The Use of Cone-Shaped Kernels for Generalized Time-Frequency Representations of Nonstationary Signals

YUNXIN ZHAO, MEMBER, IEEE, LES E. ATLAS, MEMBER, IEEE, AND ROBERT J. MARKS, II, SENIOR MEMBER, IEEE

Abstract—**Generalized time-frequency representations (GTFR's) which use cone-shaped kernels for nonstationary signal analysis are presented. The cone-shaped kernels are formulated for the GTFR's to produce simultaneously good resolution in time and frequency. Specifically, for a GFTR with a cone-shaped kernel, finite time support is maintained in the time dimension along with an enhanced spectrum in the frequency dimension, and the cross-terms are smoothed out. Experimental results on simulated data and real speech showed the advantages of the GTFR's with the cone-shaped kernels through comparisons to the spectrogram and the pseudo-Wigner distribution.**

I. Introduction

THE bilinear class of time-frequency distribution [1], or the generalized time-frequency representations (GTFR's) [2], have offered the flexibility for nonstationary signal analysis. This class of representations is characterized by 2-D kernels, where the properties of a representation are determined by the constraints imposed on its kernel.

The spectrogram and the pseudo-Wigner distribution (PWD) are both GTFR's with special kernels [2]. The spectrogram preserves nonnegativity, smooths cross-terms, but requires separate analysis for either good time resolution or good frequency resolution. This behavior of the spectrogram has been shown to cause problems in, for example, the analysis of speech. It was shown in [3] that the formant frequency changes are the most important dimensions that affect the perception of phonetic quality. When the formants change rapidly or a closure occurs, however, consistent broadenings of formant spectrum are seen in the spectrogram [4]. The pseudo-Wigner distribution has the advantage over the spectrogram of instantaneous temporal response, but it produces interfering cross-terms when used in multicomponent signal analysis. The PWD does not preserve nonnegativity. A technique which optimally smooths a Wigner distribution into a positive distribution was described in [5].

We formulate a special type of kernel for the GTFR's to simultaneously preserve the property of finite time support, enhance spectral peaks, and smooth cross-terms. As in the case of the PWD, nonnegativity is not preserved. The representations thus primarily serve as tools for analyzing the time-frequency structure of the signals, but are not energy distributions. On the 2-D time plane, the kernel takes a cone-shaped support region. On the 2-D frequency plane, the kernel takes the form of a lateral inhibition function [6] in the dimension where it convolves with the signal spectrum, and takes the form of a low-pass filter in the dimension where the cross-terms are located. We also formulate an algorithm for efficient computation of the GTFR's with the cone-shaped kernels, and present experimental results which demonstrate the potential power of the technique for speech and other nonstationary signal analysis.

II. Background and Definitions

A generalized time-frequency representation $C_x(t,f;\phi)$ of the signal $x(t)$ with kernel $\phi(t,\tau)$ is [2]

$$C_x(t,f;\phi) = \int\int_{-\infty}^{+\infty} \phi(t-t',\tau)\, x(t'+\tau/2) \cdot x^*(t'-\tau/2)\, e^{-j2\pi f\tau}\, dt'\, d\tau \qquad (1)$$

where the * denotes complex conjugation. The relation expressed in terms of the Fourier transforms of the signal and the kernel is

$$C_x(t,f;\Phi) = \int\int_{-\infty}^{+\infty} \Phi(\eta, f-f')\, X(f'+\eta/2) \cdot X^*(f'-\eta/2)\, e^{j2\pi\eta t}\, df'\, d\eta \qquad (2)$$

where the Fourier transform relations are

$$X(f) = \int_{-\infty}^{+\infty} x(t)\, e^{-j2\pi ft}\, dt$$

and

$$\Phi(\eta,f) = \int\int_{-\infty}^{+\infty} \phi(t,\tau)\, e^{-j2\pi(t\eta+\tau f)}\, dt\, d\tau.$$

Manuscript received September 4, 1987; revised July 5, 1989. This work was supported in part by the National Science Foundation and Boeing Computer Services.

Y. Zhao was with the Interactive Systems Design Lab, University of Washington, Seattle, WA. She is now with the Speech Technology Laboratory, Panasonic Technologies Inc., Santa Barbara, CA.

L. E. Atlas and R. J. Marks, II, are with the Interactive Systems Design Lab, Department of Electrical Engineering, FT-10, University of Washington, Seattle, WA 98195.

IEEE Log Number 9035650.

Equation (1) shows that a GTFR is obtained through the convolution in time t of a kernel $\phi(t, \tau)$ with the signal correlation $x(t + (\tau/2))\, x^*(t - (\tau/2))$ and a Fourier transform in time τ. Equation (2) shows that a GTFR is obtained through the convolution in frequency f of a kernel $\Phi(\eta, f)$ with the spectrum correlation $X(f + \eta/2)\, X^*(f - \eta/2)$, and followed by an inverse Fourier transform in frequency η. For a discrete time GTFR, we follow the definition of the discrete PWD given in [7] and put it analogously to (1)

$$C_x(n, m; \phi) = 2 \sum_{n'=n-L}^{n+L} \sum_{k=-L}^{L} \phi(n - n', k)\, x(n' + k) \cdot x^*(n' - k)\, e^{-j(2\pi/M)mk} \qquad |m| \le L \tag{3}$$

where $M = 2L + 1$.

