

Adaptive “chirplet” transform: an adaptive generalization of the wavelet transform

Steve Mann*

Simon Haykin, MEMBER SPIE

McMaster University

Communications Research Laboratory

1280 Main Street West

Hamilton, Ontario, L8S 4K1 Canada

Abstract. The “chirplet” transform unifies many of the disparate signal representation methods. In particular, the wide range of time-frequency (TF) methods such as the Fourier transform, spectrogram, Wigner distribution, ambiguity function, wideband ambiguity function, and wavelet transform may each be shown to be a special case of the chirplet transform. The above-mentioned TF methods as well as many new ones may be derived by selecting appropriate 2-D manifolds from within the 8-D “chirplet space” (with appropriate *smoothing kernel*). Furthermore, the chirplet transform is a framework for deriving new signal representations. The chirplet transform is a mapping from a 1-D domain to an 8-D range (in contrast to the wavelet, for example, which is a 1-D to 2-D mapping). Display of the 8-D space is at best difficult. (Although it may be displayed by moving a mesh around in a 3-D virtual world, the whole space cannot be statically displayed in its entirety.) Computation of the 8-D range is also difficult. The adaptive chirplet transform attempts to alleviate some of these problems by selecting an optimal set of bases without the need to manually intervene. The adaptive chirplet, based on expectation maximization, may also form the basis for a classifier (such as a radial basis function neural network) in TF space.

Subject terms: adaptive signal processing; chirplet; wavelet; expectation maximization; neural networks.

Optical Engineering 31(6), 1243–1256 (June 1992).

1 Introduction

The “chirplet transform,” first proposed in Ref. 1, is an extension of the well-known wavelet transform.² Informally speaking, a chirplet may be regarded as a “piece of a chirp” (windowed swept-frequency wave) in the same manner a wavelet could be loosely regarded as a “piece of a wave” (windowed tone).

The wavelet transform consists of an expansion of an arbitrary signal onto a set of bases that is affine in the physical (e.g., time) domain. Thus, for the purpose of our discussion, we will refer to the wavelet transform bases as the “physical affinities.” A particular *mother wavelet* is chosen and the other physical affinities are derived through affine coordinate transformations in the physical domain. (In 1-D there are two free parameters: translation and dilation.)

Time-frequency (TF)-affine chirplets extend the wavelet idea by requiring only that the basis functions be derivable through affine transformations in the TF domain rather than being limited to the physical (time) domain as is the case

with the wavelet transform. Since physical affinities are a subset of the TF affinities, the chirplet transform embodies the wavelet transform as a special case. Furthermore, the short-time Fourier transform (STFT) turns out to be another special case of the chirplet transform.

A second generalization of the wavelet transform, known as the “perspectives”³ (or the projective chirplet) is based on a camera metaphor. The 1-D wavelet family may be regarded as a series of 1-D pictures of the mother wavelet with two free parameters: camera shift and camera zoom. Zoom corresponds to dilation and shift to translation. The extension from the wavelet to the projective chirplet is simple: Perspectives are produced by allowing the make-believe camera to tilt, so that the mother wavelet (generating function) and the film plane (other member of the family) are no longer parallel. Thus, including this second wavelet generalization, the 1-D chirplet family embodies eight free parameters: six TF-affine parameters,⁴ and two additional nonaffine (perspective) parameters.

The high number of parameters sometimes becomes a bottleneck. Therefore, an adaptive algorithm that overcomes this limitation was formulated and is presented here. A

*Current affiliation: Massachusetts Institute of Technology, 20 Ames Street, Room E15-350, Cambridge, Massachusetts 02139.

Paper SP-011 received Dec. 24, 1991; revised manuscript received March 6, 1992; accepted for publication March 9, 1992. This is a revision of a paper presented at the SPIE conference on Adaptive Signal Processing, July 1991, San Diego, Calif. The paper presented there appears (unrefereed) in SPIE Proceedings Vol. 1565. ©1992 Society of Photo-Optical Instrumentation Engineers. 0091-3286/92/\$2.00.

³To maintain a unitary, linear time-frequency expansion, one is free to limit oneself to a five-parameter area-preserving-TF-affine space. To do otherwise requires the use of nonlinear methods based on multiple primitives. Two such methods, however, have already been discovered,⁵ thus, for the sake of completeness, we will include all six parameters in our discussion, where the reader is free to strip away the TF-area parameter if linearity is desired.

further advantage of the adaptive chirplet is its utility as a classifier in pattern recognition tasks.

Chirplets have been successfully applied to many practical applications such as radar imaging, machine vision, and characterizing the acceleration signature of objects¹ using Doppler. Chirplets have also been applied to more artistic endeavors,⁴ such as computer-enhanced "light painting" (camera parameter estimation with respect to composites of multiple exposures). In this paper, only the 1-D chirplet is presented. The extension to two or more dimensions is not particularly difficult.

The adaptive chirplet may be regarded as a new distance metric for a radial basis function (RBF) neural network. A rule for positioning the *centers*^b in TF space, which is based on the well-known expectation maximization (EM) algorithm, is also presented. The algorithm, for which we have coined the term "logon expectation maximization" (LEM) adapts a number of *centers* in the TF space in such a way that it fits the input *distribution*. Two variants of LEM are presented: LEM1, which works on the marginals of the TF distribution sequentially, and LEM2, which works in the TF domain. Thus a connection between the adaptive chirplet EM and the RBF neural network is established.

Time-frequency contours, (approximately) circular in shape, in the TF space are allowed to dilate into ellipses of arbitrary aspect ratio to embody both the STFT and physically affine (wavelet) spaces as special cases.

The ellipses may also adaptively "tilt," if desired. (In other words the time series associated with each *center* may chirp.)

An alternative chirplet space, the "bowtie" space,¹ is also presented in the context of LEM. In that space, the adaptivity appears as a number of bow-shaped *centers*, which also move about to fit the input distribution.

2 Review of the Nonadaptive Chirplet

Before discussing the adaptive chirplet, we provide here a brief outline of the nonadaptive chirplet, following Ref. 6.

The well-known wavelet transform was derived through 1-D affine transformations in the physical (e.g., time) domain. A basic primitive known as a *mother wavelet* is chosen and all other basis functions are derived from it by affine transformations in the physical domain. In one dimension, the wavelet family is given by $\psi_{ab}(t) = \psi(at + b)$. These "physical affinities" have two free parameters, a and b .

The proposed chirplet bases, however, are derived through a number of additional transformations in the TF (e.g., Wigner) plane. The transformations are actually applied in either the physical (time) or the Fourier (frequency) domain, but are interesting because of the intuitive significance each of the transformations has in the TF domain. We may characterize the chirplet by two equations, one designating the transformations in the physical domain:

$$\psi_{wpq} = w(t)\psi[p(t)] \exp[q(t)] \quad (1)$$

the other indicating the transformations in the Fourier domain:

$$\Psi_{WPQ} = W(f)\Psi[P(f)] \exp[Q(f)] \quad (2)$$

where w is the window function, p is the resampling function, and q (if pure imaginary) is the modulation function. In particular, we generally set $p(t) = (at + b)/(ct + 1)$ where a is the dilation, b is the translation, and c is the "chirpiness" due to perspective. (In terms of our hypothetical camera metaphor, a is like zoom, b is like moving the camera left or right, and c is like panning or tilting the camera.) Thus, the parameter $p(t)$ alone gives a 3-D (time-scale-perspective) transform space, as opposed to the conventional 2-D (time-scale) wavelet transform.

We generally set the functions q and Q to be quadratics with complex coefficients ($\alpha_i t^2 + \beta_i t + \gamma_i$). The imaginary part of α_i , for example, denotes the "chirpiness" in a linear FM sense (shear in frequency). Equations 1 and 2 contain some redundant parameters. We may, however, identify 8 independent parameters that each have an intuitively satisfying significance. Depending on the choice of mother chirplet, the number of free parameters may be reduced. For example, the Gaussian chirplet may be shown to have only four useful degrees of freedom: temporal center, frequency center, time-bandwidth aspect ratio, and TF-tilt (chirpiness).

Furthermore, depending on the particular application, the number of free parameters may also be reduced (by examining the physics of the problem). For example, the TF-area-preserving (unitary, linear) constraint may be imposed, leading to a maximum of seven parameters. The perspectives may be eliminated if there is reason to believe projective geometry is not relevant to the problem at hand (e.g., in Doppler radar). If, on the other hand, the physics of the problem specifically dictate projective geometry (e.g., in vision applications), then we would naturally choose to use at least the physical perspectives.

The chirplet transform thus has indexdimension^c up to eight (depending on the particular mother chirplet and the problem at hand). Chirplet theory allows for a unified framework because it embodies many other TF methods as lower dimensional manifolds in chirplet space. For example, both the wavelet transform and the STFT are planar chirplet slices, while many adaptive methods are two-indexdimensional manifolds in the chirplet transform space.

2.1 "Le Pépielette"/"The Chirplet"

Yves Meyer first coined the term *ondelette* (from the French word for *wave*, "onde," and the diminutive "ette"). The closest English translation, *wavelet*, became the accepted word within papers written in the English language.

We have coined the term "pépielette" (combining the diminutive with the French word "pépier" or "pépiement") to designate similarly a "piece of a chirp." Similarly, in English, we coin the term "chirplet." Figure 1 shows that the relationship of a chirplet to a chirp is analogous to that of a wavelet to a wave.

