A proposed High-Resolution Time-Frequency Distribution for the Analysis of Multicomponent and Speech Signals

D. Boutana, B. Barkat , and F. Marir

Abstract— In this paper, we propose a novel time-frequency distribution (TFD) for the analysis of multi-component signals. In particular, we use synthetic as well as real-life speech signals to prove the superiority of the proposed TFD in comparison to some existing ones. In the comparison, we consider the cross-terms suppression and the high energy concentration of the signal around its instantaneous frequency (IF).

Keywords—Cohen's Class, Multicomponent signal, Separable Kernel, Speech signal, Time- frequency resolution.

I. INTRODUCTION

THE spectrogram, a smoothed version of the well-known Wigner-Ville distribution (WVD), has been widely used in speech applications [1], [2], [3], [4]. The spectrogram, which is in general a cross-terms free time-frequency distribution (TFD), suffers from the undesirable trade-off between the time concentration and the frequency concentration. To address the problem of cross-terms suppression, while keeping a high time-frequency resolution, other TFDs have been proposed. Among these, one can cite the smoothed pseudo WVD (SPWVD) [9], the Zhao-Atlas-Marks distribution (ZAMD) [5] and the B-distribution (BD) [6], just to name a few. In this paper, we present a new distribution for the analysis of multicomponent signals. This distribution, inspired from the Butterworth kernel quadratic TFD [8], has the ability of suppressing the cross-terms while keeping a high-resolution in the time-frequency plane. To assess the performance of this proposed distribution, we also propose to compare it to some existing ones known for their cross-terms suppression property. In the comparison examples, we use synthetic as well as real-life data from a speech application. The comparison results show the high performance of the proposed TFD in dealing with non-stationary multicomponent signals.

The paper is organized as follows: In Section 2, a theoretical aspect of some TFR interest is presented. A proposed high time-frequency resolution quadratic TFD is introduced in Section 3. In Section 4, simulations and comparison examples as well as a discussion are presented. Section 5 concludes the paper.

II. THEORETICAL BACKGROUND FOR QUADRATIC TFDs

Quadratic, a.k.a. bilinear or Cohen's, time-frequency representations are a powerful tool in the analysis of non-stationary signals analysis such as the speech signal, ECG signal, and others biomedical signals. Many of these representations are invariant to time and frequency translations and can be considered as energy distribution in time-frequency plane. The quadratic, or Cohen's, class of TFDs can be formulated as [1,3]

$$C_x(t,f) = \iiint e^{j2\pi(\xi t - \xi f - f\tau)} \Phi_x(\xi,\tau) x\left(u+\tfrac{\tau}{2}\right) x^*\left(u-\tfrac{\tau}{2}\right) du\, d\tau\, d\xi \quad (1)$$

where x(t) represents the analytical form of signal under consideration and $\Phi_x(\xi,\tau)$ is called the kernel of the distribution. All the integrals are from $-\infty$ to $+\infty$, unless otherwise stated. A choice of a particular kernel function yields a particular quadratic TFD. A different expression of the quadratic class may be given by

$$C_x(t,f) = \iint A_x(\xi,\tau)\Phi_x(\xi,\tau) e^{-j2\pi(f\tau+\xi t)} d\xi\, d\tau \quad (2)$$

where $A_x(\xi,\tau)$, the ambiguity function, is given by

$$A_x(t,f) = \int x\left(u+\tfrac{\tau}{2}\right) x\left(u-\tfrac{\tau}{2}\right) e^{j2\pi\xi u} du \quad (3)$$

Under the terms of Equation (2), we observe that the time-frequency distribution is a two-dimensional Fourier transform of the product of the kernel and the ambiguity function. The choice of a particular kernel function defines a different distribution with its own specificities [1], [2], [3]. For instance, the spectrogram (SP), traditionally used for the time-frequency analysis of speech signals, is defined by selecting the kernel as the ambiguity function of an arbitrary window function. Another example is the Zhao-Atlas-Marks distribution kernel [5], also called cone-shaped kernel. This particular kernel, which ensures the time support property of its corresponding TFD (unlike the SP), is defined as

$$\Phi_x(\xi,\tau) = |\tau| g(\tau) \frac{\sin \pi\xi\tau}{\pi\xi\tau} \quad (4)$$

The B-distribution kernel [6], defined in the time-lag plane, can be expressed as

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:1, No:2, 2007

$$\Phi_x(t,\tau) = \left[\frac{|\tau|}{\cosh^2 t} \right]^{\beta} \qquad (5)$$

where β is a real parameter that controls the sharpness of cut-off of the 2-D filter in the ambiguity domain. The β values range between zero and unity $0 < \beta < 1$ [6].

