Nonstationary Signal Classification Using Pseudo Power Signatures : The Matrix SVD Approach

Vidya Venkatachalam, *Member, IEEE*, and Jorge L. Aravena, *Member, IEEE*

Vidya Venkatachalam is with the Computational Mathematics Laboratory, Rice University, Houston, TX 77005. Tel : (713) 527 8101-8096, Fax : (713) 737 6196, Email : vidyav@rice.edu. Dr. Venkatachalam is supported by DARPA/AFOSR and Northrop Grumman Corp.

Jorge L. Aravena is with the Dept. of Electrical & Computer Engr., Louisiana State University, Baton Rouge, LA 70803. Tel : (504) 388 5537, Fax : (504) 388 5200, Email : aravena@ee.lsu.edu

Abstract

This paper deals with the problem of classification of nonstationary signals using signatures which are essentially independent of the signal length. This independence is a requirement in common classification problems like stratigraphic analysis, which was a motivation for this research. We achieve this objective by developing the notion of an approximation to the Continuous Wavelet Transform (CWT), which is separable in the time and scale parameters, and using it to define **power signatures**, which essentially characterize the scale energy density, independent of time. We present a simple technique which uses the Singular Value Decomposition (SVD) to compute such an approximation, and demonstrate through an example how it is used to perform the classification. The proposed classification approach has potential applications in areas like moving target detection, object recognition, oil exploration, and speech processing.

I. Introduction

Signal classification is an area of great importance in a wide variety of applications. Representative applications include system identification, moving target detection, oil exploration, and pattern recognition. In most of these applications, the signals are nonstationary in nature; i.e., their statistical properties vary with time. Consequently, nonstationary signal classification is an area of active research in the signal and image processing community. This paper addresses one such classification problem, common in non-intrusive subsurface exploration, involving nonstationary multicomponent signals of unknown duration. The problem is introduced here as the following event detection situation :

There is a known class of events, $\{C_k; k = 1, \dots, n\}$, which may appear in a given scene for a variable time interval. Using a probe one collects data about the scene. The objective is to analyze the probe signal to determine what events are present and the duration of the occurrence of each of these events.

In this paper we consider a basic case. One has collected the data as a signal $x(t); t_l \leq t \leq t_h$, and it is known that only one event is present at any given time. (Note that each event here represents a signal component in the time-frequency plane.) Then there is an unknown partition $P_x = \{t_l \leq t_1 \leq t_2 \dots \leq t_r \leq t_{r+1} \dots \leq t_h\}$, of *transition times* marking the start and end times of an event. The goal is to determine the transition times and the events occurring in each time interval. This process is called *classification of the signal $x(t)$*.

In the current literature, there exist several classification schemes which use time-frequency representations to perform the above classification. These take the form of signal expansions over suitable basis function sets ([1],[2]), signal adaptive kernel function design techniques ([3],[4]), principal component analysis techniques ([5],[6]), statistical analysis techniques ([7],[8]), and neural network techniques ([9]). While each of these approaches works well for the specific problem motivating their formulation, their applicability to the classification problem under consideration, where each event may have an unknown time support, is limited ([10]). This drawback (owing to the signal length dependent nature) of conventional classification techniques using time-frequency distributions ([11]), led us to explore the possibility of obtaining representations which are intrinsically independent of time, motivating this research.

The classification approach we adopt is to extract, from a signal belonging to a class, certain attributes which have the following properties.

- They are independent of signal length, location, and magnitude.
- They are robust and reliable.
- They are discriminating.
- They have few parameters.
- They lend themselves to fast classification routines.

We call this set of attributes a *signature* for the associated signal. This signature can then be used to detect the presence of similar attributes in unknown data. Since the signals of interest in this research are nonstationary, we determine the signatures based on a time-frequency analysis of the signals. Owing to the excellent resolution of the Continuous Wavelet Transform (CWT), which is necessary to accurately determine the transition times, and its computational efficiency ([12]), the CWT is the tool selected for the classification process.

II. Pseudo power signatures

In this section, we introduce a methodology for signal classification that is essentially **independent** of the actual duration of each event. We achieve this objective by using the concept of a spectral energy distribution to develop a representation that allows us to define an "instantaneous energy distribution" which we call **pseudo power signature** (PPS).

Consider any $x \in L^2(\Re)$ with CWT, c_ψ^x, where ψ is an admissible wavelet; i.e., $C_\psi = 2\pi \int_\omega \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty$, where $\Psi(\omega)$ is the Fourier Transform of the wavelet. It is well known ([13]) that the associated scalogram, SC_ψ^x, can be interpreted as a time-scale energy density function since we can write

$$\int |x(t)|^2 dt = C_\psi^{-1} \int_b \int_a SC_\psi^x(a,b) \frac{dadb}{a^2}$$

Hence, the function

$$SC_\psi^x(a,b) = |c_\psi^x(a,b)|^2$$

can be viewed as the corresponding time-scale power density function, and the function $SC_\psi^x(\cdot, b)$ as the "scale power distribution at time b". For the stated classification purpose, an ideal situation would arise if we could define a wavelet such that, for a given class of signals, the corresponding wavelet transforms are separable[1] ; i.e.,

$$c_\psi^x(a,b) = s_\psi^x(a) r_\psi^x(b)$$

[1] If we move away from $L^2(\Re)$ signals, we can find functions whose formal CWT is separable. Consider the power signal $x(t) = Ae^{-j\theta t}$. If $\psi(t)$ is an admissible wavelet with Fourier transform, $\Psi(\omega)$, the function $c_\psi^x(a,b) = \int x(t) \frac{1}{\sqrt{a}} \overline{\psi(\frac{t-b}{a})} dt$ is defined for all values of $a \neq 0, b \in R$. Observe then that $c_\psi^x(a,b) = A\sqrt{a}\overline{\Psi(a\theta)} e^{j\theta b} = s(a)r(b)$.

Then, from the Mean Value Theorem for integrals, the scale power distribution at an arbitrary time b_0 is given by

$$SC_\psi^x(a, b_0) = |s_\psi^x(a)|^2 \int_{b_1}^{b_2} \frac{|r_\psi^x(b)|^2}{(b_2 - b_1)} db; \quad b_0 \in [b_1 \;\; b_2]$$

It is apparent that the normalized distribution would be independent of b_0, and essentially independent of the signal length. Thus, the scale function $s_\psi^x(a)$, suitably normalized, could be used as the **power signature** to characterize the corresponding signal class in a manner that is independent of duration. Unfortunately, such an ideal case is impossible. We prove below that there do not exist wavelets that admit a separable CWT for elements in $L^2(\Re)$.

