An Algorithm for the Wigner Distribution Based Instantaneous Frequency Estimation in a High Noise Environment

Igor Djurović, LJubiša Stanković

Abstract— **Estimation of the instantaneous frequency (IF) in a high noise environment, by using the Wigner distribution (WD), is considered. In this case the error is of impulse nature. An algorithm for the IF estimation, which combines the nonparametric method based on the WD maxima with the minimization of the IF variations between consecutive points, is proposed. The off-line and on-line realizations are presented. The on-line realization is an instance of the (generalized) Viterbi algorithm. Application of this algorithm on the monocomponent and multicomponent frequency modulated signals is demonstrated. For multicomponent signals, the algorithm is applied on other (reduced interference) distributions. Numerical examples, including statistical study of the algorithm performance, are given.**

I. Introduction

Time-frequency analysis is an important research area in signal analysis [1]-[7]. Instantaneous frequency (IF) estimation is one of its application fields [8]-[12]. The Wigner distribution (WD) is a widely used tool for the IF estimation of signals with fast variations of the spectral content [13]. The most common estimation technique is based on the WD maxima positions [13], [15], [16]. The error sources in this IF estimator can be divided into the following categories: (a) Bias; (b) Error due to small variations of the WD maxima within the signal's auto-term; (c) Error caused by the frequency discretization [16], [17], [18]; (d) Error caused by a high noise which can move the WD maxima outside the auto-term. The bias and small deviation of the WD maxima are considered in [16], [18]. It has been shown that the mean squared error (MSE) of the IF estimation, caused by these two sources, can be minimized by using a specific statistical approach for the adaptive window width determination. The error caused by the discrete nature of frequency can be decreased by using interpolation or displacement techniques [17], [19], [20]. An analysis of the high noise influenced error is done in [21]. This error occurs when some points outside the signals' auto-term surpass values inside the auto-term, due to the influence of a relatively high noise. It has been shown that this kind of error, when it appears, dominates over other sources of error [21].

Signal Processing, Vol. 84, No.3, March 2004.

This paper is focused on the last mentioned source of error. A new algorithm which can significantly reduce the error caused by a high noise is proposed. Monocomponent FM signal with constant amplitude in a complex, white, additive, Gaussian noise with independent real and imaginary parts (i.i.d.) is considered. The key criteria that is used in the algorithm is: *the IF should pass through as many as possible points of the WD with highest magnitudes, while the IF variation between two consecutive points should not be too fast*. The basic idea for this algorithm comes from the graph theory and algorithms for edge-following [22]. We proposed both on-line and off-line algorithm realizations. The on-line realization is an instance of the (generalized) Viterbi algorithm [23]. The algorithm can be used in the case of multicomponent signals and other types of additive or multiplicative noise. These cases will be illustrated on examples. Although we have restricted the analysis to the WD, this method can be applied on any time-frequency representation that concentrates auto-term around the IF. The proposed algorithm estimates the IF based on signal independent (non-adaptive) form of time-frequency representations. Signal adaptive time-frequency representations usually adapt parameters, based on the measures of time-frequency concentration or based on the projections in the ambiguity domain [24]-[28]. An example with these representations shows that, in a high impulse noise environment, these methods can adjust their parameter values to very high noise components (impulses), neglecting signal components with significantly lower amplitudes than those of the noise. In our numerical studies we have decided to use signal independent representations.

The paper is organized as follows. An overview of the WD based IF estimator for a signal embedded in a high noise is given in Section II. The algorithm for the IF estimation in a high noise environment is derived in Section III. Some numerical examples are given in this section, as well. Statistical study of the proposed algorithm is done in Section IV. Finally, the conclusions are presented in Section V.

II. Wigner Distribution as an IF Estimator

Consider a signal $f(t) = Ae^{j\phi(t)}$, corrupted by an additive, complex, white, Gaussian i.i.d. noise $\nu(t)$ with variance $2\sigma^2$ (variance of the real and imaginary part is σ^2). Noisy signal $x(t)$ is of the form $x(t) = f(t) + \nu(t) = Ae^{j\phi(t)} + \nu(t)$. The IF is defined as the first derivative of the phase $\omega(t) = \phi'(t)$. The WD of a discrete-time signal is given by

$$WD(t,\omega) = \sum_k w_h(kT)x(t+kT)x^*(t-kT)e^{-j2\omega kT}, \quad (1)$$

where the window function $w_h(kT)$ has the width h, $\sum_k w_h(kT) = 1$, and T is the sampling interval.

The WD is highly concentrated around the signal's IF. Thus, it has been shown that the position of the WD maxima is an appropriate tool for the IF estimation [13], [15]:

$$\hat{\omega}(t) = \arg\left[\max_{\omega} WD(t,\omega)\right]. \quad (2)$$

Influence of the high noise to the IF estimator (2) is considered in [21]. By neglecting other sources of error probability that some WD points outside IF of the linear FM signal are higher than value on the IF is equal to [21]:

$$P_E = 1 - \int_{-\infty}^{\infty}\left(1 - 0.5\mathrm{erfc}\left(\frac{\xi}{\sqrt{2}\sigma_{WD}}\right)\right)^{N-1} p(\xi)d\xi, \quad (3)$$

where: $\sigma_{WD}^2 = 4N\sigma^2(A^2+\sigma^2)$, $p(\xi) = e^{-(\xi-NA^2)/2\sigma_{WD}^2}/\sqrt{2\pi}\sigma_{WD}$ is probability density function for the WD values along the IF, and N is window width. Probabilities of the IF error in the case of the linear FM signal and $N = 256$, for various σ/A, are given in Table I. It can be seen that the probability of error is a rapidly increasing function with respect to σ/A. For values $\sigma/A > 2$ it is very close to 100%. When this error appears, there is an equal chance that any value of ω is taken as the IF estimate. Thus, we can write the IF estimation as:

$$\hat{\omega}(t) = \omega(t) \text{ with probability } 1 - P_E, \quad (4)$$

and $\hat{\omega}(t) = \omega \in Q_\omega, \omega \neq \omega(t)$ with probability P_E. The probability of error is uniformly distributed over $N-1$ samples outside the IF.

The WD value at the IF is the j-th largest WD value, at the considered instant, with the probability:

$$P(j) = \int_{-\infty}^{\infty} Q^{j-1}(\xi)\left(1 - Q(\xi)\right)^{N-j} p(\xi)d\xi, \quad (5)$$

where $Q(\Xi) = 0.5\mathrm{erfc}\left(\Xi/\sqrt{2}\sigma_{WD}\right)$. For two values of the signal to standard deviation ratio, A/σ, this probability is shown in Figure 1a,b. From these figures it can be concluded that it is highly probable that the WD value along the IF is one of the largest distribution values at the considered instant. Probability $P(j)$ almost linearly decreases with j in a high noise environment, as it is illustrated in Figure 1c. These facts will be used in the next section for development of the IF estimation algorithm.

