that it converts differentiation into a simple algebraic operation. Indeed, term-by-term differentiation of (4) yields

$$f'(t) = \sum_{-\infty}^{\infty} 2\pi i n b_n e^{2\pi int}.$$

This needs some justification, but the result is correct: *Suppose f is differentiable and f' satisfies condition (5). If the Fourier coefficients of f are denoted by b_n, then the Fourier coefficients of f' are $2\pi i n b_n$.* Similarly for higher derivatives: the Fourier coefficients of $f^{(k)}$ are $(2\pi i n)^k b_n$.

As a consequence, it is easy to read off the order of differentiability of f from its Fourier coefficients. If f has derivatives up to order k that satisfy (5), then the series $\sum (2\pi i n)^k b_n e^{2\pi int}$ must converge, and hence $n^k b_n$ must tend to zero as $|n| \to \infty$. In other words, b_n must tend to zero more rapidly than $|n|^{-k}$. Conversely, if $|b_n| \le C|n|^{-k}$, the series (4) can be differentiated at least $k-1$ times (the $(k-1)$th derivative may exist only

Box 1 The Meaning of Orthonormality

A sequence of complex-valued functions ϕ_n is said to be *orthonormal* on an interval I (which might be the whole real line) if

$$\int_I \phi_n(t)\overline{\phi_m(t)}\,dt = 0 \text{ for } n \neq m \text{(orthogonality)}, \quad \text{(a)}$$

$$\int_I |\phi_n(t)|^2\,dt = 1 \qquad \text{(normalization)}. \quad \text{(b)}$$

(The terminology comes from the analogy with an orthonormal or unit-perpendicular set of vectors $\mathbf{v}_n$; the integrals $\int_I \phi_n\overline{\phi}_m$ replace the dot products $\mathbf{v}_n \cdot \mathbf{v}_m$.) For example, it is an easy exercise to check that the functions $e^{2\pi int}$ used in Fourier analysis are orthonormal on any interval of length 1. If the functions ϕ_n are orthonormal and f can be expanded in terms of them,

$$f(t) = \sum c_n\phi_n(t) \qquad (t \in I), \quad \text{(c)}$$

the coefficients c_n must be given by

$$c_m = \int_I f(t)\overline{\phi_m(t)}\,dt. \quad \text{(d)}$$

To see this, multiply both sides of (c) by $\overline{\phi_m(t)}$ and integrate over I:

$$\int_I f(t)\overline{\phi_m(t)}\,dt = \sum_n c_n \int_I \phi_n(t)\overline{\phi_m(t)}\,dt.$$

By (a), the integrals on the right all vanish except the one with $n = m$, and that one equals 1 by (b), so the sum on the right reduces to c_m. Conversely, if (c) always implies (d), the sequence ϕ_n must be orthonormal. Indeed, the identity

$$\phi_m = \cdots + 0\cdot\phi_{m-1} + 1\cdot\phi_m + 0\cdot\phi_{m+1} + 0\cdot\phi_{m+2} + \cdots$$

is an instance of (c), and the corresponding instance of (d) is just (a) together with (b).

'almost everywhere'). In short: *The order of smoothness of a function is reflected in the rate of decay of its Fourier coefficients.*

The bad news is that it is hard to tell from a Fourier series *where* the non-smooth behavior occurs, or what it looks like. For example, the function

$$g_1(t) = \sum_{n\neq 0} \frac{e^{2\pi int}}{2in} = \sum_{1}^{\infty} \frac{\sin 2\pi nt}{n} \tag{7}$$

has jump discontinuities at the points $0, \pm 1, \pm 2,$ but is perfectly smooth (in fact, linear) elsewhere. Its close relative

$$g_2(t) = \sum_{n\neq 0} \frac{(-1)^n e^{2\pi int}}{2in} = \sum_{1}^{\infty} \frac{(-1)^n \sin 2\pi nt}{n} \tag{8}$$

