Adaptive Thresholding in Extracting Useful Information From Noisy Time-Frequency Distributions

Nicoletta Saulig
Department of Engineering
Juraj Dobrila University of Pula
Pula, Croatia
nicoletta.saulig@unipu.hr

Jonatan Lerga
Faculty of Engineering
University of Rijeka
Rijeka, Croatia
jlerga@riteh.hr

Zlatko Baracskai
Computer Science and Creative Technologies
University of the West of England
Bristol, UK
zlatko.baracskai@uwe.ac.uk

Miloš Daković
Faculty of Electrical Engineering
University of Montenegro
Podgorica, Montenegro
milos@ac.me

Abstract—This paper provides an analysis of the performance of an automatic method for extraction of useful information content from time-frequency distributions of nonstationary signals in dependence on the selected time-frequency method. The tested algorithm for the extraction of the signal components (useful information) from the noisy mixture is based on an initial segmentation of the time-frequency distribution which provides a fixed number of data classes. The normalized energies of the different classes are used as input to a statistical test which produces two outputs: 'useful information" classes and 'noise" classes, respectively. The quantity used as indicator of the class type, being the normalized energy of one class, is highly dependent on the time-frequency kernel filter. This paper reports the results of the proposed method applied to three well performing time-frequency methods, the Smoothed-Pseudo Wigner-Ville distribution, the Choi-Williams distribution, and the Modified-B distribution. The performance comparison attests the method's robustness for the different kernel filters, in various SNRs.

Index Terms—Time-frequency distributions, threshold, K-means, intersection of confidence intervals (ICI) rule.

I. INTRODUCTION

Signals encountered in various practical contexts are often mixtures of useful information (signal components) and noise. Noisy environments require to be suppressed to allow the further analysis of signal features in many engineering applications. Nonstationary noisy signals can be written as

$$
\begin{aligned}
y(t) &= \sum_{i=1}^{J} a_i(t) e^{j2\pi \int_0^t f_i(\tau)d\tau} + \nu(t) \\
&= \sum_{i=1}^{J} a_i(t) e^{j\Phi_i(t)} + \nu(t),
\end{aligned} \tag{1}
$$

with $a_i(t)$ being the time-varying amplitude, $f_i(t)$ the instantaneous frequency, and $\Phi_i(t)$ the instantaneous phase of the ith signal component, and the signal is corrupted by independent additive white Gaussian noise (AWGN) [1]. Furthermore, $x_i(n)$ is one of the signal components, where their total number J is not necessarily known to the user. When using time-frequency distributions (TFDs) $\rho(n,m)$ (with dimension $(N \times M)$), to represent such signals, the analyst's interest is focused on the contribution of the J signal components [1]. In a simplified model, often adopted in real-life applications, signal components are equated with elongated energy ridges, while the contribution of the AWGN is evenly distributed in the time-frequency (TF) plane. However, white Gaussian noise is evenly distributed over the TF plane, i.e. an additive white noise model holds for the WD only. On the other hand, if a finite number of data samples is used to compute the TFD, the variance takes a finite value, depending on the TFD kernel [2]–[5], defining the structure of the noisy TFD. The characteristic TFD structure allows the application of both hard and adaptive thresholding methods as simple denoising procedures [6]–[10]. The thresholding operation is defined as:

$$
\rho_{th}(n,m) = \begin{cases} \rho(n,m), & \text{if } \rho(n,m) > \epsilon \\ 0, & \text{otherwise} \end{cases} \tag{2}
$$

where $\rho(n,m)$ is a discrete TFD, and ϵ is an adequate hard threshold level. The hard threshold level ϵ determines the quantity of coefficients to be irreversibly removed from the TFD, and its choice is hence a crucial preprocessing step, which is usually left to the analysts' arbitrary choice.