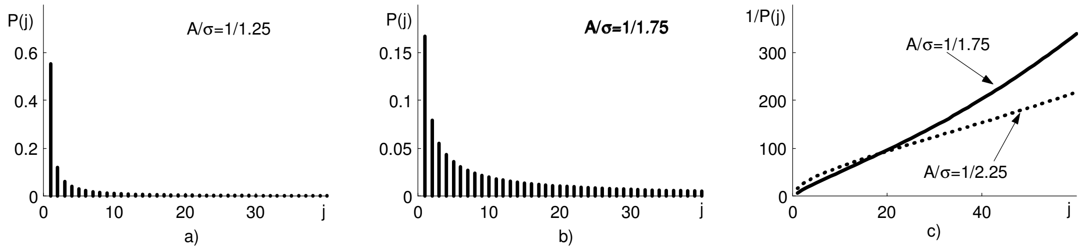

Fig. 1. Probability that the WD value at the IF is the j-th largest value at the considered instant: a) $A/\sigma = 1/1.25$; b) $A/\sigma = 1/1.75$; c) $1/P(j)$ for $A/\sigma = 1/1.75$ and $A/\sigma = 1/2.25$.

TABLE I

Error probability of the IF estimation.

σ/A	0.25	0.50	0.75	1.00	1.25
P_E	$1.90{\cdot}10^{-138}$	$5.55{\cdot}10^{-22}$	$1.98{\cdot}10^{-7}$	$4.71{\cdot}10^{-3}$	$1.37{\cdot}10^{-1}$
σ/A	1.50	1.75	2.00	2.25	2.50
P_E	$4.49{\cdot}10^{-1}$	$6.97{\cdot}10^{-1}$	$8.33{\cdot}10^{-1}$	$9.02{\cdot}10^{-1}$	$9.37{\cdot}10^{-1}$

III. Algorithm for the IF Estimation in a High Noise Environment

The IF itself is usually a slow varying function. In the case of a high noise the estimation errors are dominantly of impulse nature (4). Thus, this kind of error can be reduced by applying the median filter, directly to the estimated IF [30]. From filter theory it is known that the median filter can eliminate impulses whose occurrence frequency is up to 50%. However, in our experiments we could not get the expected results by using the median filter approach. Namely, the errors in the IF estimation are not statistically independent. If a large error occurs in the considered point, there is a high probability that the error exists in the neighboring points. This fact significantly reduces the efficiency of a direct median filter application. It clearly shows that for the IF estimation in the high noise environment we need a more accurate tool. Also, from Table I it can be seen that, for $\sigma/A > 1.7$, the probability of error is higher than 70%. In that case the median filter cannot be successfully used.

In order to develop a more sophisticated algorithm for the IF estimation in a high noise environment, we will assume the following:

(1) If the WD maximum at the considered instant is not at the IF point, there is a high probability that the IF is at a point having one of the largest WD values (for example second, third..., but not as far as, for example, the hundredth position). This has been confirmed numerically in Figure 1a,b. Thus, based on the WD values we will form a weighting function which assumes greater values for smaller values of the WD, and vice versa;

(2) The second factor is based on the assumption that the IF variation between two consecutive points is not extremely large.

According to these two assumptions, we can define the algorithm. The basic idea for this algorithm comes from the problem of edge-following in digital image processing [22]. The problem in [22] was to find out a line that passes through pixels with as high as possible values of the edge detector, and such that variations of the edge direction are as small as possible. Roughly speaking, our algorithm is similar to the algorithm for connecting points at the map such that the path length and the altitude variations are as small as possible. It can also be related to finding the most probable hidden state. That problem is solved recursively by using the well known Viterbi algorithm.

A. Algorithm

Consider time interval $n \in [n_1, n_2]$. Let all paths between n_1 and n_2 belong to a set $\mathbf{K}$. Assume that all paths from the set $\mathbf{K}$ can take only discrete frequency values which belong to the set Q_ω.

We form the IF estimate as a path that minimizes the expression:

$$\hat{\omega}(n) = \arg \min_{k(n)\in\mathbf{K}} \left[\sum_{n=n_1}^{n_2-1} g(k(n), k(n+1)) + \sum_{n=n_1}^{n_2} f(WD(n, k(n))) \right] = \arg \min_{k(n)\in\mathbf{K}} p(k(n); n_1, n_2), \quad (6)$$

where $p(k(n); n_1, n_2)$ is a sum of the penalty functions $g(x, y)$ and $f(x)$, along the line $k(n)$, from the instant n_1 to n_2. Function $g(x, y) = g(|x - y|)$ is a nonincreasing one, with respect to the absolute difference between x and y (between the IF values in the consecutive points $x = k(n)$ and $y = k(n-1)$), while $f(x)$ is a nondecreasing function of $x = WD(n, k(n))$. In this way the larger values of the WD are more important candidates for the position of the IF at the considered instant. For a considered n, the function $f(x)$ can be formed as follows. The WD values, $WD(n, \omega)$, $\omega \in Q_\omega$, are sorted into the nonincreasing sequence:

$$WD(n, \omega_1) \geq WD(n, \omega_2) \geq \ldots. \geq WD(n, \omega_j) \geq \ldots \geq WD(n, \omega_M), \ \omega_j \in Q_\omega, \ j \in [1, M], \quad (7)$$

where $j = 1, 2, ..., M$, is the position within this sequence. Then, the function $f(x)$ is formed as:

$$f(WD(n, \omega_j)) = j - 1. \quad (8)$$

Thus, we have a function which realizes our idea that the points with large WD values will be more important candidates for the IF estimates. Motivation to use this form of function $f(x)$ is in the behavior of the WD. Namely, *in an extremely high impulse noise environment, probability that the WD is on the j-th position decreases almost linearly with j,* Figure 1c. Note that the function $f(x)$ is not formed directly by using values of the WD, since the signal and noise parameters can be time-varying. It means that a particular distribution value at the considered instant may highly probably belong to the signal term, while in other points it can be influenced by noise. In the Viterbi algorithm the path penalty function is usually determined as the logarithm of the corresponding probabilities [23]. For precise determination of the WD probability, it is necessary to have an accurate information about the signal and noise parameters, which are not available.

For $g(x, y) = const$, the IF estimation (6) is reduced to the position of the WD maxima, i.e., the function $f(x)$ completely determines minimum of (6). In this paper we will use a linear form of $g(x, y)$, for the difference between two points greater than an assumed threshold Δ:

$$g(x, y) = \begin{cases} 0 & |x - y| \leq \Delta \\ c(|x - y| - \Delta) & |x - y| > \Delta. \end{cases} \quad (9)$$

The reasonable choice for Δ would be the maximal expected value of the IF variation between consecutive points. It means that there is no additional penalty due to this function for small IF variation (within Δ points, for two consecutive instants). In the realization we obtained good results by taking Δ which corresponds to a few neighboring points (for example, values around $\Delta = 3$). For $\Delta \to \infty$ estimation given by (6) will reduce to the estimation based on the WD maxima. Note that this is one possible form of the penalty functions $f(x)$ and $g(x, y)$. Proposed form of $g(x, y)$ gives also an opportunity for efficient elimination of non-optimal paths within the on-line algorithm realization. It will be demonstrated in the next subsection. Accuracy of the proposed algorithm, for the described penalty functions, is demonstrated on the examples and within the numerical study.