has the same sort of behavior, but with discontinuities at $\pm\frac{1}{2}, \pm\frac{3}{2}, \pm\frac{5}{2},$ .. And the function fashioned from g_1 by omitting all the terms whose order is not a power of 2,

$$g_3(t) = \sum_{k=0}^{\infty} \frac{\sin 2\pi 2^k t}{2^k}, \tag{9}$$

is everywhere continuous but nowhere differentiable! See *Figures 1-3.*

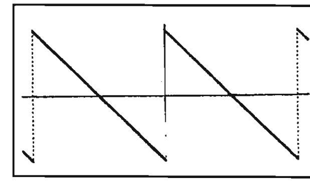

Figure 1 The function g_1 defined by (7).

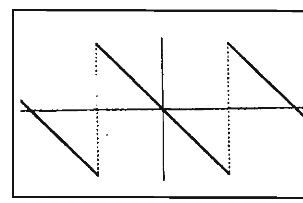

Figure 2 The function g_2 defined by (8).

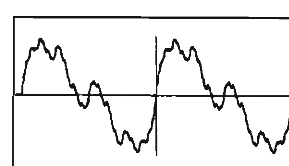

Figure 3 The function g_3 defined by (9).

We have presented Fourier series for functions that are periodic with period 1. A simple change of variable yields a corresponding expansion for functions that are periodic with period P:

$$f(t) = \frac{1}{P} \sum_{-\infty}^{\infty} b_n e^{2\pi int/P}$$

where

$$b_n = \int_{-P/2}^{P/2} f(t) e^{-2\pi int/P}\, dt.$$

Taylor series provide precise local information about a function, but only if the function has many derivatives. Fourier series or integrals can be used for functions with no smoothness properties, and they yield lots of information about the global properties of the function, but they are inefficient for analyzing the detailed behavior of a function near a point.

In the limit as $P \to \infty$, one obtains an integral rather than a series:

$$f(t) = \int_{-\infty}^{\infty} \hat{f}(\omega) e^{2\pi i \omega t}\, d\omega, \qquad (10a)$$

$$\text{where } \hat{f}(\omega) = \int_{-\infty}^{\infty} f(t) e^{-2\pi i \omega t}\, dt. \qquad (10b)$$

The formula (10a) is called the *Fourier integral* expansion of f. It is valid for (non-periodic) functions f on the real line that satisfy the analogue of (5):

$$\int_{-\infty}^{\infty} |f(t)|^2\, dt < \infty, \qquad (11)$$

and it expresses such functions as a continuous superposition of the simple sinusoidal functions $e^{2\pi i \omega t}$. The coefficients $\hat{f}(\omega)$ in this expansion are given by (10b); they constitute a function of the frequency ω called the *Fourier transform* of f. The connection between smoothness of f and the size of its Fourier coefficients, discussed above for the periodic case, continues to hold in this situation.

To summarize: Taylor series provide precise *local* information about a function, but only if the function has many derivatives. Fourier series or integrals can be used for functions with no smoothness properties, and they yield lots of information about the *global* properties of the function, but they are inefficient for analyzing the detailed behavior of a function near a point. (How could you ever tell where the discontinuities of the functions in (7) and (8) occur by examining their Fourier coefficients?) What is missing is a method for analyzing the *local irregular* behavior of functions that aren't smooth — and this is where wavelets come into play.

What is a 'wavelet'? This term is loosely used to denote a function that exhibits oscillatory behavior in some interval I, then decays rapidly to zero (or perhaps vanishes identically) outside I. If we have one such function

$\psi(t)$, we can obtain others by *dilations* or *translations* — that is, by replacing the variable t by at ($a > 0$) or $t - b$ ($-\infty < b < \infty$). The dilation $\psi(t) \to \psi(at)$ changes the frequency of the oscillations of ψ by a factor of a, and simultaneously changes the length of the interval on which ψ lives by a factor of $1/a$. The translation $\psi(t) \to \psi(t-b)$ simply shifts ψ over by the amount b. We can also combine these operations to get $\psi(at - b)$, which is the result of *first* translating by the amount b and *then* dilating by the factor a. (If we do it in the reverse order, we get something different: $\psi(t) \to \psi(at) \to \psi(a[t-b]) = \psi(at-ab)$.) See *Figure 4*.

