A New Approach for Estimation of Instantaneous Mean Frequency of a Time-Varying Signal

Sridhar Krishnan

Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, Canada M5B 2K3
Email: krishnan@ee.ryerson.ca

Analysis of nonstationary signals is a challenging task. True nonstationary signal analysis involves monitoring the frequency changes of the signal over time (i.e., monitoring the instantaneous frequency (IF) changes). The IF of a signal is traditionally obtained by taking the first derivative of the phase of the signal with respect to time. This poses some difficulties because the derivative of the phase of the signal may take negative values thus misleading the interpretation of instantaneous frequency. In this paper, a novel approach to extract the IF from its adaptive time-frequency distribution is proposed. The adaptive time-frequency distribution of a signal is obtained by decomposing the signal into components with good time-frequency localization and by combining the Wigner distribution of the components. The adaptive time-frequency distribution thus obtained is free of cross-terms and is a positive time-frequency distribution but it does not satisfy the marginal properties. The marginal properties are achieved by applying the minimum cross-entropy optimization to the time-frequency distribution. Then, IF may be obtained as the first central moment of this adaptive time-frequency distribution. The proposed method of IF estimation is very powerful for applications with low SNR. A set of real-world and synthetic signals of known IF dynamics is tested with the proposed method as well as with other common time-frequency distributions. The simulation shows that the method successfully extracted the IF of the signals.

Keywords and phrases: instantaneous frequency, nonstationary signals, positive time-frequency distributions, matching pursuit, minimum cross-entropy optimization, average frequency.

1. INTRODUCTION

The instantaneous frequency (IF) of a signal is a parameter of practical importance in situations such as seismic, radar, sonar, communications, and biomedical applications [1, 2, 3, 4]. In all these applications the IF describes some physical phenomenon associated with them. Like most other signal processing concepts, the IF of the signal was originally used in describing FM modulation in communications. In a typical radar application, the IF aids in the detection, tracking, and imaging of targets whose radial velocities change with time. When the radial velocity is not constant, the radar's Doppler induced frequency has a nonstationary spectrum, which can be tracked by IF estimation techniques. Instantaneous frequency can also be used as an analysis tool in watermarking of multimedia data such as audio and image [5, 6]. In the multimedia security application, time-frequency distribution is used as a tool to embed and to detect the watermark message of the signals of interest. Also, in biomedical signal analysis, IF is used in studying the electroencephalogram (EEG) signals to monitor key neural activities of the brain.

The importance of the IF concept arises from the fact that in most applications a signal processing engineer is confronted with the task of processing signals whose spectral characteristics (in particular the frequency of the spectral peaks) are varying with time. These signals are often referred to as nonstationary signals. A chirp signal is a simple example of a nonstationary signal, in which the frequency of the sinusoidal changes linearly with time.

It is theoretically difficult to describe the IF of a signal since most signals are multicomponent, and it is difficult to define a unique parameter for each time instant. Also, since frequency is usually defined as a number of oscillations or vibrations occurring in a unit time period, the association of the words "instantaneous" and "frequency" is still controversial.

Several authors have tried to define and estimate the IF of a signal [2]. Current research interests in IF estimation include techniques based on spectrogram [7], maximum likelihood approach [8], time-varying AR models [9], short-time Fourier transform [10], discrete evolutionary transform [11], and time-frequency distribution [12]. In this paper the IF is defined by using adaptive time-frequency distribution (TFD). The paper is organized as follows.

A review on the fundamental concepts of IF is discussed in Section 2. The proposed technique of adaptive TFD is described in Section 3. Results with synthetic signals and real-world signals are discussed and compared with those obtained from other TFDs in Section 4. A summary is given in Section 5.

2. REVIEW

The classical definition of the IF of a signal is defined as

$$\omega_i(t) = \frac{d\phi(t)}{dt}. \tag{1}$$

IF can be determined by taking the first central moment (average frequency) of the bilinear time-frequency distributions (TFDs)

$$\omega_i(t) = \frac{\int \omega\, W(t, \omega) d\omega}{\int W(t, \omega) d\omega}. \tag{2}$$

Most of the popular time-frequency distributions belong to a general class called the Cohen's class [3, 4]. Cohen's class distributions are bilinear distributions, and are generally computed as the Fourier transform of the time-varying autocorrelation function [4]. In this class, Wigner-Ville distribution is the most common and simplest instantaneous frequency (IF) tool. It yields high joint time-frequency resolution on many nonstationary monocomponent signals. However, Wigner-Ville distribution (WVD) performance is reduced significantly when applied on multicomponent signals [1, 13]. Due to its quadratic nature, WVD suffers from cross-term effect and cannot satisfy some of the requirements for TFR. Its distribution has negative values leading to imprecise IF interpretation. The effects of cross-terms can be suppressed significantly in the smoothed versions of the Wigner-Ville such as Choi-Williams distribution (CWD), pseudo Wigner-Ville distribution (PWVD), smoothed pseudo Wigner-Ville distribution (SPWVD), and reduced interference distribution (RID).