B. Implementation

There are several ways to implement algorithm (6). Here, we will describe two of them: off-line realization, and on-line realization.

III.B.1 Off-line realization

Off-line realization starts with the WD calculation for the considered interval, and the IF estimation based on the WD maxima. In this way the initial estimate $\omega^{(0)}(n)$, $n \in [n_1, n_2]$ is obtained. The second step is in determination of the path penalty functions for the initial IF estimate, and the instants with the largest path penalty function gain. Then, we look for the values of the IF estimate, in the selected points, which can decrease the path penalty function value. The procedure is performed recursively until we are no longer able to find a new point that can decrease this value. This algorithm is very efficient in the case of a relatively small number of wrongly detected points (below 20%). In the case of a large number of wrongly detected points, the algorithm convergence could be quite slow.

III.B.2 On-line realization

Let the time-frequency plane contains M frequencies and Q time-instants, $\mathbf{T} = \{(n_i, \omega_j) | i \in [1, Q], j = [1, M]\}$. The total number of paths between two ending points is M^Q. This fact makes a direct search for the optimal path impossible. Fortunately the algorithm can be realized recursively, as an instance of the generalized Viterbi algorithm [23]. Its realization can be described by the following fundamental steps.

(a) Let optimal paths, which connect the instant n_1 and all points to the instant n_i, are determined. Those paths, denoted as $\pi_i(n;\omega_j)$, $n \in [n_1, n_i]$ for $j \in [1, M]$ can be written as:

$$\pi_i(n;\omega_j) = \arg \min_{k(n) \in \mathbf{K}_{ij}} p(k(n); n_1, (n_i, \omega_j)), \;\; j \in [1, M], \qquad (10)$$

where the set $\mathbf{K}_{ij}$ contains all paths between the instant n_1 and the point (n_i, ω_j), while $p(k(n); n_1, (n_i, \omega_j))$ is a sum of the path penalty functions for the line $k(n)$. In the Viterbi algorithm terminology, paths (10) are known as the partial best paths. Current IF estimate, within the interval $[n_1, n_i]$, can be written as:

$$\hat{\omega}_{(i)}(n) = \arg \min_{\pi_i(n;\omega_j), j \in [1,M]} p(\pi_i(n;\omega_j); n_1, (n_i, \omega_j)), \qquad (11)$$

for the interval $[n_1, n_i]$.

(b) The partial best paths at the next instant n_{i+1} can be represented as concatenation of (10) with the points at the new instant $\pi_{i+1}(n;\omega_j) = [\pi_i(n;\omega_l), (n_{i+1}, \omega_j)]$, $j \in [1, M]$, for $l \in [1, M]$, that produce the minimal value of:

$$p(\pi_i(n;\omega_l); n_1, (n_i, \omega_l)) + g(\omega_l, \omega_j) + f(WD(n_{i+1}, \omega_j)), \qquad (12)$$

for each ω_j, $j \in [1, M]$. Note that the function $f(WD(n_{i+1}, \omega_j))$ is constant for the considered partial best path. Generally, for the considered point it is necessary to search over M paths, M^2 for the entire instant, and QM^2 for the entire plane.

In order to further reduce search space for partial best paths (12), procedure could be performed in the following way using the fact that $g(\omega_l, \omega_j)$ is an increasing function of the distance $|\omega_j - \omega_l|$:

(b1) Set $\rho = \Delta$;

(b2)

$$\hat{l} = \arg \min_{l \in [j-\rho, j+\rho]} [p(\pi_i(n;\omega_l); n_1, (n_i, \omega_l)) + g(\omega_l, \omega_j)],$$

$$\pi'_{i+1}(n;\omega_j) = [\pi_i(n;\omega_{\hat{l}}), (n_{i+1}, \omega_j)];$$

(b3) If

$$p(\pi_i(n;\omega_{\hat{l}}); n_1, (n_i, \omega_{\hat{l}})) + g(\omega_{\hat{l}}, \omega_j) < g(\omega_{j+\rho+1}, \omega_j) + \min[\pi_i(n;\omega_j), j = 1, M]$$

set $\pi_{i+1}(n;\omega_j) = \pi'_{i+1}(n;\omega_j)$ as the partial best path and further procedure could be stopped. Elsewhere $\rho = \rho + 1$ and goto step (b2).

Step (b) should be repeated for each point.

In this way, we got significant calculation savings, in our examples. They were even greater than 50%. Again, from the considered partial best paths, the current estimate is the one producing the smallest penalty function.

Example: In order to illustrate algorithm following synthetic example is used. Consider a time-frequency plane with $M = 3$ and $Q = 8$ points, with assumed values of $f_{ij} = f(WD(n_i, \omega_j))$, given in Figure 2a. The function $g(\omega_i, \omega_j) = g_{ij}$ is given by (9), where

$c = 2.5$ and $\Delta = 1$. Connections from each point, at the considered instant, with 3 nearest points from the previous instant, produces $g_{ij} = 0$, i.e., $g_{ij} = 0$ for $|i - j| \leq 1$. With an increase of $|i-j|$ the function g_{ij} increases for the value of $c = 2.5$. The partial best paths at the instant n_2 are marked in Figure 2a. Current IF estimate is the path which connects the points (1,5) and (2,5). It is represented by a thick line. Search for the optimal path at the next instant is shown in Figure 2b. Values written for the second instant points, f'_{2j}, $j = 1, 2, ..., 8$, represent path penalty functions of the partial best paths to those points:

$$f'_{2j} = \min_{l=1,8}[f_{1l} + g_{lj}] + f_{2j}. \tag{13}$$

Consider the point (3,4). For this point, three points from the previous instant ((2,3), (2,4) and (2,5)) produce $g_{ij} = 0$. The optimal path between these three points and (3,4) is (2,5)-(3,4), since $f'_{25} = 1$. Its value is smaller than f'_{24} and f'_{23}. Furthermore, paths between (3,4) and any other point from the previous instant produce $g_{ij} \geq 2.5$, what is greater than f'_{25}. Thus, for the considered point (3,4) search for the partial best path could be stopped after only three points. Consider now the point (3,8). Only two points, (2,7) and (2,8), give $g_{ij} = 0$ for their paths to the point (3,8). Current estimate for the partial best path to the point (3,8) is (2,7)-(3,8), since $f'_{27} = 9 < f'_{28} = 10.5$. However, the connection between (2,6) and (3,8) is better than the connection (2,7)-(3,8), since $f'_{26} + g_{68} = 4 + 2.5 = 6.5 < 9$. Therefore, the path between (2,5) and (3,8) is the best, since $f'_{25}+g_{58} = 1+5 = 6 < 6.5$. Further search could be stopped since $g_{k8} \geq 7.5$ for $k \leq 4$. This illustrates that the search procedure can be relatively fast. In the considered case, the optimal path (IF estimate) links the points (1,1), (2,1) and (3,2). Path penalty function for this path assumes the value of 3. It can be easily seen that this path is obtained by linking the partial best path to the point (2,1) and the point (3,2). Note that the new estimate could perform an update of the old ones.