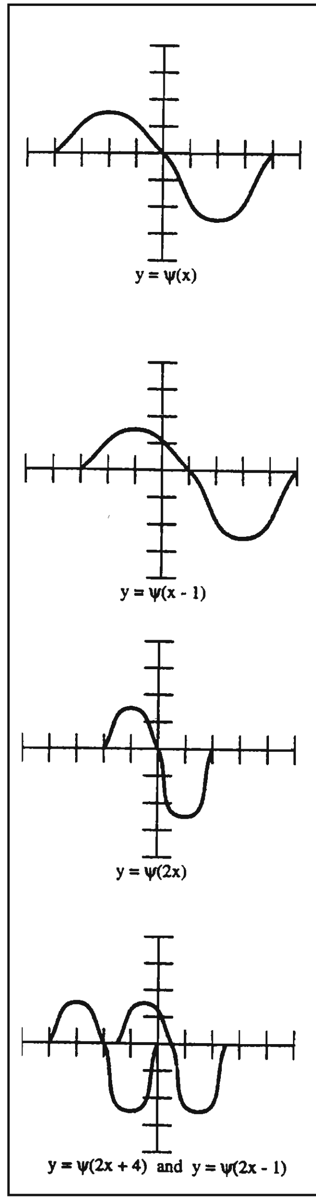

***Figure 4** The effect of translations and dilations on a function.*

Now, here is the basic problem of wavelet analysis: *Can we use the translations and dilations of a wavelet ψ to form a set of 'basic' functions for expanding general functions into infinite series as in (1)?* Let us make a formal definition: An *orthonormal wavelet* is a real- or complex-valued function $\psi(t)$ defined on the real line, with the following properties:

[1] $\psi(t)$ tends to zero faster than any power of t as $t \to \pm\infty$.

[2] ψ possesses continuous derivatives up to order N, for some positive integer N.

[3] For all integers j and n, let

$$\psi_{jn}(t) = 2^{j/2}\psi(2^j t - n). \tag{12}$$

Then every function f on the real line that satisfies (11) can be expanded uniquely in a series of the ψ_{jn}:

$$f(t) = \sum_{j,n=-\infty}^{\infty} c_{jn}\psi_{jn}(t). \tag{13}$$

[4] The coefficients c_{jn} in (13) are given by

$$c_{jn} = \int_{-\infty}^{\infty} f(t)\overline{\psi_{jn}(t)}\,dt, \tag{14}$$

where $\overline{\psi_{jn}(t)}$ is the complex conjugate of $\psi_{jn}(t)$.
As explained in *Box 1*, properties [3] and [4] imply that the functions ψ_{jn} are orthonormal; hence the name orthonormal wavelet. The factor of $2^{j/2}$ in (12) is there to make $\int_{\infty}^{\infty} |\psi_{jn}(t)|^2 \, dt$ independent of j and n.

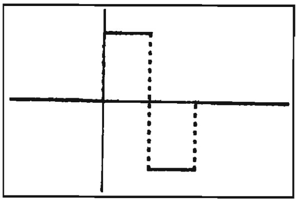

Figure 5 The Haar wavelet.

The existence of orthonormal wavelets is not at all obvious. If we omit condition [2], there is a simple example that has been known since 1910, the *Haar wavelet*

$$\psi_{\text{Haar}}(t) = \begin{cases} 1 & \text{if } 0 < t < \frac{1}{2}, \\ -1 & \text{if } \frac{1}{2} < t < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Finally, by an ingenious adaptation for some algortihms developed by enigneers for signal processing, Ingrid Daubechies constructed a family of orthonormal wavelets, with all possible finite orders of smoothness, that actually vanish outside of finite intervals.