It is known that in order to be the CWT of an $L^2(\Re)$ signal, the function, $c_\psi^x(a, b)$, must belong to a closed subspace, M, of the Hilbert space $H = L^2(\Re^2, C_\psi^{-1} \frac{da\, db}{a^2})$ [14]. Now, the space H is isomorphic to the tensor product of the spaces $S = L^2(\Re, C_\psi^{-1} \frac{da}{a^2})$, and $R = L^2(\Re, db)$; i.e., $S \otimes R$ ([15]). Hence, it trivially follows that functions of the form $s(a)r(b)$ do exist in the space H. However, the following result proves that it is impossible to have an element of this form in the closed subspace M.

Theorem II.1: Given any nontrivial function $x \in L^2(\Re)$, and an admissible (nontrivial) wavelet $\psi \in L^2(\Re)$, the space of $CWTs$ of x with respect to ψ does not contain any element of the form given by $c_\psi^x(a, b) = s(a)r(b)$ where $s \in L^2(\Re, C_\psi^{-1} \frac{da}{a^2})$, and $r \in L^2(\Re, db)$.

The proof establishes that any wavelet leading to a separable transform must be a fixed point of a given transformation. From the Contraction Mapping theorem, we can show that the only fixed point is the null function. Complete details of the proof of the theorem are presented in the Appendix. This result opens up the problem of finding a suitable separable *approximation* to the CWT of the form $s_\psi^x(a) r_\psi^x(b)$. Since this is only an approximation, the corresponding power signature given by the normalized function $\frac{s_\psi^x}{\|s_\psi^x\|}$ is termed the **pseudo power signature**(PPS) of x. Observe that this function is essentially independent of the signal length.

III. The matrix SVD approach

The approach to the generation of PPSs is based on a Principal Component Analysis technique, and is derived from the decomposition of the CWT of a signal as a sum of separable terms. This decomposition is the natural extension of the SVD analysis, and effectively determines the closest separable approximation, in the traditional least mean squares sense, to the CWT given by $c_\psi^x(a, b) \in M \subset H$. The analysis is based on the following result ([16]).

Proposition III.1: The CWT can always be expressed as

$$c_\psi^x(a, b) = \sum_{i=1}^{\infty} \sigma_i s_i(a) r_i(b)$$

where $s_i(a) \in S = L^2(\Re, C_\psi^{-1} \frac{da}{a^2})$, and $r_i(b) \in R = L^2(\Re, db)$ for each i. The function sets $\{s_i\}_i$, $\{r_i\}_i$ are complete in S and R respectively.

The principal component of c_ψ^x is given by $\sigma_1 s_1(a) r_1(b)$. The function s_1 can then be used to define the PPS for the associated signal x, and can be determined by solving a coupled integral equation system ([16]).

It is worthwhile to note that the concept of obtaining PPSs using the principal component of the SVD is not limited to the CWT. Any time-frequency distribution (TFD) can be represented as the sum of separable components of the form shown in Proposition III.1 ([16]), and hence, it is feasible to obtain PPSs using any TFD. For the specific event detection problem under consideration, better performance can be obtained using the CWT, owing to its excellent localization capability ([12]). However, for classification problems with different requirements, the use of some other TFD to generate the PPSs might be more appropriate.

A. Applicability to discrete data sets

Proposition III.1 presented a technique to determine the PPSs for signals in $L^2(\Re)$. For computational purposes, however, we usually deal with finite discrete time signals. It is thus necessary to determine the nature of the decomposition given in Proposition III.1 when applied to finite dimensional discrete signal sets. Observe that the decomposition is very similar to the more commonly known SVD [2] applied to finite dimensional matrices. If we can reduce the problem of the determination of the PPSs for finite discrete signal sets, using principal component analysis, to a standard matrix SVD problem, then we can use any one of the existing standard and efficient algorithms for the computation of the SVD. From the SVD, we can extract the principal component, and thus, determine the PPS. Since the signature is obtained from the matrix SVD analysis, it is referred to as the SVD_M signature.

In order to reduce the problem to a standard matrix SVD problem, one needs to make a link between a continuous time function in $L^2(\Re)$ whose samples are given by the finite discrete signal under consideration, the CWT of this continuous function and its discrete equivalent, and the relation between the discrete equivalent to the CWT and the discrete signal itself. Effectively, assume that for some $x(t) \in L^2(\Re)$ with CWT given by c_ψ^x, the discrete signal $x(n) \in l^2$ is obtained by sampling $x(t)$. Then, we need to find a discrete equivalent to c_ψ^x which can be obtained from $x(n)$, and which completely represents the original signal $x(t)$. Finally, we need to find a way to represent the discrete equivalent to the CWT as a finite dimensional matrix, and determine an efficient way to compute it.

[2]The SVD applied to finite dimensional matrices is defined as follows. Given a matrix $X \in \mathcal{C}^{L \times N}$, the SVD of X is given by

$$X = U \Sigma V^*$$

where $U \in \mathcal{C}^{L \times L}$ and $V \in \mathcal{C}^{N \times N}$ are unitary matrices, and $\Sigma \in \Re^{L \times N}$ is a positive semidefinite diagonal matrix. The diagonal entries of Σ, $\sigma_1 \geq \sigma_2 \ldots \geq 0$, are referred to as the singular values of X.

Using suitably defined frames and frame operators in Hilbert spaces, and a wavelet ψ that arises from a multiresolution, under certain conditions ([14]), we can define a collection $\{\psi_{l,n}\}_{l,n}$ that constitutes an orthonormal basis for $L^2(\Re)$, where $\psi_{l,n} = 2^{-\frac{l}{2}}\psi(\frac{t}{2^l}-n)$. In this case, given any $x \in L^2(\Re)$, the discretized set of CWT coefficients $\{c_{l,n} = c^x_\psi(2^l, n2^l)\}_{l,n}$ defined by

$$c_{l,n} = \langle x, \psi_{l,n} \rangle$$

provides a *complete non-redundant representation* of x in the sense that x can be recovered from this discretized set as

$$x(t) = \sum_{l,n} c_{l,n}\psi_{l,n}$$

Also, by using finitely many of the discretized coefficients, we can approximate x to any arbitrary precision. We can represent the finitely many discretized CWT coefficients $c_{l,n} = c^x_\psi(2^l, n2^l)$ as a matrix $C = [c_{l,n}]$. The problem with this representation is that the principal component of the SVD of the matrix C is not really separable in time and scale in the true sense. The time, represented by the variable n, in $c_{l,n}$, is dependent on the associated scale, represented by the variable l. In order to obtain a truly separable approximation, we must have complete independence in the time and scale parameters. This independence can be achieved with the discretization

$$\tilde{c}_{l,n} = c^x_\psi(2^l, n)$$

Such a discretization is redundant, providing an overcomplete representation of x. It has been shown by Shensa in [17], that under certain constraints, this redundant discretization also constitutes a frame.

