Improved Time-Frequency Representation of Multicomponent Signals Using Exponential Kernels

HYUNG-ILL CHOI AND WILLIAM J. WILLIAMS, SENIOR MEMBER, IEEE

Abstract—**The time-frequency distributions which are members of a generalized Cohen's class have received a considerable amount of attention in signal processing areas. However, for multicomponent signals, the inherent bilinear structure of Cohen's class causes interfering cross terms which do not permit a straightforward interpretation of the energy distribution. The amount and shape of this interference are directly related to the kernel function which identifies the specific distribution belonging to Cohen's class.**

This paper introduces another time-frequency distribution of Cohen's class and examines its properties. This distribution is called the Exponential Distribution (ED) after its exponential kernel function. First, this paper interprets the ED from the spectral density estimation point of view. This paper then shows how the exponential kernel controls the cross terms as represented in the generalized ambiguity function domain, and analyzes the ED for two specific types of multicomponent signals: sinusoidal signals and chirp signals. Next, this paper defines the ED for discrete time signals and the Running Windowed Exponential Distribution (RWED) which is computationally efficient. Finally, this paper presents numerical examples of the RWED using synthetically generated signals. It is found that the ED is very effective in diminishing the effects of cross terms while retaining most of the properties which are useful for a time-frequency distribution.

I. INTRODUCTION

FOR many years, the representation of a signal in both the time and frequency domain (or space and frequency domain) has been of interest in signal processing areas, especially when one is dealing with time-varying nonstationary signals. Examples are such signals as speech signals, seismic signals, and EEG's (electroencephalograms). The frequency and amplitude of such signals vary with time.

The most familiar representation method is the spectrogram which is based on the assumption that for a short-time basis, signals can be considered to be stationary. Using this assumption, the spectrogam utilizes a short-time window whose length is chosen so that over the length of the window a signal can be considered to be stationary. Then the Fourier transform of this windowed signal can be used to obtain the energy distribution of the signal along the frequency direction at a given time which corresponds to the center of the window. The crucial drawback of this method is that the length of the window (i.e., the length of the assumed stationarity) is directly related to the frequency resolution. To increase the frequency resolution, one must use a longer observation duration (longer window), which means that the nonstationarities occurring during this interval will be smeared in time and frequency.

The spectrogram can be considered to be a member of Cohen's class of distributions [1]. This class of distributions utilizes a bilinear transformation that depends on two variables: time and frequency. Depending on the specific kernel function used for the bilinear transformation, various different distributions have been proposed, each with its own merits and drawbacks. Some examples of these distributions are the real part of the Complex Energy Density Function (RCEDF) with a cosine kernel function, the Rule of Born and Jordan (RBJ) with a sinc kernel function, and the Wigner Distribution (WD) which makes use of 1 as a kernel function. Especially, the WD has received a great deal of attention as a convenient tool for the analysis of monocomponent signals. These are signals whose instantaneous frequency and group delay correspond to the same curve in the time-frequency plane [2] (for example, sinusoidal signals or chip signals.)

However, as was pointed out by many authors [3]–[6], in the case of signals which can be expanded as the sum of their monocomponents, the inherent bilinear structure of Cohen's class of distributions causes undesirable interfering cross terms. These cross terms could become troublesome when one wishes to use time-frequency features for pattern recognition and classification. Many different combinations of auto-terms and cross terms may occur. Cross terms may occur at the position of auto-terms and obscure the auto-terms. The amount and the shape of this interference are directly related to the particular kernel function which is used to obtain the distribution of Cohen's class specific to that kernel. Therefore, an effective kernel for the time-frequency representation should be able to diminish the effects of the interfering cross terms while inducing the desirable properties which can sustain the validity of the time-frequency distribution.

This paper introduces the time-frequency distribution of Cohen's class obtained with an exponential kernel and examines its properties. This distribution will be called the Exponential Distribution (ED) after its kernel func-

Manuscript received July 10, 1987; revised August 26, 1988.

The authors are with the Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48109.

IEEE Log Number 8927472.

tion. The organization of this paper is as follows. Section II first examines the bilinear structure of the time-frequency distributions of Cohen's class. Next it defines the ED for the continuous time signals and examines its properties. Section III presents the definition of the ED for discrete time signals and examines its properties. Next, this section presents the Running Windowed Exponential Distribution (RWED) which is computationally efficient. Section IV presents numerical examples of the RWED using synthetically generated multicomponent signals, and some applications and comparisons to similar approaches are discussed.

II. The ED for Continuous Time Signals

In order to give a particular distribution an interpretation as a distribution of a signal's energy in time and frequency, the distribution should possess certain properties. For example, the time-frequency distribution should be a real valued distribution such that the shift in time (or frequency) of a signal results in the corresponding shift of the distribution. Also, the projection of the distribution on the time (or frequency) domain should be equal to the instantaneous power (or spectral density) of the signal. Furthermore, the centroid time (or frequency) of the distribution at each frequency (or time) should be equal to the group delay (or instantaneous frequency) of the signal.

The general class of time-frequency distributions was introduced by Cohen [1]. Each member of this class is given by

$$C_f(t, \omega, \phi) = \frac{1}{2\pi} \int\int\int e^{j(\xi\mu - \tau\omega - \xi t)} \phi(\xi, \tau) \cdot f\left(\mu + \frac{\tau}{2}\right) f^*\left(\mu - \frac{\tau}{2}\right) d\mu\, d\tau\, d\xi \tag{1}$$

where $f(\mu)$ is the time signal, $f^*(\mu)$ is its complex conjugate, and ϕ is a so-called kernel function, representative of the particular distribution function. The WD has the kernel $\phi(\xi, \tau) = 1$, the RCEDF has the kernel $\phi(\xi, \tau) = \cos(\xi\tau/2)$, and the RBJ has the kernel $\phi(\xi, \tau) = \text{sinc}(\xi\tau)$ (the range of all integrations is from $-\infty$ to ∞ unless otherwise stated).

It is difficult to precisely define a multicomponent signal. The most straightforward case of a multicomponent signal is a signal which has components which, when taken individually, would produce distinct contributions at particular time or frequency values (monocomponents). This is an oversimplification, but will serve to illustrate the point.