The kernels of the spectrogram and the PWD in terms of the GTFR's can be derived from (1) and (2). The two representations are commonly defined with respect to a sliding window $h(t)$ and the signal $x(t)$ [2]. Assuming $h(t)$ is real and symmetric, i.e., $h(t) = h(-t)$, the kernels $\phi(t, \tau)$ and $\Phi(\eta, f)$ of the PWD are

$$\phi(t, \tau) = \delta(t)\, h^2(\tau/2) \tag{4.a}$$

$$\Phi(\eta, f) = 4H(2f) * H(2f) \tag{4.b}$$

where $H(f)$ is the Fourier transform of $h(\tau)$ and the $*$ denotes 1-D convolution. For the spectrogram, the kernels become

$$\phi(t, \tau) = h(t + \tau/2)\, h(t - \tau/2) \tag{5.a}$$

$$\Phi(\eta, f) = H(f + \eta/2)\, H(f - \eta/2). \tag{5.b}$$

Relating these kernels with the definitions of the GTFR's in (1) and (2) reveals the properties of the pseudo-Wigner distribution and the spectrogram. Equation (4.a) indicates that the PWD kernel does not do smoothing in time t, therefore, the PWD has good temporal resolution. Equation (4.b) indicates that the kernel is constant in η, it thus cannot suppress cross-terms which are distributed off the $\eta = 0$ axis [8]. Since $H(f)$ is a low-pass function, smearing in frequency f is introduced through the convolution of the kernel with the signal spectrum. Equation (5.a) indicates that the spectrogram kernel does temporal smoothing due to its dependency on t; (5.b) indicates that the kernel $\Phi(\eta, f)$ is low pass in both f and η, and the spectrogram thus suppresses cross-terms but also has spectral smearing.

III. Cone-Shaped Kernel Design

A. Basic Principle

A desirable kernel of a GTFR produces good resolution in both time and frequency, and at the same time suppresses cross-terms. Many efforts in improving frequency resolution have focused on the design of windows with energy contained within a small frequency band. On the other hand, it has been shown that lateral inhibition functions are used in human visual and auditory systems in enhancing perceptual signal features [9], [10]. In lateral inhibition circuitry, the output of a neuron is a weighted sum of the outputs of neurons within a small neighborhood. The neurons within the center neighborhood (excitatory neighborhood) contribute positively, and those in the surrounding neighborhood (inhibitory neighborhood) contribute negatively. Lateral inhibition functions have long been used in vision for image feature enhancement. The cone-shaped kernels for the GTFR's are constrained to take the form of lateral inhibition functions in frequency f, thus they enhance spectral peaks when convolving with the signal spectrum. In the dimension of frequency η, suppressing cross-terms requires the kernel be low pass, which is equivalent to the requirement of temporal smoothing on the signal correlation function in terms of (1). Although many forms of temporal smoothing have been introduced in the literature [8], the finite time support property of the representations has been invariantly sacrificed. For the cone-shaped kernels, we take the special effort to constrain the temporal smoothing to the extent that the finite time support property is maintained, and at the same time the cross-terms are smoothed to a negligible extent.

B. Kernel Derivation

The notion of spectral enhancement requires that a kernel be a lateral inhibition function in frequency f. Since a lateral inhibition function in one domain corresponds to a band support in its Fourier transform domain and vice versa, a constraint is imposed on the support (nonzero region) of the kernel in τ. Letting $S(\phi(t, \tau))$ denote the support region of $\phi(t, \tau)$, the constraint becomes

$$S(\phi(t, \tau)) = \begin{cases} 1 & |\hat{\tau}_1(t)| \le |\tau| \le |\hat{\tau}_2(t)| \\ 0 & \text{otherwise} \end{cases}$$

where $|\hat{\tau}_1(t)|$ and $|\hat{\tau}_2(t)|$ denote the lower and upper boundaries of the kernel support region as functions of t. Since tapering of data in τ is necessary to reduce spectral leakage, we simplify the constraint so that the kernel support region is only limited from below by the boundary $|\hat{\tau}_1(t)|$. Note that when multiplied by a taper overlapping with the lower boundary, the kernel support will automatically be limited from above. Redefine $|\hat{\tau}(t)|$ to be the lower boundary function; we then have the constraint on the kernel support region as

$$S(\phi(t, \tau)) = \begin{cases} 1 & |\tau| \ge |\hat{\tau}(t)| \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

On the other hand, the constraint on the kernel for maintaining the finite time support property was given in [2] as

$$\int_{-\infty}^{+\infty} \hat{\phi}(\eta, \tau)\, e^{j2\pi\eta t}\, d\eta = 0, \qquad |\tau| < 2|t|$$

where $\hat{\phi}(\eta, \tau)$ is the Fourier transform of $\phi(t, \tau)$ in t. On the (t, τ) plane, this constraint reads

$$\phi(t, \tau) = 0 \qquad |\tau| < 2|t|.$$

This support region of $\phi(t, \tau)$ is shown as the shaded cone area on the (t, τ) plane in Fig. 1. Combining a taper function $g(\tau)$ with the boundary constant just derived, we arrive at the kernel

$$\phi(t, \tau) = \begin{cases} g(\tau) & |\tau| \geq a|t| \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where the parameter a adjusts the slopes of the cone with the constraint that $2 \leq a < \infty$.

