

CONF-970465--1

Applications of a Fast, Continuous Wavelet Transform*

W. B. Dress

RECEIVED
FEB 12 1997
OSTI

Oak Ridge National Laboratory, P.O. Box 2008, Bldg. 3546, MS 6011
Oak Ridge TN 37831-6011; e-mail: wbd@ornl.gov

To be presented at the
SPIE Wavelets Application Conference
Orlando, Florida
April 20-25, 1997

"The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-96OR22464. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes."

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

*This research was performed at OAK RIDGE NATIONAL LABORATORY, managed by LOCKHEED MARTIN ENERGY RESEARCH CORP. for the U.S. DEPARTMENT OF ENERGY under contract DE-AC05-96OR22464.

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

Applications of a fast, continuous wavelet transform

W. B. Dress

Instrumentation and Controls Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831-6011

ABSTRACT

A fast, continuous, wavelet transform, justified by appealing to Shannon's sampling theorem in frequency space, has been developed for use with continuous mother wavelets and sampled data sets. The method differs from the usual discrete-wavelet approach and from the standard treatment of the continuous-wavelet transform in that, here, the wavelet is sampled in the frequency domain. Since Shannon's sampling theorem lets us view the Fourier transform of the data set as representing the continuous function in frequency space, the continuous nature of the functions is kept up to the point of sampling the scale-translation lattice, so the scale-translation grid used to represent the wavelet transform is independent of the time-domain sampling of the signal under analysis. Although more computationally costly and not represented by an orthogonal basis, the inherent flexibility and shift invariance of the frequency-space wavelets are advantageous for certain applications.

The method has been applied to forensic audio reconstruction, speaker recognition/identification, and the detection of micromotions of heavy vehicles associated with ballistocardiac impulses originating from occupants' heart beats. Audio reconstruction is aided by selection of desired regions in the two-dimensional representation of the magnitude of the transformed signal. The inverse transform is applied to ridges and selected regions to reconstruct areas of interest, unencumbered by noise interference lying outside these regions. To separate micromotions imparted to a mass-spring system (e.g., a vehicle) by an occupant's beating heart from gross mechanical motions due to wind and traffic vibrations, a continuous frequency-space wavelet, modeled on the frequency content of a canonical ballistocardiogram, was used to analyze time series taken from geophone measurements of vehicle micromotions.

By using a family of mother wavelets, such as a set of Gaussian derivatives of various orders, different features may be extracted from voice data. For example, analysis of the "blobs" in a low-order, Gaussian-derivative wavelet transform extracts the glottal-closing rate, while the ridges of a high-order wavelet transform give good indication of the formant frequencies and allow automatic word and phrase segmentation.

Keywords: Continuous wavelets, Hilbert transform, Fourier-space sampling, frequency-space wavelets, fractional-derivative wavelets, voice analysis, ballistocardiac analysis, audio forensics

1. INTRODUCTION

1.1 Why continuous wavelets?

The discrete wavelet transform, implemented as finite impulse response (FIR) filters, is an ubiquitous analysis tool for problems in signal processing. However, the inherent octave-scale resolution is often inadequate for situations requiring identification of features on a finer scale as might be required in speech processing. One solution is to employ wavelet voices¹; another option is to use polyphase resampling techniques² effectively altering the (discrete) scale of the original signal. The first option is rarely found in the literature and involves repeated constructions of suboctave filters while the second requires reprocessing of the original signal for each set of scale features desired. Both methods are more computationally costly than the original FIR filter implementation while allowing only limited flexibility.

The solution chosen here is to retain complete flexibility over scale by means of the continuous wavelet transform (CWT) while giving up some of the efficiency of the FIR implementation. As usually implemented in computer computations, the CWT is approximated by a set of samples of the mother wavelet and then numerically convolved with the signal, which is represented by its discrete samples. Methods of computing the CWT all seek a numerical approximation to the integral

$$\tilde{f}(a, b) = \int_{\mathcal{R}} \psi^*_{a,b}(t) f(t) dt, \quad (1)$$

Further author information—

Email: wbd@ornl.gov; Telephone: 423-574-4801; Fax: 423-574-6663

where ψ^* is the complex conjugate of ψ and \mathcal{R} is the domain of integration, usually taken as the set of reals for one-dimensional signals. Equation 1 is the general expression for the wavelet transform of a function $f(t)$ with respect to the mother wavelet $\psi(t)$. In the usual formulation³, ψ is parameterized by a scale a and a position b , as

$$\psi_{a,b}(t) = |a|^{-p} \psi\left(\frac{t-b}{a}\right). \quad (2)$$

Here, we follow Ref. 3 in deferring a choice for the normalization; the usual choices for p are either 1 or $\frac{1}{2}$, depending on whether we wish to preserve the integral of the absolute value or the L^2 norm of the wavelet under scale changes. Below, we will discuss other choices for p that can enhance the two-dimensional visual representation of the function $\tilde{f}(a, b)$.

1.2 Standard methods of approximation

There are two methods in general use to compute the integral in Eq. 1. One is a direct approach and the other makes use of the correlation theorem *a la* Wiener-Khinchin. In both approaches, the wavelet in Equation 2 is sampled on a two-dimensional lattice. The choice for the scale, a , is $a_m = \{a_o^m, m \in \mathbb{Z}\}$, for some suitable $a_o > 1$. For the translation parameter, b , the desirable choice is $b_\ell = \{\ell b_0 a_m, \ell \in \mathbb{Z}\}$ for each scale indexed by m . This choice ensures that the lattice is not overly redundant. If one is given, as is usually the case in real-world applications, the samples of the signal $f(t_k)$ for $t_k = k \Delta t$, where $k \in \mathbb{Z}$ and $\frac{1}{\Delta t}$ is the sampling rate, the usual dyadic lattice prevalent in the discrete, FIR-filter wavelet transform is reproduced in the CWT case for $b_0 = \frac{\Delta t}{2}$ and $a_0 = 2$.