For a multicomponent signal $f(t)$ which can be represented as the sum of its monocomponents,

$$f(t) = \sum_{k=1}^{n} f_k(t)$$

where $f_k(t)$ is a monocomponent signal whose instantaneous frequency and group delay correspond to the same curve in the time-frequency plane [2], due to the sesquilinear structure in (1),

$$f\left(\mu + \frac{\tau}{2}\right) f^*\left(\mu - \frac{\tau}{2}\right) = \sum_{i=1}^{n} f_i\left(\mu + \frac{\tau}{2}\right) f_i^*\left(\mu - \frac{\tau}{2}\right) + \sum_{n \neq m}\sum f_n\left(\mu + \frac{\tau}{2}\right) f_m^*\left(\mu - \frac{\tau}{2}\right),$$

the time-frequency distribution is composed of the distributions of each component (auto-terms) and the interactions of each pair of different components (cross terms) as follows.

$$C_f(t, \omega, \phi) = \sum_{i=1}^{n} C_{f_i \cdot f_i}(t, \omega, \phi) + \sum_{n \neq m}\sum C_{f_n \cdot f_m}(t, \omega, \phi)$$

where

$$C_{f_i \cdot f_j}(t, \omega, \phi) = \frac{1}{2\pi} \int\int\int e^{j(\xi\mu - \tau\omega - \xi t)} \phi(\xi, \tau) f_i\left(\mu + \frac{\tau}{2}\right) \cdot f_j^*\left(\mu - \frac{\tau}{2}\right) d\mu\, d\tau\, d\xi.$$

As can be seen, the amount and the shape of the cross terms are determined by the characteristics of the kernel function $\phi(\xi, \tau)$. These cross terms cause redundancy in the information, and they may obscure the true energy distribution over time and frequency. Thus, these cross terms seriously affect later processes such as pattern recognition and classification. However, these cross terms also contribute to form the distribution. That is, one can not consider a time-frequency distribution which does satisfy the desirable properties mentioned and does not have the cross terms at all.

Several studies have reported attempts to reduce (or remove) the cross terms through operations on the distribution. For example, Martin and Flandrin [5] suggested spreading out the cross terms over a wide range by smoothing the WD independently in the time and frequency domain, and Marinovic and Eichmann [6] suggested removing the cross terms by taking the outer product expansion of the WD and removing the suspected cross terms. But, neither of these approaches could retain the desirable properties such as the marginal energies, the first local moments. Therefore, one needs to diminish the effects of the cross terms within the range which retains the desirable properties. The exponential kernel to be defined does this job well, as will be shown.

The ED, $E_f(t, \omega)$, is a member of Cohen's class with the kernel $\phi(\xi, \tau) = e^{-\xi^2\tau^2/\sigma}$, where $\sigma (\sigma > 0)$ is a scaling factor. By substituting this kernel into (1), one can

obtain

$$E_f(t, \omega) = \int_\tau e^{-j\omega\tau}\left[\int_\mu \frac{1}{\sqrt{4\pi\tau^2/\sigma}} \exp\left(-\frac{(\mu - t)^2}{4\tau^2/\sigma}\right) \cdot f\left(\mu + \frac{\tau}{2}\right) f^*\left(\mu - \frac{\tau}{2}\right) d\mu\right] d\tau. \quad (2)$$

First, we examine how the proposed exponential kernel satisfies the properties described at the beginning of this section. The shifts in time or frequency of a signal result in corresponding shifts in the ED, since all members of Cohen's class of distributions satisfy this property. The interested reader should refer to the series of papers by Claasen and Mecklenbrauker [7]–[9], since the properties of the ED follow directly from their observations.

The ED is a real valued distribution, since for its kernel $\phi(\xi, \tau) = e^{-\xi^2\tau^2/\sigma}$

$$\phi^*(\xi, \tau) = \phi(-\xi, -\tau).$$

The integral of the ED over all times at each frequency ω is equal to the spectral density of the signal at that frequency, the integral of the ED over all frequencies at each time is the instantaneous power at that time, and the integral over the whole (t, ω)-plane is equal to the total signal energy, since for its kernel

$$\phi(\xi, 0) = 1 \quad \text{for all } \xi$$

$$\text{and } \phi(0, \tau) = 0 \quad \text{for all } \tau.$$

Furthermore, the centroid time of the ED at each frequency is equal to the group delay, and the centroid frequency of the ED at each time is equal to the instantaneous frequency of the signal, since for its kernel

$$\phi(0, \tau) = 1 \quad \text{for all } \tau$$

$$\text{and } \frac{d}{d\xi}\phi(\xi, \tau)\Big|_{\xi=0} = 0 \quad \text{for all } \tau$$

$$\phi(\xi, 0) = 1 \quad \text{for all } \xi$$

$$\text{and } \frac{d}{d\tau}\phi(\xi, \tau)\Big|_{\tau=0} = 0 \quad \text{for all } \xi.$$

However, the ED is not a nonnegative valued distribution since $\phi(\xi, \tau)$ is not a form of a linear combination of ambiguity functions.

The above definition is given for the time function. A similar definition can be obtained in the frequency domain. That is,

$$E_f(t, \omega) = \frac{1}{2\pi}\int_\xi e^{-j\xi\tau}\left[\int_\mu \frac{1}{\sqrt{4\pi\xi^2/\sigma}} \exp\left(-\frac{(\mu - \omega)^2}{4\xi^2/\sigma}\right) \cdot F\left(\mu + \frac{\xi}{2}\right) F^*\left(\mu - \frac{\xi}{2}\right) d\mu\right] d\xi \quad (3)$$

where $F(\omega)$ is the Fourier transform of $f(t)$. The derivations of these definitions may be found in [10].

We next interpret the ED from the spectral density estimation point of view. A time-frequency distribution of Cohen's class $C_f(t, \omega, \phi)$ can be considered to be the Fourier transform of the time-indexed autocorrelation function $K(t, \tau)$ which is estimated at a given time t.

$$C_f(t, \omega, \phi) = \frac{1}{2\pi}\int_\tau e^{-j\tau\omega} K(t, \tau)\, d\tau \quad (4)$$

where

$$K(t, \tau) = \int_\xi \int_\mu e^{j(\xi t - \xi\mu)}\, \phi(\xi, \tau) f\left(\mu + \frac{\tau}{2}\right) \cdot f^*\left(\mu - \frac{\tau}{2}\right) d\mu\, d\xi. \quad (5)$$

As can be seen in (5), the kernel function $\phi(\xi, \tau)$ plays an important role in determining the characteristics of the time-indexed autocorrelation which is to be estimated utilizing time averages. Then a question arises as to which properties the kernel function should have for the estimation of the time-indexed autocorrelation. Since a time average would smooth out any time-varying property of nonstationary signals, the kernel function needs to attach a large weight to $f(\mu + \tau/2) f^*(\mu - \tau/2)$ when μ is close to t and a small weight to $f(\mu + \tau/2) f^*(\mu - \tau/2)$ when μ is far away from t. On the other hand, to give a reasonable accuracy to the estimate $K(t, \tau)$, the range of the time average should be increased for a large value of τ compared to the range of the time average for a small value of τ.

For the ED, the time-indexed autocorrelation is estimated as follows:

$$K_E(t, \tau) = \int_\mu \frac{1}{\sqrt{4\pi\tau^2/\sigma}} \exp\left[-\frac{(\mu - t)^2}{4\tau^2/\sigma}\right] \cdot f\left(\mu + \frac{\tau}{2}\right) f^*\left(\mu - \frac{\tau}{2}\right) d\mu. \quad (6)$$

That is, for the ED, each value of the time-indexed autocorrelation function is obtained from a set of neighboring samples with selective weights so that $f(\mu + \tau/2) \cdot f^*(\mu - \tau/2)$ has a large weight when μ is close to t and a small weight when μ is far away from t. Furthermore, the size of the neighbor set is controlled by the variable τ so that the size is increased for a large value of τ and decreased for a small value of τ. Fig. 1 shows the variations of this weighting factor as a function of τ and $\mu - t$ for a given σ.