How well a TFD performs really depends on several factors such as the closeness and location between the signal components, the level and types of noise, and how the IF laws change with time (linear or nonlinear, rapid or nonrapid change of frequency over time). Most Cohen's class TFD derived from WVD yield the IF by correct first-moment calculation but this is often computationally expensive and is adversely affected by noise. Therefore, there is an on-going research to improve the performance of WVD in the presence of noise [14, 15].

Most TFDs such as WVD provide high signal energy concentration in time and frequency, therefore it is tempting to try to use it to measure the spread of frequencies with time. Unfortunately, the spread of the IF of the WVD is only positive for certain types of signals. Even when the spread is positive some negative distribution values may appear in the calculation, and thus its usefulness is questionable. From the literature it appears that still there are many unresolved issues regarding the IF of the signal. (A detailed review on the fundamentals of IF is available in [1].) It has been shown that the usual way of interpreting the IF as the average frequency at each time brings out unexpected results with Cohen's class of bilinear TFDs. If the IF is interpreted as the average frequency, then the IF need not be a frequency that appears in the spectrum of the signal. If the IF is interpreted as the derivative of the phase, then the IF can extend beyond the spectral range of the signal. It has been recently reported that the estimation of IF of a signal using a positive TFD brings out meaningful interpretation about the IF of the signal [16]. The motivation behind this paper is in adaptively constructing a positive TFD suitable for estimating the IF of a signal.

3. ADAPTIVE TIME-FREQUENCY DISTRIBUTIONS

The purpose of this paper is to explore the best available TFD for estimating the IF of a signal. For simple applications such as in the analysis of monocomponent signal, Cohen's class TFDs or model-based TFDs may be applied. It is widely accepted that, in case of complex signals with multiple frequency components, there is no definite TFD that will satisfy all the criteria and still will give desired performance for time-varying signal analysis and feature extraction.

Performance of Cohen's class TFDs depends totally on the kernel function. This signal-independent kernel acts as a lowpass filter on the signal's Wigner distribution to attenuate the cross-terms and retain the autoterms. In the ambiguity domain, the signal autoterms (AT) are centered at the origin while the interference terms (IT) are located away from the origin. The properties and ability to remove cross-terms of a smoothed Wigner-Ville distribution depends totally on the shape of the corresponding smoothing kernel in the ambiguity domain [17, 18].

Ideally, value of the kernel lowpass filter should be one in the autoterm region and zero in the interference term region; if the kernel is too narrow, suppression of IT also takes away some of the AT energy leading to smearing of the TFD. On the other hand, if the kernel shape is too broad, it cannot suppress all the IF terms. This reason explains why a fixed kernel design (not adaptive) cannot work well for all signal types. High joint time-frequency resolution cannot be achieved at the same time with good interference suppression. Figure 1 shows the shape of the kernel function (ambiguity domain) of different time-frequency estimators [17]. In WVD, the kernel $\Psi_T(\tau, \nu) = 1\ \forall \tau, \nu$, so it is an allpass filter, no IT suppression is allowed, AT energy is reserved. Therefore WVD has very high time-frequency resolution at the expense of cross-term presence.

It is worth noticing that there is a relation or constraint between the kernel function and the number of requirements the TFD satisfies. Strictly following the requirements of time-frequency distribution would create some limitation on IT suppression ability. For example, one of the important requirements is the marginal property which states that (in ambiguity domain) the kernel function $\Psi_T(0, \nu) = 1\ \forall \nu$ and
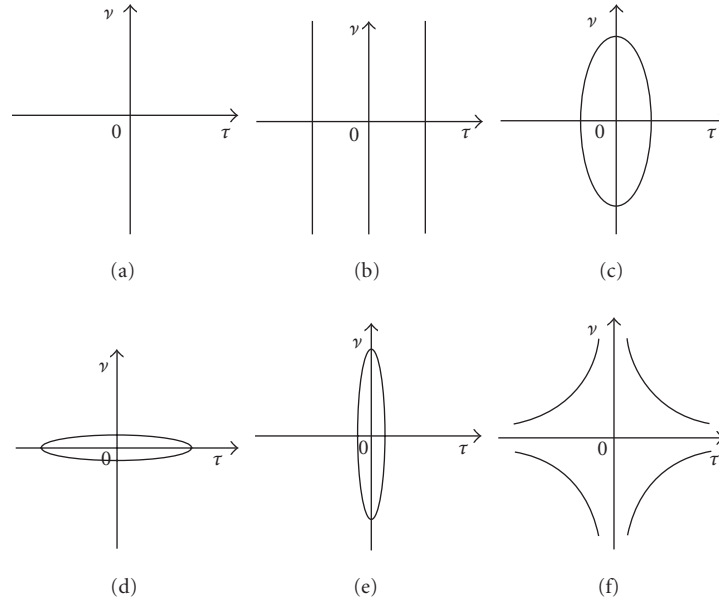
FIGURE 1: The shape of different TFDs kernel: (a) WVD, (b) pseudo-WVD, (c) smoothed pseudo-WVD, (d) spectrogram (long window), (e) spectrogram (short window), and (f) Choi-Williams distribution.