The kernel for the discrete case follows a straighforward extension and is in the form

$$\phi(n, k) = \begin{cases} g(k) & |k| \geq a|n| \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where $|k| \leq L$ and $|n| \leq L$. Since k is not scaled by a factor of 2 in (3) as is the case of τ in (1), the lower bound of a is now 1 instead of 2.

Note that the boundary function $|\hat{\tau}(t)|$ is not limited to the line family $a|t|$ as long as it falls into the shaded cone region in Fig. 1. For utilizing all lags of signal correlations, however, the function should pass through the origin; and for smoothing the cross-terms, the function should have a certain spread in t.

The analysis of the kernel is done on the 2-D frequency and 2-D time planes. On the (η, f) plane, the function of the kernel in spectral enhancement and cross-term suppression can be visualized; on the (t, τ) plane, the function of the kernel in maintaining good resolution for fast-changing spectral peaks and preserving onset time of signals can be visualized.

C. Frequency Analysis of the Kernel

In the continuous frequency case, the kernel on the (η, f) plane can be obtained through a 2-D Fourier transform on (7), and it takes the form

$$\Phi(\eta, f) = \left\{ \frac{1}{2}\delta(\eta)\,\delta(f) + \frac{1}{a\pi^2}\frac{1}{(\eta/a - f)(\eta/a + f)} \right\} * G(f)$$

where the window $G(f)$ is assumed to be Gaussian and * denotes a 1-D convolution. The kernel in the discrete case can be obtained by taking the 2-D discrete Fourier transform of (8)

$$\Phi(l, m) = \sum_{n=-L}^{L} \sum_{k=-L}^{L} \phi(n, k)\, e^{-j(2\pi/M)(km+nl)} \qquad |m| \leq L, \quad |l| \leq L$$

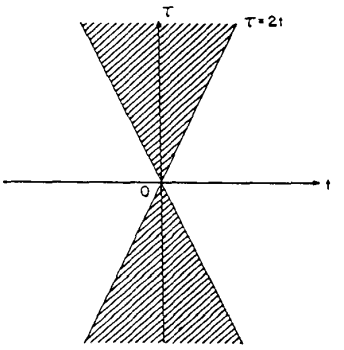

Fig. 1. The support region of the GTFR kernel $\phi(t, \tau)$ for maintaining the finite time support property.

and the kernel for the case $a = 1$ is

$$\Phi(l, m) = \left\{ M^2\delta(l)\,\delta(m) + \frac{\left(\cos\frac{\pi}{M}l - (-1)^{(l+m)}\cos\frac{\pi}{M}m\right)\cos\frac{\pi}{M}l}{\sin\frac{\pi}{M}(l-m)\sin\frac{\pi}{M}(l+m)} \right\} * G(m). \tag{9}$$

In Fig. 2 we compare the frequency responses in m of the cone-shaped kernels with angle parameters $a = 1, 2$, and the frequency responses of the spectrogram kernel and the PWD kernel, where the solid line is for the spectrogram, the medium dashed line for the PWD, the long dashed line for $a = 1$, and the short dashed line for $a = 2$, and (a), (b), (c), (d) correspond to the frequency $l = 0, 1, 2$, and 3, respectively. The window is $g(k) = e^{-2\alpha k^2}$, where $L = 64$, and α is chosen so that $g(\pm L) = 0.01$. The displayed frequency range is from $m = 50$ to 80, centered around $m = 64$. The frequency responses are normalized so that the same peak value is maintained by the 4 kernels at the origin on the 2-D frequency plane.

We observe from Fig. 2 that when $l = 0$, the frequency responses of the spectrogram kernel and the PWD kernel are identical, and the responses of the two cone-shaped kernels are very close. The positive peaks of the spectrogram and the PWD are wider than those of the cone-shaped kernels, and the cone-shaped kernels are characterized by the negative lobes on each side of the main positive peak—the shape of a lateral inhibition function. With the frequency l increasing, the response of the PWD kernel remains unchanged, but the peak values for other kernels are reducing. The peak value of the spectrogram drops the fastest, and the peak value of the cone-shaped kernels drops faster when $a = 1$ than when $a = 2$. In terms of the filtering property of the kernels in frequency l, the spectrogram kernel is low pass with a narrow band, the PWD kernel is all pass, and the cone-shaped kernels