In order to see the effectiveness of the proposed exponential kernel in diminishing the effects of interfering cross terms, we next examine the structure of the generalized ambiguity functions. An ambiguity function is another way of representing a signal in the joint time and frequency domain (time correlation with frequency shifts or frequency correlation with time shifts). Cohen and Posch [11] proposed the generalized ambiguity function by identifying an ambiguity function with the characteristic function of a time-frequency distribution.

$$A_f(\xi, \tau; \phi) = \iint C_f(t, \omega; \phi)\, e^{j(\xi t + \tau\omega)}\, dt\, d\omega \quad (7)$$

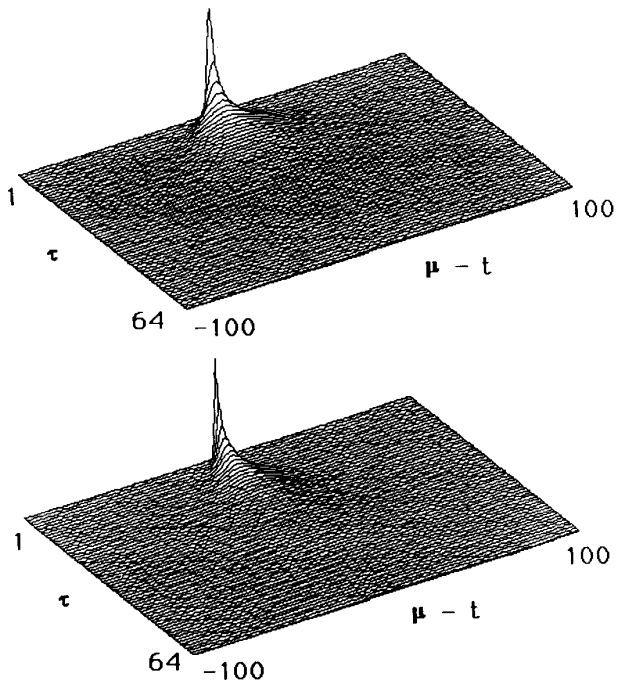

Fig. 1. Evaluations of the weighting factor which is in the time indexed autocorrelation function of the ED as a function of τ and $\mu - t$ for a given σ.

where $\phi(\xi, \tau)$ is a kernel function which identifies a specific distribution and ambiguity function. Then, by combining (1) and (7), the generalized ambiguity function has the following form:

$$A_f(\xi, \tau; \phi) = \phi(\xi, \tau) \int_\mu e^{j\xi\mu} f\left(\mu + \frac{\tau}{2}\right) \cdot f^*\left(\mu - \frac{\tau}{2}\right) d\mu. \tag{8}$$

As can be seen in (8), this ambiguity function also has a bilinear structure. That is, for a multicomponent signal which can be represented as the sum of its components

$$f(t) = \sum_{k=1}^{n} f_k(t)$$

the generalized ambiguity function can be decomposed into auto-terms and cross terms as follows:

$$A_f(\xi, \tau; \phi) = \sum_{i=1}^{n} A_{f_i \cdot f_i}(\xi, \tau; \phi) + \sum_{n \neq m} \sum A_{f_n \cdot f_m}(\xi, \tau; \phi)$$

where

$$A_{f_i \cdot f_i}(\xi, \tau; \phi) = \phi(\xi, \tau) \int_\mu e^{j\xi\mu} f_i\left(\mu + \frac{\tau}{2}\right) f_i^*\left(\mu - \frac{\tau}{2}\right) d\mu \tag{9}$$

$$A_{f_n \cdot f_m}(\xi, \tau; \phi) = \phi(\xi, \tau) \int_\mu e^{j\xi\mu} f_n\left(\mu + \frac{\tau}{2}\right) f_m^*\left(\mu - \frac{\tau}{2}\right) d\mu. \tag{10}$$

Through the relationship between an ambiguity function and a time-frequency distribution as in (7), we can note that (9) is a counterpart of an auto-term of a time-frequency distribution and (10) is a counterpart of a cross term of a time-frequency distribution. Another remarkable characteristic of this structure is, as was mentioned by Flandrin [12], that (9) has a locus passing through the origin of the ambiguity plane (ξ, τ) while the locus of (10) does not touch the origin and stays at a distance from the origin. Therefore, in order to emphasize auto-terms of a time-frequency distribution and deemphasize cross terms of a time-frequency distribution, it is quite natural to assign a large weight when ξ and τ are close to the origin of the ambiguity plane and a small weight when ξ and τ are far away from the origin as the proposed exponential kernel does.

We next analyze the ED for two types of multicomponent signals: sinusoidal signals and chirp signals. First, for a stationary signal which has two frequency components,

$$f(t) = A_1 e^{j(\omega_1 t + \theta_1)} + A_2 e^{j(\omega_2 t + \theta_2)}$$

the ED becomes

$$E_f(t, \omega) = 2\pi A_1^2 \delta(\omega - \omega_1) + 2\pi A_2^2 \delta(\omega - \omega_2) + 2A_1 A_2 \cos\left[(\omega_1 - \omega_2)t + \theta_1 + \theta_2\right] \cdot \textit{WEIGHT}$$

where

$$\textit{WEIGHT} = \sqrt{\frac{2\pi\sigma}{2(\omega_1 - \omega_2)^2}} \exp\left[-\frac{\sigma}{4(\omega_1 - \omega_2)^2} \cdot \left(\omega - \frac{\omega_1 + \omega_2}{2}\right)^2 \right]. \tag{11}$$

That is, *WEIGHT* controls the amplitude of the cross term by spreading it out in a selective manner, so that if interfering frequencies are ω_1 and ω_2, then the amplitude of the cross term at a certain frequency ω is inversely proportional to $\omega_1 - \omega_2$. In addition, the amount of interference exponentially decreases with the distance $(\omega - (\omega_1 + \omega_2)/2)^2$. Fig. 2 shows the variations of (11) for various values of ω_1(30 and 24 Hz), ω_2(34 and 40 Hz), and σ(0.1 and 1.0), where the origin of frequency axis corresponds to the average frequency of two interfering components.