This task has recently been taken over by the application of an automatic, near-to-optimal, adaptive thresholding method [11], efficient in the useful information extraction from TFDs. The method considers the set of observations $C = \{\rho(n,m)| \; n = 1,...,N, \; m = 1,...,M\}$, partitioning these $N \times M$ observations into K subsets, making use of the K-means method as $\mathbf{C} = \{C_k|k \in \mathbb{N}, \; 1 \leq k \leq K\}$, in order

to minimize the within-cluster sum:

$$\underset{\mathbf{C}}{\arg\min} \sum_{k=1}^{K} \sum_{\rho(n,m) \in C_k} \|\rho(n,m) - \mu_k\|^2, \quad (3)$$

where μ_k is the mean of each set C_k. The elements of each set are then distributed in the TF plane as

$$\rho_k(n,m) = \begin{cases} \rho(n,m), & \text{if } \rho(n,m) \in C_k \\ 0, & \text{elsewhere,} \end{cases} \quad (4)$$

to obtain K classes $\rho_k(n,m)$ from the initial TFD. Based on a reasonable sparsity constraint applicable to TFDs, implying components being narrow energy ridges, while noise being evenly spread over the TF plane, it can be expected that classes containing mainly noise coefficients will present large TF supports, while classes containing signal components will presents small supports. TF supports or amplitude-normalized energies of individual classes are intended as the class' total number of non-zero coefficients, i.e., the l_0-norm of the class [12]–[15], denoted as $E(k)$.

In order to partition the set $\mathbf{C}$ to obtain $\mathbf{C} = \{C_{noise}, C_I\}$ the individual classes C_k should be classified as one of the two subsets based on the corresponding $E(k)$.

The method for determining the pertinent neighborhood of $E(k)$ relies on the intersection of confidence intervals (ICI). K confidence intervals are calculated for each $E(k)$ together with $\hat{E}(k,i)$ being the estimate of the ideal $\tilde{E}(k)$.

In order to determine the affine regions of $E(k)$ the proposed procedure tracks the overlapping of confidence intervals resulting in the value $i^+(k)$ (where $i^+(k)$ defines the largest index $i(k)$ of $\hat{E}(k,i)$ for which the intersection of all consecutive confidence intervals is nonempty and estimation error $e(k) = \tilde{E}(k) - \hat{E}(k,i)$ is minimal [16], [17]).

It is considered that the ideal $\tilde{E}(k)$ belongs to the interval $D(k,i)$ with the probability $1 - \alpha$, where the confidence intervals $D(k,i)$ are defined as

$$D(k,i) = [\hat{E}(k,i) - \Gamma\sigma(k,i), \hat{E}(k,i) + \Gamma\sigma(k,i)]. \quad (5)$$

The ICI rule introduces a set of growing class indices for each $E(k)$ ($i(k) = 1, 2, \cdots, K - k + 1$). Next, it calculates the corresponding confidence intervals $D(k,i)$ (confidence intervals are reducing in size as $i(k)$ increases). The smallest upper and largest lower confidence limits are $\underline{U}(k,i)$ and $\overline{L}(k,i)$. In the next step it tracks the intersection of the confidence intervals as long as it exists (in other words, as long as it is nonempty) [16]. Finally, the class index $i^+(k)$ is obtained as the largest one for which it is still true that the intersection of all confidence intervals up to, and including $i^+(k)$, is nonempty. In particular, $i^+(k)$ is

$$i^+(k) = \underset{i(k)}{\arg\max} \{\cap_{i(k)=1}^{K-k+1} D(k,i) \neq \emptyset\}. \quad (6)$$

Assuming a sufficiently large number of classes was chosen, $i^+(k)$ can be defined as the integer value of $i(k)$ close to the ideal one $i^*(k)$ resulting in an estimation of $\hat{E}(k,i^+)$ as close

TABLE I
KERNEL FILTERS IN THE DISCRETE TIME-LAG DOMAIN.

Distribution	$G(n,l)$
CWD	$\frac{\sqrt{\pi\sigma_{\text{CWD}}}}{\|2l\|} \exp\left(\frac{-\pi^2\sigma_{\text{CWD}}n^2}{4l^2}\right) * *[\text{sinc } n \text{ sinc } l]$
MBD	$\frac{\cosh^{-2\beta} n}{\sum_n \cosh^{-2\beta} n}$
SPWVD	$g(n)w(l)$

as possible to the optimal $\hat{E}(k,i^*)$. The useful information content can then be recovered as

$$\rho_I(n,m) = \sum_{k=k^++1}^{K} \rho_k(n,m). \quad (7)$$

Fig. 1 shows TFD of the first tested noisy signal with its segmentation to classes, as well as the extracted information. Fig. 2 reports the described procedure of segmentation and the ICI criterion applied to the spectrogram of a three-component noisy signal with $\Gamma = 0.8$. However, since TF techniques with high representation quality are often preferred to the basic spectrogram, in the following section the performance of the adaptive thresholding method is tested on several advanced TF methods.