^bThe term *center*, in italics, is used loosely, as is common practice in the EM literature, to designate all of the parameters (including both mean and variance) of a particular basis function.

^cIndexdimension refers to the number of indices that a discrete sampling of the space would have, if it were stored in an array. Indexdimension is analogous to *tensor rank*. *Vectordimension*, however, refers to the size of the array.

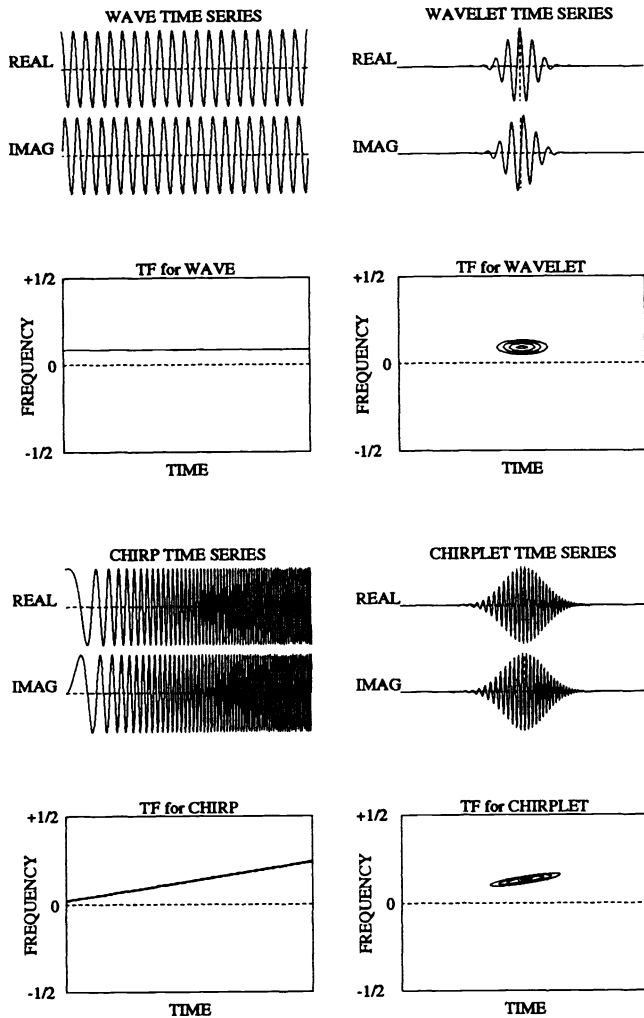


Fig. 1 Relationship between wave and wavelet and chirp and chirplet, in terms of time series and magnitude TF distributions. The physical affinity of the wavelet is first extended to TF affinity by adding up and down translation and shear in TF space. These extra degrees of freedom are achieved by multiplication of the wavelet by a chirp, hence the term chirplet. (Modulation, which is a multiplication by a pure tone, is just a special case of chirping.)

This work was first published^d in Mann and Haykin¹; the term chirplet also appears later in the literature,⁸ although with a slightly different meaning.

The TF-affine chirplet consists of all the members of a particular time-domain signal, which are affine transformations of each other when viewed in TF space. Philosophically, there are two ways to think of this basis function:

1. Using the "piece-of-a-chirp" framework.
2. Thinking in terms of affine transformations in the TF space, which consist of dilations and "chirpings" (in both time and frequency). We note that translations (modulations and delays) are just special cases of chirpings (in time and frequency) where the chirp rate is zero.

^dSome of the work of Mann and Haykin¹ was presented briefly as part of "Radar Vision,"⁷ prior to publication.

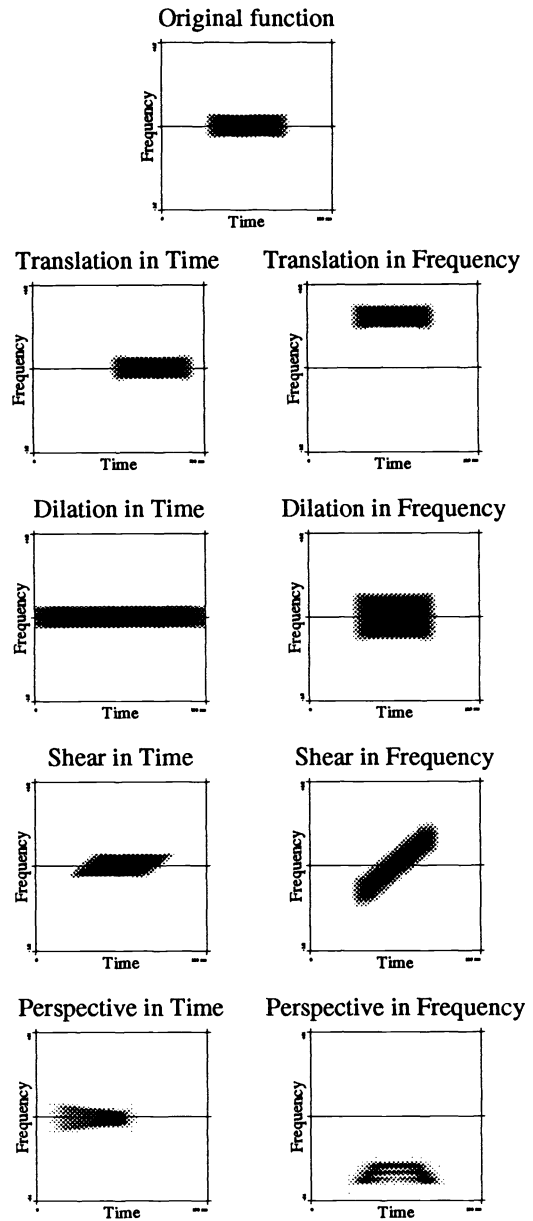










Fig. 2 The full gamut of TF perspectives. Actual TFDs of multiple-primitive-prolate-chirplets, based on multiple discrete prolate spheroidal sequence data windows.

2.2 Prolate Chirplet

We illustrate the chirplet concept by a simple example. We use a function that is somewhat rectangular in TF space, the discrete prolate spheroidal sequence (DPSS). These functions are of special interest in the signal processing community⁹⁻¹³ and are commonly referred to as *prolates* or *Slepians*. When we apply the 8 transformations to the prolate, we obtain a specific class of chirplets, which we refer to as prolate chirplets. Our tendency to favor the prolate over other equally valid functions is mainly for illustrative purposes; in choosing a mother chirplet, one must consider each problem individually. We have, however, concentrated on the prolate in developing an extension of Thomson's method of multiple windows.²³ The prolate chirplet has all eight degrees of freedom (illustrated in Fig. 2):

1.  Translation in time [This delay operator is equivalent to Fourier transformation, followed by multiplication by a complex exponential (wave), followed by inverse Fourier transformation.]
2.  Translation in frequency: multiplication by a complex exponential
3.  Dilation in time
4.  Dilation in frequency (often inversely related to time dilation)
5.  Shear in time: Fourier transformation, followed by multiplication by a chirp, followed by inverse Fourier transformation. Shear in time may be thought of as a frequency-dependent delay and is also equivalent to convolution with a chirp.
6.  Shear in frequency: multiplication by a chirp (linear frequency modulation).
7.  Perspective projection in the time domain.
8.  Perspective projection in the frequency domain.

2.3 Warblets

Initially, our choice of mother chirplets was based on various optimization criteria, for example, maximum concentration in parallelogram “boxes” (using multiple prolate chirplet data windows), or maximum concentration near a center point. All of these chirplets were monotonic-increasing or monotonic-decreasing in frequency.

Recently, however, we have become interested in a set of bases that matches phenomena whose time-frequency evolution is periodic in nature.

We have identified a particular class of chirplet⁵ that has a very profound significance, as well as practical applications in fields such as marine radar. To emphasize this special class of chirplets, we have coined the word “warblet.” Warblets are chirplets where the mother chirplet is a single tone FM signal (like the sound produced by either a police siren or the bird known as a warbler). The indexdimensionality may be reduced to four by making use of the fact that only translations and dilations produce meaningful sinusoidal FM bases. This warblet paradigm is best understood by pretending that the time-frequency plane is just an ordinary oscilloscope, so that warblets appear as waves.

Expansions onto warblets have been found to perform very well in marine radar applications for detecting small iceberg fragments, which are hazardous to navigating vessels. The reason warblets give such good performance is based on the fact that they very closely match the underlying physics of floating objects.

In practical applications, adaptivity is suited well to the warblet; with only one basis function, we may describe several seconds of radar data from a floating object. It has been found that adaptivity is almost essential; nonadaptive two-indexdimensional manifolds in warblet space do not always provide meaningful results, because a problem exists with phase alignment of the sinusoidal “squiggle” in the TF space, which cannot be resolved by the analytic signal (no counterpart exists to the analytic signal in the TF-oscilloscope duality).

3 LEM: An Expectation Maximization Paradigm in Time-Frequency Space: The Adaptive Chirplet

3.1 Introduction

We propose^e a signal-dependent TF method, logon expectation maximization (LEM), based on expectation maximization (EM).