$\Psi_T(\tau, 0) = 1 \; \forall \tau$. Choi-Williams distribution (CWD) satisfies this condition while pseudo-WVD (PWVD) does not. This results in the presence of cross-terms in CWD if the signal components have the same time or frequency values.

In [19], a solution to the multicomponent problem was given by proposing an algorithm to select an optimal TFD from a set of TFDs for a given signal. The purpose of this section is to address the same problem by constructing TFDs according to the application in hand, that is, to tailor the TFD according to the properties of the signal being analyzed. It is appropriate to call such TFDs as *adaptive TFDs*. In the present work, the concept of adaptive TFDs is based on signal decomposition.

In practice, no TFD may satisfy all the requirements needed for instantaneous feature extraction and identification for nonstationary signal analysis. In the method proposed in this section, by using constraints, the TFDs are modified to satisfy certain specified criteria. It is assumed that the given signal is somehow decomposed into components of a specified mathematical representation. By knowing the components of a signal, the interaction between them can be established and used to remove or prevent crossterms. This avoids the main drawback associated with Cohen's class TFDs; numerous efforts have been directed to develop kernels to overcome the cross-term problem [12, 20, 21, 22, 23, 24].

The key to successful design of adaptive TFDs lies in the selection of the decomposition algorithm. The components obtained from a decomposition algorithm depend largely on the type of basis functions used. For example, the basis function of the Fourier transform decomposes the signal into tonal (sinusoidal) components, and the basis function of the wavelet transform decomposes the signal into components with good time and scale properties. For TF representation, it will be beneficial if the signal is decomposed using basis functions with good TF properties. The components obtained by decomposing a signal using basis functions with good TF properties may be termed as TF atoms. An algorithm that can decompose a signal into TF atoms is the MP algorithm described in the next section.

3.1. Matching pursuit

The MP algorithm [25] decomposes the given signal using basis functions that have excellent TF properties. The MP algorithm selects the decomposition vectors depending upon the signal properties. The vectors are selected from a family of waveforms called a dictionary. The signal $x(t)$ is projected onto a dictionary of TF atoms obtained by scaling, translating, and modulating a window function $g(t)$:

$$x(t) = \sum_{n=0}^{\infty} a_n g_{\gamma_n}(t), \tag{3}$$

where

$$g_{\gamma_n}(t) = \frac{1}{\sqrt{s_n}} g\left(\frac{t - p_n}{s_n}\right) \exp\left[j\left(2\pi f_n t + \phi_n\right)\right], \tag{4}$$

and a_n are the expansion coefficients. The scale factor s_n is used to control the width of the window function, and the parameter p_n controls temporal placement. $1/\sqrt{s_n}$ is a normalizing factor that restricts the norm of $g_{\gamma_n}(t)$ to 1. The parameters f_n and ϕ_n are the frequency and phase of the exponential function, respectively. γ_n represents the set of parameters (s_n, p_n, f_n, ϕ_n).

In the present work, the window is a Gaussian-type function, that is, $g(t) = 2^{1/4} \exp(-\pi t^2)$; the TF atoms are then called Gabor atoms, and they provide the optimal TF resolution in the TF plane.

In practice, the algorithm works as follows. The signal is iteratively projected onto a Gabor function dictionary. The first projection decomposes the signal into two parts:

$$x(t) = \langle x, g_{\gamma_0} \rangle g_{\gamma_0}(t) + R^1 x(t), \tag{5}$$

where $\langle x, g_{\gamma_0} \rangle$ denotes the inner product (projection) of $x(t)$ with the first TF atom $g_{\gamma_0}(t)$. The term $R^1 x(t)$ is the residue after approximating $x(t)$ in the direction of $g_{\gamma_0}(t)$. This process is continued by projecting the residue onto the subsequent functions in the dictionary, and after M iterations:

$$x(t) = \sum_{n=0}^{M-1} \langle R^n x, g_{\gamma_n} \rangle g_{\gamma_n}(t) + R^M x(t), \tag{6}$$

with $R^0 x(t) = x(t)$. There are two ways of stopping the iterative process: one is to use a prespecified limiting number M of the TF atoms, and the other is to check the energy of the residue $R^M x(t)$. A very high value of M and a very small value for the residual energy will decompose the signal completely at the expense of increased computational complexity.