Next, a similar analysis will be made for a multicomponent chirp signal which is defined as follows:

$$f(t) = A_1 e^{j\alpha_1 t^2/2} + A_2 e^{j\alpha_2 t^2/2}. \tag{12}$$

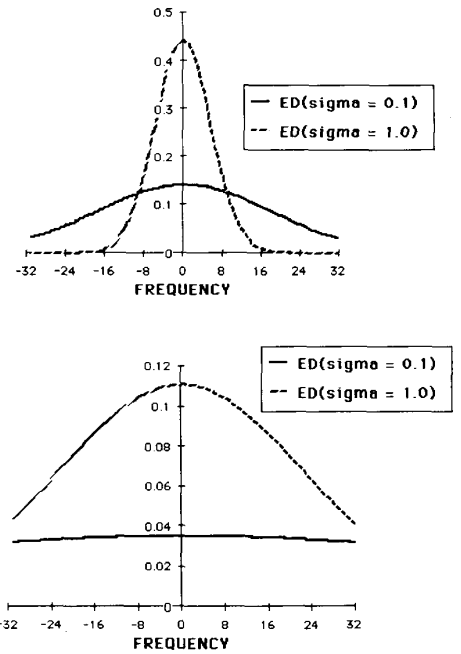

Fig. 2. Evaluations of the weighting factors of the ED which control the cross term of multicomponent sinusoids as a function of ω for given ω_1, ω_2, and σ.

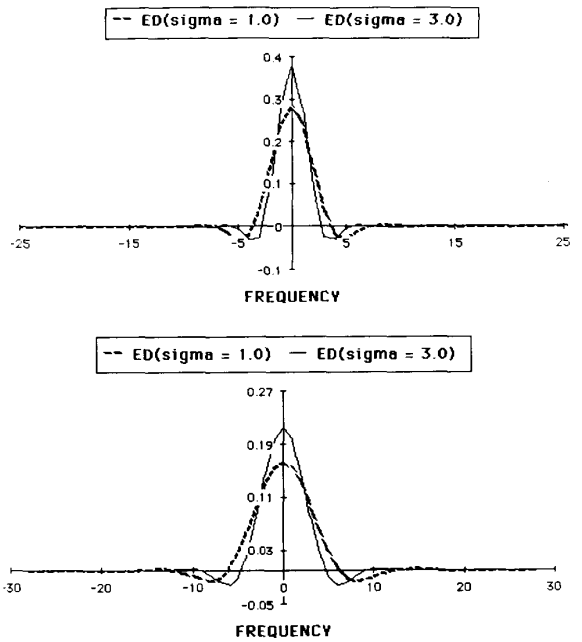

Fig. 3. Evaluations of auto-terms of the ED of multicomponent chirp signal as a function of ω for various values of α and σ.

That is, $f(t)$ is a chirp signal which has two frequency components at each time instant.

For this signal, ED becomes

$$E(t, \omega) = AUTO + CROSS$$

where

$$AUTO = A_1^2 \int_\tau e^{-j(\omega-\alpha_1 t)\tau} \exp\left[2\alpha_1^2\tau^4/(2\sigma)\right] d\tau + A_2^2 \int_\tau e^{-j(\omega-\alpha_2 t)\tau} \exp\left[2\alpha_2\tau^4/(2\sigma)\right] d\tau \tag{13}$$

$$CROSS = A_1A_2 \int_\tau \exp\left[-j(\omega - (\alpha_1+\alpha_2)t/2)\tau\right] \cdot \int_\mu e^{j(\alpha_1+\alpha_2)\mu\tau/2} \cdot \frac{1}{\sqrt{4\pi\tau^2/\sigma}} \exp\left[-\frac{\mu^2}{4\tau^2/\sigma}\right] \cos\left[(\alpha_1-\alpha_2) \cdot (\mu+t)^2/2 + (\alpha_1-\alpha_2)\tau^2/8\right] d\mu\, d\tau. \tag{14}$$

Since the integrations in (13) and (14) cannot be expressed as explicit forms, the auto- and cross terms of the ED will be examined numerically using a Discrete Fourier transform. Fig. 3 shows the auto-terms of the ED of the signal defined by (12) for various values of α(1.0 and 3.0) and σ(1.0 and 3.0), where the origin of frequency axis corresponds to frequency $\omega = \alpha t$ and a scale of an amplitude axis is adjusted so that the corresponding auto-term of the WD has a value of 1. As can be seen in this figure, the auto-term of the ED has a bell shape which becomes more concentrated as σ increases and more spread out as σ decreases. Fig. 4 shows the cross terms of the ED of the signal defined by (12) for various values of t(10 and 100) and σ(1 and 3) given $\alpha_1 = 1$ and $\alpha_2 = 3$, where the origin of the frequency axis corresponds to an average frequency of two interfering components. As can be seen in this figure, the cross terms of the ED are functions of the variable t. That is, the maximum magnitudes of these cross terms decrease as t increases, and the overall shapes of these cross terms become more spread as t increases. Furthermore, these cross terms are also controlled by σ so that a large σ concentrates the cross term around the origin and a small σ spreads out the cross term around the origin.

From this analysis, one can note that there should be a tradeoff between the auto-term resolution and the cross-term suppression. That is, in order to obtain a sharp auto-term resolution, σ should be large. On the other hand, in order to reduce the effects of the cross terms, σ should be small. Therefore, a large $\sigma(>1.0)$ is recommended for signals whose amplitude and frequency are changing relatively quickly and a small $\sigma(\leq 1.0)$ is recommended for signals whose amplitude and frequency are changing relatively slowly. The optimal choice of σ for a given situation needs more study, but a good choice of σ will be found in the range from 0.1 to 10 as will be shown later.

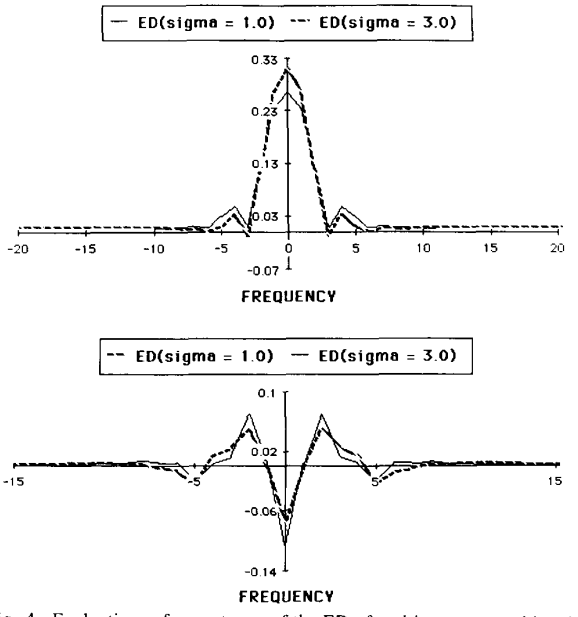

Fig. 4. Evaluations of cross terms of the ED of multicomponent chirp signal as a function of ω for various values of α_1, α_2, n, and σ.

III. The ED for Discrete Time Signals

When one is dealing with sampled signals or digital signals, it is necessary to consider a discrete version of the time-frequency distribution. The transition from the ED of continuous time signals to the ED of discrete time signals is not a trivial problem. Several cautions and approximations have to be made for this transition. This section gives the definition of the ED for discrete time signals and examines its properties. This section also defines the Running Windowed Exponential Distribution (RWED) which is the computationally efficient version of the ED.