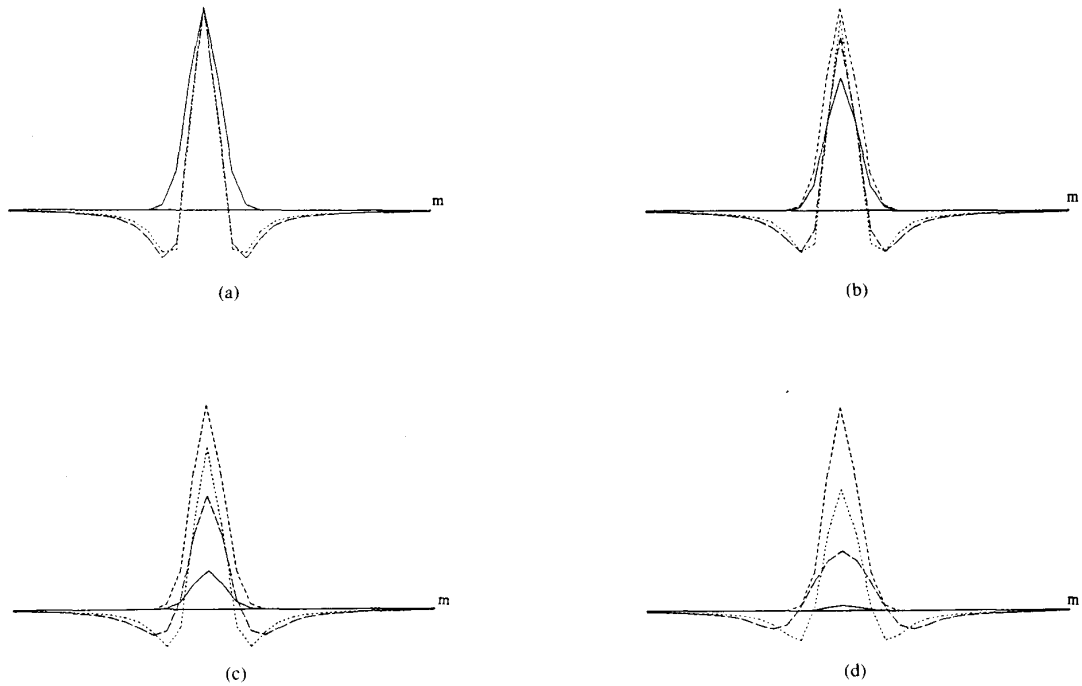

Fig. 2. The frequency responses of the cone-shaped kernels and the kernels of the spectrogram and the PWD in frequency m. The response points are jointed by lines, where the solid line is for the spectrogram, the medium dashed line for the PWD, the long dashed and the short dashed lines are for the cone-shaped kernels with $a = 1$ and $a = 2$, respectively. (a) Frequency $l = 0$; (b) frequency $l = 1$; (c) frequency $l = 2$; (d) frequency $l = 3$.

are in between the two cases with the angle parameter a controlling the low-pass bandwidth. To approach the power of the spectrogram kernel in suppressing cross-terms, $a = 1$ is chosen for the cone-shaped kernel in the later experiment.

D. Time Analysis of the Kernel

To analyze the effect of the cone-shaped kernel on fast-changing spectral peaks, we consider a signal which changes its tonal frequency from f_1 to f_2 at the time instance t_0, shown in Fig. 3(a). Let the signal be analytic for simplicity; we have

$$x(t) = e^{j2\pi f_1 t}U(t_0 - t) + e^{j2\pi f_2 t}U(t - t_0)$$

where $U(t)$ is the step function

$$U(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0. \end{cases}$$

The correlation function for $x(t)$ is then

$$\begin{aligned} &x(t + \tau/2)\, x^*(t - \tau/2) \\ &\quad = e^{j2\pi f_1 \tau}U[t_0 - (t + \tau/2)]\, U[t_0 - (t - \tau/2)] \\ &\qquad + e^{j2\pi(f_1 + f_2/2)\tau}\Big(e^{j2\pi(f_2 - f_1)t}U[(t + \tau/2) - t_0] \\ &\qquad \cdot U[t_0 - (t - \tau/2)] \\ &\qquad + e^{-j2\pi(f_2 - f_1)t}U[t_0 - (t + \tau/2)] \\ &\qquad \cdot U[(t - \tau/2) - t_0]\Big) \\ &\qquad + e^{j2\pi f_2 \tau}U[(t + \tau/2) - t_0]\, U[(t - \tau/2) - t_0]. \end{aligned}$$

The support region of this function on the (t, τ) plane is shown in Fig. 3(b), where the single hatched area to the left of t_0 is for the f_1 term, the right one for the f_2 term, and the double hatched area for the cross-terms. It is easy to visualize that when the cone-shaped kernel is moving along the time axis t, the f_1 and the f_2 terms are not smoothed together. Although the cross-terms are smoothed into the f_1 component when $t < t_0$, and into the f_2 component when $t > t_0$, these cross-terms are attenuated by the smoothing of the kernel along time t on the components $e^{j2\pi(f_2 - f_1)t}$ and $e^{-j2\pi(f_2 - f_1)t}$. Equivalently, the low-pass filtering of the kernel in η can filter out the cross-terms since these terms are displaced off the $\eta = 0$ axis by $\pm(f_2 - f_1)$. On the other hand, when a spectrogram kernel is moving along the time axis t, the f_1 and f_2 components are smoothed together. The boundary of the frequency change is thus smeared. In Fig. 3(c) we show a comparison of the spectrogram and the GTFR with a cone-shaped kernel for the signal $x(t) = \sin 2\pi f_1 t U(t_0 - t) + \sin 2\pi f_2 t U(t - t_0)$, where f_1 is 2 kHz and f_2 is 4.5 kHz. Shown from top to bottom is the signal, the spectrogram, and the GTFR with the cone-shaped kernel.