II. ALGORITHM'S PERFORMANCE ASSESSMENT FOR DIFFERENT KERNEL FILTERS

The structure of TFDs is highly influenced by the selection of the kernel filter, which controls the tradeoff between components' concentration and cross-terms minimization, but also rules distortions of the TFD introduced by the presence of noise [18]–[22]. In order to determine whether the proposed thresholding method performance is stable for different kernel filters three different TFDs have been considered. The tested TFDs are namely: the Smoothed-Pseudo Wigner-Ville distribution (SPWVD) [1], the Modified B-distribution (MBD) [23], both from the class of the separable-kernel TFDs, and the Choi-Williams distribution (CWD), known as the exponential distribution and exponent of the Reduced interference distributions [24], [25]. Kernels of tested TFDs in the discrete time-lag domain are given in Table I. Note that these distributions were chosen to demonstrate the method's performances. However, it may also be successfully applied to various other TFDs.

The performance of the procedure is evaluated by means of a reference TFD, i.e. the TFD of the noise-free test signal. The error rate is calculated using the total number of residual non-zero coefficients obtained by the subtraction of the thresholded TFD from the reference, noise-free TFD of the same test signal. This measure considers both the residual noise in the TFD and involuntarily removed parts of components with low energy. In other words, each non-zero coefficient can be classified either as a false positive (residual noise) estimate, or a false negative (involuntarily removed parts of components) estimate. The error rate, represented as the total number of
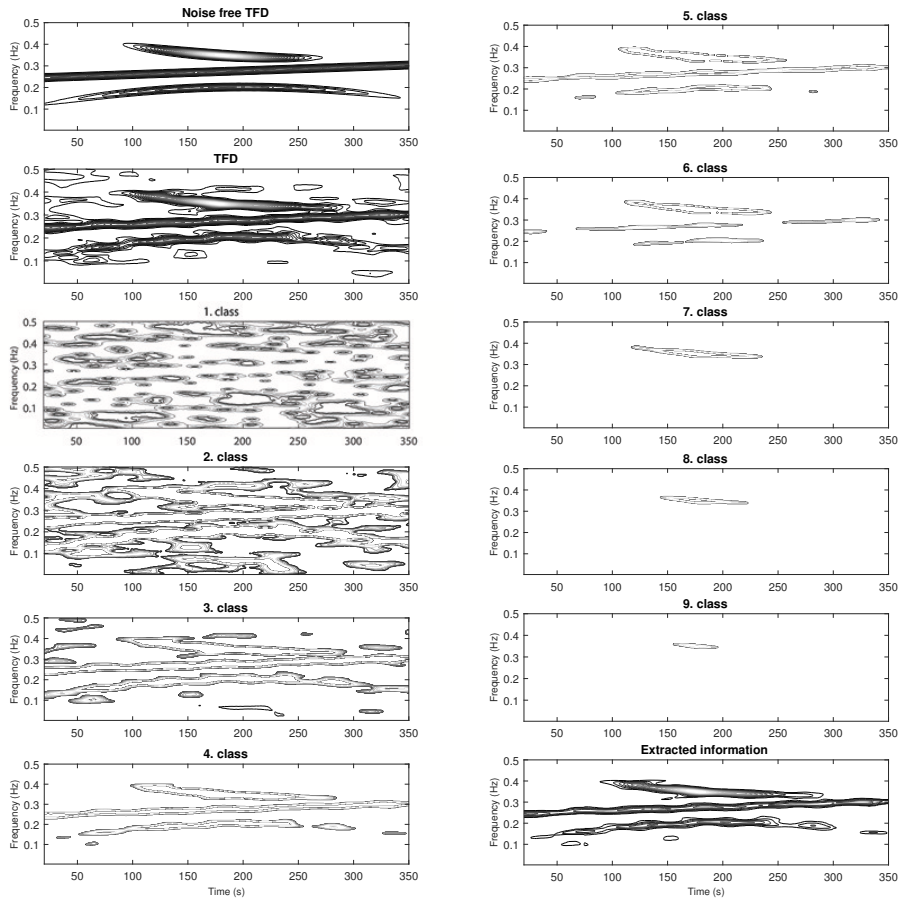
Fig. 1. TFD of the noise-free signal and in $SNR = 3$ dB, TFD segmentation in 9 classes, and extracted information.