For discrete time signals, the definition of the general time-frequency distribution can be given as follows:

$$C_f(n, \theta; \phi) = \frac{2}{2\pi} \sum_{\tau} \sum_{\mu} \int_{\xi=-\pi}^{\pi} e^{j(\xi\mu - 2\tau\theta - \xi n)} \phi(\xi, \tau) \cdot f(\mu + \tau) f^*(\mu - \tau)\, d\xi \qquad (15)$$

where n, τ, and μ are discrete variables, and ξ and θ are continuous variables (the range of summations is from $-\infty$ to ∞ unless otherwise stated). By substituting $\phi(\xi, \tau) = e^{-\xi^2\tau^2/\sigma}$ in the equation, one can derive the following definition of the ED for discrete time signals:

$$E_f(n, \theta) = 2 \sum_{\tau=-\infty}^{\infty} e^{-j2\theta\tau} \left[\sum_{\mu=-\infty}^{\infty} \frac{1}{\sqrt{4\pi\tau^2/\sigma}} \exp\left(-\frac{(\mu - n)^2}{4\tau^2/\sigma}\right) \cdot f(\mu + \tau) f^*(\mu - \tau) \right]. \qquad (16)$$

In this derivation, it is assumed that one can ignore the values of $(1/2\pi)\, e^{-\xi^2\tau^2/\sigma}$ outside the range $[-\pi, \pi]$ of ξ when $\tau \neq 0$, since they become very small (usually less than 0.005). The complete derivation of this definition can be found in [10].

The definition is given in the time domain. A similar definition can be obtained in the frequency domain. That is,

$$E_f(n, \theta) = \frac{1}{\pi} \int_{\xi=-\pi/2}^{\pi/2} e^{j2\xi n} \cdot \left[\int_{\nu=\xi-\pi}^{\xi+\pi} P\left(\frac{\exp\left[-(\nu - \theta)^2/(4\xi^2/\sigma)\right]}{\sqrt{4\pi\xi^2/\sigma}} \right) \cdot F(e^{j(\nu+\xi)})\, F^*(e^{j(\nu-\xi)})\, d\nu \right] d\xi \qquad (17)$$

where $F(e^{j\theta})$ is the discrete Fourier transform of $f(n)$ and $P(G(\theta))$ is the periodic function of $\boldsymbol{G}(\theta)$ which is the truncated version of $G(\theta)$ in the range $[-\pi/2, \pi/2]$. That is,

$$P\left(\frac{\exp\left[-(\nu - \theta)^2/(4\xi^2/\sigma)\right]}{\sqrt{4\pi\xi^2/\sigma}} \right) = \sum_{m=-\infty}^{\infty} \boldsymbol{G}(\nu - \theta + m\pi)$$

where $\boldsymbol{G}(\nu - \theta)$ is the truncated version of $G(\nu - \theta)$ such as

$$G(\nu - \theta) = \left(\frac{\exp\left[-(\nu - \theta)^2/(4\xi^2/\sigma)\right]}{\sqrt{4\pi\xi^2/\sigma}} \right)_{\nu-\theta \in (-\pi/2, \pi/2)}.$$

In this derivation, it is assumed that one can ignore the values of $G(\nu - \theta)$ outside the range of $(-\pi/2, \pi/2)$ of $\nu - \theta$, since they become very small (usually less than 0.005 [10]).

So far, the definitions of the ED for discrete time signals have been derived. The next topic of this section is to examine the properties of the ED of discrete time signals. The ED is a function of the discrete variable n and the continuous variable θ. While all spectra of discrete time signals are periodic with period 2π with respect to the continuous variable θ, the ED is periodic with a period of π as is the WD (assuming a sampling interval is 1 second).

$$E_f(n, \theta) = E_f(n, \theta + \pi).$$

This periodicity in π could cause aliasing effects, which prevent several of the properties of the continuous ED from carrying over in a straightforward manner. However, as was discussed in [6], one can consider the situations in which the aliasing does not cause any problems. That is, if the spectrum $F(e^{j\theta})$ has nonzero values only for the interval $[0, \pi]$ of θ, then the discrete ED does not suffer from the aliasing effects. In order to fulfill this requirement, one can increase the sampling frequency larger

than twice the Nyquist rate or one can interpolate a given discrete signal by a factor of 2. As another approach, one can consider the analytic signal $f(n)$ associated with a given real valued signal $f_r(n)$. For an analytic signal, one can use the usual sampling frequency according to the Nyquist rate, since the frequency spectrum of an analytic signal vanishes for negative frequencies. The analytic signal $f(n)$ is a complex valued signal in which the real part is equal to $f_r(n)$ and the imaginary part is equal to the Hilbert transform of $f_r(n)$. For the discrete ED, the latter approach will be taken.

Other properties of the discrete ED are as follows: the ED is a real valued distribution such that shifts in time or frequency of a signal results in the corresponding shifts in the distribution. Also an integral of the ED over all frequencies at each time n is equal to the instantaneous power of the signal at time n, and the summation of the ED over all time at each frequency θ is equal to the spectral density of the signal at that frequency. Furthermore, the centroid time of the ED is equal to the group delay of the signal. However, the centroid frequency of a periodic signal needs a careful definition [8], and further study is needed on the relationship between the centroid frequency of the ED and the instantaneous frequency of the discrete signal.

The next topic of this section is to define the Running Windowed Exponential Distribution (RWED). For signals consisting of many samples, it is no longer possible to calculate the time-frequency distribution for the complete signal. That is, for computational purposes it is necessary to apply the weighting windows, $W_N(\tau)$ and $W_M(\mu)$, for summations in (16) before evaluating the ED at each time index n. Then, by sliding these windows along the time axis, one can obtain the RWED which is defined as follows:

$$RWE_f(n, \theta) = 2 \sum_{\tau=-\infty}^{\infty} W_N(\tau)\, e^{-j2\theta\tau} \left[\sum_{\mu=-\infty}^{\infty} W_M(\mu) \frac{1}{\sqrt{4\pi\tau^2/\sigma}} \cdot \exp\left(-\frac{\mu^2}{4\tau^2/\sigma}\right) f(n+\mu+\tau) \cdot f^*(n+\mu-\tau) \right] \quad (18)$$

where $W_N(\tau)$ is a symmetrical window which has nonzero values for the range of $-N/2 \le \tau \le N/2$, and $W_M(\mu)$ is a rectangular window which has a value of 1 for the range of $-M/2 \le \mu \le M/2$. Denoting the summation in the parenthesis of (18) by $\boldsymbol{K}(n, \tau)$, (15) can be rewritten as follows:

$$RWE_f(n, \theta) = 2 \sum_{\tau=-\infty}^{\infty} W_N(\tau)\, e^{-j2\theta\tau} \boldsymbol{K}(n, \tau). \quad (19)$$

From this formula, it is clear that as long as M is large enough, the RWED is a smoothed version of the ED in the frequency domain, since multiplication in the time domain corresponds to convolution in the frequency domain. That is, the parameter N, the length of the window $W_N(\tau)$, and the shape of this window determines the frequency resolution of the RWED, while the parameter M, the length of the window $W_N(\mu)$, determines the range from which the time indexed autocorrelation function is to be estimated. Through experimental observations, it is found that oscillatory fluctuations of the cross terms can be reduced by decreasing the size of the window $W_N(\tau)$. On the other hand, it is also noted that the frequency resolution of the auto-terms decreases as the size of the window $W_N(\tau)$ decreases. That is, the size of the window $W_N(\tau)$ determines a tradeoff between the high frequency resolution of the auto-terms and the smoothed cross terms.