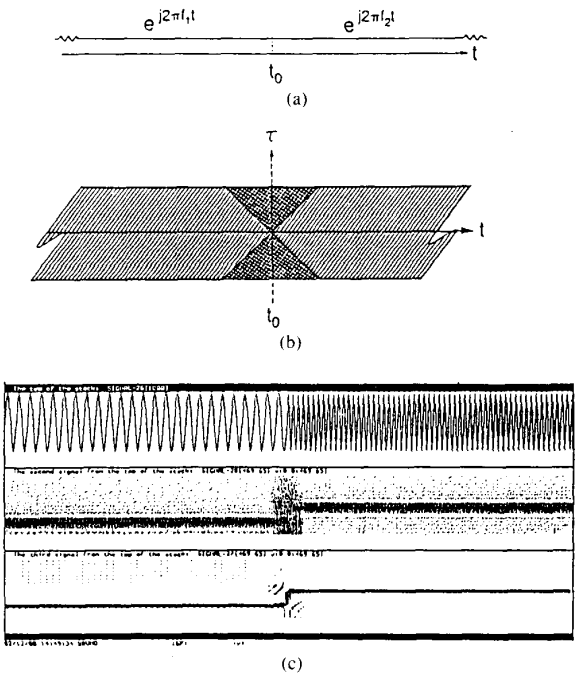

Fig. 3. Tracking a signal frequency change: the GFTR with a cone-shaped kernel does not smooth the two frequencies disjoint in time, whereas the spectrogram does. (a) The signal $x(t)$ with frequency f_1, $t < t_0$ and f_2, $t > t_0$. (b) The support region of $x(t + \tau/2)\, x^*(t - \tau/2)$, where the horizontal strip illustrates the tapering in τ. The single hatched area to the left of t_0 has the f_1 term, the right one has the f_2 term, and the double hatched area has the cross-terms. (c) The comparison of the spectrogram to the GTFR with a cone-shaped kernel, where the signal has $f_1 = 2$ kHz, $t < t_0$ and $f_2 = 4.5$ kHz, $t > t_0$. From top to bottom are the signal, the spectrogram, and the GTFR with a cone-shaped kernel. The horizontal axis is time and the vertical axis is frequency. The sampling frequency is 20 kHz, the window is Gaussian with length of 64 samples, and the analysis interval is 2 samples. The DFT size of the GTFR with a cone-shaped kernel is 128 and that of the spectrogram is 256.

For signals with imbedded zero intervals, it is sometimes important that finite time support holds in its GTFR. Let a zero interval of a signal be of length Z, and the kernel window length be M; the sufficient condition to maintain the zero interval in the GTFR with a cone-shaped kernel is $M - 1 \le Z$. An equivalent statement is that the GTFR at time n is zero as long as $x(n + k)\, x^*(n - k)$'s are zero for $|k| \le L$, where $L = (M - 1)/2$.

E. Computation Algorithm

Spectrograms are commonly computed by short time FFT, and the number of frequency bins is usually a power of 2. Using the definition of the discrete GTFR in (3), however, the window length M needs to be odd to maintain a real GTFR, thus, an FFT with radix 2 cannot be used directly to compute the GTFR's. Here we present an algorithm for the GTFR's with the cone-shaped kernels to maintain real transforms while utilizing an FFT of radix 2 in the transform computation.

Let $p = n - n'$; using a continuous frequency variable θ in (3) gives

$$C_x(n, \theta; \phi) = 2 \sum_{p=-L}^{L} \sum_{k=-L}^{L} \phi(p, k) \cdot x(n - p + k)\, x^*(n - p - k)\, e^{-jk\theta}.$$

Using (8) with $a = 1$, we have

$$\begin{aligned} C_x(n, \theta; \phi) &= 2 \sum_{k=-L}^{L} g(k)\, e^{-jk\theta} \sum_{p=-|k|}^{|k|} x(n - p + k) \cdot x^*(n - p - k) \\ &= 4 \sum_{k=0}^{L} \hat{g}(k) \cos(k\theta) \cdot \left\{ \sum_{p=-|k|}^{|k|} x(n - p + k)\, x^*(n - p - k) \right\} \\ &= 4 \sum_{k=0}^{L} \hat{g}(k)\, y(n, k) \cos(k\theta) \end{aligned} \tag{10}$$

where

$$\hat{g}(k) = \begin{cases} 0.5\, g(k) & k = 0 \\ g(k) & \text{otherwise} \end{cases}$$

and

$$y(n, k) = \sum_{p=-|k|}^{|k|} x(n - p + k)\, x^*(n - p - k).$$

Note that (10) can now be formulated as the real part of a standard DFT which can be computed with an FFT of radix 2 without affecting the realness of the GTFR, i.e.,

$$C_x(n, \theta; \phi) = 4 \operatorname{Re} \left(\sum_{k=0}^{L} \hat{g}(k)\, y(n, k)\, e^{-jk\theta} \right).$$

For example, with $M = 2^q - 1$, we can quantize θ into intervals of length $2\pi/2^q$.