	Signal 1				Signal 2		
SNR	CWD	SPWVD	MBD	SNR	CWD	SPWVD	MBD
-3dB	**4.90**	4.85	7.49	-3dB	8.65	**6.43**	6.93
0dB	3.81	**3.04**	4.45	0dB	5.97	**4.05**	4.37
3dB	3.36	2.00	**1.95**	3dB	5.27	2.82	**1.52**
6dB	3.11	**1.67**	1.98	6dB	5.39	1.73	**1.41**

residual non-zero coefficients, is considered as percentage of the $N \times M$-dimensional set of observations. Table II reports the results obtained for two test signals with multiple components. The TFDs of the noisy test signals and adaptively thresholded TFDs are shown in Figs. 3 and 4. The simulation results tend to show that the error rate stays reasonably stable for the three TFDs for two tested signals; however better performance of the SPWVD and MBD can be observed. Note that the method may be also efficiently applied to signals with fast varying instantaneous frequency. By visual evaluation of Figs. 3 and 4 better representation quality of the SPWVD and MBD compared to the CWD can be noticed, which justifies the better results achieved by the proposed method in terms of error rate.

III. CONCLUSION

In this paper the robustness of adaptive TFDs' thresholding with respect to the kernel filter has been investigated. The thresholding method relies on the differences in the TF structures of noise and signal components, while these structures are themselves determined by the kernel filter. For the three tested TFDs, being the SPWVD, the CWD and the MBD, the method has shown tendency to perform better when applied to the SPWVD and MBD rather than to the CWD, which is consistent with the representation quality achieved by the TFDs. The reported results show that the method performance improves as the representation quality of the TFD is enhanced by advanced TF methods, expanding the possibility form its current application to the spectrogram to TFDs with high
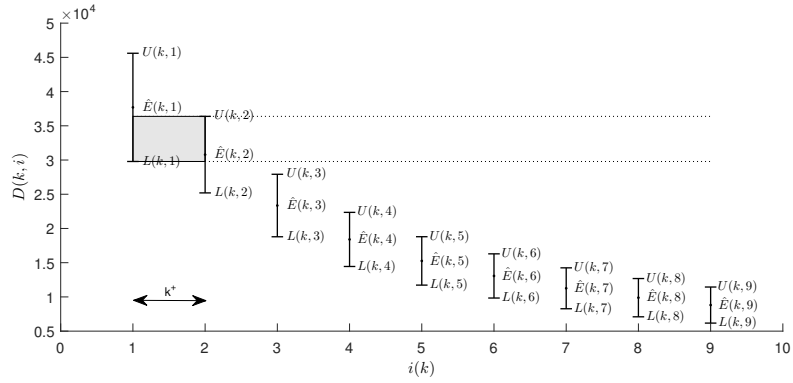
Fig. 2. Intersection of confidence intervals (for $k = 1$) for the TF supports of the 9 classes of the TFD from Fig. 1.
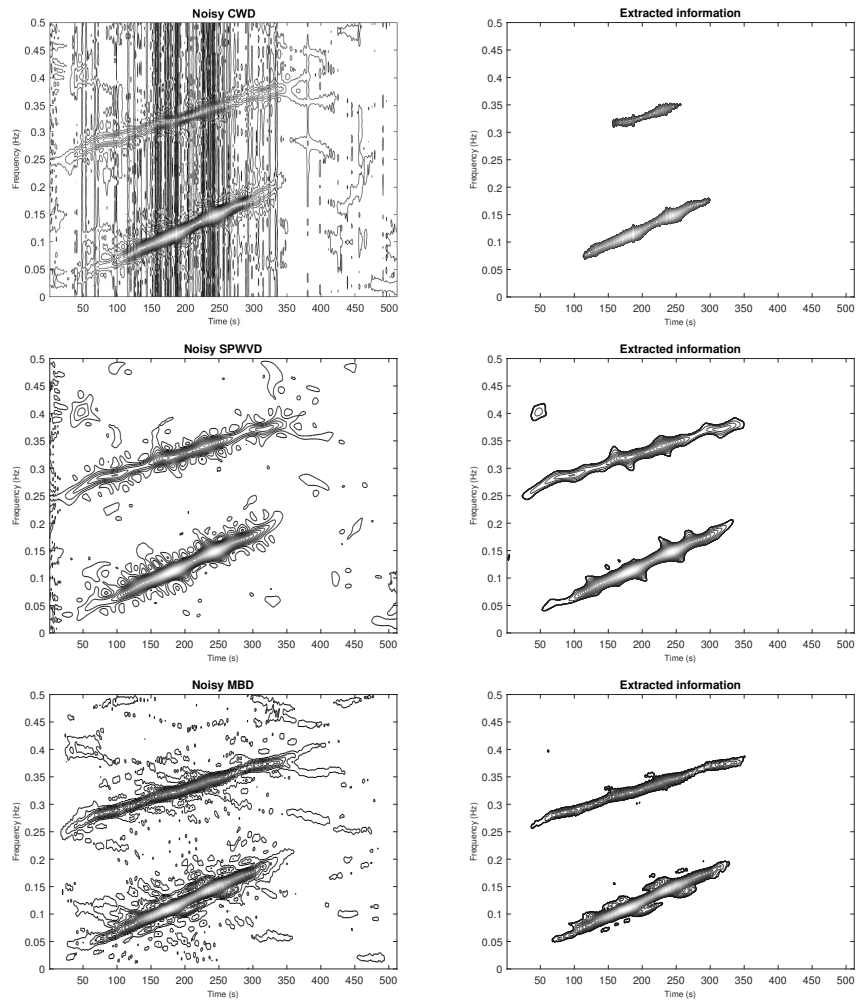