For the first method, once we have the samples of the signal and have sampled the wavelet on the two-dimensional lattice, $\{a_m, b_\ell\}$, it is a simple matter to approximate the convolution integral by a set of dot products or simple sums of products, one set for each choice of scale parameter. There remains a question of end effects: do we pad the signal's samples with enough zeros to avoid discontinuities—we will have sampled the continuous wavelet function out to where its magnitude is below a characteristic error of the problem; or do we use a circular convolution wherein the function is wrapped around on itself? Of course, these are the perennial choices facing users of the discrete Fourier transform (DFT).

If we are willing to employ the DFT for both the signal samples and the samples of the wavelet on the chosen lattice, then we can speed the process up greatly by computing the inverse DFT of the product of two Fourier-coefficient series, one such for each choice of scale made. The only restriction is that the number of samples along the translation axis must equal the number of samples of the signal. This becomes particularly wasteful for large scales, since b_ℓ is now $\ell \Delta t$ instead of the sparse $\ell b_0 a_m$. A way around this objection is to resample the signal—not a difficult undertaking in this case. This approach also automatically determines which method of end-effect treatment will be used as the DFT is a circular transform, whether or not padding is used. It is a straightforward matter to count the number of operations needed by each of these methods once the evaluation lattice is specified. Details, such as obtaining an optimal number of samples over the nominal support of the wavelet as its scale changes, are quite simple, but beyond the scope of this paper.

1.3 Better methods of approximation

Other authors have discussed methods of implementing the CWT that go beyond the standard approaches discussed above. Cassasent and Shenoy⁴ proposed a method of effecting a fast, continuous, Gabor transform primarily to overcome the problem of shift noninvariance endemic to discrete-wavelet-transform analysis. In addition, their method retains scale flexibility and capability to use a wide variety of mother wavelets. In these respects, they have accomplished the goals of this work. However, our requirements include that of reconstruction for an interactive wavelet filter in addition to that of detection of signals that may be masked by noise or other contamination, so our solution is somewhat different. Another recent discussion of continuous wavelets is by Antoine and Murenzi⁵ who present a general view of the transform without emphasis on the implementation issues that were of concern in Ref. 4 and the present paper.

The goal is to find a method of computing an approximation to Eq. 1 while retaining the “flavor” of continuous functions for both the signal and the mother wavelet as long as possible, keeping the constraint of rapid computations in view as well as retaining flexibility of scale and preserving shift invariance. At the point where sampling for display, representation, and interactive reconstruction purposes becomes necessary, the $\{a_m, b_\ell\}$ -lattice must be fixed, thereby losing some semblance of shift invariance to within the discrete period of the translation grid.

We accomplish our goal by using a Fourier-domain version of Shannon's sampling theorem to reconstruct the continuous Fourier transform of the signal from its Fourier coefficients obtained via the DFT from its time-domain samples. There are four benefits so this approach: (1) more accurate samples of the wavelet family are easily computed, (2) separation of the wavelet into two factors,

one dependent on a and the other on b , makes the computations easier and faster, (3) additional flexibility in choosing the range of scales make for a more efficient description of the transformed signal, and (4) shift invariance. The actual computations involve only matrix operations and discrete Fourier transforms and their inverses (for reconstruction). The complexity of the algorithm is then $O(N \ln N)$, where N is the number of signal samples. There is an overall multiplicative factor that depends on the number of lattice points needed to display or manipulate the transformed signal, also the number of elements in the matrix is proportional to N , not N^2 . If the lattice is altered by choosing a different translation grid, the frequency-space portions of the chosen mother wavelet do not need to be resampled. The shift invariance of the signal is maintained in the frequency representation up to an overall phase, as is shown below. This implies that the magnitude of the wavelet coefficient, $|\tilde{f}(a, b)|$, is invariant with respect to shifts in the time axis—as long as the overall energy of the signal in the time window is not altered by the shift.

2. APPROXIMATING THE CWT

2.1 Continuous wavelets in the time domain

Suppose that $f(t)$ is a continuous function in the time domain where $f \in \mathcal{L}^2$ with $t \in \mathcal{R}$. Further, suppose that f satisfies the requirements of Shannon's sampling theorem, namely that f is band-limited in that the highest frequency component in the Fourier transform of f is less than some $\Omega < \infty$. (In some sense, continuity is equivalent to limiting the bandwidth in frequency space, but we will not worry about such niceties here.) Suppose there are N samples of f taken in the usual manner where the Nyquist condition, $\Delta t \leq \frac{1}{2\Omega}$, has been met. Then the sampling theorem, discussed in Chapter 5 of Ref. 3, lets us recover the function from its samples $f_k = f(t_k)$ where $t_k = k \Delta t$, according to

$$f(t) = \sum_{k \in \mathbb{Z}} f_k \frac{\sin 2\pi \Omega (t-t_k)}{2\pi \Omega (t-t_k)} = \sum_{k \in \mathbb{Z}} f_k s_\Omega(t_k), \quad (3)$$

where $s_\Omega(\cdot)$ is the sinc function with width $2 \Delta t$. The wavelet transform of f from Eq. 1 is then,

$$\tilde{f} = \langle \psi, f \rangle_{\mathcal{R}} = \sum_{k \in \mathbb{Z}} f_k \langle \psi, s_\Omega(t_k) \rangle_{\mathcal{R}}, \quad (4)$$

where the notation $\langle f, g \rangle_{\mathcal{R}} = \int_{\mathcal{R}} f^* g$ indicates an inner product in the appropriate space, denoted here by \mathcal{R} . If one could let Ω increase without bounds, the wavelet transform of the sinc function would become increasingly peaked at $t = t_k$ and the inner product would approach the value of the wavelet sampled at t_k . This is the justification for the direct method discussed in the Introduction. However, the signal $f(t)$ is band-limited, with Ω determined by the *a priori* sample rate, so we are *not* at liberty to increase the bandwidth. In practice, selecting samples of the wavelet instead of computing the integrals indicated in Equation 4 introduces errors of perhaps 5 to 10% or more. The actual error depends on how rapidly the wavelet is changing between $t_{k-1} < t < t_{k+1}$, and thus gets larger the smaller the scale becomes. For large scales, the wavelet is slowly varying, so the error is usually negligible. However, much of the important information about a signal is often found at small scales where the error becomes quite noticeable. Also note that if the behavior of \tilde{f} on a and b is made explicit, there is no convenient separation of $\langle \psi, s_\Omega(t_k) \rangle$ into a function of a and a function of b that would simplify the numerical computations. That is, if one decides to shift the scale or the position of the wavelet, the entire three-dimensional array $\psi_{a,b}(t_k)$ must be recomputed.