As a final example of the kernel, we can cite the Butterworth kernel given by [7,8]

$$\Phi_x(\xi,\tau) = \frac{1}{1 + \left(\frac{\xi}{\sigma_\xi}\right)^{2N} \left(\frac{\tau}{\sigma_\tau}\right)^{2M}} \qquad (6)$$

This kernel function is regarded as a general form of the exponential kernel representations and can be considered as a low-pass filter in the ambiguity domain. A suitable choice of the various parameters N, M, σ_ξ and, σ_τ helps to remove the cross-terms that appear in the time-frequency domain in a multicomponent signal analysis.

III. PROPOSED TIME-FREQUENCY DISTRIBUTION

The adoption of a separable kernel function defines the pseudo smoothed Wigner-Ville (SPWV) [9] which has the advantage of reducing the effects of the interferences (ITs) or cross-terms and, in the same time, having a high time-frequency resolution. The general expression of separable kernel is written as

$$\Phi(\xi,\tau) = G(\xi)h(\tau) \qquad (7)$$

and its corresponding TFD can be expressed as

$$SPWVD(t,f) = \int h(\tau) \left[\int g(s-t) x\left(s+\frac{\tau}{2}\right) x^*\left(s-\frac{\tau}{2}\right) ds \right] e^{-j2\pi f\tau} d\tau \qquad (8)$$

Here, we propose a separable kernel, inspired from the Butterworth kernel, to have a good trade-off between cross-terms suppression and high time-frequency resolution. We define the kernel as follows

$$\Phi_x(\xi,\tau) = \frac{1}{1 + \left(\frac{\xi}{\sigma_\xi}\right)^{2N}} \cdot \frac{1}{1 + \left(\frac{\tau}{\sigma_\tau}\right)^{2M}} \qquad (9)$$

We opt for this separable configuration because it gives us the flexibility of controlling both the time and the frequency resolutions independently. Using the inverse Fourier Transform and fixing N equal to unity, N=1, we obtain the time-lag kernel expression given by

$$\Phi_x(t,\tau) = \frac{1}{1 + \left(\frac{\tau}{\sigma_\tau}\right)^{2M}} \pi\sigma_\xi \exp(-\pi\sigma_\xi |t|) \qquad (10)$$

Now, by substituting expression (9) in Equation (8), we obtain the proposed TFD expression as

$$SPWVD(t,f) = \int \frac{1}{1 + \left(\frac{\tau}{\sigma_\tau}\right)^{2M}} \cdot R_t(\tau) e^{-j2\pi f\tau} d\tau \qquad (11)$$

Where

$$R_t(\tau) = \int \pi\sigma_\xi e^{-\pi\sigma_\xi|s-t|} x\left(s+\frac{\tau}{2}\right) x\left(s-\frac{\tau}{2}\right)^* ds \qquad (12)$$

IV. EXAMPLES AND DISCUSSION

A speech signal, formed by multiple frequency components (formants) generates a great number of undesired cross-terms in time-frequency domain. It is very desirable to eliminate these cross-terms in order to facilitate the analysis of such signal. By selecting an appropriate kernel we can obtain less cross-terms in the time-frequency plane; however, the design of such kernels is, in general, not easy. To facilitate the task, we prefer to use separable kernels, as discussed earlier. The choice of the functions g(t) and h(t) is based on the a priori knowledge of the speech signal itself. Usually, the length of g(t) must be less or equal to the time of stationarity of the signal. In [10], it was shown that for a speech signal, in the ambiguity plane, the interference terms can be reduced by limiting the extent of the variable ξ (i.e., $|\xi| > 2f_0$) where f_0 is the bound between the auto terms and cross terms .It must satisfy $f_0 \leq \frac{1}{4} f_{pitch}$ (f_{pitch} being the pitch frequency or fundamental frequency of the speech signal). Also, It is known that pitch frequency of a speech signal for men varies between 80Hz< f_{pitch} < 200 Hz. Based on these observations, in the simulation examples, we fixed the value $\sigma_\tau = 0.004$ (refer to Equations (10)-(12) and the window length to a quarter of the signal duration [1].The input speech signals considered in the examples were spoken in Arabic and digitalized with 8 bits at 11KHz sampling frequency.