See *Figure 5*. However, the Haar wavelet is too crude to be very useful. Finding orthonormal wavelets that satisfy [2] is much trickier, and it was not done until 1986–87 when a group of mathematicians led by Yves Meyer took some ideas that had been used in signal processing and other areas of applied mathematics and turned them into a rigorous mathematical theory. Meyer constructed an orthonormal wavelet that has derivatives of all orders, but its decay at infinity is relatively slow (a computational disadvantage). See *Figure 6*. Guy Battle and Pierre-Gilles Lemarié constructed a family of orthonormal wavelets, one for each order of smoothness N in [2], with exponential decay at infinity. Finally, by an ingenious adaptation of some algorithms developed by engineers for signal processing, Ingrid Daubechies constructed a family of orthonormal wavelets, with all possible finite orders of smoothness, that actually vanish outside of finite intervals. See *Figure 7*.

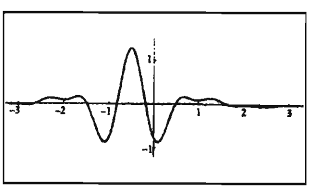

Figure 6 A wavelet of Meyer type.

Let us pause to say a few more words about these constructions. Meyer's method is Fourier-analytic; he gives an explicit formula not for ψ but for its Fourier transform. The Battle-Lemarié wavelets are *splines*, that is, piecewise polynomial functions. Both these types of functions are well known in mathematical analysis.

But Daubechies's wavelets are something new. They are given not by analytic formulas but as limits of recursive transformations. That is, a Daubechies wavelet ψ is obtained as the limit of a sequence of functions ϕ_k, where ϕ_0 is a simple piecewise linear function and ϕ_k is defined in terms of its predecessor ϕ_{k-1} by a recursion formula of the form

$$\phi_k(t) = \sum_{n=-N}^{N} C_n \phi_{k-1}(2t-n), \qquad (15)$$

and the whole secret is to find the appropriate coefficients C_n to make this process work. The resulting limit function ψ is a rather peculiar beast from the standpoint of classical mathematics, but it is lovely from the computational point of view, because the formula (15) involves only simple arithmetic which a computer is quite happy to perform over and over again!

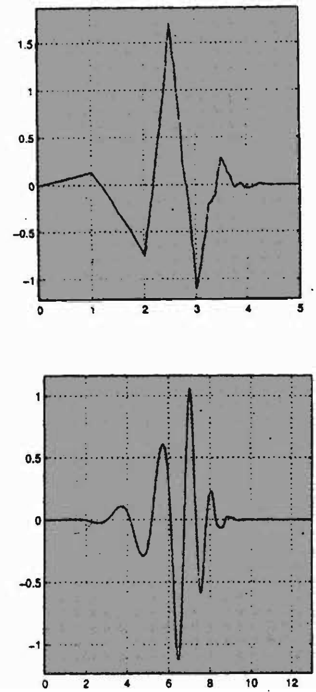

Figure 7 Two of the Daubechies wavelets.

We have suggested that wavelets should have some oscillatory behavior, and the accompanying graphs also indicate this, but there seems to be nothing in properties [1]–[4] about oscillation. In fact, it is not too hard to show that properties [1]–[4] automatically imply the following additional property of wavelets:

[5] Let N be the integer in property [2]. Then

$$\int_{-\infty}^{\infty} \psi_{jn}(t) P(t)\, dt = 0 \qquad (16)$$

for all integers j and n and all polynomials P of degree $\leq N$

This property implies that ψ must be quite oscillatory in order to produce the cancellations that make all the integrals in (16) vanish. For example, by taking $P(t) \equiv 1$ (and $j = n = 0$) we see that $\int_{-\infty}^{\infty} \psi(t)\, dt = 0$, which means that positive values of ψ in one region must be balanced by negative values in another.

We now examine the meaning and applications of wavelet expansions of the form (13). To simplify the discussion, we shall assume that we are working with a Daubechies wavelet ψ with N continuous derivatives that vanishes outside a bounded interval I. (Almost everything we say remains true for other wavelets if one modifies the statements to take account of the rapidly decaying tails of the wavelet.)