The second method mentioned in the Introduction also suffers from the above problems and requires additional computations for the inverse Fourier transform of the frequency-space convolutions. Nevertheless, the second method is computationally more efficient since convolutions computed via the DFT can be much faster than the many time-domain convolutions required for the first method.

The correct basis set to use for analyzing the signal is then not $\psi_{a,b}$ but rather the wavelet transform of the sinc function as indicated in Eq. 4. The implied extra computations make time-domain approximations to the CWT particularly inefficient. However, there is a way to overcome both problems of introducing errors and inverse DFT calculations, while avoiding the integrals over the sinc functions.

2.2 Continuous wavelets in the frequency domain

Suppose, for the moment, that f is time-limited to $|t| \leq T$ instead of band-limited. Then an analog of the Shannon sampling theorem can be established in the frequency domain using samples of $\hat{f}(\omega)$ at $\omega_k = k \Delta \omega$, where $\Delta \omega \leq \frac{1}{2T}$, to reconstruct the continuous function. Thus,

$$\hat{f}(\omega) = \sum_{k \in \mathbb{Z}} \hat{f}_k \frac{\sin 2\pi T(\omega-\omega_k)}{2\pi T(\omega-\omega_k)} = \sum_{k \in \mathbb{Z}} \hat{f}_k s_T(\omega_k), \quad (5)$$

where the Fourier transform of a function, f , is defined as

$$\hat{f}(\omega) = \langle e_\omega, f \rangle_{\mathcal{R}} = \int_{\mathcal{R}} e_\omega^* f(t) dt = \int_{\mathcal{R}} e^{-2\pi i \omega t} f(t) dt, \quad (6)$$

and the Fourier basis set is given by $e_\omega(t) = e^{2\pi i \omega t}$. Following Ref. 3, we choose the Fourier integration constant to be 2π , which preserves symmetry between the transform and its inverse, allowing frequency to be measured in cycles per unit time or distance. The wavelet transform of f , using Parseval's identity to work in the frequency domain, becomes

$$\tilde{f} = \langle \psi, f \rangle_{\mathcal{R}} = \langle \hat{\psi}, \hat{f} \rangle_{\mathcal{R}} = \sum_{k \in \mathbb{Z}} \hat{f}_k \langle \hat{\psi}, s_T(\omega_k) \rangle_T. \quad (7)$$

The inner product of $\hat{\psi}$ with the frequency-domain sinc function may be rewritten using Parseval's identity once more. The result is

$$\langle \hat{\psi}, s_T(\omega_k) \rangle_T = \langle \psi, \hat{s}_T \rangle_T, \quad (8)$$

where the inverse Fourier transform of the frequency-domain sinc function is given by

$$\hat{s}_T = \frac{1}{2T} e^{2\pi i t \omega_k}, \quad (9)$$

for $|t| < T$, and zero otherwise. Expressing the inner product in its integral form, gives

$$\langle \hat{\psi}, s_T(\omega_k) \rangle_T = \langle s_T(\omega_k), \hat{\psi} \rangle_T^* = \frac{1}{2T} \left(\int_{-T}^T \psi(t) e^{-2\pi i t \omega_k} dt \right)^*. \quad (10)$$

Our signal is band-limited rather than time-limited, so we may let T increase as much as necessary to compute accurate values of the wavelet transform. This retains the spirit of Equation 5 and lets us replace the inner product with a sample of the frequency-space function, achieving as much accuracy as required. For the time-domain case, this desire to sample the wavelet was prohibited by the predetermined value of Ω inherent in the set of samples, $\{f_k\}$. The inner product in Equation 10 becomes

$$\langle \hat{\psi}, s_T(\omega_k) \rangle_T = \langle s_T(\omega_k), \hat{\psi} \rangle_T^* \approx \left(\int_{-\infty}^{\infty} \psi(t) e^{-2\pi i t \omega_k} dt \right)^* = \hat{\psi}^*(\omega_k), \quad (11)$$

so the wavelet transform of the signal $f(t)$ with respect to the mother wavelet $\psi(t)$ is closely approximated by

$$\tilde{f} = \sum_{k \in \mathbb{Z}} \hat{f}_k \hat{\psi}^*(\omega_k). \quad (12)$$

3. A FAMILY OF GAUSS-HERMITE WAVELETS

3.1 Time-domain functions

Our preferred example, used to illustrate the above ideas and for many of the calculations presented below, is the Gauss-Hermite family of wavelets. The generator of this family is the unadorned Gaussian

$$g_0(t) = e^{-t^2}, \quad (13)$$

and family members g_n , for $n = 1, 2, \dots$, are given by the n^{th} derivative of g_0 as

$$g_n(t) = \partial_t^{(n)} g_0(t). \quad (14)$$

Since the generator of the Hermite polynomials is g_0 , the family of wavelets we are interested in is given by

$$g_n(t) = (-1)^n g_0(t) H_n(t), \quad (15)$$

where $H_n(t)$ is the n^{th} Hermite polynomial in t ; for example, $H_3(t) = 12t - 8t^3$. The preference is that each family member should "look like" any other over its domain and range. Of course, the support or domain of these functions is the real number line, but for practical reasons we will later restrict it to values of t for which $g_0 \leq \varepsilon_p$, where ε_p is some problem-dependent error such as the least-significant bit of the digitizer used to obtain the samples of the signal. We can ensure that the range of the family is approximately $[-1, 1]$ by requiring that g_n for n even, which are symmetric functions, be equal to 1 at the origin and that g_n for n odd, which are antisymmetric functions, have positive slope at the origin. These specifications determine an overall normalization constant that depends on n . The normalized Gauss-Hermite wavelet family is then

$$g_n(t) = (-1)^{\lfloor \frac{n}{2} \rfloor} \frac{\Gamma(\frac{n+2}{2})}{\Gamma(n+1)} g_0(t) H_n(t), \quad (16)$$

where $\lfloor \cdot \rfloor$ represents the “floor” operation or the integer part of the argument and $\Gamma(\cdot)$ is the gamma function. The first 4 members of the wavelet family are shown in Figure 1, with the Gaussian generator superimposed as a dashed line.