3.2. Matching pursuit TFD

A signal-decomposition-based TFD may be obtained by taking the WVD of the TF atoms in (6), and is given as [25]

$$W(t, \omega) = \sum_{n=0}^{M-1} | \langle R^n x, g_{\gamma_n} \rangle |^2 W g_{\gamma_n}(t, \omega)$$
$$+ \sum_{n=0}^{M-1} \sum_{\substack{m=0 \\ m \neq n}}^{M-1} \langle R^n x, g_{\gamma_n} \rangle \langle R^m x, g_{\gamma_m} \rangle^* W_{[g_{\gamma_n} g_{\gamma_m}]}(t, \omega), \tag{7}$$

where $W g_{\gamma_n}(t, \omega)$ is the WVD of the Gaussian window function. The double sum corresponds to the cross-terms of the WVD indicated by $W_{[g_{\gamma_n} g_{\gamma_m}]}(t, \omega)$, and should be rejected in order to obtain a cross-term-free energy distribution of $x(t)$ in the TF plane. Thus only the first term is retained, and the resulting TFD is given by

$$W'(t, \omega) = \sum_{n=0}^{M-1} | \langle R^n x, g_{\gamma_n} \rangle |^2 W g_{\gamma_n}(t, \omega). \tag{8}$$

This cross-term-free TFD, also known as matching pursuit TFD (MPTFD), has good signal representation and is appropriate for analysis of nonstationary, multicomponent signals. The extraction of coherent structures makes MP an attractive tool for TF representation of signals with unknown SNR.

3.3. Minimum cross-entropy optimization of the MPTFD

One of the drawbacks of the MPTFD is that it does not satisfy the marginal properties. If a TFD is positive and satisfies the marginals, it may be considered to be a proper TFD for extraction of time-varying frequency parameters such as IF. This is because positivity coupled with correct marginals ensures that the TFD is a true probability density function, and the parameters extracted are meaningful [26]. The MPTFD may be modified to satisfy the marginal requirements, and still preserve its other important characteristics. One way to optimize the MPTFD is by using the cross-entropy minimization method [27, 28]. Cross-entropy minimization is a general method of inference about an unknown probability density when there exists a prior estimate of the density, and new information in the form of constraints on expected values is available. The minimum cross-entropy optimization was first applied to TFDs, namely, spectrograms, by Loughlin et al. [29]. A similar approach could be applied to MPTFD, and the resulting TFD would qualify as a positive TFD. If the optimized MPTFD or OMP TFD (an unknown probability density function) is denoted by $M(t, \omega)$, then it should satisfy the marginals [29]

$$\int M(t, \omega) d\omega = |x(t)|^2 = m(t), \tag{9}$$

$$\int M(t, \omega) dt = |X(\omega)|^2 = m(\omega). \tag{10}$$

Equations (9) and (10) may be treated as constraint equations (new information) for optimization. Now, $M(t, \omega)$ may be obtained from $W'(t, \omega)$ (a prior estimate of the density) by minimizing the cross-entropy between them, given by [29]

$$H(M, W') = \iint M(t, \omega) \log \left(\frac{M(t, \omega)}{W'(t, \omega)} \right) dt \, d\omega. \tag{11}$$

As we are interested only in the marginals, OMP TFD may be written as [28]

$$M(t, \omega) = W'(t, \omega) \exp \{ -(\alpha_0(t) + \beta_0(\omega)) \}, \tag{12}$$

where the α's and β's are Lagrange multipliers which may be determined using the constraint equations. In the minimum cross-entropy optimization, an iterative algorithm to obtain the Lagrange multipliers and solve for $M(t, \omega)$ is presented next.

At the first iteration, we define

$$M^1(t, \omega) = W'(t, \omega) \exp \left(-\alpha_0(t) \right). \tag{13}$$

As the marginals are to be satisfied, the time-marginal constraint has to be imposed in order to solve for $\alpha_0(t)$. By imposing the time-marginal constraint given by (9) on (13), we obtain [29]

$$\alpha_0(t) = \ln \left(\frac{m'(t)}{m(t)} \right), \tag{14}$$

where $m(t)$ is the desired time marginal and $m'(t)$ is the time marginal estimated from $W'(t, \omega)$. Now, (13) can be rewritten as

$$M^1(t, \omega) = W'(t, \omega)\, \frac{m(t)}{m'(t)}. \tag{15}$$

At this point, $M^1(t, \omega)$ is a modified MPTFD with the desired time marginal; however, it need not necessarily have the desired frequency marginal $m(\omega)$. In order to obtain the desired frequency marginal, the following equation has to be solved [29]:

$$M^2(t, \omega) = M^1(t, \omega) \exp\left(-\beta_0(\omega)\right). \tag{16}$$

Note that the TFD obtained after the first iteration $M^1(t, \omega)$ is used as the incoming estimate in (16). By imposing the frequency marginal constraint given by (10) on (16), we obtain [29]

$$\beta_0(\omega) = \ln\left(\frac{m'(\omega)}{m(\omega)}\right), \tag{17}$$

where $m(\omega)$ is the desired frequency marginal, and $m'(\omega)$ is the frequency marginal estimated from $W'(t, \omega)$. Now, (16) can be rewritten as [29]

$$M^2(t, \omega) = M^1(t, \omega)\frac{m(\omega)}{m'(\omega)}. \tag{18}$$

By incorporating the desired marginal constraint, the $M^2(t, \omega)$ TFD may be altered and need not necessarily give the desired time marginal. Successive iteration could overcome this problem and modify the desired TFD to get closer to $M(t, \omega)$. This follows from the fact that the cross-entropy between the desired TFD and the estimated TFD decreases with the number of iterations [28].