For most practical applications, c^x_ψ has near compact support in the time-frequency plane. For a signal of finite time support, and a suitably chosen ψ, (where ψ has compact time support), it can be well approximated using finitely many discretized CWT coefficient values. This implies that there exists L, N such that $c_{l,n} \approx 0, \forall\ \ l > L$ and $\forall\ \ n > N$. We can represent this using a finite dimensional matrix $C^x_\psi = [c_{l,n}]$ of dimension $L \times N$. Applying the SVD to this finite matrix C^x_ψ, we obtain

$$C^x_\psi = U\Sigma V^*$$

and hence,

$$C^x_\psi(l,n) = \sum_i \sigma_i u_i(l)\overline{v_i(n)}$$

The principal component is then obtained by extracting the *rank one* matrix $\sigma_1 u_1 v_1^*$, where the vectors $u_1,\ v_1$ are truly separable in time n and scale l. It is shown below that, under certain approximations, the unit vector u_1 is the discrete approximation to the PPS of x.

The simplest discrete approximation to the PPS would be the vector obtained from its samples. However, for a general measurable function, there is no guarantee that its samples are bounded, and offer a stable reconstruction. In order to ensure boundedness, and guarantee a stable reconstruction, we can define an approximation to the ideal sampling operator using as model the actual operation of A/D converters. Given a number ϵ, *suitably small*, let $\mathcal{T}$ denote the operator

$$\mathcal{T}[c](l,n) = \frac{2^{2l}}{4\epsilon^2}\int_{2^l-\epsilon}^{2^l+\epsilon}\int_{n-\epsilon}^{n+\epsilon} c(a,b)\frac{dbda}{a^2}, \quad c \in H, \quad \epsilon > 0$$

With this definition, observe that for any given $\epsilon > 0$,

$$\begin{aligned}
\frac{16\epsilon^4}{(2^{2l})^2} \mid \mathcal{T}[c](l,n) \mid^2 &= \mid \int_{2^l-\epsilon}^{2^l+\epsilon}\int_{n-\epsilon}^{n+\epsilon} c(a,b)\frac{dbda}{a^2} \mid^2 \\
&\leq \int_{2^l-\epsilon}^{2^l+\epsilon}\int_{n-\epsilon}^{n+\epsilon} \mid c(a,b) \mid^2 \frac{dbda}{a^2} \\
\Rightarrow \quad 16\epsilon^4 \mid \frac{\mathcal{T}[c](l,n)}{2^{2l}} \mid^2 &\leq \int_{2^l-\epsilon}^{2^l+\epsilon}\int_{n-\epsilon}^{n+\epsilon} \mid c(a,b) \mid^2 \frac{dbda}{a^2} \\
\Rightarrow \quad 16\epsilon^4 \sum_{l,n} \mid \frac{\mathcal{T}[c](l,n)}{2^{2l}} \mid^2 &\leq \sum_{l,n}\int_{2^l-\epsilon}^{2^l+\epsilon}\int_{n-\epsilon}^{n+\epsilon} \mid c(a,b) \mid^2 \frac{dbda}{a^2} \\
\Rightarrow \quad 4\epsilon^2 \left\| \frac{\mathcal{T}[c](l,n)}{2^{2l}} \right\|_2 &\leq C_\psi \, \|c\|_H \, < \, \infty
\end{aligned}$$

Since $\epsilon > 0$, this result implies that the weighted sequence $\frac{\mathcal{T}[c](l,n)}{2^{2l}} \in l^2(\mathcal{Z}^2)$. Hence $\mathcal{T} : H \to l^2(\mathcal{Z}^2, \frac{1}{2^{4l}})$, where $l^2(\mathcal{Z}^2, \frac{1}{2^{4l}}) = \{x(l,n) : \sum_{l=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} \frac{|x(l,n)|^2}{2^{4l}} < \infty\}$ is a weighted Hilbert space of two dimensional square summable sequences. Moreover, if $c \in H$ is continuous at $(2^l, n)$, and $\epsilon > 0$ is sufficiently small, $\mathcal{T}[c](l,n) \approx c(2^l, n)$. We can thus view the operator $\mathcal{T}$ as an approximation to the sampling operator, and denote $\mathcal{T}[c](l,n) = c(2^l,n)$ for all $c \in H$, for all l, n. [3] Then, from Proposition III.1,

$$\begin{aligned}
\mathcal{T}[c_\psi^x](l,n) &= \mathcal{T}[\sum_i \sigma_i s_i r_i](l,n) \\
&= \sum_i \sigma_i \mathcal{T}[s_i r_i](l,n) \\
&= \sum_i \sigma_i s_i(2^l) r_i(n)
\end{aligned}$$

The PPS of x is given by the function s_1. The elements of the matrix C_ψ^x are precisely the elements $\mathcal{T}[c_\psi^x](l,n)$. Thus, the discrete vector $s_{d_1} = [\frac{s_1(2^l)}{2^{2l}}]$ of dimension L, can be directly related to the vector u_1 of dimension L obtained from the SVD of the matrix C_ψ^x, as $s_{d_1} = u_1$. Hence, the vector u_1 denotes the discrete approximation to the PPS of x.

The above analysis implicitly assumes that the functions $s_i \in S$, and $r_i \in R$ are piecewise constant. In general, it does not follow that a rank one matrix is always associated with a separable element in H. However, the analysis makes the assumption that the rank one matrix obtained from the samples of s_i and

[3] It is to be understood that the actual value of $\mathcal{T}[c](l,n)$ is obtained using the definition of the operator $\mathcal{T}$. The expression $\mathcal{T}[c](l,n) = c(2^l,n)$ is written for convenience of notation, to emphasize that $\mathcal{T}$ is an approximation to the sampling operator.

r_i, maps directly to the separable element $\sigma_i s_i r_i \in H$. This mapping is valid if we impose the condition that the elements s_i, r_i be piecewise constant. In order to see this more clearly, we can define the following maps :