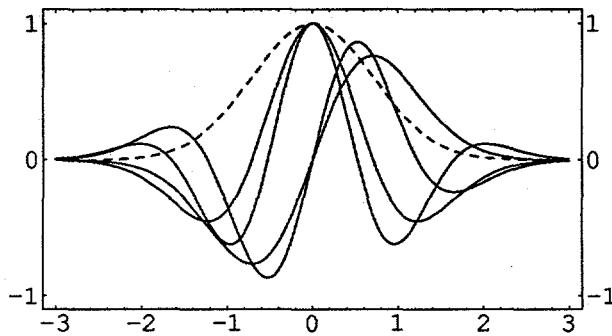


Figure 1. First four members of the Gauss-Hermite family of mother wavelets ($n=1,2,3,4$). The “grandmother” function, g_0 , is shown as a dashed line. As n increases, either the slope at zero increases (n odd) or the width around zero decreases (n even).

Nothing done so far restricts us to Gaussian functions; the only requirement is that the admissibility criterion be met. The simplest statement of this criterion is that the wavelet have no dc component or that the average of the function over its support is zero.

3.2 Frequency-domain functions

We now wish to examine the Fourier transform of the wavelet family guided by the analysis given in Section 2. The Fourier transform of g_n , represented by the “hat” notation, is

$$\hat{g}_n(\omega) = \frac{(2\pi\omega)^n}{2} \frac{\Gamma(\frac{n}{2})}{\Gamma(n)} \hat{g}_0(\omega) \quad (17)$$

where $\hat{g}_0 = \sqrt{\pi} e^{-\pi^2 \omega^2}$ is the Fourier transform of the Gaussian generating function. The normalization has been altered to ensure that these functions are real, and hence represent analytic or Hardy functions. This is primarily a matter of convenience that allows an extension to complex wavelets via the Hilbert transform. It also changes the normalization for the time-domain versions by making the n -odd functions pure imaginary. This causes no problems in the computer calculations—simply ignore the factor of $\sqrt{-1}$ should we have need to consider the time-domain.

Figure 2 shows Fourier transforms of several of the mother wavelets. The function \hat{g}_0 is shown in the left plot as a dashed line. Like their time-domain counterparts, these functions exhibit even and odd symmetry depending on n . Also note that they all are zero at $\omega = 0$, so they are all admissible as mother wavelets.

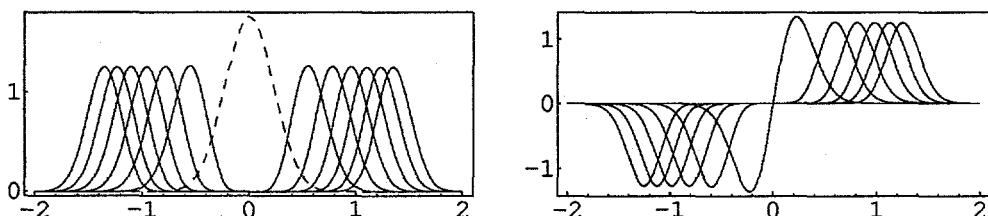


Figure 2. Gauss-Hermite wavelets in frequency space. The curves on the left are the values of \hat{g}_n for $n = 6, 12, 18, 24, 30, 36$ with \hat{g}_0 shown as a dashed line; on the right, for $n=1, 7, 13, 19, 25, 31$ and 37. The distance of each graph from the origin increases with increasing n .

This illustration is meant to show the utility of the Gauss-Hermite family in providing a tight cover of the frequency axis. This family of mother functions is able to analyze the behavior of a wide variety of signals, having a broad range of frequencies, accomplished by choosing an appropriate value of n as well as a range of scales, a . Examples of this versatility are given below.

There is an added benefit to the frequency-space formulation: fractional derivatives are now possible. The fractional-derivative Gaussian functions are all admissible as wavelets and excel in analyzing the behavior of signals near zero frequency. As an example, consider Figure 3 where the frequency-space behavior for several fractional wavelets is plotted for $0 \leq \omega \leq 0.1$. The top three curves are for derivative orders of 0.01, 0.1, and 0.5, while the bottom curve is for the usual Mexican-hat wavelet (second order

derivative). The fractional-derivative wavelets with derivative orders less than 1 are clearly more sensitive to low-frequency behavior than the integral-derivative orders. In particular, the slope of these functions at zero is much larger than 1 while that of the usual functions is much smaller than 1. The only drawback to this family of functions is a minor one in that they have no vanishing moments beyond the zeroth.

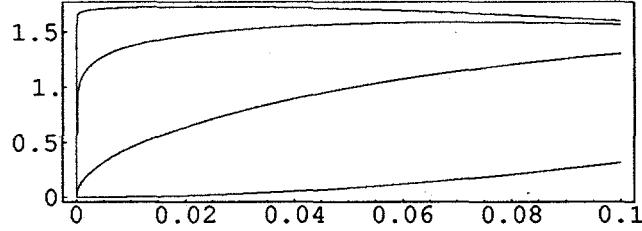


Figure 3. Several members of the set of fractional-derivative wavelets in frequency space. The ordinate is frequency, from 0 to 0.1 normalized units. The functions, from top to bottom, are $\hat{g}_{0,01}$, $\hat{g}_{0,1}$, $\hat{g}_{0,5}$, and \hat{g}_2 .