We first generalize Gabor’s notion of logons¹⁵ (a logon is a “tile” in the tiling of the TF plane) to depict both the Weyl-Heisenberg (e.g., constant bandwidth spectrogram) as well as the physically affine (wavelet) transforms, each as a specific case of this generalization. We then take ideas from the well-known EM algorithm and apply them to the TF distribution of some arbitrary signal. Thus, we fit the TF distribution by a number of translates and dilates of some scaling function in TF space. For each *center* of the scaling function in this joint space there is a corresponding “wavelet” in the physical (time) domain. Thus, we have abstracted each adaptive “wavelet” by looking at it as a scaling function in the TF distribution space. Recall that we use the term “wavelet” (in quotes) whenever we refer to our generalization of the physically affine wavelets. “Wavelets” include time-domain functions corresponding to the modulated envelope bases of the scalogram.

The use of this EM paradigm leads to a novel variant of the radial basis function (RBF) neural network classifier. Rather than classifying on the basis of Mahalanobis distance from the *centers* as in conventional RBF networks, we classify on the basis of the receptive field output in TF space. Thus we treat each input time series not as a feature vector that lies somewhere “in” a probability density function (PDF), but rather by how it lies in the TF space. In other words, LEM may be thought of as providing a new distance metric for an RBF network. Classically, pattern recognition involves a *feature vector* whose length is equal to the vectordimension of the feature space. [Recall our distinction between indexdimension (a vector is a one-indexdimensional array) and vectordimension (a 512-sample time series may be viewed to lie at a point in 512-dimensional vector space).]

In the case of a two-dimensional feature space, for example, the input vectors should be of length two. In our case, however, the input to the network becomes the time series itself (a vector of, say, 512 samples), even though the time series lies in a space of indexdimension two (time and frequency). The RBF network then becomes an interpolator in this space rather than in the usual vector space.

3.2 Similarities Between Time-Frequency Distributions and Probability Density Functions

We know that a probability function of time and frequency is disallowed,¹⁶ but nevertheless, we consider ways to approximate PDF behavior in a TF distribution, leading eventually to treatment of the time-frequency distribution (TFD) as a distribution to be itself approximated by a number of *centers* of some self-similar interpolation function. The parallel between TFDs and PDFs is as follows:

^eIt has recently been observed that our LEM paradigm is, in some ways, similar to the earlier work of Jones et al.,¹⁴ although our approach is quite different.

- Complex-valued functions whose energy has probability-like characteristics are often useful:
 1. Time-frequency decompositions, like frequency spectra, may be complex, but often we wish to establish measures of concentration or position of their power (or magnitude) spectral components.
 2. In the physics literature "wavefunctions," which are complex, are chosen so only their squared magnitudes are meaningful; their squared magnitudes are PDFs. We are free to choose a number of possible wavefunctions to represent one specific energy function. The well-known Weyl correspondence leads to one choice from among a large family of possible wavefunctions. The energy function is what we measure or observe, while the wavefunction is chosen to fulfill some purpose in terms of the mathematical theory.

Thus, TFDs and PDFs may both have an associated complex-valued function whose magnitude or squared magnitude fulfills the intended criteria. Thereby an expansion of a signal, which has some arbitrary TFD, onto a basis of adaptive Gabor logons is analogous to expansion of some unknown PDF by an adaptive sum of Gaussians (as in the classic EM algorithm).

Criteria desirable (not necessarily possible) for a TFD or PDF:

- Marginals obtainable by integration (e.g., to get the spectrum, simply integrate the TFD with respect to time).
- The area under the distribution should be constant. Probability density functions are usually normalized to a unit area. Time-frequency distributions are usually normalized [provided the upper and lower frame bounds are equal (tight frame)], so the integrated squared magnitude is equal to the energy in the original time-domain signal, thus achieving *isometry*. When we deal with energy distributions, we will normalize them to have a unit area or L^1 norm (in other words, the underlying functions will be normalized to a unit L^2 norm).
- In TF estimation, as in spectral estimation, we cannot measure the spectrum exactly [for reasons such as the minimum time-bandwidth product (equivalent to Heisenberg's uncertainty relation in physics)], but the more data we have, the more certain we are of the spectrum. Similarly, the more data we have, the better estimate we can get of the conjoint TF density function. A similar result holds with PDFs; the more data we have, the better estimate we have of the statistics.

Thus an approximate parallel between TFDs and PDFs may be constructed, both in terms of the underlying problem space and in terms of the desirable properties of each.

Gabor functions, because of their Gaussian envelopes, have elliptically shaped equiprobability contours in TF space, as shown in Fig. 3. Imagine a distribution close to the shape of a bivariate Gaussian, rising up out of the page, with a curve drawn where its height is 1σ . This curve is drawn as an ellipse through the scatter points.

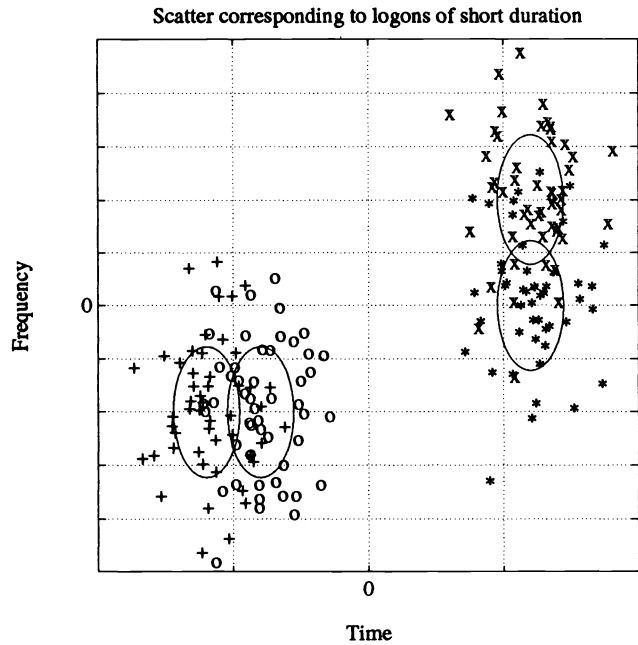


Fig. 3 Scatter and equiprobability contours for four examples of tone packets of relatively short duration and broad bandwidth.

3.3 Measures of Concentration

In any TF basis, we generally wish to attain compact support simultaneously in both the physical domain (e.g., time) and the Fourier domain. These two constraints are, however, conflicting requirements. This constraint is often referred to in the literature as the *uncertainty principle*.

We wish to derive a measure of concentration in TF space. There are a number of such measures in the literature. One such example (Baraniuk and Jones¹⁷ and Vakman¹⁸) considers the normalized volume of a TF distribution to be a quantity that we wish to minimize. In other words, Baraniuk and Jones determine the maximum of the TF distribution, then normalize the whole distribution so that the maximum is one. They then compute the volume.

We take a slightly different approach, starting with Gabor's¹⁵ definition of the effective duration of a pulse:

$$\sigma = \left\{ 2\pi \left[\frac{m_2}{m_0} - \left(\frac{m_1}{m_0} \right)^2 \right]^{1/2} \right\}, \quad (3)$$

where m_k denotes the k 'th moment of the energy or power of the signal.

In terms of the Dirac notation, the energy moments of some arbitrary signal ψ are given by

$$m_k = \langle \psi | t^k | \psi \rangle . \quad (4)$$

When $k=0$, we may write Eq. (2) as

$$m_0 = \langle \psi | \psi \rangle , \quad (5)$$

which is the energy in the signal.

The triple inner product, $\langle a | b_r | c \rangle$, is defined as $\int_{-\infty}^{\infty} dt a^*(t) b_r(t) c(t)$. Note that the center variable, b_r , will always be real, and that the first variable a is the one that we

conjugate. (Note also our use of the vertical bar as opposed to the comma, often used in this notation.)

The effective duration in frequency (bandwidth) is defined in the same manner. In the electrical engineering field, the bandwidth of a signal is generally given in terms of the width at the half power points. This definition has very poor mathematical properties, however. The rms deviation from the central epoch, as given in Eq. (3), is much more meaningful. What this measure does is fit a Gaussian distribution to the energy in the signal. The moments given in Eq. (4) are akin to the mean and variance of a Gaussian distribution.

The m_0 in Eq. (3) simply makes the signal have unit L_2 norm (unit energy) before computing the variance.

Note that, for example, a rectangular pulse of duration T has an effective duration of $(\pi/6)^{1/2}T$, or about $0.7236T$.

We extend this support measure to the TF plane by simply fitting whatever TFD we have by a bivariate Gaussian distribution.

The equation for an ellipse simply represents a contour of a bivariate Gaussian distribution. The bivariate Gaussian, in terms of time and frequency, is given by:

$$\text{TF}(t,f) = \exp \left[- \left([t-t_c, f-f_c] \mathbf{S}^{-1} \begin{bmatrix} t-t_c \\ f-f_c \end{bmatrix} \right) \right], \quad (6)$$

where \mathbf{S} is given by

$$\mathbf{S} = \begin{bmatrix} \sigma_t^2 & 0 \\ 0 & \sigma_f^2 \end{bmatrix}.$$

We wish to generalize this notion to include chirping Gabor functions, so we redefine \mathbf{S} as follows:

$$\mathbf{S} = \begin{bmatrix} \sigma_{tt}^2 & \sigma_{tf}^2 \\ \sigma_{ft}^2 & \sigma_{ff}^2 \end{bmatrix}. \quad (7)$$

We define the four quantities, $\sigma_{\square\square}$, to be positive and real; for the time being, we do not consider negative or complex bandwidths or temporal extents. Since $\sigma_{\square\square}$ are all positive, \mathbf{S} is always real. When \mathbf{S} is diagonal, the distribution is aligned along the Cartesian axes, t and f . With off-diagonal elements, however, some tilt or shear is present. We may diagonalize the matrix of variances \mathbf{S} , with an eigendecomposition as follows¹⁹:

$$\mathbf{S} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^{-1}, \quad (8)$$

where \mathbf{V} is a unitary matrix, and $\mathbf{\Lambda}$ is a diagonal matrix, given by

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}. \quad (9)$$

Because \mathbf{V} is unitary, it is equal to \mathbf{V}^\dagger , so the eigendecomposition becomes

$$\mathbf{S} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^\dagger. \quad (10)$$

Since \mathbf{S} is real, \mathbf{V} is also real, and the \dagger operator simply performs a row-column transposition. (Our use of the \dagger notation, rather than simply T , is for consistency with respect to complex signal bandwidths, which are not addressed

in this paper. Assume \mathbf{V} is always real, and thus \dagger is equivalent to T .)