In order to perform a real-time analysis, a fast Fourier transform (FFT) technique can be utilized for the evaluation of the RWED. For this purpose, one can set $\theta = \pi k/N$. Then, (18) can be rewritten as follows:

$$RWE_f(n, k) = 2 \sum_{\tau=-\infty}^{\infty} W_N(\tau)\, e^{-j2\pi k\tau/N} \cdot \left[\sum_{\mu=-\infty}^{\infty} W_M(\mu) \frac{1}{\sqrt{4\pi\tau^2/\sigma}} \exp\left(-\frac{\mu^2}{4\tau^2/\sigma}\right) \cdot f(n+\mu+\tau) f^*(n+\mu-\tau) \right] \quad (20)$$

where n is the time index and k is the frequency index of the RWED. In order to evaluate (20), one first needs to obtain an analytic signal of a given real valued realization. The analytic signal can be obtained using two FFT's; a forward FFT of a given real valued realization, multiplication of resulting positive harmonics by 2, and negative harmonics by 0, and then an inverse FFT. Then, after obtaining the time indexed autocorrelation function corresponding to the inner summation of (20), an N-point FFT can be used to evaluate $RWE_f(n, k)$ at each time instant n.

IV. Experimental Results and Discussion

This section presents numerical examples of the RWED with the multicomponent signals which are synthetically generated. The list of signals to be considered is as follows:

$$g_1(n) = 4 \cdot \cos\left[2\pi(n/8) \cdot n/256\right] + 4 \cdot \cos\left[2\pi((512 - n)/8 + 40) \cdot n/256\right] \quad (21)$$

$$g_2(n) = 4 \cdot \cos\left[2\pi(n/4) \cdot n/256\right] + 4 \cdot \cos\left[2\pi(n/8) \cdot n/256\right] \quad (22)$$

$$g_3(n) = p(n) \cdot 4 \cdot \cos\left[2\pi \cdot 20n/256\right] + p(n) \cdot 4 \cdot \cos\left[2\pi \cdot 50n/256\right] \quad (23)$$

where

$$p(n) = \begin{cases} 1.0 - \frac{1}{400}(n - 128)^2 & \text{if } 108 \le n \le 148 \\ 0 & \text{elsewhere.} \end{cases}$$

Fig. 5 shows the signals defined by (21)–(23). That is, $g_1(n)$ is a chirp signal which has two frequency components at each time instant; the frequency of one component is increasing with time and the frequency of the other is decreasing with time, but the rate of change is the same in both cases. Signal $g_2(n)$ is another chirp signal which has two frequency components at each time instant. The frequencies of both components of $g_2(n)$ are increasing with time, but the rate of change is different. Signal $g_3(n)$ is a concentrated signal whose amplitude is controlled by the parabolic window $p(n)$, and this signal has two separated groups of frequency components in the time interval of $108 \le n \le 148$. The analytic signals of $g_1(n)$, $g_2(n)$, and $g_3(n)$ will be denoted by $f_1(n)$, $f_2(n)$, and $f_3(n)$, respectively. In this experiment, the windows $W_N(n)$ and $W_M(n)$ are set to be rectangular windows whose sizes are 128 and 29, respectively, and σ is set to be 2.0.

The first plots of Figs. 6–8 show the RWED's of each signal, and the second and third plots show the marginal energies of each RWED together with the instantaneous power of each signal. By comparing Fig. 6(a) and Fig. 7(a), one can note that the cross terms of $f_1(n)$ are much smaller than the cross terms of $f_2(n)$. That is, as the difference between α_1 and α_2, the frequency changing rates of each component, becomes larger, the magnitude of the cross term becomes larger, and vice versa. These plots also show that the time marginal energy of the RWED is exactly the same as the instantaneous power of the signal. That is, at the time instant where the instantaneous power of the signal is large, the RWED has a large value at the desired frequency. At the time instant where the instantaneous power of the signal is small, the RWED has a small value at the desired frequency. Such a property can also be seen in the auto-terms of the concentrated signal $f_3(n)$, Fig. 8(a), where the markers on the first row indicate starting and ending time bins of the concentrated signal defined by (23). On the other hand, one can note that the RWED has a smoothed frequency marginal energy compared to the spectral density of the signal at each frequency. That is, even though there are fluctuations of the spectral density of the signal along the frequency direction, the RWED exhibits fairly smooth changes along the frequency direction. This property results from the fact that the RWED is obtained using the finite rectangular windows $W_N(\tau)$ and $W_N(n)$.

This paper would not be complete without some comparison of the ED to the Wigner transform. Fig. 9 accomplishes this in effect, and shows that a choice of σ near 1 may be optimal. Values of 100, 10, 1, and 0.1 are used for analysis of the $g_2(n)$ signal (illustrated for $\sigma = 2$ in

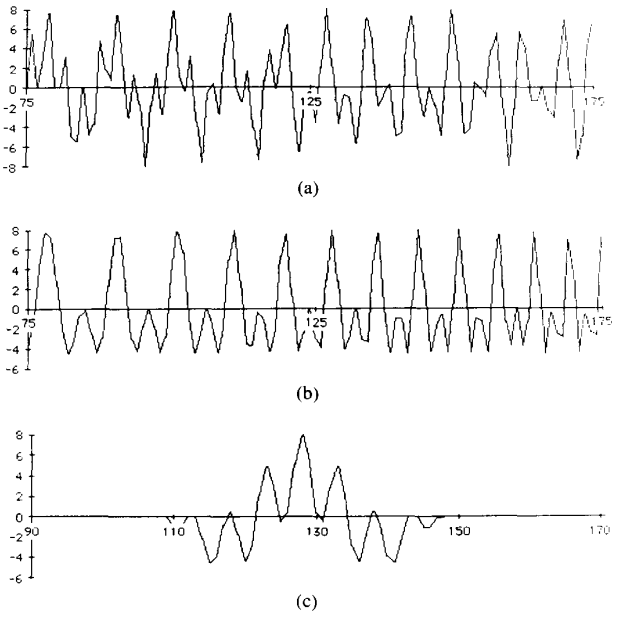

Fig. 5. (a) Plot of chirp signal which has two frequency components at each time: one is increasing and the other decreasing with same rate. (b) Plot of chirp signal which has two frequency components at each time: both of them are increasing with different rates. (c) Plot of the concentrated signal whose energy is concentrated in certain ranges of time and frequency.