3.3 Wavelets on the translation-scale lattice

The modification to the expressions for g and \hat{g} that include wavelet scale and position follows directly from Eq. 2. The family of Gauss-Hermite wavelets in frequency space is a function of n , a , and b as

$$\hat{g}_{n,a,b}(\omega) = \frac{(2\pi a \omega)^n}{2} \cdot \frac{\Gamma(\frac{n}{2})}{\Gamma(n)} |a|^{1-p} e^{-2\pi i \omega b} \sqrt{\pi} e^{-\pi^2 a^2 \omega^2}. \quad (18)$$

Again, suppose there are N frequency samples of the signal function from the Nyquist frequency, $\omega_N = \Omega$ down to zero in steps of $\Delta\omega = \frac{1}{N} \omega_N$ as computed from the DFT of the N time-domain samples, and we wish to analyze the signal's behavior in the range, $\omega_{\min} \leq \omega_k \leq \omega_{\max}$, where $\omega_{\min} \geq 0$ and $\omega_{\max} \leq \omega_N$. By the analysis of Section 2.2, it is sufficient to sample the mother wavelet at each frequency sample of interest, as well as on a lattice of scale and translation samples indicated in Equation 18. If the translation grid is not restricted to be functionally dependent on scale, we can separate the frequency-space wavelet into two independent factors at the expense of greater translation redundancy, but allowing faster computations due to computer-optimized matrix operations. The separation is shown in Equation 19 as

$$\hat{g}_{n,k,m,j} = \Phi(b_j, \omega_k) \Psi^{(n)}(\omega_k, a_m). \quad (19)$$

Here, Φ is the "phase" matrix with dimensions $n_b \times n_\omega$ and $\Psi^{(n)}$ is the frequency-space "bump" matrix shown in Fig. 2; $\Psi^{(n)}$ has dimensions $n_\omega \times n_a$. The subscripts on the dimensions refer to the lattice variables. The frequency samples, $\{f_k\}$ can be represented as a diagonal matrix, Λ , with dimensions $n_\omega \times n_\omega$, so the discrete approximation to the CWT of a signal becomes

$$\tilde{f}_n = \Phi \Lambda \Psi^{(n)} = \sum_{j,m,k} \Phi_{j,k} \Lambda_k \Psi_{k,m}^{(n)}, \quad (20)$$

for the n^{th} member of the mother wavelet family. In light of the analysis of Section 2.2, Equation 20 is a "discrete" approximation only for the translation-scale lattice, and retains the full continuous nature of the signal and the wavelet functions as far as frequency is concerned, and by implication, time as well. A typical problem may have 8192 signal samples, require only 500 frequency steps, and be adequately represented on a translation-scale lattice of 64×32 grid points. Once the wavelet transform has been initialized by the sampling indicated in Eq. 19, the computation of the transform of the signal requires an FFT, which is quite fast, and two matrix multiplications indicated in Eq. 20. This latter computation takes a small fraction of a second on most any PC or workstation. Larger problems, requiring the full frequency range available in the sampled signal, take proportionally longer.

In principle, this formulation is valid for all $a, b \in \mathbb{R}$; in practice, there are certain limitations that depend on the precision required and the bandwidth of the function being analyzed. A convenient parametrization for the scale is the octave-based one shown in Equation 21,

$$a_m = a_0 2^{\lambda m}, \quad (21)$$

where $\lambda = \frac{q}{M}$, M is the maximum value of m , and there are q octaves spanned by the set $\{a_m\}$. If q divides M , there are M/q voices in each octave. The scale of the scale, so to speak, is controlled by a_0 . For small a_0 , we have a voiced octave scale for analyzing small features of the signal, and analogously for large a_0 . This control over a_0 helps us avoid those scales that make no physical

sense. For example, the smallest value of a_0 that has physical meaning is $\frac{1}{\Omega}$. Likewise, the largest scale that one should consider is the scale of the sampled segment of the signal under analysis, or $a_0 = \frac{N}{2\Omega}$. In actual applications, signal features of interest will most probably lie between these two extremes, resulting in a practical range of $1 < \Omega a_0 < \frac{N}{2}$ with the choice of λ and M determined by the particular problem being analyzed. If the translation grid is parametrized as

$$b_\ell = \ell \Delta b, \quad (22)$$

for $\ell = 0, \dots, n_b - 1$, corresponding to each value of the scale, a_m , there is an optimum range for the translation spacing—namely $\Delta b = a_m$. This value is optimum in the sense that if it were smaller, the lattice would have more redundancy than necessary, and if it were larger, there would be too few coefficients for accurate reconstruction. The range of b is likewise limited by properties of the set of signal samples. For example, if we choose to analyze N samples of a signal, any other choice for the span of b than $N \Delta t = \frac{N}{2\Omega}$ would make little sense. This implies that the number of samples along b , is limited by

$$n_b \leq \frac{N 2^{-\lambda m}}{2 \Omega a_0}, \quad (23)$$

with additional limitations imposed by the allowable range of m given an a_0 . Although the matrix formulation presented above requires that b_0 be independent of scale, the linear expression, given by Eq. 20, does not. We can retain the efficiency of the frequency-domain representation, with its separability in a and b , while using the optimal lattice where samples of the translation variable get sparser as the scale increases. For the octave-based scale presented above, this reduces the computations by a factor of 2 at the cost of keeping track of the vectors along b that no longer present themselves in a convenient matrix representation.

The computer code written to implement this version of the CWT incorporates full flexibility over the $\{a_m, b_\ell\}$ -lattice as well as allowing the user to select the frequency range. Each change in sample rate or scale or frequency range necessitates a resampling of Φ or Ψ or both. The result is a graphical method of “zooming” in on scale and frequency range independently, within the physical constraints imposed by the data.

4. THREE APPLICATIONS

4.1 ESDS

The enclosed-space detection system, or ESDS, was developed at Oak Ridge for the Department of Energy to provide automated security at sensitive portals.⁶ The goal of the project was a system for reliable detection of persons hiding in vehicles passing through a security portal. A commercial device, incorporating the wavelet algorithm discussed here, is presently offered to installations requiring access security. Figure 4 shows a typical wavelet transform of a geophone measurement, illustrating the sensitivity of the CWT to a sequence of motions of the test vehicle. The ballistocardiac effect due to an “intruder” hiding in the vehicle induces shock motions of the order of 1 to 2 microns at the exterior surfaces (e.g., bumpers or undercarriage) where the geophones are temporarily placed. The prototype system used a conventional FFT method and could not always distinguish coherent heart beats taking place over the 20-second measurement from grosser motions due to wind and passing traffic. The wavelet coefficients from a properly chosen Gauss-Hermite wavelet clearly show about 15 to 18 heart-beat events during the measurement period, and make the search—no-search decision quite simple: search!