The eigenvalues indicate the major and minor axes of the elliptical contours of the distribution. In other words, premultiplying by \mathbf{V} and postmultiplying by \mathbf{V}^\dagger essentially "dechirp" the signal in terms of the TF-affine space. For a chirping Gaussian signal, these eigenvalues simply indicate the physical extent along oblique directions. Thus, their product still provides the same value. In other words, for Gaussian enveloped chirps, no matter what the chirp rate, their concentration is still one-half. Thus we define an inverse-concentration (spread) measure equal to the product of the "oblique bandwidth" (bandwidth after dechirping in time) and "oblique duration" (duration after dechirping in frequency). This measure of spread is proportional to the area of the "one sigma equiprobability contour." We may apply this concentration to any distribution such as a rectangular block. The distribution need not be Gaussian, we are simply modeling it as Gaussian; the modeling is an artifice. We simply want a method to measure the concentration without "penalizing" a distribution for being oblique. Thus, for any distribution, the matrix \mathbf{S} may be computed, and an eigendecomposition performed, leading to a useful measure of concentration.

This concentration measure is simply the product of the two eigenvalues $\lambda_1 \lambda_2$, which is just the determinant of the matrix \mathbf{S} . If we want the oblique bandwidth and oblique duration to be positive, we must impose the following constraint:

$$\lambda_1 > 0 \text{ and } \lambda_2 > 0. \quad (11)$$

In other words, we must impose the constraint that \mathbf{S} be positive definite.

Our concentration measure also provides a measure of the "chirpiness" of the signal, but does not explicitly provide the shears in each of the two directions. Another method has been derived for independently providing a measure of the shear in each of the two directions.

If \mathbf{S} is not diagonal (s_{tf} and s_{ft} are nonzero) then the *effective duration* σ_{tt} and *effective bandwidth* σ_{ff} have a product greater than one-half. This increase is due to the slant, or *chirp*. Thus, regardless of the slope of a distribution, we may determine its concentration in a particularly meaningful way by fitting a bivariate Gaussian to it and computing the determinant.

Our concentration measure has a number of nice properties. For example, if there are two peaks in the distribution, as they move farther apart, they contribute more severely to the degradation in concentration as reported by our measure. Thus a distribution cannot "get away with" drifting away from the desired analysis point without affecting the cost function. Thus, when performing a TF analysis about a particular point in TF space, contributions farther from that point are "penalized" more severely than those that are close. Area-based measures do not possess this desirable property. Similarly, a "brick-wall" measure, such as that used in Thomson's method, does not address this issue, but rather it simply treats all contributions equally within a specified rectangular block in TF space. Furthermore, we may either estimate the mean epoch from the TF distribution or specify it. When we specify μ_t and μ_f , we obtain a measure of concentration about that specific point.

When we obtain μ_t and μ_f from the distribution, we obtain a measure of concentration that is independent of any of our TF-affine operators. Thus, we do not penalize a distribution for being slender, nor do we penalize it for being slanted (chirped).

3.4 The Adaptive "Wavelet"

We use a modified version of the well-known expectation maximization (EM) algorithm. Expectation maximization is generally used to approximate a PDF. It essentially fits a mixture distribution by a sum of Gaussian (or other) random processes.

In our case we are not interested in fitting a PDF, but rather we would like to approximate a TFD. Initially, suppose we wish to use only one *center* to fit the TF distribution of an arbitrary time series s . A simple selection of the *center* location follows:

$$t_c = \frac{\langle s|t|s \rangle}{\langle s|s \rangle}, \quad (12)$$

$$f_c = \frac{\langle S|f|S \rangle}{\langle S|S \rangle}, \quad (13)$$

where t_c and f_c are the coordinates of the *center* in time and frequency and S is the Fourier transform of s , given by $S(f) = \langle \exp(+j2\pi ft)|s(t) \rangle$.

Note that by Plancherel's theorem (conservation of energy in the transform domain), the denominators in Eqs. (12) and (13) are both equal to the L^2 norm (energy) of the signal g .

When more than one *center* is used to fit the distribution, we propose a variant of the EM algorithm. Chapter 6 in Duda and Hart²⁰ is a good standard reference for this algorithm, although it does not explicitly refer to the algorithm by the commonly used term EM. Another standard reference is Dempster et al.²¹ Our variant of EM follows directly from Hinton,²² where he outlines three steps to fitting a distribution by a sum of Gaussians:

- For each data point d , compute the probability density $p(d|i)$ for each unit i , using the current mean and variance for that unit.
- Normalize these probability densities to get the probability that each Gaussian gives rise to each data point (the "blame" assigned to that Gaussian):

$$p(d \leftarrow i) = \frac{p(d|i)}{\sum_j p(d|j)}. \quad (14)$$

- Using these normalized probabilities as weighting factors, compute a new mean and variance for each Gaussian (i.e., find the maximum likelihood fit to the weighted data points).

Also note that no learning rate is required (it is quite unlike steepest descent).

In our case, since we do not have discrete points, we use a batch process on the whole distribution. We also do not limit ourselves to Gaussian distributions. There are some problems though with functions such as Slepian's, which have sharp transitions in time and frequency support. The

Gaussian, because it is smooth in both time and frequency, seems to work best. It also has some other nice properties.

Consider a simple two-*center* example: let the *centers* be designated by Gaussian functions c_1 and c_2 and their shifted Fourier transforms by C_1 and C_2 . These *centers* may be thought of as functions of time, of frequency, or as elliptical "blobs" in TF space. Suppose they are discrete and consist of 512 samples.

We update the location of *center* 1, iteratively, according to:

$$t_1 \leftarrow \frac{\sum_{t=1}^{512} t b_1(t) s^*(t) s(t)}{\sum_{t=1}^{512} b_1(t) s^*(t) s(t)} \quad (15)$$

where $b_1(t)$ is the "blame" (or credit) associated with c_1 . This blame (relative likelihood that each sample was produced by Gaussian *center* c_1) is given by:

$$b_1(t) = \frac{c_1^*(t) c_1(t)}{c_1^*(t) c_1(t) + c_2^*(t) c_2(t)} \quad (16)$$

and

$$f_1 \leftarrow \frac{\sum_{f=1}^{512} f B_1(f) S^*(f) S(f)}{\sum_{f=1}^{512} B_1(f) S^*(f) S(f)}, \quad (17)$$

where B_1 is the "blame-spectrum" of C_1 , given by:

$$B_1(f) = \frac{C_1^*(f) C_1(f)}{C_1^*(f) C_1(f) + C_2^*(f) C_2(f)} \quad (18)$$

The location, in time, of *center* 2 is updated in the same manner:

$$t_2 \leftarrow \frac{\langle s|b_2 t|s \rangle}{\langle s|b_2|s \rangle}, \quad (19)$$

where, for simplicity, we have dropped the time dependence in the variables s and c . This temporal adaptation stage is shown, hypothetically, in Figs. 4 (TF-domain) and 5 (time domain).

The spectral adaptation stage for *center* 2 is given by:

$$f_2 \leftarrow \frac{\langle S|B_2 f|S \rangle}{\langle S|B_2|S \rangle}, \quad (20)$$

where we have again simplified by dropping the dependency (this time the frequency dependency) of the variables. Also recall that uppercase characters denote the Fourier transforms of the corresponding lowercase characters. Figure 6 (TF domain) shows the spectral adaptation stage for a hypothetical 2 *center* example.