With the operator $\mathcal{T} : H \to l^2(\mathcal{Z}^2, \frac{1}{2^{4l}})$ defined as before, the adjoint operator $\mathcal{T}^* : l^2(\mathcal{Z}^2, \frac{1}{2^{4l}}) \to H$ is given by

$$\mathcal{T}^*[h](a,b) = \sum_{l,n} h(l,n) p_l(a) q_n(b), \quad h \in l^2(\mathcal{Z}^2, \frac{1}{2^{4l}})$$

where the functions $p_l \in S$, and $q_n \in R$ are defined as

$$p_l(a) = \begin{cases} 1, & 2^l \leq a < 2^{l+1} \\ 0, & elsewhere \end{cases}$$

$$q_n(b) = \begin{cases} 1, & n \leq b < n+1 \\ 0, & elsewhere \end{cases}$$

Clearly, if $h(l,n) = s(2^l)r(n)$, then $\mathcal{T}^*[h]$ maps to a separable element in H. Note that $\mathcal{T}^*\mathcal{T}[c_\psi^x]$ is an approximation to the CWT function c_ψ^x which is separately piecewise constant in both variables. Then,

$$\begin{aligned} \mathcal{T}^*\mathcal{T}[c_\psi^x](a,b) &= \sum_{l,n} \mathcal{T}[c_\psi^x](l,n) p_l(a) q_n(b) \\ &= \sum_{l,n} c_\psi^x(2^l, n) p_l(a) q_n(b) \\ &= \sum_{l,n} \sum_i \sigma_i s_i(2^l) r_i(n) p_l(a) q_n(b) \\ &= \sum_i \sigma_i \sum_l s_i(2^l) p_l(a) \sum_n r_i(n) q_n(b) \\ &= \sum_i \sigma_i \tilde{s}_i(a) \tilde{r}_i(b) \end{aligned}$$

If we assume that the functions $s_i \in S$, and $r_i \in R$ are piecewise constant; i.e., $s_i = \tilde{s}_i$, and $r_i = \tilde{r}_i$, for all i, then $\mathcal{T}^*\mathcal{T} = I$, and the vector $s_{d_1} = [\frac{s_1(2^l)}{2^{2l}}]$ of dimension L, suitably normalized, is given by the vector u_1 obtained from the SVD of the matrix $C_\psi^x = [c_\psi^x(2^l, n)]$. Under these assumptions, it follows that, for a given $x \in L^2(\Re)$, the discrete representation $\sigma_1 u_1 v_1^*$ obtained from the SVD of the discretized CWT matrix C_ψ^x, corresponds to the discrete approximation to the principal component $\sigma_1 s_1 r_1$ of c_ψ^x.

IV. Simulation results

This section presents three sets of computer results. Some of these results were presented in [18]. The first result validates the claim that the PPSs indeed do not depend on the length or the position of data points on the time plane, *as long as the underlying signal is monocomponent.* The second result analyzes the robustness of the SVD_M signatures in the presence of noise, and the third result serves to demonstrate the applicability, and limitations, of the SVD approach to the classification of some artificially generated signals.

The first case considered is the chirp signal shown in Figure 1a. This is the Gaussian amplitude modulated chirp signal given by $e^{-.1t^2+j.05t^2+j50t}$. The signature using the above matrix SVD analysis is shown in Figure 1b (the axis is expressed as a logarithmic function of the scale on a dyadic grid). Figures 1c, 1e, and 1g, show different arbitrarily picked samples of the same chirp signal, varying in length and location on the time plane. Their SVD_M signatures are shown in Figures 1d, 1f, and 1h. These signatures were generated using the $Db4$ wavelet [4]. We used Shensa'a algorithm ([17]) to compute the discretized CWT coefficients with the scale varying on a dyadic grid. The PPSs were then readily obtained from the principal component of the SVD of the coefficient matrix. Observe that there is no significant variation in the signature for each sample considered. We performed this test on several different sample signals with similar results. This example is a representative one, used to justify the claim that the concept of using PPSs to characterize signals independent of time (duration and location) is valid, and applicable to whole classes of nonstationary signals.

However, there do exist signals which show noticeable variations in their signatures when one considers different sampled data points. A possible explanation for this phenomenon is that these signals are essentially multicomponent; i.e., have several localized disjoint peaks in the time-frequency plane. This hypothesis is currently being researched on multicomponent data sets. The following experiment supports the concept that variability can be attributed to the presence of several components in the same signal.

We generated two signals, x_1, x_2, with clearly different spectral densities. The signals and their SVD_M signatures (using $Db4$) are displayed in Figure 2. We then created a family of 11 signals of the same length, by concatenating segments of both x_1 and x_2, in length increments of 10%. Thus, the first element is only x_1, the second is 90% x_1 followed by 10% x_2, and so on. The last element of the family is only x_2. The corresponding family of SVD_M signatures is displayed in Figure 3 ($2-D$ and mesh plots). The signatures show a certain robustness when one component is clearly dominant, but the situation is variable when the two modes contribute almost equal energy to the combination.

It is important to note that the entire exercise of representing a signal in a class using one signature pattern is based on the premise that the signal is essentially *monocomponent*. For signals which do not satisfy this premise, and are multicomponent, one needs to extract each component, and apply the above process to it. These signals would then be represented by a *set* of signatures, improving the accuracy of their classification. If the various components have different time localization (as in the example illustrated in the previous paragraph), a sliding window approach to signature computation provides a practical solution.

[4]This is one of Daubechies' compact support wavelets, and is defined through a two scale equation with 8 coefficients.

If the SVD analysis shows the existence of several singular values close to the principal value, one might consider using approximations with more terms. However, before considering that alternative, we have some theoretical results ([19]) which suggest that it may be more efficient to develop alternative techniques to generate signatures.

The second example shown here is a chirp signal corrupted with white Gaussian noise. The SVD_M signature of the pure and corrupted signals is shown in Figure 4. Observe that the SVD_M signatures are quite robust in the presence of noise, and do not exhibit a significant deviation. This is a very important property when dealing with real data, which is most often corrupted by noise.

The third example considers the signals shown in Figure 5. These signals are the simple modulated sinc ($sinc(x) = \frac{sin(\pi x)}{\pi x}$) functions $\{x1, x2, x3\}$ given by

$$
\begin{aligned}
x_1(t) &= e^{j.5\pi t} sinc(\frac{t}{3}) \\
x_2(t) &= e^{j.55\pi t} sinc(\frac{t}{3}) \\
x_3(t) &= e^{j1.55\pi t} sinc(\frac{t}{3})
\end{aligned}
$$

Their frequency spectra $\{f1, f2, f3\}$ (the axis is expressed as a fraction of π) and their PPSs $\{S1, S2, S3\}$ are also shown in the same figure. As before, these signatures were generated using the $Db4$ wavelet. Now consider a signal created by concatenating segments of each signal class: $x1$ over the interval [-125:-50], $x2$ over the interval [-50:50] and $x3$ over the interval [50:115]. The composite signal, its $STFT$, and its discretized CWT are shown in Figure 6. Observe that merely examining the signal, its Short Time Fourier Transform ($STFT$), or the CWT is not sufficient to identify either the component signals or the transition points. Furthermore, direct comparison of the $CWTs$ of each signal class with the CWT of the composite signal is also not feasible because the CWT support is dependent on the signal duration which is, in general, unknown. For classification purposes, we need a representation which is more intrinsic to each signal class, and is independent of the signal support. These conditions are satisfied by the PPSs shown in Figure 5.