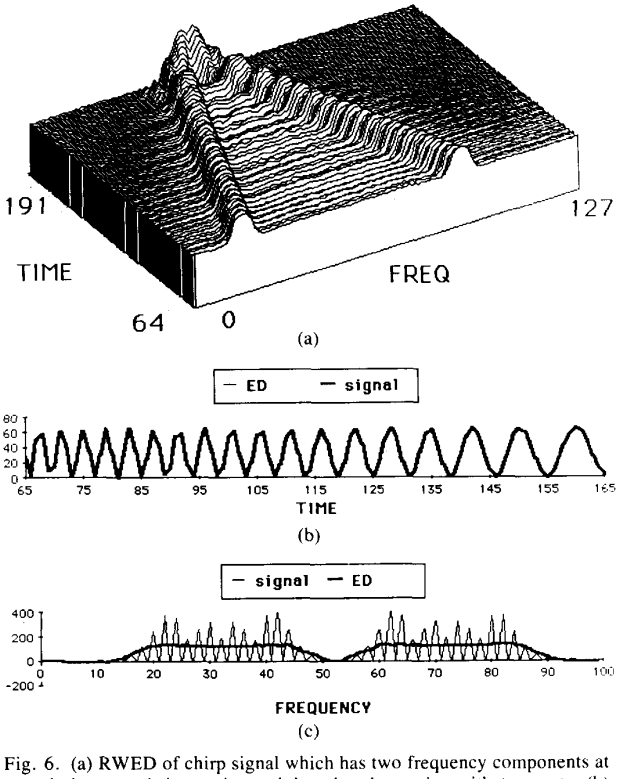

Fig. 6. (a) RWED of chirp signal which has two frequency components at each time: one is increasing and the other decreasing with same rate. (b) Instantaneous power of multicomponent chirp signal and time marginal energy of the corresponding RWED. (c) Frequency energy density of multicomponent chirp signal and frequency marginal energy of the corresponding RWED.

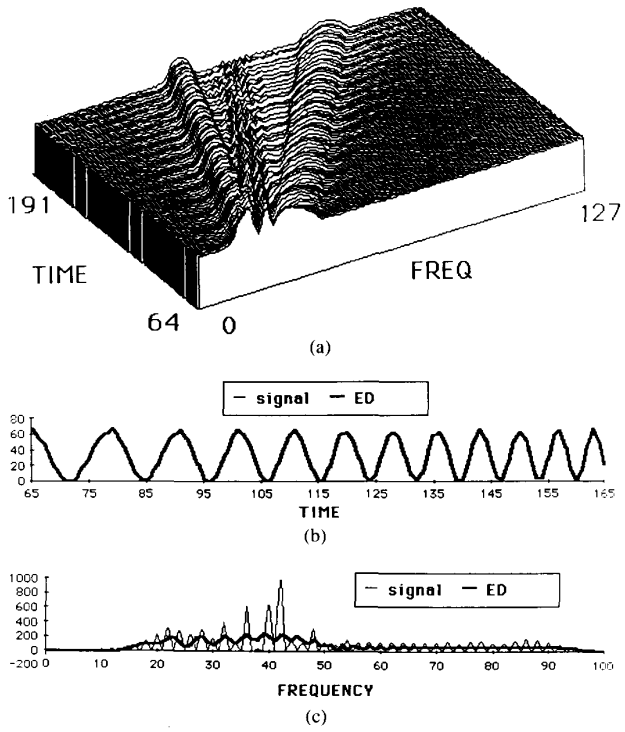

Fig. 7. (a) RWED of chirp signal which has two frequency components at each time: both of them increasing with different rate of changing. (b) Instantaneous power of multicomponent chirp signal and time marginal energy of the corresponding RWED. (c) Frequency energy density of multicomponent chirp signal and frequency marginal energy of the corresponding RWED.

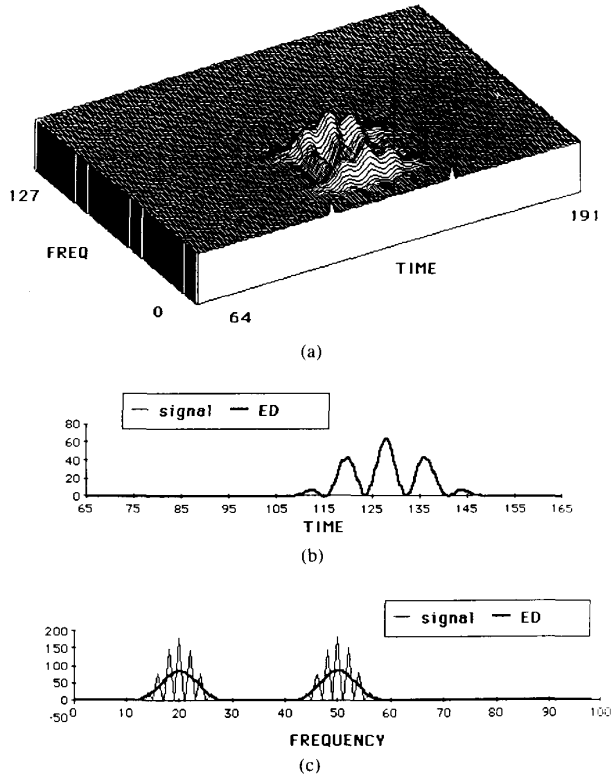

Fig. 8. (a) RWED of concentrated signal in both of time and frequency. (b) Instantaneous power of concentrated signal and time marginal energy of the corresponding RWED. (c) Frequency energy density of concentrated signal and frequency marginal energy of the corresponding RWED.

Fig. 7). The σ value of 100 causes the ED to approach the Wigner transform, since the kernel is nearly constant. It can be seen in Fig. 9(a) that the cross terms are large as one might expect. As σ is reduced, the cross terms diminish in size and seem to be quite small for $\sigma = 1$ and $\sigma = 0.1$. The small "bumps" between the auto-terms may be partially due to legitimate summation of the tails of the auto-terms themselves. This figure also illustrates the price to be paid to decreasing σ. The width of the auto-term component spreads and resolution becomes poorer. The loss of resolution seems to be quite mild, however. It also seems that a σ value of 1 may be sufficient to gain the benefits without paying much of a penalty. Whether or not this will be true, in general, remains to be discovered with further research.

These experiments clearly show the effectiveness of the Exponential Distribution for an analysis of multicomponent signals. The proposed exponential kernel controls the effects of the cross terms while preserving a relatively sharp resolution of the auto-terms. Furthermore, with the control parameter σ, one can adjust the resolution of the auto-terms and the effects of the cross terms according to the characteristics of the signal to be analyzed.