IV. Experimental Results

For comparing the GTFR with the cone-shaped kernels (abbreviated below as GTFR-CK) to the spectrogram and the PWD, experiments were performed on simulated data and real speech. The computing facility was a Symbolics 3640 with Zetalisp running ISP [11]. Parameters used in the experiments were a sampling frequency of 20 kHz; an angle parameter $a = 1$ for the cone-shaped kernel; and the Gaussian windows $h(k) = e^{-\alpha k^2}$ and $g(k) = e^{-2\alpha k^2}$, where $g(\pm L) = 0.01$. The window length and analysis interval for the GTFR-CK, the PWD, and the spectrogram were the same within each experiment, and the FFT bin sizes of the GTFR-CK and PWD were half of that of the spectrogram since the folding frequencies of the former two are at $\pi/2$, and that of the spectrogram is at π. In the time-frequency displays, the horizontal axis repre-

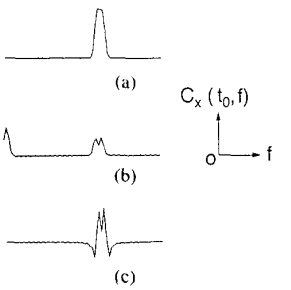

Fig. 4. Spectral profiles of a signal with two tones at 3.0 kHz and 3.08 kHz for illustrating the spectral enhancement capability of the GTFR-CK. The window length was 128 samples and the FFT bin size was 128 for the GTFR-CK. (a) Spectrogram. (b) PWD. (c) GTFR-CK.

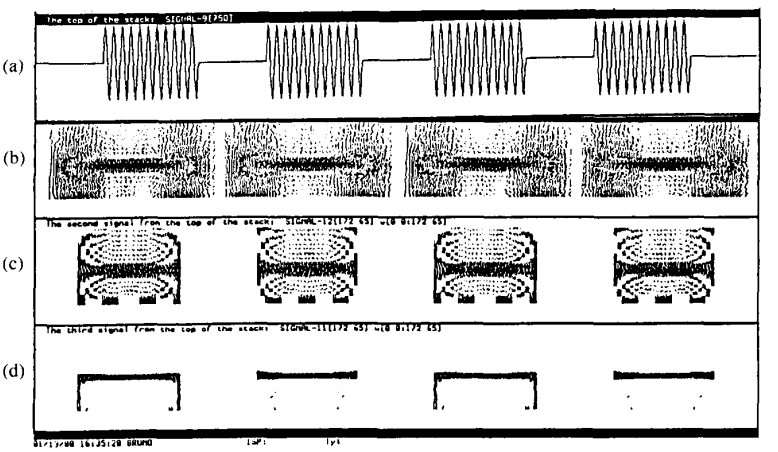

Fig. 5. The time-frequency representation of a single-tone pulse series for illustrating the capability of the GTFR-CK of simultaneously enhancing the signal spectrum and preserving the signal onset times. The frequency was 2.2 kHz, the pulse width was 100 samples, and the zero interval was 70 samples. The window length was 64 samples and the FFT bin size was 128 for the GTFR-CK. (a) Signal. (b) Spectrogram. (c) PWD. (d) GTFR-CK.

sents time and the vertical axis represents frequency from 0 to 5 kHz. For display purpose, negative values of the GTFR-CK and PWD were set to zero.

Experiment 1: This experiment illustrates the spectral enhancement capability of the GTFR-CK. The data were simulated two-tone signals of 3.0 and 3.08 kHz. The window length was 128 samples, and the FFT bin size was 128 for the GTFR-CK. In Fig. 4(a)–(c) are shown the spectral profiles of a time frame. We observe that the spectrogram does not resolve the two spectral peaks, the PWD is better but has a spike of interference at dc, and the GTFR-CK gives two distinct spectral peaks. The effect of lateral inhibition is also seen in the GTFR-CK where the small negative sidelobes sharpened the prominent positive peaks.

Experiment 2: This experiment illustrates the simultaneous spectral enhancement and finite time support of the GTFR-CK. The signal was a simulated single tone pulse series shown in Fig. 5(a). The frequency was 2.2 kHz, the pulse width was 100 samples, and the zero interval was 70 samples. The window length was 64 samples and the FFT bin size was 128 for the GTFR-CK. The time-frequency representations are shown in Fig. 5(b)–(d) for the spectrogram, the PWD, and the GTFR-CK, respectively. The spectrogram gives a smoothed picture in both time and frequency, the PWD maintains the time support of the signal but produces complicated patterns, and the GTFR-CK maintains the time support of the signal and at the same time gives a clear-cut spectrum line.