C. Examples

Example 1. The linear FM signal $f(t) = e^{jat^2/2}$ with $N = 256$ samples within interval $t \in [-0.5, 0.5]$ will be used as a statistical model. The value of parameter $a = 64\pi$ is chosen such that the IF lies at the discrete frequency grid on the diagonal of the time-frequency plane. The IF variation between consecutive points is 1. We chose the parameter Δ in (9) to be equal to 3. Three values of the standard deviation $\sigma = \sqrt{1.5}$, $\sqrt{2.5}$ and 2, are considered, while $A = 1$ is used in all examples. The input SNR, $A^2/2\sigma^2$, is equal to 1/3, 1/5 and 1/8, i.e., $-4.8[dB]$, $-7[dB]$ and $-9[dB]$, respectively. In the first case the number of errors is approximately 10% (Table I), in the second one it is greater than 50%, while in the third case it is greater than 80%. The IF estimates by using: the WD, the median filter of lengths 3 and 5 applied directly to the IF estimation, and the proposed algorithm for $\sigma = \sqrt{1.5}$ are shown in Figures 3a-d. Results for higher noise, $\sigma = \sqrt{2.5}$ and $\sigma = 2$, are depicted in Figures 3e,f. It can be seen that the median filter in this case depicted by dash-dot lines cannot perform any improvement in this case. The proposed algorithm performs well in all trials in the first two cases, while in the third case, for $\sigma = 2$, it performs well in 75% of trials. More details on the statistical performance of the algorithm are given in the next section.

Example 2. The proposed algorithm is successfully applied on the nonlinear FM signals. Signal with the IF of a sinusoidal shape is used in this example. The inner interference effect, that appears in the case of a nonlinear FM signal, is reduced by using the Hanning window of the length 256 samples. Any reduced interference distributions from the Cohen class can be used for reducing this effect. This smoothing of the WD values prevents the algorithm from taking values significantly influenced by the inner interferences. However, smoothing of the signal auto-term can disturb the performance of the algorithm with respect to the noise influence, as it will be seen in the next section. Signal to noise ratio was $A^2/2\sigma^2 = 1/2$, i.e., $-3[dB]$. The IF estimations based on: the WD maxima, the proposed algorithm, and the true IF are shown in Figure 4. The maximal variation between consecutive

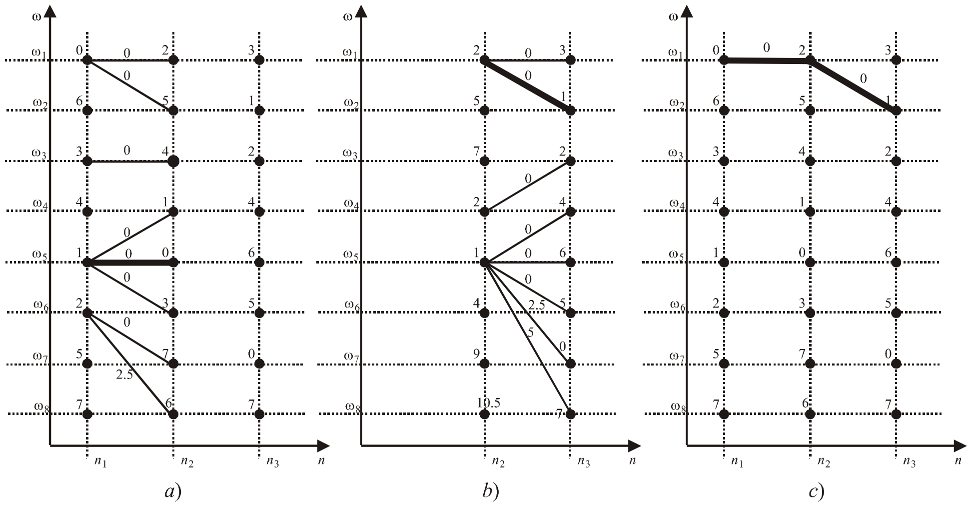

Fig. 2. Illustration of the instantaneous frequency estimation in 3x8 time-frequency plane. a) Partial best paths for n_2; b) Partial best paths for n_3; c) The IF estimate.

IF samples was 2. We assumed $\Delta = 3$ in the algorithm. Because the WD of this signal is not ideally concentrated, the ratio of the WD maxima and the distribution standard deviation is lower than in the case of the linear FM signals. This fact causes that our algorithm performs worse for this signal than for the linear FM signal.

Example 3. Consider the signal from Example 1, embedded in a noise that exhibits impulse nature:

$$\nu(t) = 0.65(\nu_1^3(t) + j\nu_2^3(t)), \qquad (14)$$

where $\nu_i(t)$, $i = 1, 2$ are Gaussian white noises with unitary variances. The noisy signal is shown in Figure 5a. The IF estimation in an impulse noise environment is considered in [30], [33]-[35]. The IF estimates, performed by using the proposed algorithm and the WD, are shown in Figure 5b. It is obvious that the WD can be used as a basis for the IF estimation in the case of the impulse noise environment, as well. Note that many excellent signal-adaptive methods are proposed for the IF estimate, even in the noisy environment [24]-[28]. However, in a very high noise environment, being the topic of this paper, these methods can recognize high noise values as the main components and adjust their parameters to them, ignoring the signal. This effect is demonstrated within this example, where the signal-adaptive method based on the radially Gaussian kernel is applied [26]. The normalized kernel volume $\alpha = 2$ is used. The adaptive time-frequency representation recognizes noise impulses as the signal components, Figure 5c. The corresponding IF estimate is shown in the Figure 5d.

Example 4. In this example we considered real-life underwater biological signal representing sounds produced by marine fauna. Currently very active research filed is autonomous passive tomography for tracking natural underwater activities (tectonic motion, underwater fauna) or artificial signals (submarines). This approach has numerous advantages but it requires very sophisticated signal processing methods for signal detection and parameters estimation. We considered 6600 samples long signal representing sound of an underwater mammal (approximately 0.14sec)[1]. The WD of this signal is shown in Figure 6a. The WD is calculated by using the Hanning window of the length 256 samples with additional

[1]Signal is provided is provided by ENSIETA (Ecole Nationale Superieure d'Ingenieurs des Etudes et Techniques d'Armaments), Brest-FRANCE . It is recorded during an underwater biological signal recording campaign, organized in 2000 by GESMA (Groupe d'Etudes Sous-Marinnes d'Atlantique) in cooperation with ENSIETA.