As the iterative procedure is started with a positive distribution $W'(t, \omega)$, the TFD at the nth iteration $M^n(t, \omega)$ is guaranteed to be a positive distribution. Such a class of distributions belongs to the Cohen-Posch class of positive distributions [26]. The OMP TFDs may also be taken to be adaptive TFDs because they are constructed on the basis of the properties of the signal being analyzed.

As mentioned before, a method for constructing a positive distribution using the spectrogram as a priori knowledge was developed by Loughlin et al. [29]. The major drawback of using the spectrogram as a priori knowledge is the loss of TF resolution; this effect may be minimized by taking multiple spectrograms with different sizes of analysis windows as initial estimates of the desired distribution. The method proposed in this section starts with the MPTFD, overcomes the problem of using multiple spectrograms as initial estimates, and produces a high-resolution TFD tailored to the signal properties. The OMP TFD may be used to derive higher moments by estimating the higher-order Lagrange multipliers. Such measures are not necessary in the present work, and are beyond the scope of this paper.
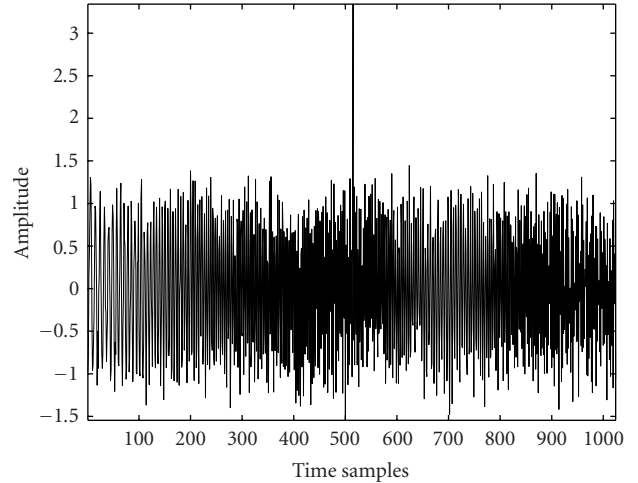

FIGURE 2: Monocomponent, nonstationary, synthetic signal "syn1" consisting of a chirp, an impulse, and a sinusoidal FM component (SNR = 10 dB).

The IF of a signal can be computed as the first moment of $\mathrm{TFD}(t, \omega)$ along each time slice, given by

$$\mathrm{IF}(t) = \frac{\sum_{\omega=0}^{\omega_m} \omega \mathrm{TFD}(t, \omega)}{\sum_{\omega=0}^{\omega_m} \mathrm{TFD}(t, \omega)}. \tag{19}$$

IF characterizes the frequency dynamics of the signal.

4. RESULTS

The proposed method of extracting the IF of a signal was applied to a set of synthetic signals with known IF laws, and a real-world example of knee joint sound signal.

4.1. Synthetic signal

The first simulation demonstrates the proposed technique's adaptivity by decomposing signals into atoms with known TF properties. The synthetic signal "syn1" is composed of nonoverlapping chirp, transient, and sinusoidal FM components, and is shown in Figure 2. "syn1" is an example of a monocomponent signal with linear and nonlinear frequency dynamics. To simulate noisy signal conditions, the signal was corrupted by adding random noise to an SNR of 10 dB ("syn1" in Figure 2) and 0 dB ("syn2" in Figure 3). The frequency behavior of the signals is shown in Figure 4.

The MP method has given a clear picture of the IF representation: the three simulated components are perfectly localized in the TFDs shown in Figures 5 and 6. This is because the OMP TFD provides adaptive representation of signal components, and is due to the possibility that each high-energy component is analyzed by the TF representation independent of its bandwidth and duration. The good localization of transients produced by MP is because of the good TF localization properties of the basis functions, whereas with
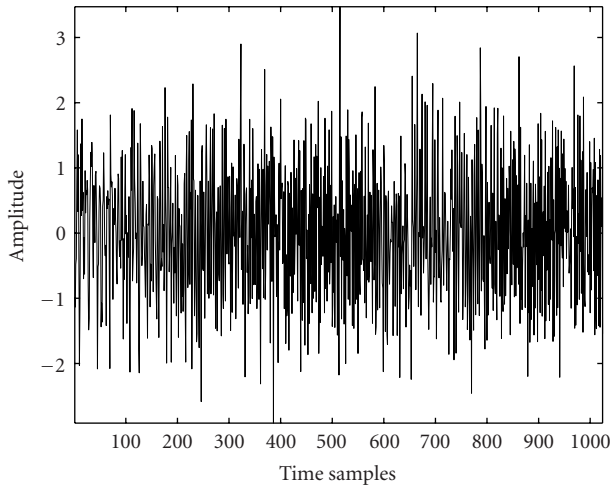
FIGURE 3: Monocomponent, nonstationary, synthetic signal "syn2" consisting of a chirp, an impulse, and a sinusoidal FM component (SNR = 0 dB).
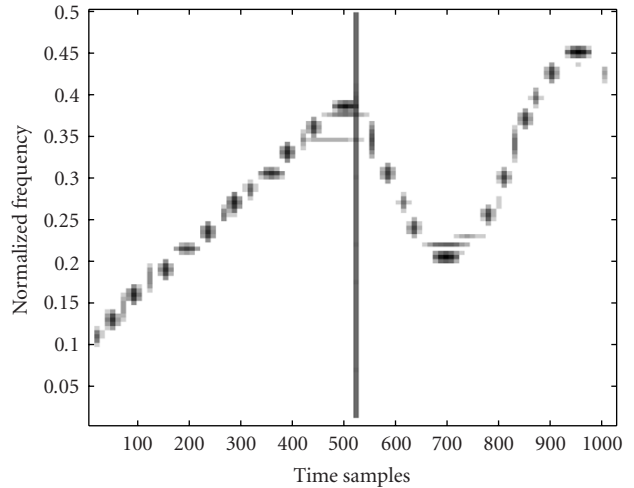