The results of the classification using a correlation approach are shown here. Two assumptions are made in performing this classification.

- All the signal classes are present.
- Only one signal class is present at any given time.

It was established that the PPS of a signal represents the normalized scale power distribution, and is independent of time. Thus, we can get an accurate picture of the signal composition, with particular reference to the location of the transition points, if we determine the correlation of each Si with the discretized CWT of the composite signal for each b. The results are presented in Figure 7. Observe that the results show quite clearly that there are 2 transition points in the signal, (the first around -50, and the second around 50), a situation which is not very evident upon examination of the signal. Here, we can make the legitimate assumption that the correlation values must remain fairly constant, and relatively large, over a range for the signal to be classified as having support in that range. Thus, we can conclude from the graphs that the support of $x1$ is $[-125 : -50]$, that of $x2$ is $[-50 : 50]$, and that of $x3$ is $[50 : 115]$. Based on the underlying assumptions, we disregarded the high correlation values of $S1$ in the range $[-50 : 50]$ because $S2$ has a higher correlation in that range than $S1$, and is more likely to be present in the range $[-50 : 50]$ than anywhere else.

It is clear from the results presented that the simplistic process of taking the principal component of the SVD matrix obtained from the discretized c_ψ^x as the PPS of a signal class can, at times, lead to ambiguous interpretations. The examples presented show that the PPSs indeed do satisfy the requirement of being time independent signatures, and are more discriminating than the Fourier spectra, and more robust than the CWT. However, they lack the ability to capture fine distinctions between different signal classes, and hence, are not capable of separating signals belonging to two closely spaced signal classes. This suggests that we need to determine a more sophisticated technique to find PPSs with better discriminating capability.

V. Conclusions

In this paper, we introduced the idea of signal signatures which are essentially independent of the signal length. The determination of such signatures was based on using separable approximations to the CWT of the signal. We presented a simple approach using the SVD of the matrix of discretized CWT coefficients, to generate these signatures. We tested this approach on several examples with good results. However, the SVD_M signatures were limited by a lack of fine discriminating capability as was demonstrated through the example shown in this paper.

Note that the representation is dependent on the wavelet used, and so a related problem is the determination of the wavelet that provides the most discriminating signatures for a given class of signals. A well-known result, borne out here by empirical observations, is that better performance is obtained when the analyzing wavelet used to generate the SVD_M signatures matches the signals as closely as possible.

We conjecture that the principal component of the matrix SVD does not create the best signature because the traditional matrix SVD analysis does not determine the best separable approximation in the weighted Hilbert space H. Hence, we propose to create signatures by finding *the separable approximation* *in H whose projection onto M is the closest to a given CWT*. This technique essentially involves solving an inverse projection problem, and will be addressed in a future work.

In conclusion, we have formulated a concept which is very useful for classification problems involving signals of unknown duration. Moreover, the actual classification can be done quickly because the signatures are vectors of small dimension. This concept has potential applications in areas like oil exploration, speech analysis, moving target detection, object recognition, and system identification.

References

[1] Stephane Mallat and Zhifeng Zhang, " Matching Pursuit With Time-frequency Dictionaries," *IEEE Trans. on Signal Processing*, Vol. 41, No. 12, pp 3397-3415, Dec. 1993.

[2] F. Hlawatsch and W. Kozek, " Time-frequency Projection Filters And Time-frequency Signal Expansions," *IEEE Trans. on Signal Processing*, Vol. 42, No. 12, pp 3321-3334, Dec. 1994.

[3] Y. Zhao, Les E. Atlas and R. Marks, " The Use Of Cone-shaped Kernels For Generalized Time-frequency Representations Of Nonstationary Signals," *IEEE Trans. ASSP*, Vol. 38, No. 7, pp 1084-1091, July 1990.

[4] E. J. Zalubas and M. G. Amin, " Time-frequency Kernel Design By The Two-dimensional Frequency Transformation Method," *IEEE Trans. on Signal Processing*, Vol. 43, No. 9, pp 2198-2202, Sep. 1995.

[5] D. Beyerbach and H. Nawab, " Principal Components Analysis Of The Short Time Fourier Transform," in Proc. IEEE ICASSP, vol. 3, pp 1725-1728, Toronto, Ont., CA, May 1991.

[6] V. Gouaillier and L. Gagnon, " Ship Silhouette Recognition Using Principal Components Analysis," in Proc. SPIE Conf. on Applications of Digital Image Processing XX, vol. 3164, pp 59-69, San Diego, July 1997.

[7] Mark L. Fowler and Leon H. Sibul, " Signal Detection Using Group Transforms," in Proc. IEEE ICASSP, vol. 3, pp 1693-1696, Toronto, Ont., CA, May 1991.

[8] X. Xu, I.S. Reed, W. Kraske, and A.D. Stocker, " A Robust Multi-spectral Object Detection By Using Wavelet Transform," in Proc. IEEE ICASSP, vol. V, pp 141-144, San Francisco, CA, USA, March 1992.

[9] Mackay A. E. Okure and Michael A. Peshkin, " Quantitative Evaluation Of Neural Networks For NDE Applications Using The ROC Curve," *Proc. of the 21st Annual Review of Progress in Quantitative Nondestructive Evaluation*, Snowmass, CO, 1994.

[10] V. Venkatachalam, " *Pseudo Power Signatures for Nonstationary Signal Analysis and Classification*," Ph.D Dissertation, Louisiana State University, Baton Rouge, LA, Dec. 1998.

[11] J. McLaughlin and L. E. Atlas, " Applications of Operator Theory to Time-Frequency Analysis and Classification," *accepted by IEEE Transactions on Signal Processing*, 1998.

[12] Stephane Mallat, " *A Wavelet Tour of Signal Processing,*" Academic Press, 1998.

[13] O. Rioul and M. Vetterli," Wavelets and Signal Processing," *Signal Processing Magazine*, vol.8, pp. 14-38, Apr. 1991.

[14] I. Daubechies, " *Ten Lectures on Wavelets,*" Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1992.

[15] M. Reed and B. Simon, " *Methods of Modern Mathematical Physics 1 : Functional Analysis,*" Academic Press, Inc., San Diego, 1980.

[16] L. Cohen, " *Time-Frequency Analysis,*" Prentice Hall Signal Processing Series, New Jersey, 1995.

[17] M. J. Shensa, " The Discrete Wavelet Transform: Wedding the À Trous and Mallat Algorithms," *IEEE Transactions on Signal Processing*, vol. 40, pp. 2464-2482, Oct. 1992.