The function ψ_{jn} defined by (12) then vanishes outside the interval I_{jn} obtained by translating I by the amount n and then dilating by the factor 2^{-j}; thus, if I has length l then I_{jn} has length $2^{-j}l$. At the same time, we may think of the frequency of the oscillations of ψ as being on the order of magnitude of 1. (This is just a manner of speaking, as the oscillations of ψ are irregular and don't have a definite frequency.) Then the frequency of the oscillations of ψ_{jn} is on the order of magnitude of 2^j. In short, for $j \ll 0$, ψ_{jn} represents *low-frequency* oscillations over *long* intervals, and for $j \gg 0$, ψ_{jn} represents *high-frequency* oscillations over *short* intervals.

We can now see how to interpret (13) and (14):

$$f = \sum_{j,n=-\infty}^{\infty} c_{jn}\psi_{jn}, \qquad c_{jn} = \int_{-\infty}^{\infty} f(t)\overline{\psi_{jn}(t)}\,dt$$
$$= \int_{I_{jn}} f(t)\overline{\psi_{jn}(t)}\,dt.$$

Let us rewrite the expansion of f as

$$f = \sum_{j=-\infty}^{\infty} S_j, \text{ where } S_j = \sum_{n=-\infty}^{\infty} c_{jn}\psi_{jn}. \tag{17}$$

The sum $\sum_{j<0} S_j$ represents a 'blurred' version of f that shows its large features but omits the fine details of size 1 or less. The sum S_0 adds in oscillations of frequency roughly 1 on intervals of length roughly 1; the sum S_1 adds in oscillations of frequency roughly 2 on intervals of length roughly 1/2; and so forth. Each additional sum S_j adds another level of detail at the length scale 2^{-j}.

Moreover, the individual terms in these sums are *local* in nature. The coefficient c_{jn} depends only on the values of f on the interval I_{jn}, and the term $c_{jn}\psi_{jn}$ lives on this same interval. Since these intervals only overlap to a finite extent as n varies (with j fixed), the sum S_j provides an efficient description of the local details of the behavior of f at length scale 2^{-j}. As j increases, it is like looking at pieces of f under increasingly strong microscopes.

(On the other hand, when $j \ll 0$ S_j might be said to represent a view of f through the wrong end of a telescope. If f is negligibly small outside an interval of length L, it is easy to see that S_j will also be negligibly small for $2^{-j} \gg L$, because things on the scale of L are 'invisible' to the long-range waves ψ_{jn}. Hence the terms in (13) with $j \ll 0$ are not of much importance for most purposes.)

The smoothness of f in the neighbourhood of a point $t = a$ is reflected in the decay as $j \to +\infty$ of those wavelet coefficients c_{jn} for which the interval I_{jn} contains a.

We mentioned earlier that smoothness of a periodic function is reflected in the decay of its Fourier coefficients. The same is true for wavelet expansions, but now on a local level: *The smoothness of f in the neighbourhood of a point $t = a$ is reflected in the decay as $j \to +\infty$ of those wavelet coefficients c_{jn} for which the interval I_{jn} contains a.* The reason for this fact (of which there are several much more precise versions) can be briefly explained as follows. If f has N continuous derivatives near $t = a$, we can use Taylor's formula (3) to write $f = P_N + R_N$ where P_k is the Taylor polynomial of degree N about $t = a$ and $R_N(t)$ is very small for t near a. Hence, by property [5],

$$c_{jn} = \int_{I_{jn}} f(t)\overline{\psi_{jn}(t)}\,dt = \int_{I_{jn}} [P_N(t) + R_N(t)]\overline{\psi_{jn}(t)}\,dt$$

$$= \int_{I_{jn}} R_N(t)\overline{\psi_{jn}(t)}\,dt.$$

Wavelets have been found to be particularly useful in problems that require separating out the parts of a signal that pertain to different scales of time or length, and in processing signals or images whose most significant features lie in the regions where rapid variations occur.