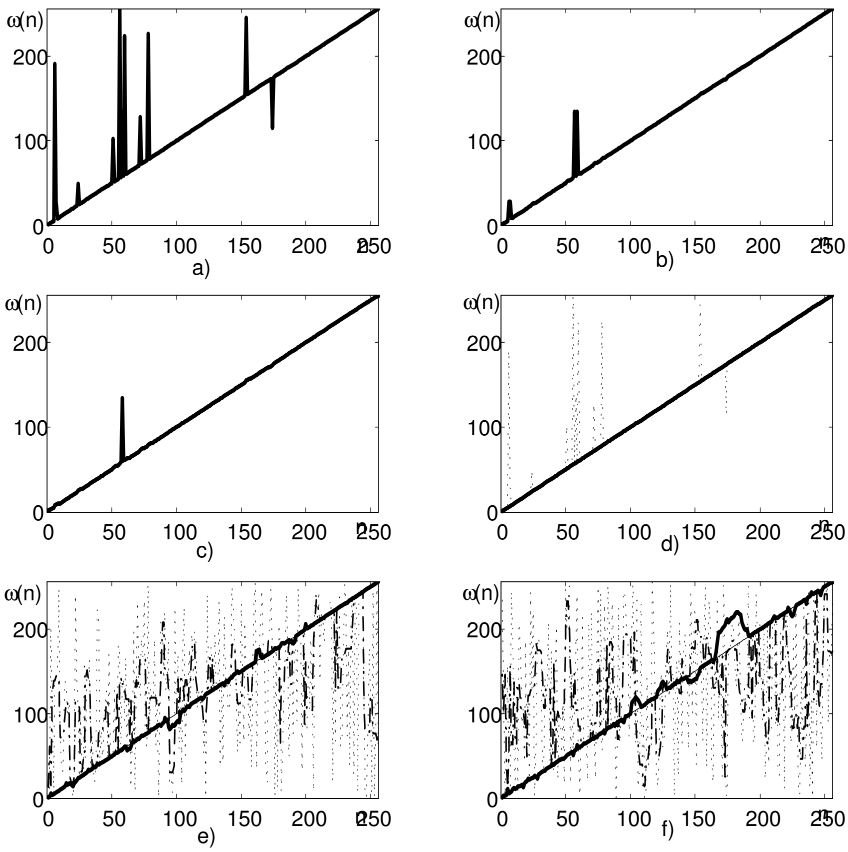

Fig. 3. The IF estimation for the linear FM signal: a) Based on the WD maxima SNR=-4.8[dB]; b) Median filter of length 3 applied to the IF estimate; c) Median filter of length 5 applied to the IF estimate; d) Proposed algorithm (solid line) and WD maxima (dotted line); e) SNR=-7[dB], Thick line - proposed algorithm; Dotted line - WD maxima; Dash-dot line - Median filter of length 5; f) SNR=-9[dB], Thick line - proposed algorithm; Dotted line - WD maxima; Dash-dot line - Median filter of length 5.

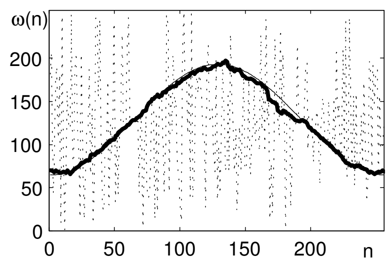

Fig. 4. IF estimation for the sinusoidal FM signal based on: WD maxima - dotted line; Proposed algorithm - thick line; True IF - thin line.

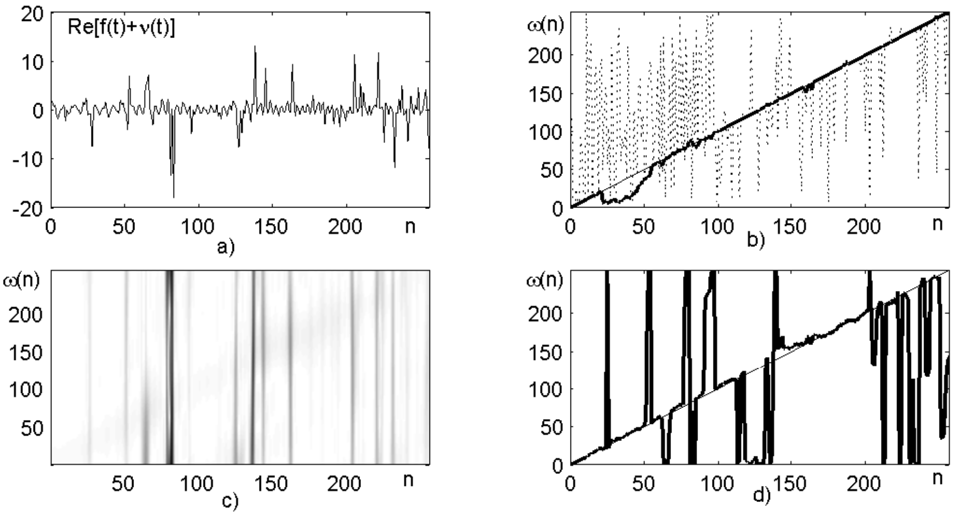

Fig. 5. IF estimation of the signal in an impulse noise environment: a) Noisy signal; b) IF estimation based on the proposed algorithm - thick line; WD maxima - dotted line; c) Signal-adaptive distribution; d) IF estimation based on maxima of the signal adaptive distribution.

smoothing along frequency coordinate. The IF estimators based on the proposed algorithm and the WD maxima are applied on this signal (Figure 6b). Dotted line represents the WD maxima while solid line is for the proposed estimator. Strong short component in the zone around $t = 0.13$sec moves the IF estimate outside the auto-term of the useful signal. The proposed algorithm neglected this component and tracked accurately the IF of useful signal. In the second experiment the artificial Gaussian noise with the same variance as the power of the original signal is added. The WD of this signal is given in Figure 6c while the IF estimates are given in Figure 6d. The WD maxima based estimator is accurate only in the range where considered signal is very strong ($t \in [0.04, 0.09]$sec) while outside this region it is behaves poorly. The proposed estimator is accurate in the entire considered interval.

D. Algorithm Application on Multicomponent Signals and Other Time-Frequency Representations

The algorithm can be applied, in a straightforward manner, on any other time-frequency representation. Note that out of the Cohen class of distributions, the WD produces the best auto-term to distribution standard deviation ratio for signals with varying IF [21]. This is the reason why the proposed algorithm, based on the WD for monocomponent signals, performs best. However, the WD exhibits very emphatic cross-terms in the case of multicomponent signals. They can make the IF estimation impossible. This is the reason why distributions with reduced interferences should be used [31], [32]. Here, the weighted pseudo-WD, known as the S-method [32], [36], [37], will be used. It can produce the auto-terms close to those in the WD, but with significantly reduced cross-terms.

Algorithm for the IF estimation of the multicomponent signals can be summarized as follows.

(a) The IF estimation by using the proposed algorithm $\hat{\omega}^{(0)}(n)$, $i = 0$. This IF corresponds to the highest signal component.

(b) Forming a new time-frequency representation by taking zero-values in the region around determined IF estimate $[\hat{\omega}^{(i)}(n) - \delta, \hat{\omega}^{(i)}(n) + \delta]$.