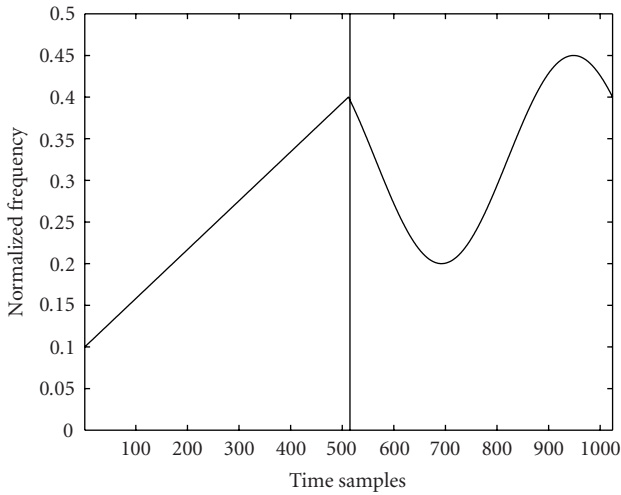FIGURE 5: OMP TFD of the signal "syn1" in Figure 2.

FIGURE 4: Ideal TFD depicting the frequency laws of the signals "syn1" and "syn2" in Figures 2 and 3.
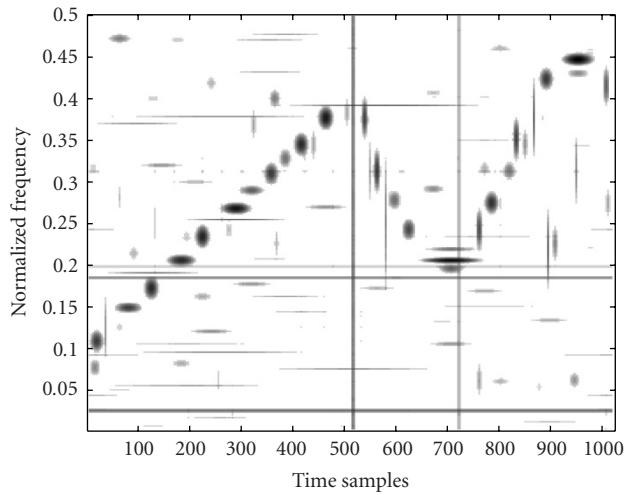

FIGURE 6: OMP TFD of the signal "syn2" in Figure 3.

other techniques such as Fourier and wavelets, the transient information gets diluted across the whole basis and the collection of basis functions is not as large compared to that in the MP dictionary.

The second simulation involves a group of two monocomponent signals and two multicomponent signals. The monocomponent signals have IF of a line and a sine wave; the multicomponent signals are made of two linear chirps either in parallel or in crossing positions.

Commonly known TFD estimators such as Wigner-Ville (WV) and pseudo Wigner-Ville (PWV) are used to calculate the TFD of the testing signals and estimate their IF (first moment in time). The results are then compared to those obtained using matching pursuit decomposition technique. Estimated IF values from TFD are compared to the known IF laws of the signal using the cross-correlation coefficient.

TABLE 1: Correlation coefficients between the estimated and referenced IF.

Correlation coefficients	WV	PWV	OMP
Linear (monocomponent)	0.9609	0.9611	0.9569
Sine (monocomponent)	0.9997	0.9998	0.9956
Linear (multicomponent 1)	0.1004	0.0994	0.7936
Linear (multicomponent 2)	0.6924	0.6868	0.9120

In the case of multicomponent signals, it is assumed that at any time instance, the energy of the individual components is equal, therefore IF law is the average of the individual IFs.