[18] V. Venkatachalam and J. L. Aravena, " Nonstationary Signal Classification Using Pseudo Power Signatures," in Proc. ISCAS 98, vol. 5, pp 449-452, Monterey, CA, USA, May31-June3 1998.

[19] V. Venkatachalam and J. L. Aravena, " Enhanced signatures for event classification : The projector approach," in Proc. IEEE-SP Time-Frequency and Time-Scale Analysis, pp 493-496, Pittsburgh, PA, Oct. 1998.

Appendix

Theorem .1: Given any nontrivial function $x \in L^2(\Re)$, and an admissible (nontrivial) wavelet $\psi \in L^2(\Re)$ (i.e., $C_\psi = 2\pi \int_a \frac{|\Psi(a\omega)|^2}{|a|} da < \infty$), the space of Continuous Wavelet Transforms (CWTs) of x with respect to ψ does not contain any element of the form given by $c_\psi^x(a,b) = s(a)r(b)$ where $s \in L^2(\Re, C_\psi^{-1}\frac{da}{a^2})$, and $r \in L^2(\Re, db)$.

Proof: Let $H = L^2(\Re^2, C_\psi^{-1}\frac{da db}{a^2})$. Let $s \in S = L^2(\Re, C_\psi^{-1}\frac{da}{a^2})$, and $r \in R = L^2(\Re, db)$. Then the space $H = S \otimes R$ ([15]), and hence, it trivially follows that $s \otimes r \in H$. However, $S \otimes R$ is isomorphic with the space $(S \times R, C_\psi^{-1}\frac{da}{a^2} \otimes db)$, which implies that there exists a unitary operator $\overline{U}$ such that $\overline{U}(s \otimes r) = sr$. Hence, the element $sr \in H$. Let M be the space of the Continuous Wavelet Transforms (CWTs). Then, M is a closed subspace of H. Let $c_\psi^x \in M$ denote the CWT of $x \in L^2(\Re)$, where $\psi \in L^2(\Re)$ is an admissible wavelet. Using a proof by contradiction, we can show that this function cannot be of the form $s(\cdot)r(\cdot)$.

By the definition of the CWT,

$$c_\psi^x(a,b) = \int_t x(t) \frac{1}{\sqrt{a}} \overline{\psi(\frac{t-b}{a})} dt$$

Assume that $c_\psi^x(a,b) = s(a)r(b)$. Then,

$$s(a)r(b) = \int_t x(t) \frac{1}{\sqrt{a}} \overline{\psi(\frac{t-b}{a})} dt$$

Keeping a fixed, and taking the Fourier Transform on both sides, we get

$$\begin{aligned} s(a)\int_b r(b)e^{-j\omega b} db &= \int_b \int_t x(t) \frac{1}{\sqrt{a}} \overline{\psi(\frac{t-b}{a})} dt e^{-j\omega b} db \\ s(a)R(\omega) &= \int_t x(t) \int_b \frac{1}{\sqrt{a}} \overline{\psi(\frac{t-b}{a})} e^{-j\omega b} db dt \quad (Fubini) \\ &= \int_t x(t)\sqrt{a}\,\overline{\Psi(a\omega)} e^{-j\omega t} dt \\ &= \sqrt{a}\,\overline{\Psi(a\omega)} X(\omega) \end{aligned}$$

By Fubini's Theorem [5], the interchange of integrals is allowed. Fix $\omega \in \Omega$, where Ω is the support of $X(\omega)$, and let $p(\omega) = \frac{R(\omega)}{X(\omega)}$. Then,

$$\begin{aligned} p(\omega)\frac{s(a)}{\sqrt{a}} &= \overline{\Psi(a\omega)} \\ \int_a | p(\omega)\frac{s(a)}{\sqrt{a}} |^2 \frac{da}{|a|} &= \int_a \frac{|\Psi(a\omega)|^2}{|a|} da \\ | p(\omega) |^2 \| s \|_S &= \frac{1}{2\pi} \end{aligned}$$

[5]**Fubini's Theorem :** If $\int_x [\int_y | f(x,y) | dy] dx < \infty$, then, $\int\int f(x,y) dx dy = \int_x [\int_y f(x,y) dy] dx = \int_y [\int_x f(x,y) dx] dy$.

The above implies that $| p(\omega) |$ is a constant function $\forall \omega \in \Omega$. This results in the condition $| X(\omega) |= L | R(\omega) |$, where $L = \sqrt{2\pi \parallel s \parallel_S}$. Thus,

$$\frac{| s(a) |}{L\sqrt{a}} =| \Psi(a\omega) |, \ \ \forall \omega \in \Omega, \ \ \forall a \tag{1}$$

which implies that $| \Psi(a\omega) |$ is constant for all $\omega \in \Omega$. Consider $\omega_1, \omega_2 \in \Omega$ with $\omega_1 < \omega_2$. From Equation 1,

$$| \Psi(a\omega_1) |=| \Psi(a\omega_2) |, \ \ \forall a \tag{2}$$

Let $\alpha = a\omega_1$. Let $\lambda = \frac{\omega_2}{\omega_1} > 1$. Equation 2 can then be rewritten as

$$| \Psi(\alpha) |=| \Psi(\lambda\alpha) |, \ \ \forall \alpha$$

Define a map $T[\Psi](\alpha) = \Psi(\lambda\alpha)$. Then,

$$\parallel T[\Psi] \parallel = \frac{1}{\lambda} \parallel \Psi \parallel$$

Thus, $\parallel T \parallel= \frac{1}{\lambda} < 1$, i.e. there is a strict contraction here. By the Contraction Mapping Theorem [6], the only fixed point of this transformation is $\Psi(\alpha) = 0$ *a.e.* By the Parseval's Identity,

$$\parallel \Psi \parallel_2^2 = 2\pi \parallel \psi \parallel_2^2$$

which implies that $\psi(t) = 0$ *a.e.*, which provides the contradiction. ■

[6] Let $T : X \to X$ be defined on a complete metric space X with domain X, and metric d. Let α satisfy $0 < \alpha < 1$, and $d(T(x), T(y)) \leq \alpha d(x, y)$, for all $x, y \in X$. Then, T has a unique fixed point $\hat{x}$.