(c) Repeating the algorithm for this time-frequency representation, and obtaining the next IF estimate $i = i + 1$, $\hat{\omega}^{(i)}(n)$. Steps (b)

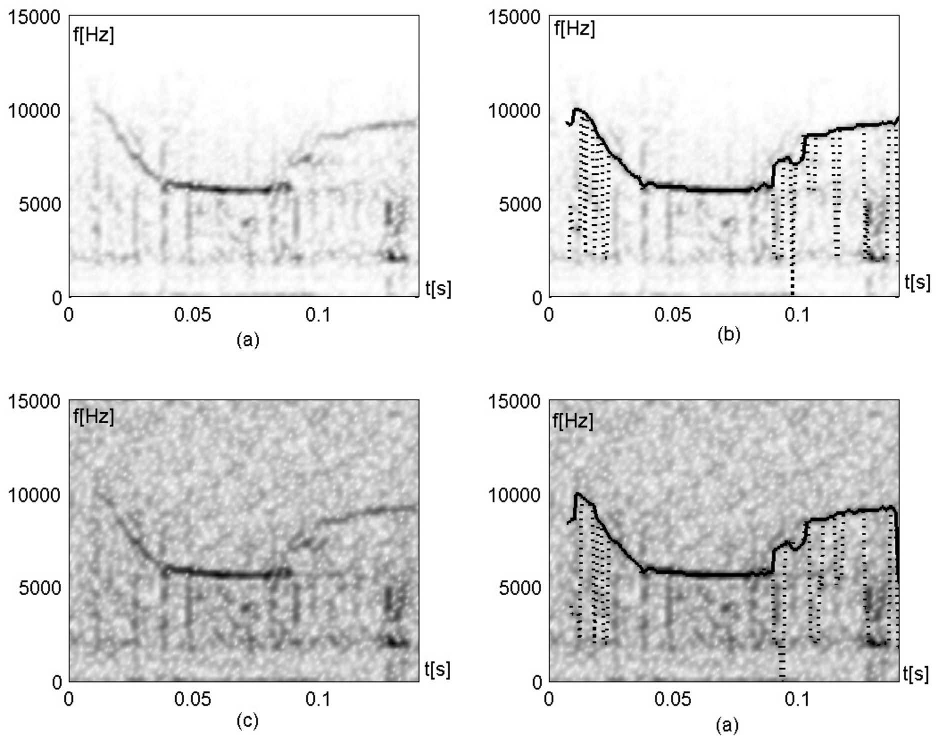

Fig. 6. IF estimation of signal produced by underwater mammal: (a) Wigner distribution of recorded signal; (b) IF estimation of recorded signal: dotted line - WD maxima; solid line: proposed algorithm; (c) Wigner distribution of signal embedded in artificial noise; (d) IF estimation of signal embedded in artificial noise: dotted line - WD maxima; solid line - proposed algorithm.

and (c) should be repeated for each component.

This procedure works well for the components separated in the time-frequency plane. For signal components which intersect in the time-frequency plane, the proposed algorithm can switch components after the crossing point. For the illustration, consider two multicomponent signals that consist of a linear and a sinusoidal FM signal. We assume that both signal components have the same amplitude $A = 1$. In the first example, Figures 7a,b, the signal components are well separated, while in the second example, Figures 7c,d, the components intersect. In both instances we considered two cases: noiseless and noisy signal with variance $2\sigma^2 = 2$. After the IF estimation of the first component is performed, we neglect the zone of five samples around it, and identify the optimal path in the remaining part of the time-frequency plane. In both cases we estimated the IF of the signal components. In the last example (Fig.7d), the algorithm, in the first pass, tracks the sinusoidal FM signal and after components intersection it follows the linear FM signal (it is depicted with solid line). This component switching can be avoided for signals with smooth IFs by introducing additional constraint that minimizes the first derivative of the IF estimate.

IV. Statistical Analysis

Statistical analysis of the algorithm performance with respect to the amount of noise, and the considered time intervals for the IF estimation, is presented next. We have compared performance of the WD with the Born-Jordan distribution, as basic distributions for the proposed IF estimation algorithm.

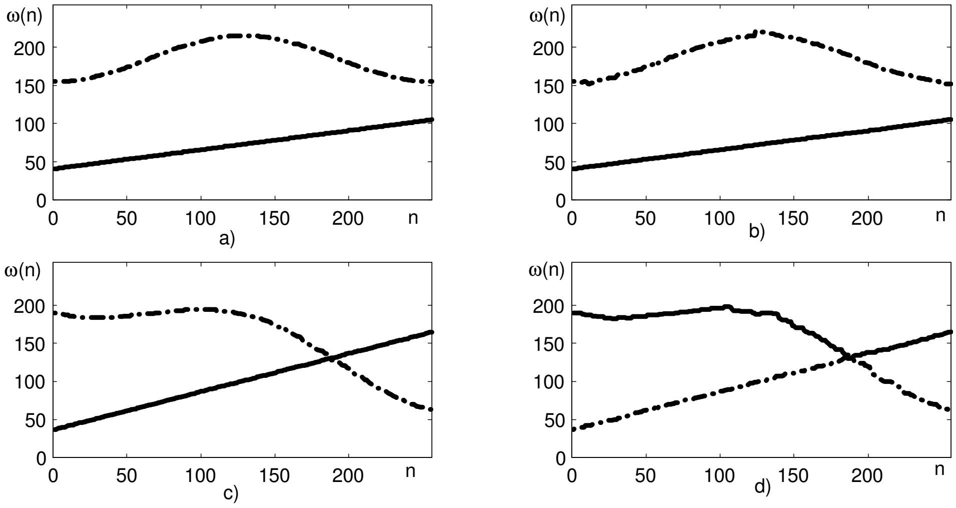

Fig. 7. IF estimation for the multicomponent signals: a) Nonnoisy signal with separated components; b) Noisy signal with separated components; c) Nonnoisy signal with intersected components; b) Noisy signal with intersected components. Thick line - First estimated signal component; Dotted line - Second component.

A. *Noise Influence*

Consider the linear FM signal from Example 1 embedded in a Gaussian additive noise. The standard deviation of noise from the range $\sigma \in [0, 3]$ is considered, with a step of 0.05. For each σ we have performed 25 trials. The WD with $N = 256$ samples is used. The IF estimation is obtained in $N = 256$ instants. We assume that the algorithm works successfully when the resulting mean absolute error (MAE) of the IF estimation is at least 10 times lower than the MAE in the estimation based on the WD maxima. In Example 1, when $P_E \to 1$, for the WD based estimator, 10% of the expected MAE is approximately 9.6.