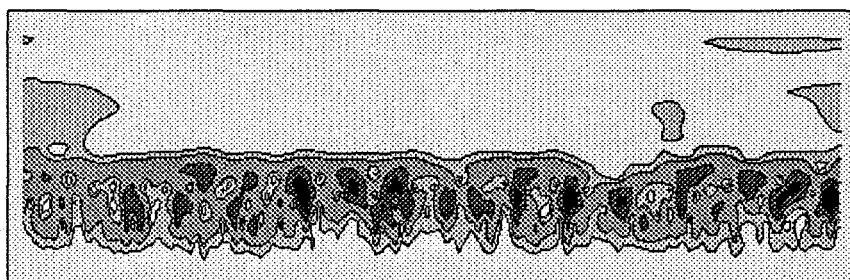


Figure 4. Gauss-Hermite wavelet transform of 20 sec of geophone measurement, sampled at 100 Hz, taken from a tractor-trailer rig with a person concealed in the front section of the trailer. The ordinate is the translation axis, representing time, and the abscissa is the wavelet scale, from about .03 sec at the bottom to about 9 sec at the top. The transform was computed with derivative order 6 to emphasize the compactness of the heart-beat blobs; p was set to 0.5 to give a balanced appearance across the scales shown; the frequency range is from 0.05 Hz to 12.0 Hz; and the scale spans a little over 8 octaves with 8 voices in each octave. The intensity of the wavelet coefficients is indicated by the shading, with black being the most intense.

A similar measurement taken with no intruder present, but with the usual background motions, shows activity in the same regions, but it is of a disjointed nature having no band structure over the duration of the measurement so evident above.

4.2 Voice analysis

Voice analysis, for compression, word recognition, and speaker identification, has been of interest to the wavelet community since its inception. Most of the wavelet effort in voice processing has been concerned with compression and coding techniques. There has been some voice analysis with continuous wavelets, most notable being the glottal-rate analysis by Kadambe and Boudreaux – Bartels⁷ who showed the glottal closing rate could be tracked by comparing the peak positions of main peaks across several scales. Here, we extend their notion to a large number of scales that may be thought of as generating a “continuous” feature or “blob”. Figure 5 shows the CWT of a portion of a phrase taken from the TIMIT Speech Corpus⁸, from the sentence designated therein as “abc0_sa2”.

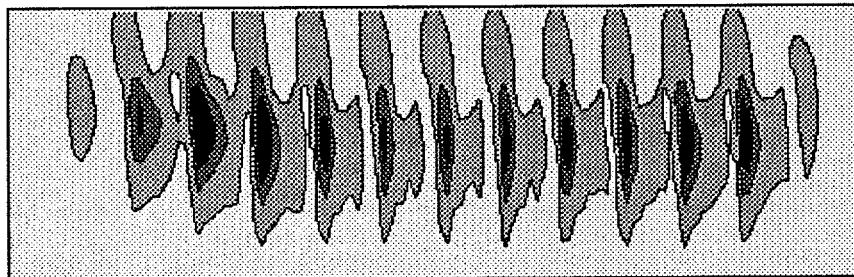


Figure 5. Gauss-Hermite transform of 0.128 sec of speech, sampled at 16000 Hz, taken from the TIMIT database. Shown is the vowel portion of “dark”. The scale ranges from about 0.20 ms at the bottom to about 2.60 ms at the top. The transform was computed with derivative order 2 to emphasize the glottal closures; p was set to zero to emphasize the large scale where glottal closure is important; the frequency range is from 100 Hz to 1000 Hz; and the scale spans about 3.6 octaves with 18 voices in each octave.

The time between centroids of each blob is a measure the glottal closing period and may be tracked as function of time. The (vertical) position in scale indicates the global frequency content of the glottal events as modified by the vocal tract.

Increasing the derivative order to bring out the events that are coherent over the duration of a vowel, as in Figure 6, emphasizes behavior that is associated with vocal formant frequencies. In particular, the ridges⁹ so evident in the transform coefficients may be followed in time¹⁰, giving an indication of how the formant frequencies are changing. The total picture, formant behavior and blob shape, gives a means of identifying the individual uttering the phrase as well as possible means of recognizing the set of phones comprising the word.

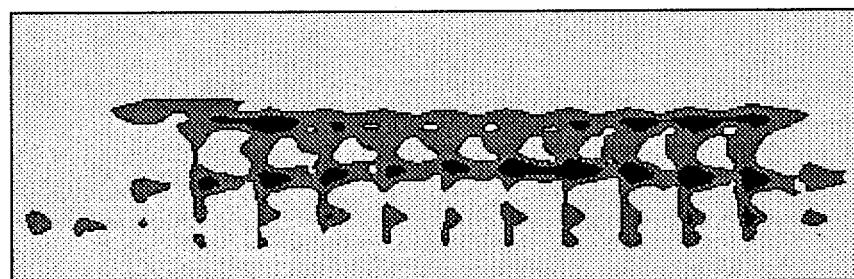


Figure 6. Gauss-Hermite transform of the same speech segment shown in Fig. 5. The scale ranges from about 0.03 ms at the bottom to about 13 ms at the top. The transform was computed with derivative order 50 to bring out the formant frequencies that are coherent over the time of utterance; p was set to 2 to emphasize the smaller scales; the frequency range is from 100 Hz to 3200 Hz; and the scale spans a little over 5 octaves with 12 voices in each octave.

A particularly interesting observation of Fig. 6 is that the three main formants seem to have quite different behaviors in scale as a function of time. Top-most ridge (at large scales) decreases, then rises slightly as time increases; the middle ridge has the opposite behavior. The bottom ridge seems to cross the middle ridge, drop to a minimum, and then level off. It is such behavior, coupled with

the shape of the regions shown in Fig. 5, that may provide information concerning the speaker identity, the phones uttered, and the inflection or emotional state of the person speaking.