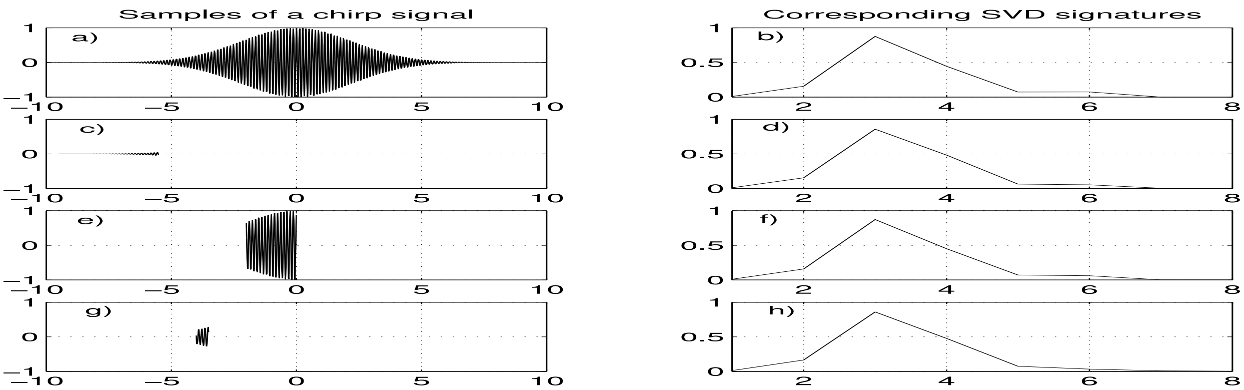

Fig. 1. Samples of a chirp signal and the corresponding SVD_M signatures

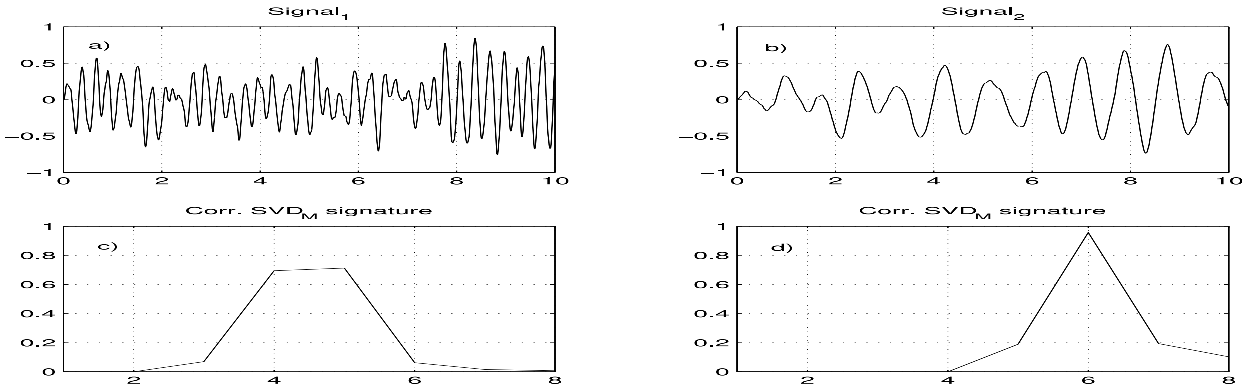

Fig. 2. The test signals and their SVD_M signatures

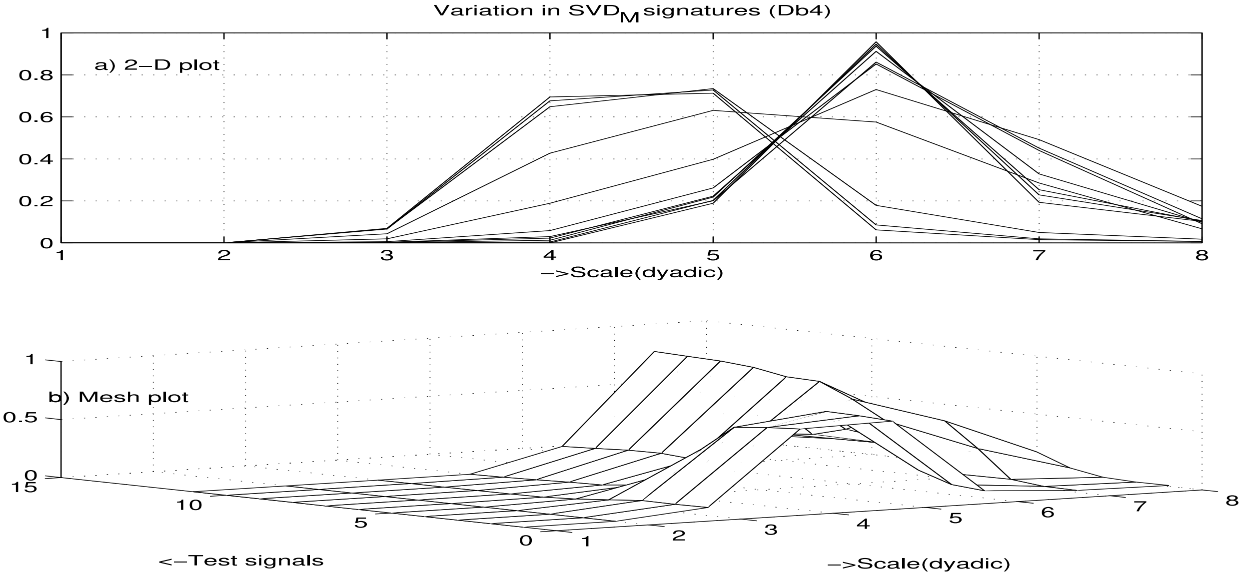

Fig. 3. Variability in the SVD_M signatures

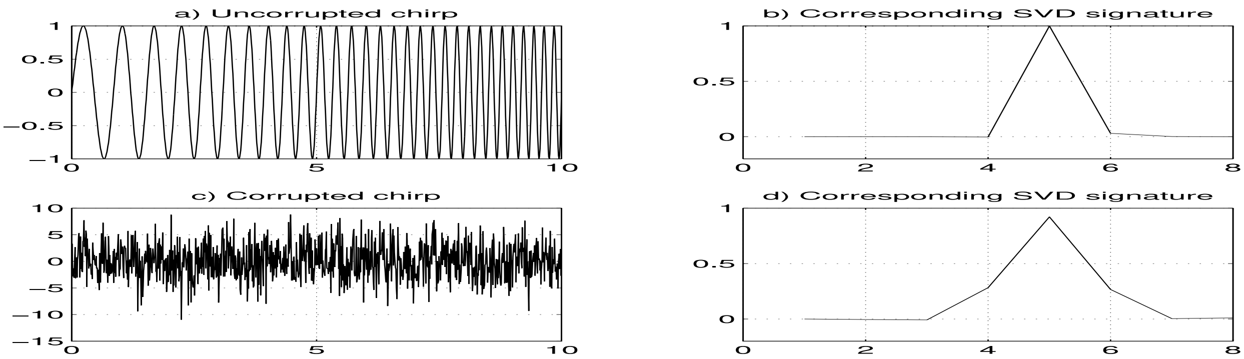