In general, we use the following iterative update rule for the coordinates of *center* k :

$$(t_k, f_k) \leftarrow \left(\frac{\langle s|b_k t|s \rangle}{\langle s|b_k|s \rangle}, \frac{\langle S|B_k f|S \rangle}{\langle S|B_k|S \rangle} \right). \quad (21)$$

The above is a maximum-likelihood estimate of the mean for each basis function c_k , weighted by the signal s . Similarly, we may find the maximum-likelihood estimate of the

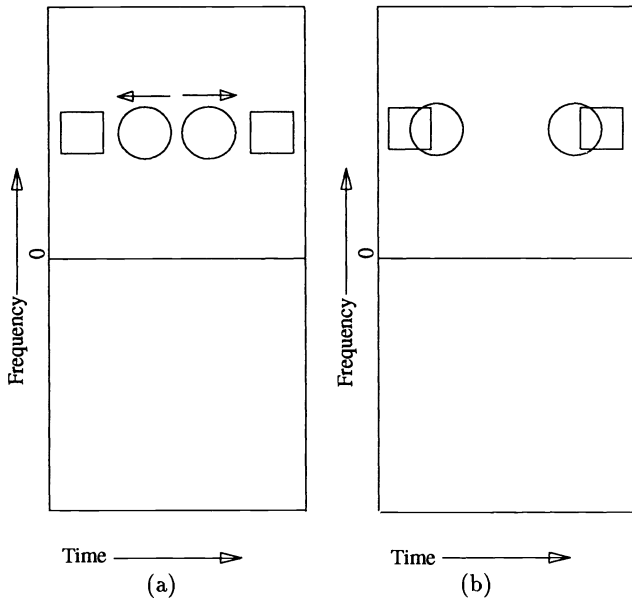


Fig. 4 Temporal adaptation portion of LEM as viewed in the TF domain: The energy distribution of the wavelets iteratively adapts to the energy distribution of the signal (temporal locations of maximum energy). In this simple hypothetical example, we are trying to fit a signal, which is a sum of two Slepian functions, by two Gabor wavelets. Note that zero frequency is in the center, since the wavelets are analytic (complex, and lying within the Hardy space). (a) Initially and (b) eventually the two centers are pulled onto temporal maxima.

variance for each c_k by:

$$\sigma_k^2 = \frac{N-1}{N} \left[\frac{\sum t^2 b_k s^* s}{\sum b_k s^* s} - \left(\frac{\sum t b_k s^* s}{\sum b_k s^* s} \right)^2 \right] \quad (22)$$

where $N = 512$ samples.

Figures 7 and 8 show two different examples of LEM. Figure 7 was chosen to emphasize the process from the temporal adaptation perspective [as in Eqs. (15) and (19)], while Fig. 8 was chosen to highlight the spectral adaptation process [Eqs. (16) and (20)].

In our simulations, c_1 and c_2 were initialized to analytic Gabor functions, with arbitrarily chosen center frequencies and arbitrary center temporal epochs.

Figure 4 is a somewhat idealized illustration, showing how the *centers* migrate along the time axis only, and Fig. 5 shows the corresponding *centers* plotted in the time domain, initially, and after a few iterations.

Likewise, Fig. 6 is a hypothetical illustration of how the *centers* move along the frequency axis only, to match the distribution in that direction.

By alternately applying Eqs. (15), (17), (19), and (20) (cycling through the four equations), *centers* move to local maxima in an arbitrary TF distribution, moving around in the 2-D space. If the signal s has a distribution with two distinct peaks, then the algorithm works quite well. If, however, there are two peaks very close together, then the convergence is very slow. There must also be some reasonable amount of disjointness in the starting values of the *centers* in both time and frequency.

Figure 7 shows an example illustrating how the time and frequency update rules are applied together. The four subplots on the left are the real parts of the time-domain signals corresponding to each of the four TFD contour plots on the right. The upper subplot pair is of the signal, two *rf pulses* (portions of sinusoids). The next subplot pair shows the initial starting guess of the *centers* (two Gabor functions). We can see that the *centers* move toward the peaks in the distribution, while the time-domain signals look more and more like the desired function leading to as good a fit as can be expected using Gabor functions to approximate *rf pulses*.

In Fig. 7 we saw the LEM process in both TF space and in the time domain. We may also observe LEM in the frequency domain. In Fig. 8 we see how the marginals, in frequency, adapt.

If the initial values of the *centers* are chosen poorly, the convergence will be extremely slow. For example, if the *centers* ever lose their disjointness in either domain (time or frequency), they will become “locked” together in that domain. Figure 9 illustrates what happens when the *centers* become “frequency-locked.” In Fig. 7, if we had chosen the starting values so that the left *center* was lower in frequency than the right one, the *centers* would have become “locked” in at least one of the two domains. In other words, LEM1 will not permit the *centers* to pass over one another. Should our initial guess be such that they need to cross over, then they will, at some time during that crossover, lose disjointness in either time or frequency; thus LEM1 will often fail.

We therefore propose another algorithm, LEM2, which operates in the two dimensions simultaneously, rather than alternately. The coordinates for each *center* are updated from the first moments of the signal-weighted “blame” function. The moments are computed in the 2-D TF space:

$$t_k = \frac{\langle \mathbf{B}_k \circ \mathbf{S} | \mathbf{1} \rangle \langle \mathbf{t} | \mathbf{S} \rangle}{\langle \mathbf{S} | \mathbf{B}_k | \mathbf{S} \rangle}, \quad (23)$$

$$f_k = \frac{\langle \mathbf{B}_k \circ \mathbf{S} | \mathbf{f} \rangle \langle \mathbf{1} | \mathbf{S} \rangle}{\langle \mathbf{S} | \mathbf{B}_k | \mathbf{S} \rangle}$$

where boldface sans serif uppercase characters indicate TF distributions, $\mathbf{1}$ is the vector of unity (in the discrete case) or a constant value (in the continuous case), $(|\mathbf{1}\rangle \langle \mathbf{t}|, |\mathbf{f}\rangle \langle \mathbf{1}|)$ is the outer product space of time and frequency, and \circ indicates the Hadamard product [also known as the Schur product, or simply element-by-element (component-wise) multiplication, as given by the symbols “.*” in Matlab.] if the functions are discrete, and just function multiplication if they are continuous.

Figure 10 shows how the use of LEM2 breaks locked *centers*. Disjointness between *centers* is now only required in one domain.

Another subtle advantage of LEM2 over LEM1 is that the compactness in TF space may be estimated. LEM1 limits us to area-preserving-TF-affine transformations, while LEM2 allows us the possibility of a change in the mathematical description of the basis, particularly useful when, for example, dealing with a mother chirplet, which is itself a family of (DPSSs).

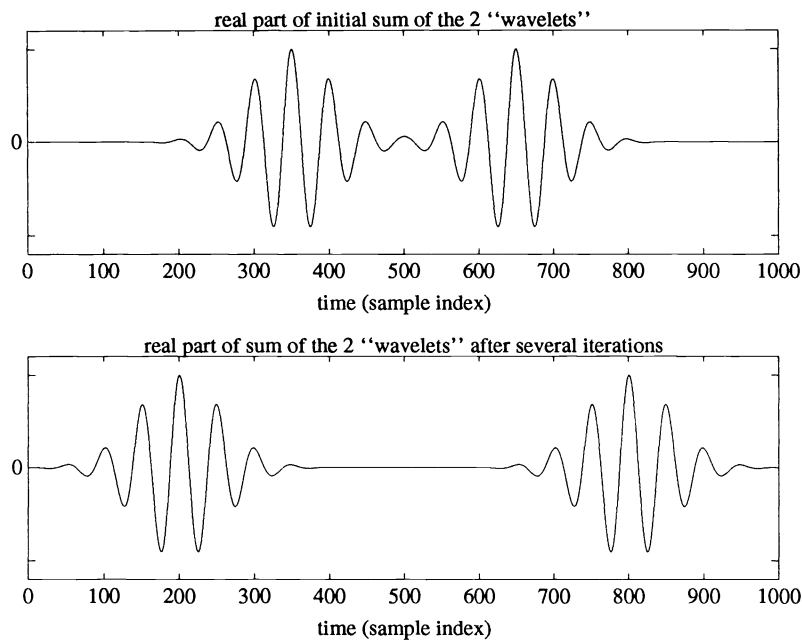


Fig. 5 Temporal adaptation portion of LEM as viewed in physical space (the time domain): In the time domain, the wavelets undergo translation to match locations of maximum energy within their bandwidths, which is equivalent to the way the distributions are moving in Fig. 4.

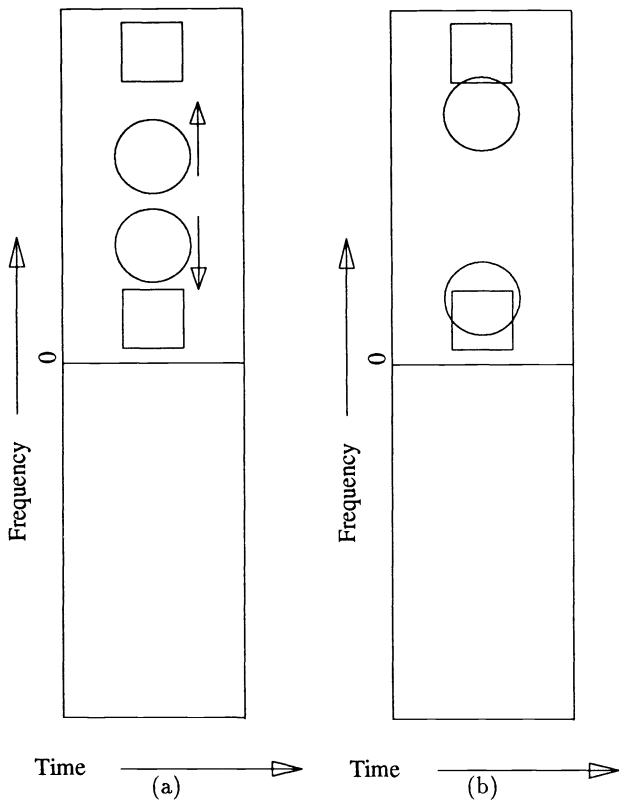


Fig. 6 Frequency adaptation portion of LEM as viewed in the time-frequency domain: Even though the signal contains two components on top of one another in time, they can be resolved by the adaptive wavelets since they are disjoint in Fourier support (separated in frequency). (a) Initially and (b) eventually the two centers are pulled onto frequency maxima.