The experiment outcomes are summarized in Table II. For $\sigma \leq 0.9$ we did not get a single error in the estimation based on the WD maxima. For $\sigma \in [0.95, 1.75]$ we did not get a single wrong result in the IF estimation by using the proposed algorithm. Note that for $\sigma = 1.75$ the noise variance is equal to $2\sigma^2 = 6.125$, i.e., the SNR is $-7.87[dB]$. For $\sigma \in [1.8, 1.95]$ we got $23-24$ trials where the proposed algorithm produces good results. For $\sigma \in [2, 2.15]$ algorithm works successfully in $18 - 19$ trials, or 75% of trials. For $\sigma \in [2.2, 2.3]$ algorithm works well in 50% cases. We can conclude that the limit for the algorithm application is roughly $\sigma \approx 2.3$, i.e., SNR$=-10.24[dB]$.

As it has been mentioned earlier, this approach can be applied to any other time-frequency representations concentrated along the IF. The reduced interference distributions would reduce the noise influence but, at the same time, they would disturb the auto-term quality [38]. Therefore, in general, behavior of these distributions with respect to the noise influence would not be better than in the WD case. It is illustrated by the seventh column in Table II, where the MAE values are given for the case of the algorithm application to the Born-Jordan distribution. It can be seen that in this case algorithm preforms significantly worse than in the case of the WD as the algorithm base. Importance of the reduced interference distributions application for this algorithm is primarily in reduction of the cross-terms and inner interferences as it is shown within the examples. Detailed comparison of various non-parametric techniques for IF estimation is presented in [14]. The cross-Wigner distribution (XWD) outperforms other techniques considered in [14] in term of the noise amount for which produces satisfactory results. In order to prove efficiency of our approcah we compared our technique with the XWD (for details on realization of the XWD based IF estimator please refer to [14]). The MAE of the XWD is given in the last column of Table I. It can be seen that this technique works well for SNR$\geq -5[dB]$ but worse than the proposed algorithm. However, for SNR$\leq -5[dB]$ it cannot produce satisfactory results while our approach produces accurate results in each trial until SNR$\approx -8[dB]$.

B. MAE and Interval Length

The interval of $N = 256$ is divided into nonoverlapped subintervals and the algorithm is applied on each subinterval separately. The MAE obtained with different interval lengths is calculated. Results for different lengths of subintervals are given in Table III. It can be seen that for the subinterval length of 16 instants we got the error of the same order of magnitude as for the interval of 256 instants. Note that the memory demand significantly decreases when narrower subintervals are used. Namely, for the original algorithm we should keep in memory $N = 256$ with 256 instants, i.e., 65536 in total. For subinterval of length 16, one path with $(256 - 16)$ instants and 256 paths with 16 instants, i.e., 4336 in total, should be memorized. Note that the time-frequency representations can be memory demanding and any savings can be helpful. Furthermore, on the shorter segment, paths with extremely high value of the penalty function can be neglected from further consideration, with a very small probability to produce an error. This is not possible for a wide segment. In this way the calculation complexity can be slightly reduced.

V. Discussion and Concluding Remarks

In this paper we have presented a new approach for the IF estimation based on the WD (or any other time-frequency distribution) in a high noise environment. Algorithm uses two assumptions: the IF is placed at the position of the time-frequency representation with a high magnitude; and the IF is a slow-varying function. The proposed algorithm is relatively resistive to the variations of its parameters. It has been shown that the considered time interval can be divided into nonoverlapped subintervals without significant increase of error. This reduces memory demand. Other possible approaches and penalty functions are open for further research. The algorithm can be applied not only on the WD, but on any other time-frequency distribution. Procedures for realization of the proposed algorithm are presented. Recursive, on-line realization is based on the generalized Viterbi algorithm. The algorithm can also be used for the IF estimation of signal components in multicomponent signals.

VI. Acknowledgment

This work is supported by the Volkswagen Stiftung, Federal Republic of Germany.

References

[1] Proceedings of IEEE, Special issue on time-frequency analysis, Vol. 84, No. 9, Sept. 1996.

[2] M. Sun, S. Qian, X. Yan, S. B. Baumann, X. -G. Xia, R. E. Dahl, N. D. Ryan and R. J. Sclabassi: "Localizing functional activity in the brain through time-frequency analysis and synthesis of the EEG," *Proc. IEEE*, Vol. 84, No. 9, Sept. 1996, pp. 1302-1311.

[3] Y. Lu and J.M. Morris: "Gabor expansion for adaptive echo cancellation," *IEEE Sig. Proc. Mag.*, Vol. 16, No.2, Mar. 1999, pp. 68-80.

[4] J.J. Wang and J. Zhou: "A seismic design based on artificial simulations," *IEEE Sig. Proc. Mag.*, Vol. 16, No.2, Mar. 1999, pp. 94-99.

[5] D. König and J.F.Böhme: "Wigner-Ville spectral analysis of automotive signals captured at knock," *Appl. Sig. Proc.*, Vol. 3, 1996, pp. 54-64.

[6] M.G. Amin and W.J. Williams: "High spectral resolution time-frequency distribution kernel," *IEEE Trans. Sig. Proc.*, Vol. 46, No.10, Oct. 1998, pp. 2796-2804.

[7] G.F. Boudreaux-Bartels and T.W. Parks: "Time-varying filtering and signal estimation using the Wigner distribution," *IEEE Trans. ASSP*, Vol. 34, No.6, June 1986, pp. 442-451.

[8] S. Barbarossa and O. Lemoine: "Analysis of nonlinear FM signals by pattern recognition of their time-frequency representations," *IEEE Sig. Proc. Let.*, Vol. 3, No.4, Apr. 1996, pp. 112-115.

[9] A. Francos and M. Porat: "Analysis and synthesis of multicomponent signals using positive time-frequency distributions," *IEEE Trans. Sig. Proc.*, Vol. 47, No.2, Feb. 1999, pp. 493-504.

[10] C. Wang and M.G. Amin: "Performance analysis of instantaneous frequency-based interference excision techniques in spread spectrum communications," *IEEE Trans. Sig. Proc.*, Vol. 46, No.1, Jan. 1998, pp. 70-82.

[11] C. Wang and M. G. Amin: "Zero-tracking time-frequency distributions," in *Proc. of IEEE ICASSP*, Munchen 1997.

TABLE II

MAE for the IF estimation based on: WD maxima (Max); median of length 3 applied directly to the IF (Med 3); proposed algorithm (Prop); BJD Prop - proposed algorithm applied to the BJD; XWD - cross Wigner distribution. Suc. is number of correct results out of the 25 considered trials.

σ/A	$A^2/2\sigma^2[dB]$	Max	Med 3	Prop	Suc.	BJD Prop	XWD
≤ 0.9	$\leq 0dB$	0	0	0	25	≤ 1.23	$\leq 5 \cdot 10^{-3}$
1	$-3dB$	$3.36 \cdot 10^{-1}$	$1.38 \cdot 10^{-2}$	0	25	4.34	$1.68 \cdot 10^{-1}$
1.25	$-4.9dB$	12.79	4.74	$1.66 \cdot 10^{-2}$	25	7.24	2.30
1.5	$-6.5dB$	39.21	25.08	0.24	25	16.80	19.12
1.75	$-7.9dB$	59.65	45.65	1.07	25	23.63	32.34
2	$-9dB$	71.45	60.52	11.74	18	41.02	46.20
2.25	$-10dB$	77.84	68.42	22.36	12	46.17	63.40
2.5	$-11dB$	80.68	71.71	33.61	11	51.20	74.35
2.75	$-11.8dB$	82.07	72.99	46.88	4	65.49	80.23

TABLE III

MAE of the IF estimation based on the WD and the proposed algorithm for different subinterval lengths.