Fig. 4. Robustness of the SVD_M signatures

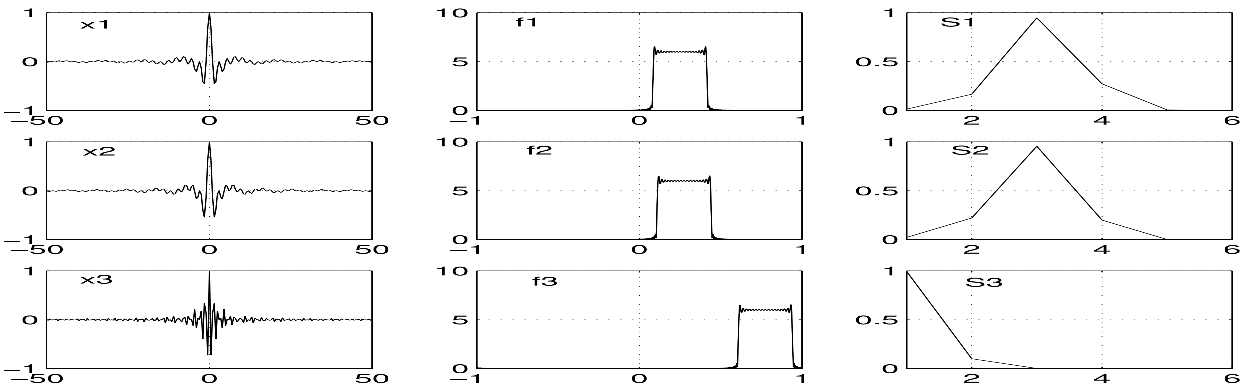

Fig. 5. The 3 signal classes and their corresponding signatures

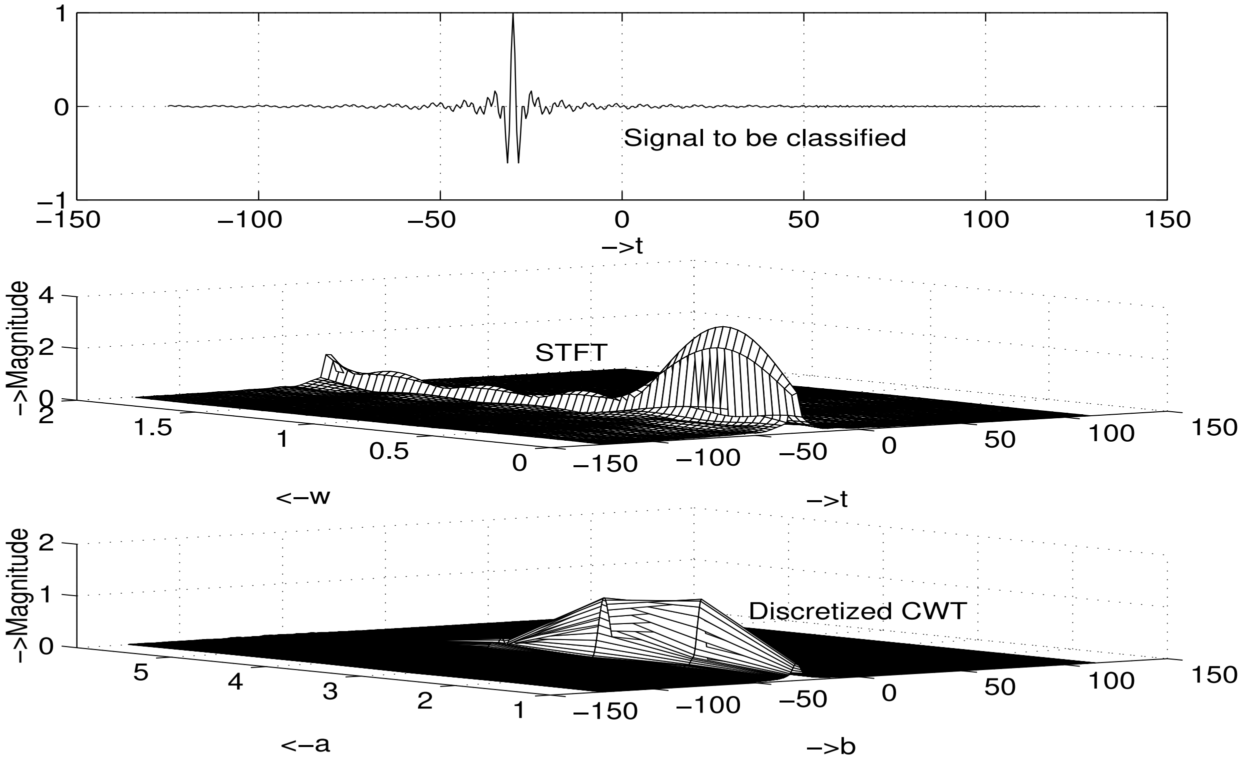

Fig. 6. The signal, its $STFT$, and its CWT

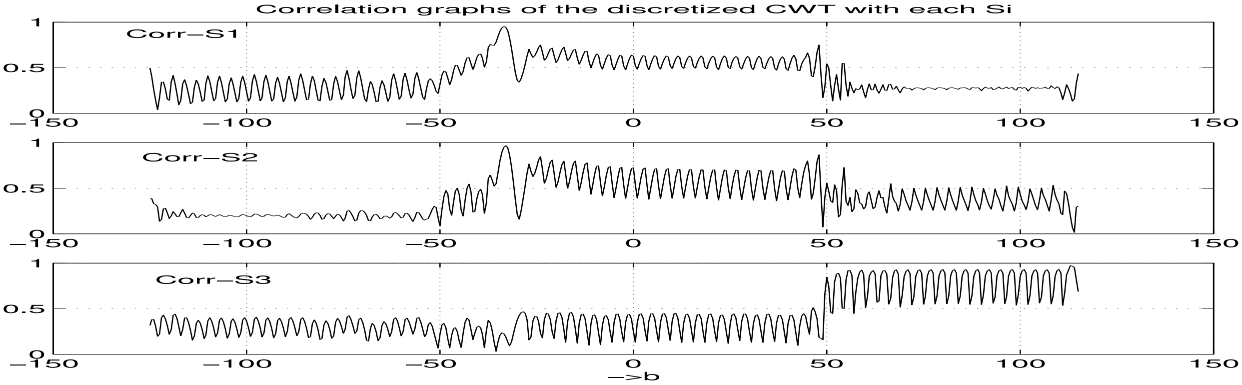

Fig. 7. Correlation graphs of the discretized CWT with each S_i