A. Example 1: Signal with close components

In this first example, the synthetic signal consists of four sinusoids, very close in pairs. The first two sinusoids have frequencies equal to 1000 Hz and 1200 Hz, whereas the second pair of sinusoids have frequencies equal to 3000Hz and 3200 Hz, respectively. The signal length is fixed at N=256 and the sampling frequency is fixed equal to f_s=1Hz.This model with very closely components is discussed in order to prove the superiority of the proposed technique in the possibility of resolving close components in the time-frequency domain. As can be seen from Fig. 1 , the simulations results obtained for the ZAMD, the BD and the proposed TFD show a better performance, in terms of frequency resolution, than the SP. Moreover, we can also observe that the highest performance is achieved for the proposed TFD.
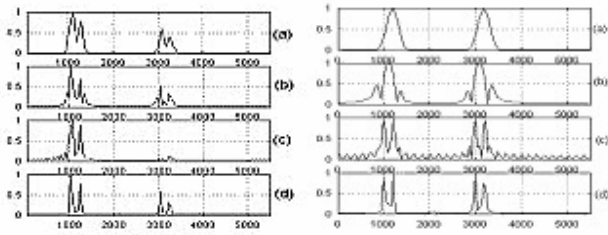
World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:1, No:2, 2007

Fig 1. Slices taken at the same instant n=65 (left) and n=129 (right)
(a) SP, (b) ZAMD, (c) BD, (d) Proposed TFD.

In this example, we use two different window lengths in the evaluation of the TFDs. Namely; we use a medium size window length (65 samples) and a large size window length (129 samples). For each window length, we take slices of the TFDs at time instants n= 64 and n=129 (recall that n= 0,1,2, ..., 255). We plot the normalized amplitudes of these slices, for each window length, in Figures 1 and 2, respectively. Once again, we can observe that the proposed TFD not only can successfully separate the components but it has the best resolution   (i.e., narrower main-lobe and smaller side-lobes) compared to all the other considered distributions.

B. Example 2: Two crossed chirps and a constant frequency

In this example, the synthetic signal consists of two crossed chirps (one linearly increasing from 0.1Hz to 0.3Hz, the second linearly decreasing from 0.3Hz to 0.1Hz) and a constant frequency at 0.4 Hz. A unity sampling frequency is considered here with a signal length equal to 256. The same TFDs, for a window length equal to nh=65, are represented in Figure 2.
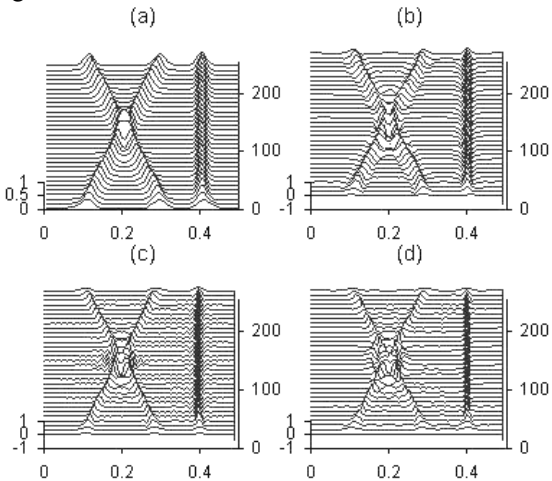

Fig 2. (a) SP, (b) ZAMD, (c) BD, (d) Proposed TFD. The analysis window length nh=65.

Slices of these TFDs, taken at the same time instant n=64, are shown in Figure 3 for various analysis window lengths (i.e.,33, 65 and 129 samples). Here again, we see that the proposed TFD does better than the other ones. Now, some additive white Gaussian noise is added to the signal.

We consider three different cases: signal-to-noise ratio (SNR)

equal to 0 dB, 5 dB and 10 dB. In Figure 4 we plot the same slices as above. The same conclusion can be drawn here again with the exception that at low SNR, all the TFDs start to show some distortions.
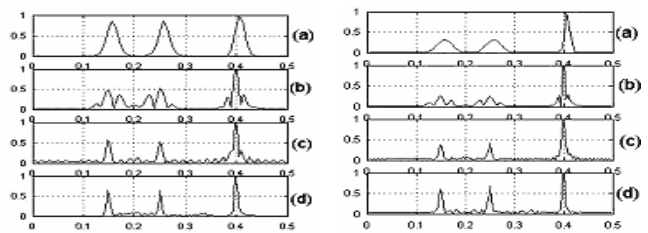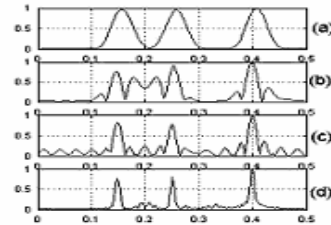