Experiment 3: This experiment demonstrates the power of the GTFR-CK in tracking fast-changing spectral peaks of a simulated signal $x(t) = \sin[A(\beta t - \nu \cos \beta t)] + \sin[A(\beta t + \nu \cos \beta t)]$, i.e., a mixture of two FM components with the instantaneous frequencies $\omega_1(t) = A\beta(1 + \nu \sin \beta t)$ and $\omega_2(t) = A\beta(1 - \nu \sin \beta t)$. The parameters were $\beta = 100\pi$, $A = 50.0$, and $\nu = 0.6$. The signal duration was 100 ms. The window length was 32 samples, analysis interval 2 samples, and the FFT size was 128 for the GTFR-CK. From top to bottom of Fig. 6(a)–(d) are shown the signal, the spectrogram, the PWD, and

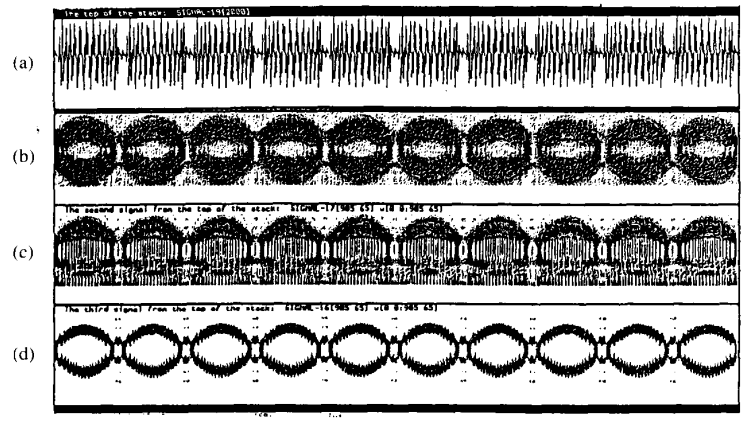

Fig. 6. The time-frequency representations of a mixture of two FM components for illustrating the capability of the GTFR-CK in tracking fast changing spectral peaks. The signal was $x(t) = \sin[A(\beta t - \nu \cos \beta t)] + \sin[A(\beta t + \nu \cos \beta t)]$. The parameters were $\beta = 100\pi$, $A = 50.0$, and $\nu = 0.6$. The signal duration was 100 ms. The data window length was 32 samples, analysis interval 2 samples, and the FFT size was 128 for the GTFR-CK. (a) Signal. (b) Spectrogram. (c) PWD. (d) GTFR-CK.

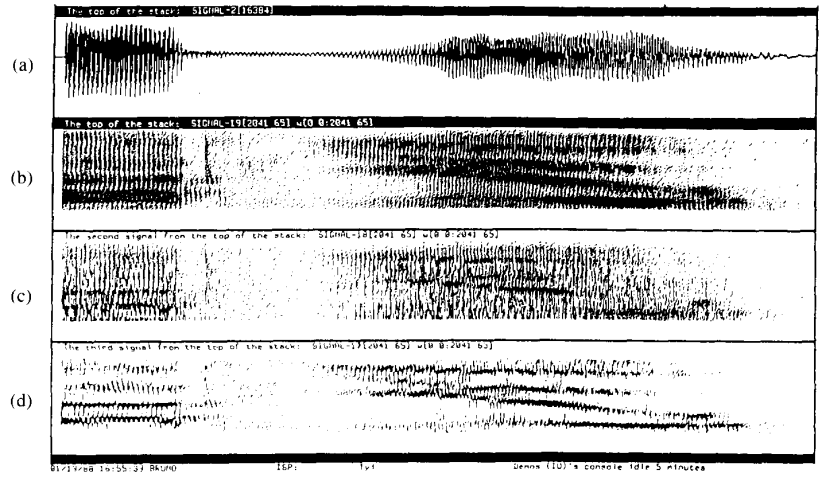

Fig. 7. The time-frequency representations of speech for illustrating the potential power of the GTFR-CK in application to speech analysis. The speech signal was "that you" spoken by a female speaker. The signal duration was 154 ms and was low-pass filtered at 5 kHz. The window length was 64 samples, analysis interval 8 samples, and FFT size was 128 for the GTFR-CK. (a) Signal. (b) Spectrogram. (c) PWD. (d) GTFR-CK.

the GTFR-CK. The GTFR-CK consistently maintains clear tracks of the spectral peaks whether or not the signal frequency is changing slowly or rapidly, the spectrogram smears the spectrum heavily, and the PWD suffers from a distorted display due to the interfering cross-terms.