Fig. 3. TFDs of the noisy signal (left column), and thresholded TFDs (right column) for a two-component signal with linear frequency modulations.
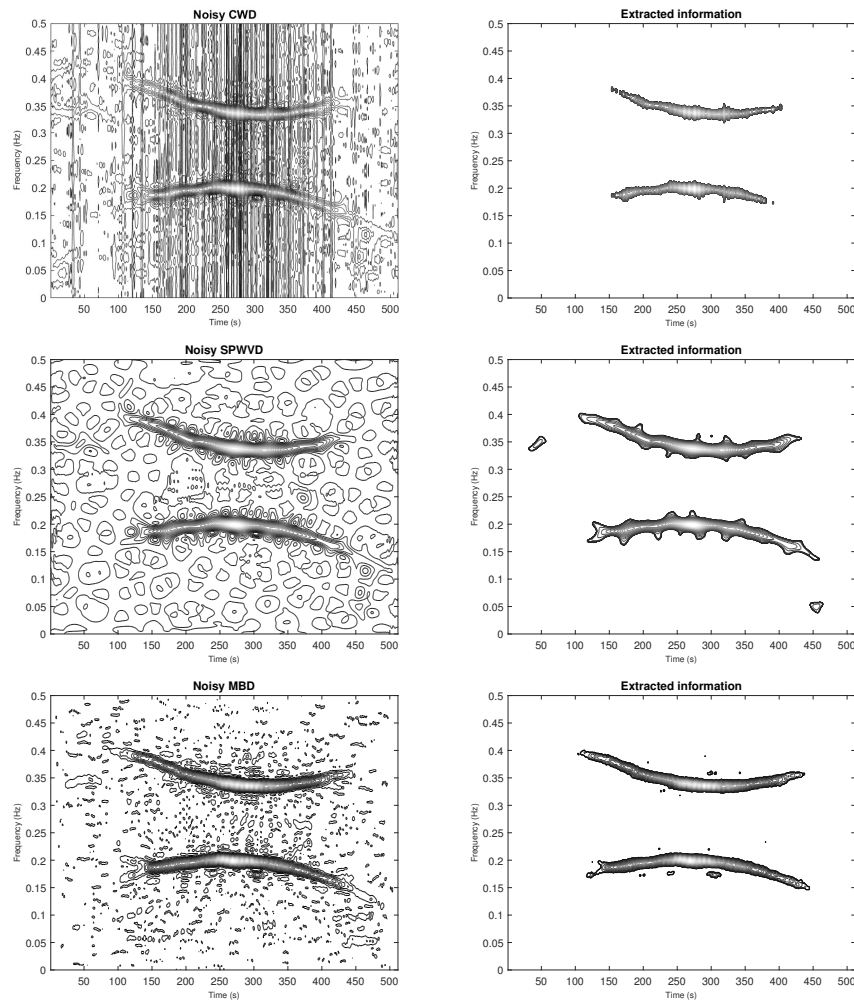
Fig. 4. TFDs of the noisy signal (left column), and thresholded TFDs (right column) for a two-component signal with parabolic frequency modulations.