Fig 3. Slices taken at the same instant n=64 for different window lengths n=33 (top) , nh=65 (right), and nh=129 (left) are displayed for (a) SP, (b) ZAMD, (c) BD, (d) Proposed TFD.
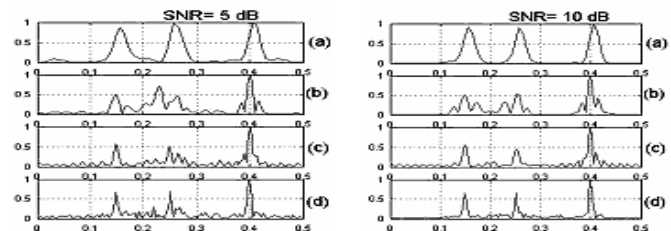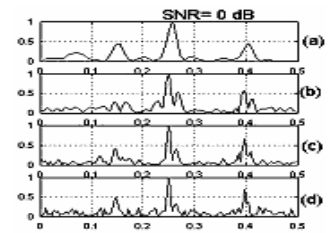

Fig 4.   Slices taken at the same time instant n= 64 for an analysis window length nh=65 samples and SNR equal to 0, 5, 10 dB, (a) SP, (b) ZAMD, (c) BD, (d) Proposed TFD.

C. Example 3: Signal of vowel / a/

A real-life speech signal as vowel /a/ of length N=256 and sampled at frequency f_s=11 kHz is analyzed using SP, ZAMD, BD and proposed SPWVD. The results are displayed in Figures 5 and 6. The figures show four formants located at F1= 688Hz, F2=1167 Hz, F3=2707 Hz, and F4=3654 Hz. It can be seen in Figure 6 that the BD and the proposed TFD have done better frequency resolution.

World Academy of Science, Engineering and Technology
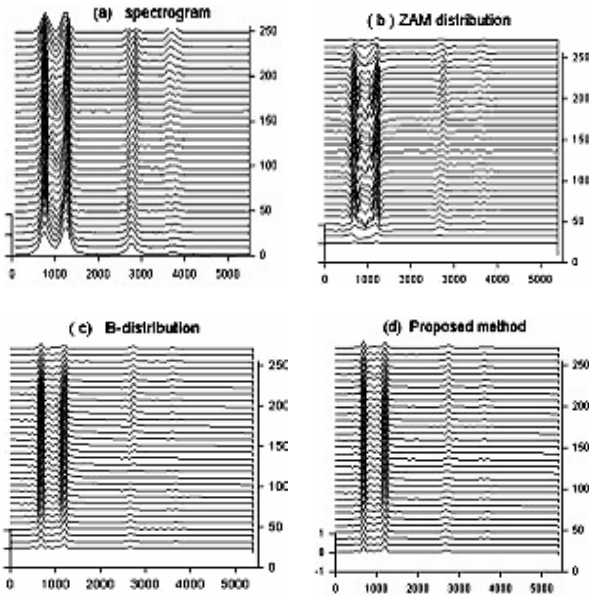International Journal of Electronics and Communication Engineering
Vol:1, No:2, 2007

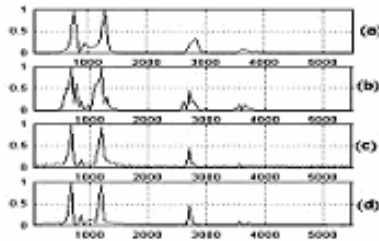Fig 5. Vowel /a/ analyzed by (a) SP, (b) ZAMD,(c) BD, (d) Proposed TFD.

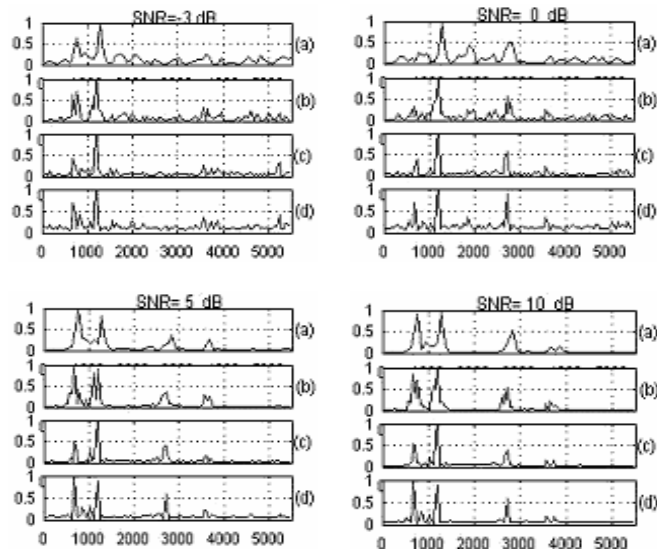Fig 6. Slices taken at the same instant n=128 , (a) SP,(b) ZAMD, (c) BD, (d) Proposed TFD, for vowel / a/.