4.3 Audio forensics

In general terms, audio forensics deals with extracting evidence from tape recordings that have been made during interview processes, calls to police and 911, home answering machines, and similar situations where information concerning a felony may have been obtained. The practitioner of audio forensics is often given an analog tape (the newer digital equipment is usually too expensive for local law enforcement organizations) that may have been damaged or recorded under less than ideal conditions.

Figure 7 shows a segment of an audio tape submitted for forensic analysis. The microphone used was located near a noisy ceiling fan and the events of interest occurred in an adjoining room. The segment contains a sound event not associated with the background noise.

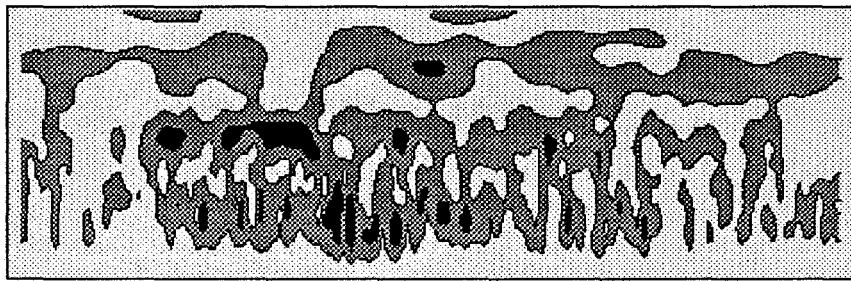


Figure 7. Gauss-Hermite transform of about one-tenth sec of a noisy audio tape, sampled at 22050 Hz. The scale ranges from about 0.14 ms at the bottom to about 5.7 ms at the top. The transform was computed with derivative order 10 to enhance the coherence of events that last longer than the background noise; p was set to 1.4 to emphasize the smaller scales; the frequency range is from 100 Hz to 3200 Hz; and the scale spans a little over 5 octaves with 12 voices in each octave.

A listener can perceive an event by listening to the tape, perhaps a footstep or a door closing, but the background makes identification impossible. The dark blob in the middle of the scale range, located near the center of the first half of the figure, has a coherence time and a scale that distinguish it from the background. Reconstruction of this sound fragment from the wavelet coefficients in the surrounding region provides a considerably clearer representation of the event than if taken directly from the original recording. Of course, this is merely an application of a digital filter, but of a type whose response is interactively guided by the user who visually examines the results of the transform and selects the regions for reconstruction.

5. COMPLEX WAVELETS

The Gauss-Hermite wavelet family described above consists of real functions, therefore its wavelet transform of a real signal has constant phase. In certain applications, the behavior of the phase of the transformed function is useful in locating scale-dependent phase "discontinuities" as might occur during speed changes in rotating machinery or during vowel transitions in the vocal tract. We can restore the phase-analysis capability to the Gauss-Hermite family, and to any real mother wavelet, by computing the imaginary part of an analytic function whose real part is the wavelet itself. The idea is a simple one from the theory of functions of a complex variable and may be found in the standard references. Here, we follow the treatment of Morse and Feshback¹¹ and suppose $f(z) = u(z) + i v(z)$ is an analytic function with complex argument $z = x + i y$. The canonical example of such a function is e^{iz} where $u(x) = \cos x$ and $v(x) = \sin x$ are the real and imaginary parts of the function on the real axis. Identify the real part of $f(z)$ on the real axis with the Gauss-Hermite wavelet as $u(x) \longleftrightarrow \hat{g}_n(\omega)$. An application of the Cauchy integral theorem allows us to compute the imaginary part of the function, also on the real axis. The Hilbert transform is the name given to this calculation, and is defined as

$$v(x) = -\frac{1}{\pi} \mathcal{P} \int \frac{u(\xi)}{\xi - x} d\xi, \quad (24)$$

where the integration is over the real axis and \mathcal{P} indicates the principal value of the integral (there is a similar transform for u in terms of v). The Hilbert transform has the property that it is orthogonal to the original (real-valued) function in analogy with the cosine and sine pair. Denote the Hilbert transform of a real function, f , by f^\dagger , then the quadrature functions to the Gauss-Hermite wavelet family are

$$\hat{g}_n^\dagger(\omega) = (-1)^{\frac{n+1}{2}} \pi^2 e^{-\pi^2 \omega^2} \frac{\Gamma(\frac{n+2}{2})}{\Gamma(\frac{n+1}{2})} L_{\frac{n-1}{2}}^{-\frac{n}{2}}(\pi^2 \omega^2) \quad (25)$$

for n odd, and

$$\hat{g}_n^\dagger(\omega) = (-1)^{\frac{n}{2}} \pi^2 (\pi \omega) e^{-\pi^2 \omega^2} L_{\frac{n-1}{2}}^{-\frac{n-1}{2}}(\pi^2 \omega^2) \quad (26)$$

for n even. Here, $L_n^a(\cdot)$ is the associated Laguerre polynomial¹² of order n . Note that $\hat{g}_n(\omega)$ is odd (even) for n odd (even) and that $\hat{g}_n^\dagger(\omega)$ is even (odd) for n odd (even); thus, the product is always an odd function and the inner product is zero showing that the functions are orthogonal. These closed-form expressions have limited practical use, other than in plotting graphs for illustration purposes. In the actual computer calculations, the standard FFT-based method of computing the sequence of Hilbert-transform values from samples of the frequency-space wavelet was used, with the caution that we must extend the wavelet to regions where it goes to zero smoothly since “discontinuities” in the samples can cause large errors in the numerical algorithm. Figure 8 shows two examples of the wavelet (dashed) and its Hilbert transform (solid).