The drawback of LEM2 is that it requires a number of TF evaluations, but this drawback can be overcome by using the following strategy:

- Select initial guess: position the *centers* at their starting points.
- Apply LEM2 once with very coarse TF resolution.
- If necessary apply LEM2 again with finer resolution in TF space . . . (apply in coarse to fine manner if necessary).
- Apply LEM1 at full resolution.

Typically a single coarse run of LEM2 will be sufficient, although a coarse to fine strategy will also help to refine the initial estimate. For example, a 512-point time series may be started with LEM2 at 32×32 point resolution, followed by a few iterations of LEM1.

We may also wish to adapt the variances of the distributions. Unlike the general probability distributions, we often desire to maintain a constant area logon. Thus, rather than changing the diameter of the circle, as is done in approximating PDFs²², we allow the aspect ratio to vary. Such a hypothetical example is illustrated in Fig. 11, with the corresponding time-domain representation in Fig. 12. These functions are neither affine wavelets (wavelets of constant shape), nor are they modulated versions of a single function as in the spectrogram basis. They are free to vary in both dilation, temporal center, and modulation.

In a similar manner, the chirpiness of each *center* may be adapted using LEM2.

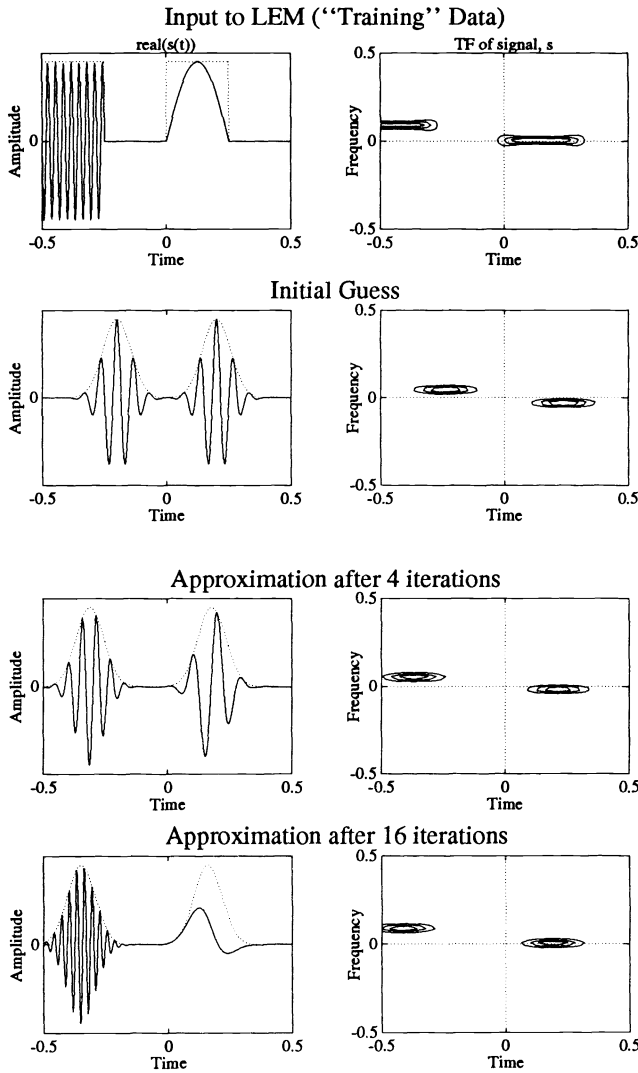


Fig. 7 Using LEM to approximate two rf pulses with two Gabor functions. Our adaptive Gabor bases are complex; we have shown only the real parts for simplicity. Here phase is treated as a nuisance parameter and we are simply interested in moving the peaks of the bases to the peaks of energy in each domain (time and frequency). The dotted line indicates the magnitude of the distributions in the time-domain plots.

3.5 RBF-TF Neural Network

Our results so far enable us to “learn” a TF distribution. The signal s need not be fixed, but rather, the system may be presented with a series of signals s_1, s_2, \dots, s_T . It will then “learn” the average trends. For example, LEM may be presented with radar data, which have a slowly varying TF distribution. Some of the *centers* will track the clutter component, and some will track the objects of interest. Computing the inner product of each of the *centers* with the current radar signal will then provide a feature vector that may be passed on to a classifier. We propose the use of a neural network similar to the RBF neural network. We suggest simply modifying the input layer to become these inner products and leaving the other layers as they normally are.

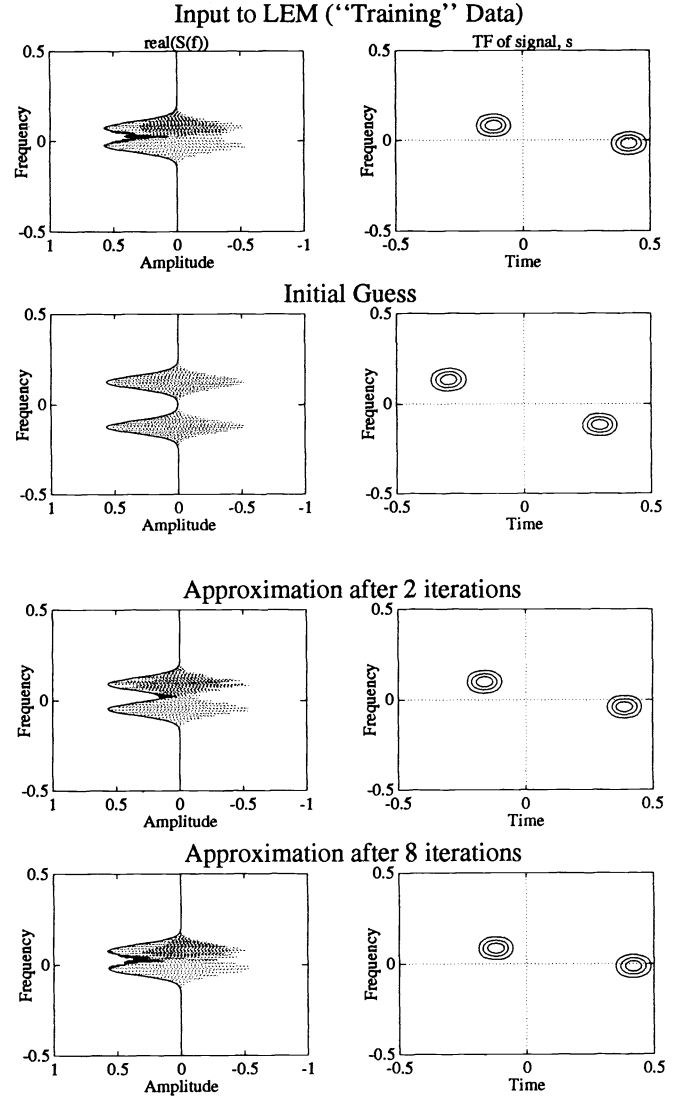


Fig. 8 Example of LEM shown in both the frequency domain (solid line indicates magnitude spectrum) and the TF domain. The scale parameter and initial training data were chosen to make the distribution interesting in the frequency domain. The direction of the frequency plots is sideways so they line up with the frequency axis of the TF space.

Relationship between LEM, the chirplet transform and the RBF neural network. In this section, we make reference to LEM as a new distance metric for a RBF network. We will now briefly elaborate on this connection. Generally the RBF network consists of a radially symmetric nonlinear layer, followed by a linear layer. (Here we refer to symmetry from the perspective of the Mahalanobis metric; clearly an ellipse is not radially symmetric in the Euclidean metric.)

The nonlinear layer may be trained by, for example, the EM algorithm, while the linear layer may be trained separately.

Figure 13 shows the RBF network architecture. Here we consider a two-vector dimensional input space and an arbitrary number of outputs. The outputs of the network y_k are

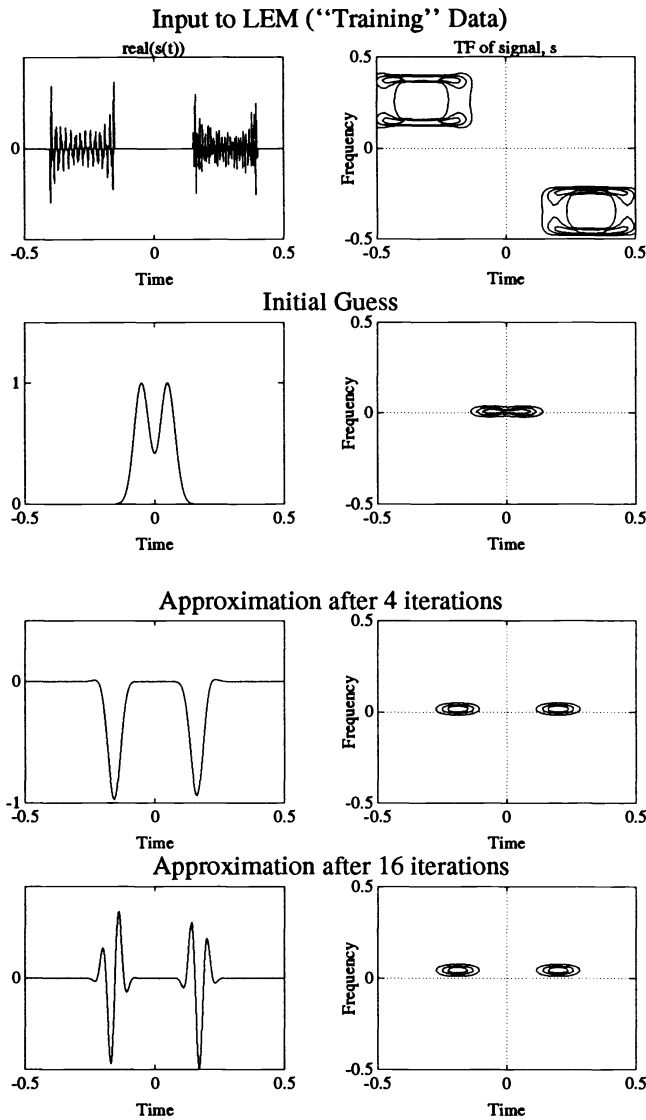


Fig. 9 We illustrate frequency-locking, a failure mode of LEM, by deliberately selecting starting values not disjoint in Fourier space. Here we have two Gabor wavelets trying to fit two Slepian functions. The fit is successful only in the time domain, while in the frequency domain, the logons become stuck. (LEM will also similarly fail in the time domain if the centers ever lose their disjointness in time.)