Experiment 4: This experiment demonstrates the potential power of the GTFR-CK applied to speech analysis. The speech signal was "that you" spoken by a female speaker, and its waveform is shown in Fig. 7(a). The signal duration was 154 ms and was low-pass filtered at 5 kHz. The window length was 64 samples, analysis interval 8 samples, and FFT size was 128 for the GTFR-CK. Fig. 7(b)–(d) displays the spectrogram, the PWD and the GTFR-CK from top to bottom. Again, the PWD is interfered by cross-terms, and the GTFR-CK shows located formant tracks much more precisely than the spectrogram. A very prominent demonstration is in the low energy coarticulation between "that" and "you", where the for-

mants linking "that" and "you" are clearly visible in the GTFR-CK, but are almost smeared out in the spectrogram.

V. Summary

Experiments have shown the advantages of the GTFR with cone-shaped kernels in resolving close spectral peaks, maintaining zero intervals of signals, displaying clear tracks of fast-changing spectral peaks for the FM signals, and clear tracks of formants for speech signals. The technique therefore seems applicable to speech analysis and other nonstationary signal analysis.

Acknowledgment

The authors wish to thank R. Lyon of Apple Computers, Inc., Prof. R. D. Martin of the University of Washington Department of Statistics, Prof. J. Ritcey of the University of Washington Department of Electrical Engineering, and the anonymous reviewers for their helpful comments and suggestions.

References

[1] L. Cohen, "Generalized phase-space distribution functions," *J. Math. Phys.*, vol. 7, pp. 781–786, 1966.

[2] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The Wigner distribution—A tool for time-frequency signal analysis. Part 3: Relations with other time-frequency signal transformations," *Philips J. Res.*, vol. 35, pp. 373–389, 1980.

[3] D. H. Klatt, "Speech processing strategies based on auditory models," in *The Representation of Speech in the Peripheral Auditory System.* New York: Elsevier Biomedical, 1982, pp. 181–196.

[4] H. F. Silverman and Y.-T. Lee, "On the spectrographic representation of rapidly time-varying speech," *Comput. Speech Language*, vol. 2, pp. 63–86, 1987.

[5] J. C. Andrieux, M. R. Feix, G. Mourgues, P. Bertrand, B. Izrar, and V. T. Nguyen, "Optimum smoothing of the Wigner–Ville distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 764–769, June 1987.

[6] T. Kohonen, *Self-Organization and Associative Memory.* Berlin: Springer-Verlag, 1984.

[7] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The Wigner distribution—A tool for time-frequency signal analysis. Part 2: Discrete-time signals," *Philips J. Res.*, vol. 35, pp. 277–300, 1980.

[8] P. Flandrin, "Some features of time-frequency representations of multicomponent signals," in *Proc. ICASSP*, San Diego, CA, Mar. 1984, pp. 41B.4.1–41B.4.4.

[9] T. N. Cornsweet, *Visual Perception.* New York: Academic, 1970.

[10] L. Atlas, "Auditory coding in higher centers of the CNS," *IEEE Eng. Med. Biol. Mag.*, pp. 29–32, June 1987.

[11] G. Kopec, "The integrated signal processing system ISP," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 842–851, Aug. 1984.

Yunxin Zhao (S'86–M'88) received the B.S. degree in 1982 from Beijing Institute of Posts and Telecommunications, Beijing, China, and the M.S.E.E. and Ph.D. degrees in 1984 and 1988, respectively, from the University of Washington, Seattle.

She has done research on computer network performance analysis, speech and image processing, and recognition. She is currently with Speech Technology Laboratory, Panasonic Technologies Inc., Santa Barbara, CA, working on speech recognition.

Les E. Atlas (S'78–M'82) received the B.S.E.E. degree from the University of Wisconsin and the M.S. and Ph.D. degrees from Stanford University.

He joined the University of Washington in 1984 and is currently an Associate Professor of Electrical Engineering. He cofounded the Interactive Systems Design Laboratory at the University of Washington, and he is currently doing research in speech processing and recognition, neural network classifiers, and biologically inspired signal processing algorithms and architectures. His research in these areas is funded by the National Science Foundation, the Office of Naval Research, and the Washington Technology Center. He is also the past Chair of the Pacific Northwest Chapter of the Acoustical Society of America and has organized several special conference sessions on the applications of artificial neural networks and their comparison to more conventional techniques.

Dr. Atlas was a 1985 recipient of the National Science Foundation's Presidential Young Investigator Award.

Robert J. Marks, II (S'71–M'72–SM'83) received the Ph.D. degree in 1977 from Texas Tech University, Lubbock.

He joined the Faculty of the Department of Electrical Engineering at the University of Washington, Seattle, in December 1977, where he currently holds the title of Professor.

He was also the cofounder and first President of the Puget Sound Section of the Optical Society of America and was recently elected that organization's first honorary member.

Dr. Marks was awarded the Outstanding Branch Councilor Award in 1982 by the IEEE and, in 1984, was presented with an IEEE Centennial Medal. He is Chair of IEEE Neural Networks Council and was the cofounder and first Chair of the IEEE Circuits and Systems Society Technical Committee on Neural Systems and Applications. He is a Fellow of the Optical Society of America. He is a cofounder of Multidimensional Systems Corporation and the Christian Faculty Fellowship at the University of Washington. He is a member of Eta Kappa Nu and Sigma Xi.