Table 1 gives the cross-correlation value between estimated IFs and known IF laws with different IF estimators.
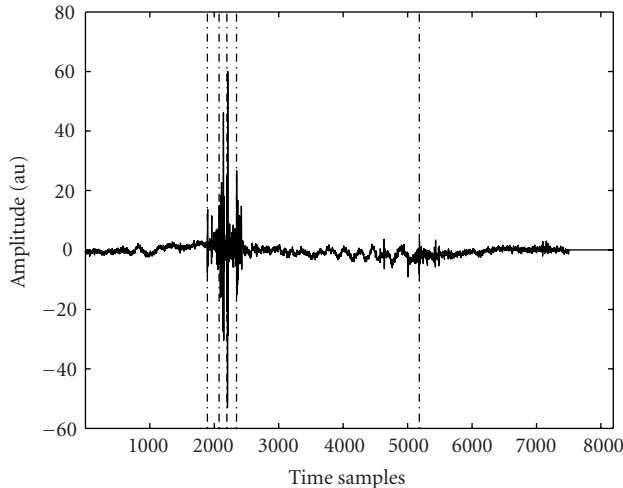
FIGURE 7: Knee sound signal of a normal subject. Broken lines are the adaptive segment boundaries denoting points of nonstationarity (*au* denotes arbitrary acceleration units).

FIGURE 8: IF estimated from the OMP TFD of the normal knee sound signal in Figure 7. Note that the unit of the frequency parameter is in Hz.

Performance of the TFD estimators varies depending on the input signals' characteristics such as linearity, rate of frequency change, being mono- or multicomponent, and the proximity of the frequency components in the signal. In the case of monocomponent signals, all estimated IFs are highly correlated with the corresponding IF reference. For multicomponent signals, WV and PWV became unreliable for estimating instantaneous mean frequency (IMF). Performance of WV degrades when there are more than one frequency component at a time instance and especially when the distance between the components is close. The matching pursuit decomposition technique (OMP TFD) has stable scores throughout the test. In general, matching pursuit can adapt well to different signal types because it can decompose the signals into known atoms and become cross-term free.

4.2. Real-world example

The proposed technique was applied to real-world signals, namely, the knee sound signals. Due to the differences in the cartilage surface between normal and abnormal knees, sound signals with different IFs are produced [30]. Figure 7 shows the knee sound signal of a normal subject. The IF of the same signal is shown in Figure 8. Automatic classification of the sound signals using IF as a feature for pattern classification has produced good results in screening abnormal knees from normal knees [30].

5. CONCLUSION

A novel method of extracting the IF of a signal is proposed in this paper. The extraction of IF is based on constructing an adaptive positive TFD that satisfies marginal properties and extracting the IF as a first central moment for each time slice. The method was tested on synthetic signals with known IF, and the results were found to be satisfactory even for low SNR cases.
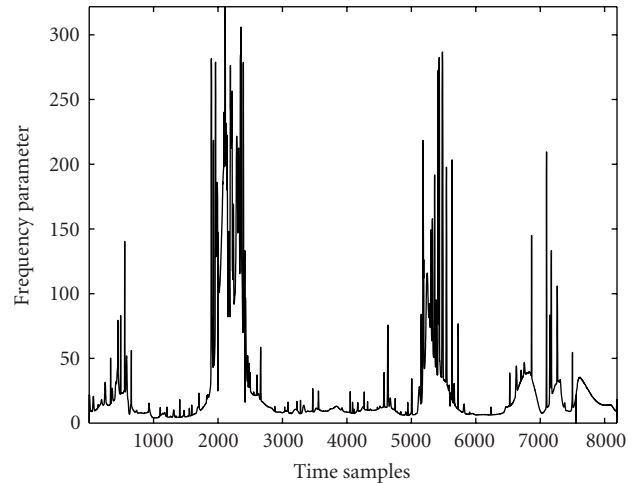
REFERENCES

[1] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals," *Proc. IEEE*, vol. 80, no. 4, pp. 520–538, 1992.

[2] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal. II. Algorithms and applications," *Proc. IEEE*, vol. 80, no. 4, pp. 540–568, 1992.

[3] L. Cohen, "Time-frequency distributions—a review," *Proc. IEEE*, vol. 77, no. 7, pp. 941–981, 1989.

[4] L. Cohen, *Time-Frequency Analysis*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1995.

[5] S. Erkucuk, S. Krishnan, and M. Zeytinoglu, "Robust audio watermarking using a chirp based technique," in *Proc. IEEE International Conference on Multimedia and Expo (ICME '03)*, vol. 2, pp. 513–516, Baltimore, Md, USA, July 2003.

[6] A. Ramalingam and S. Krishnan, "A novel robust image watermarking using a chirp based technique," in *Proc. IEEE Canadian Conference on Electrical and Computer Engineering (CCECE '04)*, vol. 4, pp. 1889–1892, Ontario, Canada, May 2004.

[7] I. Djurovic, V. Katkovnik, and L. Stankovic, "Instantaneous frequency estimation based on the robust spectrogram," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '01)*, vol. 6, pp. 3517–3520, Salt Lake City, Utah, USA, May 2001.

[8] F. A. C. de Bastos and M. L. R. de Campos, "A new method for estimating the instantaneous frequency based on maximum likelihood," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '00)*, vol. 2, pp. 677–680, Istanbul, Turkey, June 2000.

[9] A. A. Beex and P. Shan, "A time-varying Prony method for instantaneous frequency estimation at low SNR," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS '99)*, vol. 3, pp. 5–8, Orlando, Fla, USA, May–June 1999.