given by

$$y = Wz, \tag{24}$$

where y represents the output vector, z represents the hidden (middle) layer, and W represents the matrix of weights (connection strengths) in the linear portion of the network. The linear portion of the network may be solved very easily, using least-mean-squares (LMS), or linear-least-squares (pseudo-inverse) as follows:

$$W^\# = W^\dagger (WW^\dagger)^{-1}, \tag{25}$$

where the symbol \dagger denotes the conjugate transpose.

In Fig. 14 we provide an alternate visualization of the first layer where we draw in the hypothetical equiprobability

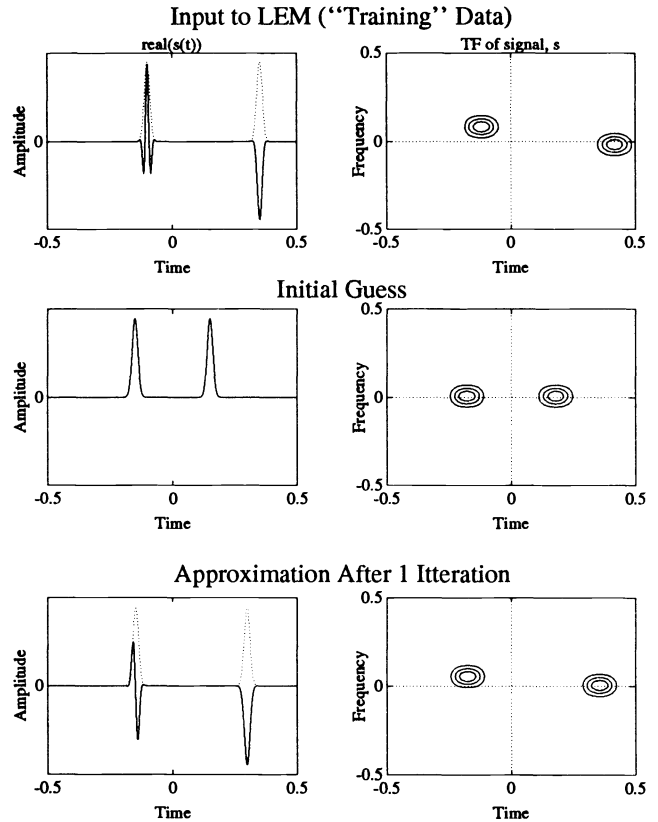


Fig. 10 LEM2 should be used for the first one or two iterations to prevent loss of disjointness. Here we show how it can specifically break lockup, allowing the centers to pass over each other in one domain if necessary.

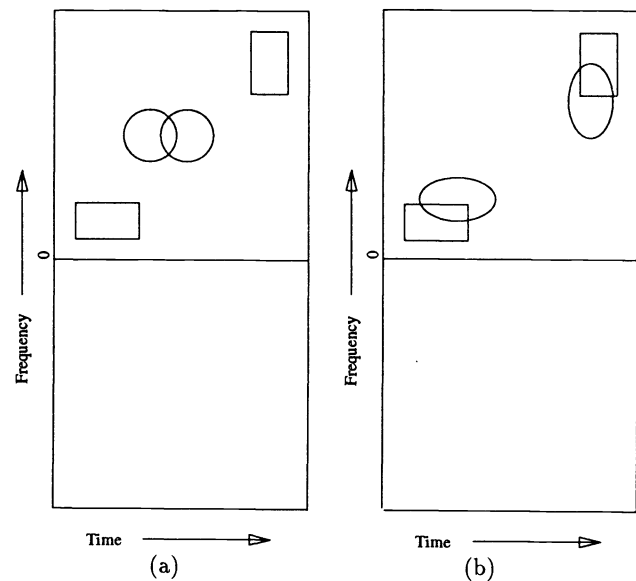


Fig. 11 Here we have a fully adaptive wavelet, which includes both the wavelet and sliding window Fourier bases as special cases. Hence, we have a mixed Weyl-Heisenberg and affine space, where translation, dilation, and modulation are free to vary independently to match the input distribution. (a) Initially, then (b) after a few iterations, the centers translate and dilate (while preserving constant area) to match the input distribution.

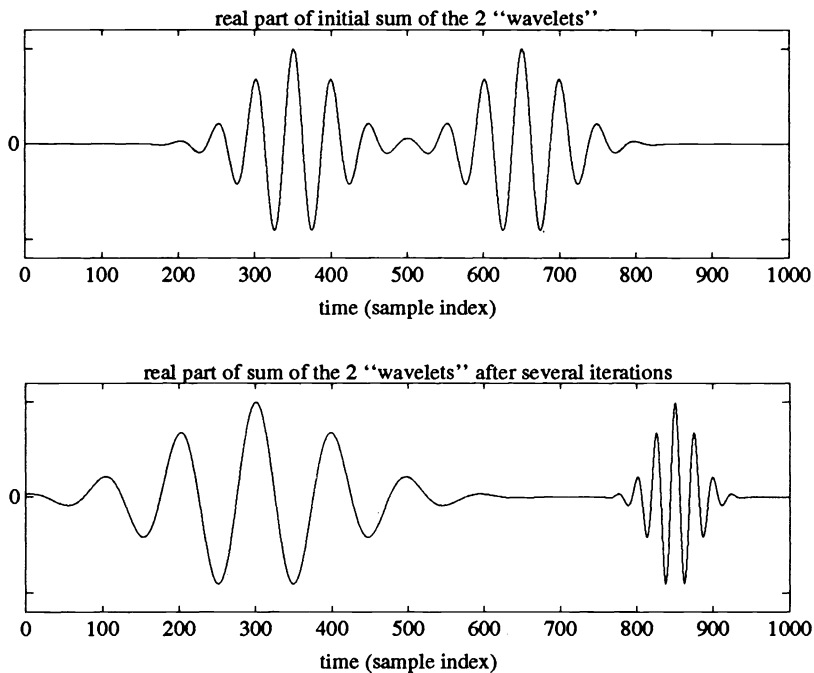


Fig. 12 Simultaneous time and frequency adaptations of LEM as seen in the time domain. The fully adaptive nature of our wavelet is explicitly visible here. Again, the time-domain description of the signal shows how the wavelet centers are translating in time, while at the same time dilating to match the bandwidth and temporal extent of the signal as well as modulating (acquiring more or less squiggles or waves) to match the center frequency independent of the extent of dilation.

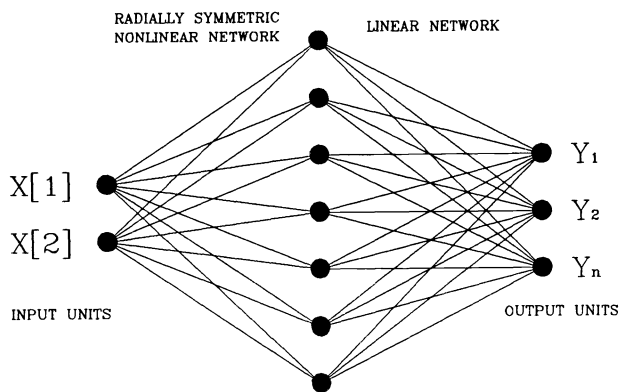


Fig. 13 The radial basis functions neural network in its usual form. The first layer is a radially symmetric nonlinear network. It is often realized by computing a Mahalanobis distance between the incoming vector \mathbf{y} and each of the centers. This network layer may be trained independently of the linear network that follows: the linear network may be trained using a pseudo-inverse (least-squares).

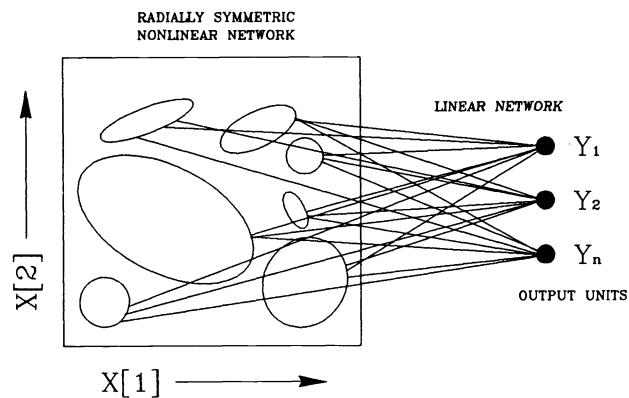


Fig. 14 Alternative visualization of the radial basis function neural network. The radially symmetric nonlinear layer is represented in the vector space. The connections are shown from seven hypothetical centers, but we may note that any arbitrary function could be used.

contours of the *centers* (in this case, seven). We show elliptical contours to represent a Gaussian nonlinearity in the nonlinear portion of the network.