Fig 7. Slice taken at the same instant for vowel /a/ buried in noise (a) SP, (b) ZAMD, (c) BD,(d) Proposed TFD (horizontal axis shows frequency in Hz.)

D. Example 4: Signal of vowel /a/ buried in additive noise.

The precedent real signal of vowel /a/ is corrupted by white Gaussian noise. Slices, taken at the same time instant, of the considered TFDs are displayed in Figure 7 for different

SNR values. It can be seen that for –3 dB and 0 dB all distribution can't resolve the fourth formant. we see that the BD and proposed TFD does better than the other ones for   5 and 10 dB.

V. CONCLUSION

In   this paper, we presented a novel member of the quadratic TFD. The kernel of this new TFD, inspired from the Butterworth kernel, was designed in such a way to have separable quantities for the time and lag variables in it. It was shown, using generated noisy data and real-life speech signals, that the proposed distributions achieves a high suppression of the cross-terms generally encountered in the analysis of non-stationary signals. Moreover, it was noted that this new TFD has a good resolution in the time-frequency plane. We have also presented a qualitative comparative study with other distributions known for their high cross-terms suppression property. The simulation results have confirmed the superiority of the proposed technique in terms of trade-off between   cross-terms   suppression   and   high   energy concentration in the time-frequency domain.

REFERENCES

[1]   B. Boashash," Time Frequency signal analysis and Processing: A Comprehensive Reference," *in Elsevier Publications,* First edition 2003, ISBN 0- 08- 0443354.
[2]   T.A.C.M. Claasen and W.F.G Mecklenbrauker, "The Wigner Distribution: A tool for time-frequency signal analysis," PART I II and III, *Philips Journal of Research*, Vol. 35, No 4/5, 1980.
[3]   L. Cohen, " Time frequency distribution: A review," in *Proc. of IEEE*, Vol. 77, No. 7, pp. 941-981, July 1989.
[4]   P. Flandrin, "Temps-fréquence, " 2 e édition revue et corrigée, *Paris : Hermes*, 1998.
[5]   L. Atlas, Y. Zhao and R. J. Marks, "the use of cone-shaped kernels for generalized time–frequency representations of non-stationary signals," *IEEE Trans. Acoust., Speech, signal Processing*. Vol 38, pp.1084-1091, 1990.
[6]   B. Barkat and B. Boashash "A High-Resolution   Quadratic Time– Frequency Distribution for Multicomponent   Signals Analysis," *IEEE Transaction on signal processing* ,Vol. 49, N0. 10, October 2001.
[7]   A. Papandreou and G. Faye Boudreaux-Bratels, "Distributions for time frequency analysis: A generalisation of Choi–Williams and the Butterworth distribution," *IEEE Int conf. On   Acoust. speech and signal Proc. ICASSP-92*, Victoria, Canada, 1992.
[8]   A. Papandreou and G. Faye Boudreaux-Bartels," Generalisation of the Choi–Williams and the Butterworth Distribution ,for time frequency analysis," *IEEE Trans. on signal processing*, Vol 41,N0 1, January 1993.
[9]   E. F. Velez , R. G. Absher,   "Transient analysis of speech signal using the Wigner time frequency representation," *ICASSP May 89,Glasgow*, pp. 22242-2245.
[10]   J. Leibrich, H. Puder, "A TF distribution for disturbed and undisturbed speech signals and its application to   noise reduction," *Signal Processing 80(2000),* pp.1761-1776.

Daoud Boutana, Department of Electronic, Faculty of Sciences, University of Jijel, Algeria , e–mail : daoud.boutana@mail.com

Braham Barkat , Nanyang Technological University, School of Electrical and Electronic Engineering, Singapore ,e–mail : ebarkat@ntu.edu.sg

Farid Marir, Department of Electronic, Faculty of Science, University of Constantine, Algeria ,e –mail: F.Marir@yahoo.fr