Estimation of Instantaneous Mean Frequency

[10] H. K. Kwok and D. L. Jones, "Improved instantaneous frequency estimation using an adaptive short-time Fourier transform," *IEEE Trans. Signal Processing*, vol. 48, no. 10, pp. 2964–2972, 2000.

[11] A. Akan, M. Yalcin, and L. F. Chaparro, "An iterative method for instantaneous frequency estimation," in *Proc. 8th IEEE International Conference on Electronics, Circuits and Systems. (ICECS '01)*, vol. 3, pp. 1335–1338, Malta, September 2001.

[12] Z. M. Hussain and B. Boashash, "Multi-component IF estimation," in *Proc. 10th IEEE Workshop on Statistical Signal and Array Processing (SSAP '00)*, pp. 559–563, Pocono Manor, Pa, USA, August 2000.

[13] P. Flandrin, "Some features of time-frequency representations of multicomponent signals," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '48)*, vol. 9, pp. 266–269, San Diego, Calif, USA, March 1984.

[14] I. Djurovic and L. Stankovic, "Robust Wigner distribution with application to the instantaneous frequency estimation," *IEEE Trans. Signal Processing*, vol. 49, no. 12, pp. 2985–2993, 2001.

[15] L. Stankovic and V. Katkovnik, "The Wigner distribution of noisy signals with adaptive time-frequency varying window," *IEEE Trans. Signal Processing*, vol. 47, no. 4, pp. 1099–1108, 1999.

[16] P. J. Loughlin and B. Tacer, "Comments on the interpretation of instantaneous frequency," *IEEE Signal Processing Lett.*, vol. 4, no. 5, pp. 123–125, 1997.

[17] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Processing Mag.*, vol. 9, no. 2, pp. 21–67, 1992.

[18] F. Hlawatsch, T. G. Manickam, R. L. Urbanke, and W. Jones, "Smoothed pseudo-Wigner distribution, Choi-Williams distribution, and cone-kernel representation: Ambiguity-domain analysis and experimental comparison," *Signal Processing*, vol. 43, no. 2, pp. 149–168, 1995.

[19] V. Sucic and B. Boashash, "Optimisation algorithm for selecting quadratic time-frequency distributions: performance results and calibration," in *Proc. 6th International Symposium on Signal Processing and Its Applications (ISSPA '01)*, vol. 1, pp. 331–334, Kuala Lumpur, Malaysia, August 2001.

[20] H.-I. Choi and W. J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 6, pp. 862–871, 1989.

[21] Z. Guo, L.-G. Durand, and H. C. Lee, "The time-frequency distributions of nonstationary signals based on a Bessel kernel," *IEEE Trans. Signal Processing*, vol. 42, no. 7, pp. 1700–1707, 1994.

[22] R. G. Baraniuk and D. L. Jones, "A signal-dependent time-frequency representation: optimal kernel design," *IEEE Trans. Signal Processing*, vol. 41, no. 4, pp. 1589–1602, 1993.

[23] V. Sucic, B. Barkat, and B. Boashash, "Performance evaluation of the B-distribution," in *Proc. 5th International Symposium on Signal Processing and Its Applications (ISSPA '99)*, vol. 1, pp. 267–270, Queensland, Brisbane, Australia, August 1999.

[24] P. J. Loughlin and K. L. Davidson, "Modified Cohen-Lee time-frequency distributions and instantaneous bandwidth of multicomponent signals," *IEEE Trans. Signal Processing*, vol. 49, no. 6, pp. 1153–1165, 2001.

[25] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[26] L. Cohen and T. Posch, "Positive time-frequency distribution functions," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, no. 1, pp. 31–38, 1985.

[27] J. Shore and R. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inform. Theory*, vol. 26, no. 1, pp. 26–37, 1980.

[28] J. Shore and R. Johnson, "Properties of cross-entropy minimization," *IEEE Trans. Inform. Theory*, vol. 27, no. 4, pp. 472–482, 1981.

[29] P. J. Loughlin, J. W. Pitton, and L. E. Atlas, "Construction of positive time-frequency distributions," *IEEE Trans. Signal Processing*, vol. 42, no. 10, pp. 2697–2705, 1994.

[30] S. Krishnan, R. M. Rangayyan, G. D. Bell, and C. B. Frank, "Adaptive time-frequency analysis of knee joint vibroarthrographic signals for noninvasive screening of articular cartilage pathology," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 6, pp. 773–783, 2000.

Sridhar Krishnan received the B.E. degree in electronics and communication engineering from Anna University, Madras, India, in 1993, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Calgary, Calgary, Alberta, Canada, in 1996 and 1999, respectively. He joined Ryerson University, Toronto, Ontario, in 1999 as an Assistant Professor in electrical and computer engineering, and was promoted to an Associate Professor in 2003. Sri Krishnan's research interests include adaptive signal processing, biomedical signal/image analysis, and multimedia processing and communications. He is a registered Professional Engineer of the Province of Ontario. He is a Senior Member of the IEEE and serves as the Chair of the Signals and Applications Chapter of the IEEE Toronto Section.