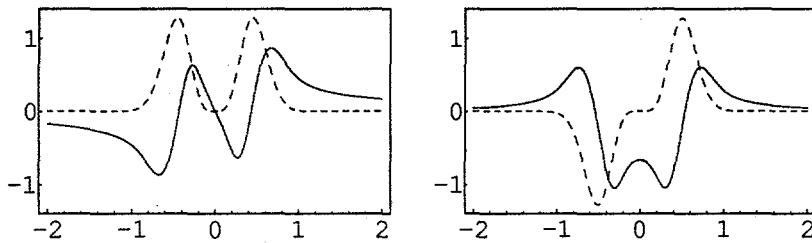


Figure 8. Real and imaginary parts of the complex Gauss-Hermite wavelets in frequency space. The ordinate is normalized frequency. The dashed curves show the real parts and the solid, the imaginary parts. The curves on the left are for $n=4$, and for $n=5$ on the right.

Equations 25 and 26 display closed-form expressions for the Hilbert transform of the mother wavelet, that is, for $a = 1$ and $b = 0$. The general expression for arbitrary a and b is vastly more complicated and correspondingly less illuminating. Since we use a numerical method for the actual computations, it is not necessary to be aware of the closed-form expressions, and will say no more about them. The essential point to this digression, is that we can extend a real-wavelet function to a complex function of a real variable; the resulting complex function has both a magnitude and a phase, which has great utility in many applications.

6. SUMMARY

There are numerous advantages to the frequency-space formulation presented above. Here, we summarize the most important ones, noting that both the development presented above and the applications of the CWT shown are ongoing projects. The most direct advantage, stemming from the shift invariance of the Fourier transform, is that coefficients of the transformed signal are invariant to shifts along the time axis. The problem of shift invariance was a key feature to the exposition of Ref. 4 and is crucial to pattern-recognition applications, as the authors noted.

The standard methods of computing the CWT by sampling wavelet functions in the time domain introduces errors in the transform; the smaller the scale, the larger these errors become, defeating one of the reasons for using the wavelet transform. The frequency-domain representation avoids this problem without needing to compute the “correct” basis set—namely the wavelet transform of the sinc function. Additionally, the frequency-space representation allows separation of the scale and translation variables, greatly facilitating the numerical algorithms. In the case of the Gauss-Hermite family, an additional bonus is that the part of the function depending on the order, n , is independent of the Gaussian “bump” part. Thus, changing the order does not entail recomputing the two-dimensional array of exponential functions—merely raising the existing array representing the factor $(2\pi a\omega)$ to the new power of n and multiplying by the existing Gaussian array.

Another advantage of the frequency-space formulation is that fractional-derivative wavelet analysis—for any mother wavelet—becomes possible. This is a powerful analysis technique for derivatives of order $0 < x < 1$. Such wavelets excel at getting close to the origin in frequency space and so have a distinct advantage for a particular class of problems.

We also showed how to compute the complex part of a real wavelet, enabling phase analysis for a large class of mother wavelets. The extension from a real wavelet to its complex representation via the Hilbert transform is easier to compute numerically in

frequency space, primarily since the DFTs necessary are already in hand. Phase-analysis capability for such an important class of wavelets as the Gauss-Hermite family greatly extends its utility.

ACKNOWLEDGMENTS

This research was performed at Oak Ridge National Laboratory, managed by Lockheed Martin Energy Research, Inc., for the U. S. Department of Energy under Contract No. DE-AC05-96OR22464. The enthusiastic support of Vivian Baylor, Leo Labaj, and Tom McCoig and their help with the application areas is gratefully recognized. Special thanks to colleagues Steve Kercel and Raymond Tucker for many useful suggestions, improvements, comments, and moral support during the major part of this work.

REFERENCES

1. Ingrid Daubechies, "The Wavelet Transform, Time-Frequency Localization and Signal Analysis", *IEEE Trans. Inform. Theory*, 36(5), 961-1005, 1990.
2. Ronald E. Crochiere and Lawrence R. Rabiner, "Interpolation and Decimation of Digital Signals—A Tutorial Review", *Proc. IEEE*, 69, pp. 300-331, 1981.
3. Gerald Kaiser, *A Friendly Guide to Wavelets*, Birkhäuser, Boston, MA, 1994.
4. David Cassasent and Rajesh Shenoy, "New Gabor wavelets with shift-invariance for improved time-frequency analysis and signal detection", *Proc. SPIE Wavelet Applications*, 2762, pp. 244-255, 1996.
5. Jean-Pierre Antoine and Romain Murenzi, "The continuous wavelet transform, from 1 to 3 dimensions", Chapter 5 in *Subband and Wavelet Transforms*, edited by Ali N. Akansu and Mark J. T. Smith, Kluwer Academic Press, Norwell, MA, pp. 149-187, 1996.
6. Stephen W. Kercel, Vivian M. Baylor, William B. Dress, Tim W. Hickerson, W. Bruce Jatko, Leo E. Labaj, Jeff D. Muhs, and Richard M. Pack, "Application of the smart portal in transportation", *Proc. SPIE*, 2902, edited by Alan Chachich, in press, 1996.
7. S. Kadamb and G. F. Boudreux-Bartels "Application of the Wavelet Transform for Pitch Detection of Speech Signals," *IEEE Trans. on Inf. Th.*, 38(2), pp. 917-924, 1992.
8. John S. Garofolo, Lori F. Lamel, William M. Fisher, Jonathan G. Fiscus, David S. Pallett, and Nancy L. Dahlgren, *DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM*, NIST Speech Disc 1-1.1, National Institute of Standards and Technology, 1990.
9. Nathalie Delprat, "Global Frequency Modulation Laws Extraction from the Gabor Transform of a Signal", *IEEE Trans. Speech and Audio Proc.*, 5(1), pp. 64-71, 1997.
10. N. Delprat, B. Escudié, P. Guillemain, R. Kronland-Martinet, Ph. Tchamitchian, and B. T. Torrésani, "Asymptotic Wavelet and Gabor Analysis: Extraction of Instantaneous Frequencies", *IEEE Trans. Inf. Theor.*, 38(2), pp. 644-664, 1992.
11. Philip M. Morse and Herman Feshbach, *Methods of Theoretical Physics*, McGraw-Hill, New York, pp. 370-374, 1953.
12. Urs W. Hochstrasser, *Handbook of Mathematical Functions*, 7th printing, U. S. Government Printing Office, edited by Milton Abramowitz and Irene A. Stegun, pp. 772-792, 1968.