But if I_{jn} contains a and j is large (so that the length of I_{jn} is small), I_{jn} will be contained in the region where R_N is very small, and hence c_{jn} will be very small!

For example, let us return to the functions g_1, g_2, and g_3 defined by (7), (8), and (9). The Fourier series for g_1 and g_2 look almost alike, but their wavelet expansions are very different because the sets of indices (j, n) for which I_{jn} contains one of the discontinuities of g_1 or g_2 are very different. It is only for those values of j and n that the coefficients c_{jn} for g_1 or g_2 will be significant for j large. (In fact, if I_{jn} does not contain a discontinuity of g_1 or g_2, c_{jn} will actually be zero by property [5], since g_1 or g_2 is linear on I_{jn}!) On the other hand, g_3 is more or less equally rough at all points, so its coefficients c_{jn} will all decay slowly as $j \to +\infty$.

Wavelet expansions have significant applications in a number of areas, particularly in the analysis and processing of electric and acoustic signals and of two-dimensional images such as photographs. (The latter require wavelets in two variables, whose theory is similar to the one-dimensional theory sketched above.) Wavelets have been found to be particularly useful in problems that require separating out the parts of a signal that pertain to different scales of time or length, and in processing signals or images whose most significant features lie in the regions where rapid variations occur. (For example, in an ordinary photograph the most important features may be the edges that indicate the boundaries of the objects in the photograph; in an astronomical photograph it is important to distinguish between distant stars and even more distant galaxies, whose details show up at different length scales.) There is considerable evidence that the human brain uses something akin to wavelet analysis in processing the information it receives from the eyes and ears, and researchers in human vision were using wavelets in an informal way long before the present mathematical theory was developed.

This is not to say that wavelet expansions of the sort we have discussed are good for everything, however. For example, they are not particularly efficient for encoding musical signals because they do not provide enough frequency resolution. (Roughly speaking, the expansion (17) only provides frequency resolution to within the nearest octave, because the frequency scales specified by the index j differ by factors of 2.) However, the discovery of orthonormal wavelets has stimulated the development of a variety of related techniques for expanding functions (some of which go under the names of bi-orthogonal wavelets, wavelet packets, local Fourier bases, etc.) that offer much greater flexibility in adapting the expansion to the problem at hand. For example, one can sacrifice some precision in spatial or temporal resolution to obtain better frequency resolution, or sacrifice the perfect efficiency of the orthogonal wavelet expansion for expansions that have some built-in redundancy and hence are less sensitive to encoding errors. But this is another story.

There is considerable evidence that the human brain uses something akin to wavelet analysis in processing the information it receives from the eyes and ears.

Suggested Reading

- C K Chui (ed.). *Wavelets: a Tutorial in Theory and Applications*. Academic Press. Boston, 1992. It is a collection of expository articles about wavelets and related topics.
- Y Meyer. *Wavelets: Algorithms and Applications*. Society for Industrial and Applied Mathemtatics. Philadelphia, 1993. It covers the wavelet theory and its applications on a somewhat higher mathematical level.
- R S Strichartz. How to make wavelets. *American Mathematical Monthly*. Vol.100. pp.539-556,1993. It is a good short exposition of the techniques for constructing waves.
- J J Benedetto and M W Frazier (eds.). *Wavelets: Mathematics and Applications*. CRC Press. Boca Raton, Florida, 1994. It is a collection of expository articles about wavelets and related topics.
- E Hernandez and G Weiss. *A First Course on Wavelets*. CRC Press. Boca Raton, Florida, 1996. It is a fine introductory treatise on wavelet theory.
- B B Hubbard. *The World According to Wavelets*. A K Peters. Wellesley. Mass., 1996. It is a delightful non-technical account of the development of wavelet theory and its applications.

Address for Correspondence
Gerald B Folland
Department of Mathematics
University of Washington
Seattle, WA 98195-4350 USA