σ	$A^2/2\sigma^2[dB]$	Max	4	16	64	Prop
1	$-3dB$	$3.36 \cdot 10^{-1}$	0	0	0	0
1.25	$-4.9dB$	12.79	0.20	$2.2 \cdot 10^{-2}$	$1.69 \cdot 10^{-2}$	$1.66 \cdot 10^{-2}$
1.5	$-6.5dB$	39.21	4.44	0.27	0.25	0.24
1.75	$-7.9dB$	59.65	24.83	1.35	1.12	1.07
2	$-9dB$	71.45	47.50	20.64	12.54	11.74

[12] R.G. Baraniuk, M.Coates and P.Steeghs: "Hybrid linear/quadratic time-frequency attributes," *IEEE Trans. Sig. Proc.*, Vol. 49, No.4, Apr. 2001, pp. 760-766.

[13] B. Boashash: "Estimating and interpreting the instantaneous frequency of a signal - Part 1: Fundamentals," *Proc. IEEE*, Vol. 80, No.4, Apr. 1992, pp. 519-538.

[14] B. Boashash: "Estimating and interpreting the instantaneous frequency of a signal - Part 2: Algorithms and applications," *Proc. IEEE*, Vol. 80, No.4, Apr. 1992, pp. 540-568.

[15] P. Rao and F.J. Taylor: "Estimation of the instantaneous frequency using the discrete Wigner distribution," *El. Let.*, Vol. 26, 1990, pp. 246-248.

[16] V. Katkovnik and LJ. Stanković: "Instantaneous frequency estimation using the Wigner distribution with varying and data-driven window length," *IEEE Trans. Sig. Proc.*, Vol. 46, No.9, Sept. 1998, pp. 2315-2326.

[17] G. Andria, M. Savino and A. Trotta: "Application of Wigner-Ville distribution to measurements on transient signals," *IEEE Trans. Ins. Meas.*, Vol. 43, No.2, Apr. 1994, pp. 187-193.

[18] LJ. Stanković and V. Katkovnik: "Instantaneous frequency estimation using higher order distributions with adaptive order and window length," *IEEE Trans. Inf. Th.*, Vol. 46, No.1, Jan. 2000, pp. 302-311.

[19] B.G. Quin: "Estimating frequency by interpolation using Fourier coefficients," *IEEE Trans. Sig. Proc.*, Vol. 42, No.5, May 1994, pp. 1264-1268.

[20] I. Djurović and LJ. Stanković: "A virtual instrument for time-frequency signal analysis," *IEEE Trans. Instr. Meas.*, Vol. 48, No.6, Dec. 1999, pp. 1086-1092.

[21] I. Djurović and LJ. Stanković: "Influence of high noise on the instantaneous frequency estimation using time-frequency distributions," *IEEE Sig. Proc. Let.*, Vol. 7, No.11, Nov. 2000, pp. 317-319.

[22] A. Martell: "Edge detection using heuristic search methods," *Computer Graphics and Image Processing*, Vol. 1, No. 2, Aug. 1982, pp. 169–182.

[23] G.D. Forney: "The Viterbi algorithm," *Proc. IEEE*, Vol.61, No.3, Mar. 1973, pp. 268-278.

[24] D.L. Jones and R.G. Baraniuk: "An adaptive optimal-kernel time-frequency representation," *IEEE Trans. Sig. Proc.*, Vol. 43, No.10, Oct. 1995, pp. 2361-2371.

[25] F.S. Cohen, G.F. Boudreaux-Bartels and S. Kadambe: "Tracking of unknown non-stationary chirp signals using unsupervised clustering in the Winger distribution space," in *Proc. of IEEE ICASSP'1998*, Vol. 4, Apr. 1988, New York, USA, pp. 2180-2183.

[26] R.G. Baraniuk and D.L. Jones: "Signal-dependent time-frequency analysis using a radially Gaussian kernel," *Sig. Proc.*, Vol. 32, No.2, June 1993, pp. 263-284.

[27] T.H. Sang and W.J. Williams: "Renyi informa-

tion and signal dependent optimal kernel design," in *Proc. IEEE ICASSP'95*, Vol. 2, 1995, pp. 997-1000.

[28] B. Ristić and B. Boashash: "Kernel design for time-frequency signal analysis using the Radon transform," *IEEE Trans. Sig. Proc.*, Vol. 41, No.5, May 1995, pp. 1996-2008.

[29] LJ. Stanković and S. Stanković: "On the Wigner distribution of the discrete-time noisy signals with application to the study of quantization effects," *IEEE Trans. Sig. Proc.*, Vol. 42, No.7, July 1994, pp. 1863-1867.

[30] I. Djurović and LJ. Stanković: "Robust Wigner distribution with application to the instantaneoues frequency estimation," *IEEE Trans. Sig. Proc.*, Vol. 49, No. 12, Dec. 2001, pp. 2985-2993.

[31] J. Jeong and W.J. Williams: "Kernel design for reduced interference distributions," *IEEE Trans. Sig. Proc.*, Vol. 40, No.2, Feb. 1992, pp. 402-412.

[32] LJ. Stanković: "A method for time-frequency signal analysis," *IEEE Trans. Sig. Proc.*, Vol. 42, No.1, Jan. 1994, pp. 225-229.

[33] V. Katkovnik: "Robust $M-$estimates of the frequency and amplitude of a complex-valued harmonic," *Sig. Proc.*, Vol. 77, No.1, Aug. 1999, pp. 71-84.

[34] V. Katkovnik, I. Djurović and LJ. Stanković: "Instantaneous frequency estimation using robust spectrogram with varying window length," *AEU - International Journal of Electronics and Communications*, Vol. 54, No.4, 2000, pp. 193-202.

[35] LJ. Stanković, I. Djurović and S. Stanković: "The robust Wigner distribution," in *Proc of IEEE ICASSP'2000*, Istanbul, Vol. I, pp. 77-80.

[36] B. Boashash and B. Ristić: "Polynomial time-frequency distributions and time-varying higher order spectra: Applications to analysis of multicomponent FM signals and to treatment of multiplicative noise," *Sig. Proc.*, Vol. 67, No.1, May 1998, pp. 1-23.

[37] L.L. Scharf and B. Friedlander: "Toeplitz and Hankel kernels for estimating time-varying spectra of discrete-time random process," *IEEE Trans. Sig. Proc.*, Vol. 49, No.1, Jan. 2001, pp. 179-189.

[38] LJ. Stanković: "The auto-term representation by the reduced interference distributions; The procedure for a kernel design," *IEEE Trans. Sig. Proc.*, Vol. 44, No.6, June 1996, pp. 1557-1564.