The exponential distribution was applied for the analysis of complex, multicomponent waveforms with successful results. The specific problem of interest to us involves the classification of brain waves (evoked potentials) evoked by words presented briefly to subjects by means of a tachistoscope. The goal is to classify the evoked potential in terms of the semantic class of the word being presented. Further details about the experimental goals and approach are available elsewhere [13].

Finally, we consider the difference between the inherent smoothing operation of the ED and the operation which smooths the WD with a Gaussian weight function as was suggested in [14] and [15]. Janssen and Claasen [14] suggested convolving the WD with a double Gaussian weight function so that the resulting distribution becomes nonnegative. Andrieux *et al.* [15] showed that this smoothing operation is optimal in a sense that the Gaussian weight function is best in smoothing and localizing the distribution. From the spectral density estimation point of view, one can easily conclude that this smoothing operation corresponds to estimating the time indexed autocorrelation function from a set of neighboring sample points with selective weights, but the size of this neighbor set is fixed instead of varying according to the autocorrelation index as in the ED. Thus, despite the superficial similarities concerning the Gaussian weighting function, the results are quite different.

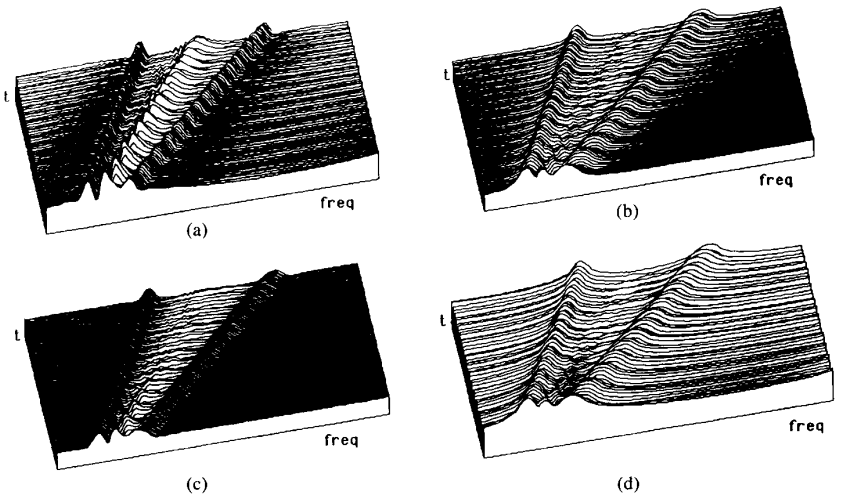

Fig. 9. Effects of σ on the ED. Signal $g_2(n)$ is analyzed using σ values of (a) $\sigma = 100$ (essentially the Wigner result), (b) $\sigma = 10$, (c) $\sigma = 1$, and (d) $\sigma = 0.1$.

REFERENCES

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

[2] K. Kodera, R. Gendrin, and C. Villedary, "Analysis of time-varying signals with small BT values," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 64-76, Feb. 1978.

[3] F. Hlawatsch, "Interference terms in the Wigner distribution," in *Digital Signal Processing—84*, V. Cappellini and A. G. Constantinides, Eds. New York: North-Holland, 1984, pp. 363-367.

[4] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "Time-frequency signal analysis by means of Wigner distribution," in *Proc. IEEE Int. Conf. ASSP*, vol. 3, 1981, pp. 69-72.

[5] W. Martin and P. Flandrin, "Wigner-Ville spectral analysis of nonstationary processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1461-1470, Dec. 1985.

[6] N. N. Marinovic and G. Eichmann, "An expansion of Wigner distribution and its application," in *Proc. IEEE Int. Conf. ASSP*, vol. 3, 1985, pp. 27.3.1-27.3.4.

[7] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The Wigner distribution—A tool for time-frequency signal analysis, Part I: Continuous time signals," *Philips J. Res.*, vol. 35, pp. 217-250, 1980.

[8] ——, "The Wigner distribution—A tool for time-frequency signal analysis, Part II: Discrete time signals," *Philips J. Res.*, vol. 35, pp. 276-300, 1980.

[9] ——, "The Wigner distribution—A tool for time-frequency signal analysis, Part III: Relations with other time-frequency signal transformations," *Philips J. Res.*, vol. 35, pp. 372-389, 1980.

[10] H. I. Choi, "Pattern analysis for nonstationary signals: Representation and Feature selection methods," Ph.D. dissertation, Univ. Michigan, 1987.

[11] L. Cohen and T. E. Posch, "Generalized ambiguity functions," in *Proc. IEEE Int. Conf. ASSP*, vol. 3, 1985, pp. 27.6.1-27.6.4.

[12] P. Flandrin, "Some features of time-frequency representations of multicomponent signals," in *Proc. IEEE Int. Conf. ASSP*, vol. 3, 1984, pp. 41B.4.1-41B.4.4.

[13] H. I. Choi, W. J. Williams, and H. Zaveri, "Analysis of event related potentials: Time-frequency energy distribution," *Biomed. Sci. Instrument.*, vol. 23, pp. 251-258, 1987.

[14] J. E. M. Janssen and T. A. C. M. Claasen, "On positivity of time-frequency distributions," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1029-1032, Aug. 1985.

[15] J. C. Andrieux *et al.*, "Optimum smoothing of the Wigner-Ville distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 764-769, June 1987.

Hyung-Ill Choi was born in Seoul, Korea, on August 16, 1956. He received the B.E. degree in electronics from Yonsei University, Seoul, Korea, in 1979, and the M.S.E. and Ph.D. degrees in computer, information and control engineering from the University of Michigan, Ann Arbor, in 1983 and 1987, respectively.

Since September 1987 he has been an Assistant Professor in the Department of Computer Science at Soong-Sil University, Seoul, Korea. His major interest areas are computer vision, artificial intelligence, and image processing.

William J. Williams (S'63-M'64-SM'73) was born in Ohio. He received the B.E.E. degree from Ohio State University, Columbus, in 1958, the M.S. degree in physiology from the University of Michigan, Ann Arbor, in 1966, and the M.S. and Ph.D. degrees in electrical engineering from the University of Iowa, Iowa City, in 1961 and 1963, respectively.

From 1958 to 1960 he was a Research Engineer at Battelle Memorial Institute, Columbus, OH. He was a Senior Engineer Specialist in communication systems with Emerson Electric from 1963 to 1964. He joined the Faculty of the (now) Electrical Engineering and Computer Science Department and was appointed Professor in 1974. During a sabbatical leave in 1974 he was a Visiting Scientist in Physiology at Johns Hopkins University. He has been active in the Bioengineering Program at the University of Michigan for a number of years. His interests include the theory and application of signal processing and communication techniques especially, but not exclusively, to biological problems.

Dr. Williams was the Guest Editor for a Special Issue of the PROCEEDINGS OF THE IEEE on Biological Signal Processing and Analysis which appeared in May of 1977. He is a member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi, and the Society for Neuroscience.