If we follow our LEM algorithm by a linear network, we obtain a similar structure, which we show in Fig. 15.

3.6 LEM in frequency-frequency ("bowtie") space

The frequency-frequency space¹ is a particular slice through the chirplet transform space. It is equivalent to a Hough

(radon) transform of the Wigner distribution. Points in this bowtie chirplet transform (BCT) space correspond to straight lines in time-frequency space.

Bowtie chirplet transform space provides an alternative, sometimes more insightful manner of viewing the LEM process. In particular, we view LEM from BCT space, while it fits a Doppler radar return of a scene with three objects present. The nonadaptive BCT is shown in Fig. 16. Using the adaptive BCT, we are able to adapt the three bowties

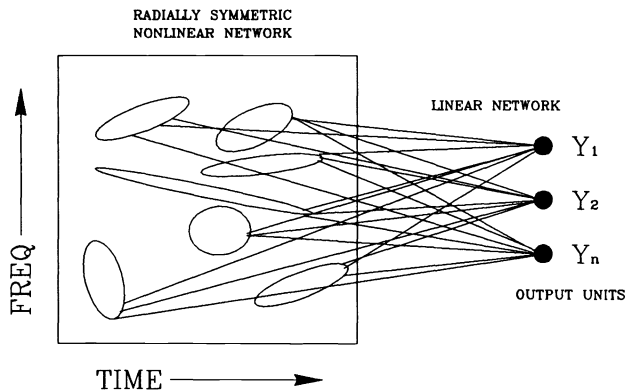


Fig. 15 RBF-TF: The proposed variant of the RBF neural network. The radially symmetric nonlinear layer is no longer a vector space, but rather a time-frequency distribution. The connections are shown from seven hypothetical centers. These centers are now any affine scaling function in TF space. For example, we may use a Gabor function (Gaussian enveloped sinusoid). We obtain the slant by using a Gaussian enveloped chirp. In this case, we have four degrees of freedom; t_c , f_c , aspect ratio (slenderness), and rotation angle.

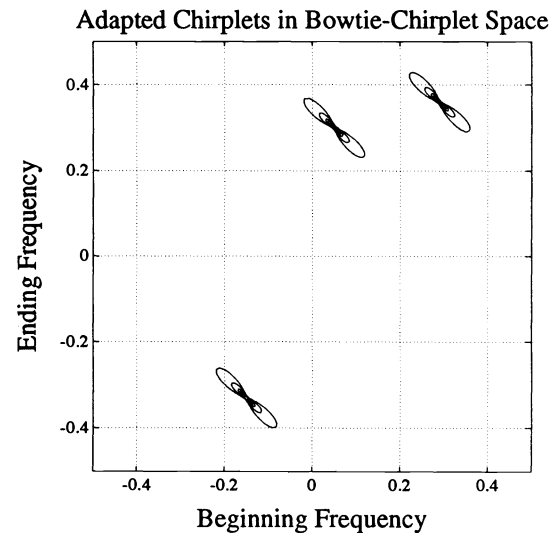


Fig. 17 Three LEM centers in the bowtie space, which have adapted to fit the distribution of a radar return from a scene with three independently moving objects.

Bowtie-Chirplet Transform with three objects present

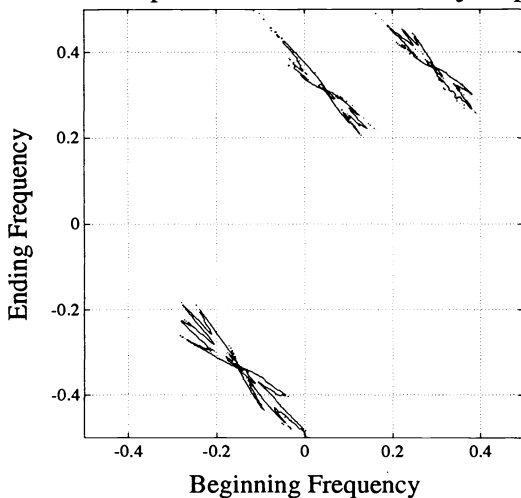


Fig. 16 BCT of radar return from three objects undergoing uniform acceleration.

(see Fig. 17). They move into the three positions corresponding to each of the three acceleration values.

4 Conclusion

We have extended the method of TF perspectives (the chirplet transform), proposed by Mann and Haykin,^{1,3} into an adaptive domain.

While the nonadaptive chirplet has been successfully applied to many problems, such as Doppler processing,¹ "radar vision,"⁷ machine vision and camera parameter estimation,⁴ and image processing, adaptivity provides two

distinct advantages over the nonadaptive chirplet by providing:

- a compact description of the signal
- a possible means of classification.

Furthermore, adaptivity enables us to manage the large number of free parameters by adapting some of them, and exploring specific manifolds within the chirplet space.

Acknowledgments

We would like to thank Douglas Jones and Richard Baraniuk of the University of Illinois at Urbana-Champaign, as well as Richard Mann of the University of Toronto, for carefully reading the manuscript and making many valuable suggestions.

References

1. S. Mann and S. Haykin, "The generalized logon transform (GLT)," *Vision Interface '91*, June 3-7, 1991.
2. I. Daubechies, A. Grossmann, and Y. Meyer, "Painless nonorthogonal expansions," *J. Math. Phys.*, **27**(5), 1271-1283 (1986).
3. S. Mann and S. Haykin, "Time-frequency perspectives: The chirplet transform," *IEEE ICASSP-92*, March 23-26, 1992.
4. S. Mann, "Wavelets and chirplets: Time-frequency perspectives, with applications," *Advances in Machine Vision, Strategies and Applications*, in P. Archibald, Ed., pp. 99-128, World Scientific, New York (1992).
5. S. Mann (Supervisor, Simon Haykin), "The chirplet transform: A new time-frequency paradigm, with applications in Doppler radar," Master's Thesis, McMaster University, Communications Research Laboratory (Aug. 1991).
6. S. Mann and S. Haykin, "Chirplets and warplets: novel time-frequency representations," *Electron. Lett.* **28**(2) (Jan. 1992).
7. Simon Haykin, "Radar Vision," Second International Specialist Seminar on Parallel Digital Processors, Portugal, April 1991.
8. D. Mihovilovic and R. N. Bracewell, "Adaptive chirplet representation of signals on time-frequency plane," *Electron. Lett.* **27**(13), 1159-1161 (1991).
9. D. Slepian and H. O. Pollack, "Prolate spheroidal wave functions, Fourier analysis and uncertainty, I," *Bell System Tech. J.* **40**, 43-64 (Jan. 1961).

10. H. J. Landau and H. O. Pollack, "Prolate spheroidal wave functions, Fourier analysis and uncertainty, II," *Bell System Tech. J.* **40**, 65–84 (Jan. 1961).
11. D. Slepian and H. O. Pollack, "Prolate spheroidal wave functions, Fourier analysis and uncertainty, III: The dimension of essentially time-and-band-limited signals," *Bell System Tech. J.*, **41**, 1295–1336 (July 1962).
12. D. Slepian, "Prolate spheroidal wave functions, Fourier analysis and uncertainty, IV: Extensions to many dimensions; generalized prolate spheroidal functions," *Bell System Tech. J.*, **43**, 3009–3058 (Nov. 1964).
13. D. Slepian, "Prolate spheroidal wave functions, Fourier analysis and uncertainty, V: The discrete case," *Bell System Tech. J.*, **57**, 1371–1430 (May–June 1978).
14. D. L. Jones and T. W. Parks, "A high resolution data-adaptive time-frequency representation," *IEEE ICASSP-87*, 681–684 (Apr. 1987).
15. D. Gabor, "Theory of communication," *J. Inst. Elec. Eng.* **93** (Part III), 429–457 (1946).
16. D. Lowe, "Joint representations in quantum mechanics and signal processing theory: why a probability function of time and frequency is disallowed," *Royal Signals and Radar Establishment, 1986*, Report 4017.
17. R. G. Baraniuk and D. L. Jones, "A radially-Gaussian, signal-dependent time-frequency representation," *IEEE ICASSP-91*, 3181–3184 (May 1991).
18. D. Ye. Vakman, "Optimal signals which maximize partial volume under an ambiguity surface," *Radio Engineering and Electronic Physics*, p. 1160 (Aug. 1967).
19. G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins Series in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD (1983).
20. R. Duda and P. Hart, *Pattern Recognition and Scene Analysis*, John Wiley and Sons, New York (1973).
21. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the {EM} algorithm," presented at a Royal Statistical Society meeting (Dec. 8, 1976).
22. G. E. Hinton, "Neural networks for industry," in *Seminar Proceedings*, University of Toronto, Department of Computer Science (Dec. 1989).
23. D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE* **70**(9), 1055–1096 (Sep. 1982).



Steve Mann received degrees in both physics and electrical engineering from McMaster University, where he developed the chirplet theory for his MEng thesis. He is currently with the Media Laboratory of the Massachusetts Institute of Technology where he is working toward a PhD degree. His current research interests include image science and technology, machine vision, computer graphics, structured imaging, and developing a unified theoretical framework for image processing.

Simon Haykin: Biography and photograph appear with the special section guest editorial.