quality performance, fitted to the users' requirements.

ACKNOWLEDGMENT

REFERENCES

[1] B. Boashash, *Time frequency signal analysis and processing: A comprehensive reference. Elsevier Academic Press*, 2016.

[2] A. H. Nuttall, Signal processing studies, NUSC, New London, Connecticut, Technical Report, 1989.

[3] M. G. Amin, Minimum variance time-frequency distribution kernels for signals in additive noise, in IEEE Transaction on Signal Processing, vol.44, no.9, 1996, pp. 2352-2356, doi: 10.1109/78.536695.

[4] LJ. Stanković, and V. Ivanović, Further results on the minimum variance time-frequency distribution kernels, IEEE Transactions on Signal Processing, vol.45, no.6, 1997, pp. 1650-1655, doi:10.1109/78.600007.

[5] LJ. Stanković, and V. Katkovnik, The Wigner distribution of noisy signals with adaptive time-frequency varying window, IEEE Transactions on Signal Processing, vol.47, no.4, 1999, pp. 1099-1108, doi:10.1109/78.752607.

[6] A. Aissa-El-Bey, N. Linh-Trung, K. Abed-Meraim, A. Belouchrani, and Y. Grenier, Underdetermined blind separation of nondisjoint sources in the time-frequency domain, in IEEE Transactions on Signal Processing, vol.55, no.3, 2007, pp.897-907, doi: 10.1109/TSP.2006.888877.

[7] N. Saulig, N. Pustelnik, P. Borgnat, P. Flandrin, and V. Sucic, Instantaneous counting of components in nonstationary signals, in Proceedings of European Signal Processing Conference (EUSIPCO), Marrakesh, Marocco, 2013, pp. 1-5.

[8] B. Barkat, and K. Abed-Meraim, Algorithms for blind components separation and extraction from the time-frequency distribution of their mixture, in EURASIP Journal of Applied Signal Processing, vol.2004, no.1, 2004, pp. 2025-2033, doi: 10.1155/S1110865704404193.

[9] N. Linh-Trung, A. Belouchrani, K. Abed-Meraim, and B. Boashash, Separating more sources than sensors using time-frequency distributions, in EURASIP Journal on Applied Signal Processing, vol.2005, no.17, 2005, pp. 2828-2847, doi: 10.1155/ASP.2005.2828.

[10] N. Saulig, Ž. Milanović, and C. Ioana, A Local Entropy-Based Algorithm for Information Content Extraction from Time-frequency Distributions of Noisy Signals, in Digital Signal Processing, vol.70, 2017, pp. 155-165.

[11] N. Saulig, J. Lerga, Ž. Milanović and C. Ioana, Extraction of Useful Information Content From Noisy Signals Based on Structural Affinity of Clustered TFDs Coefficients, in IEEE Transactions on Signal Processing, vol. 67, no. 12, pp. 3154-3167, 15 June15, 2019. doi: 10.1109/TSP.2019.2912134

[12] L. Stankovič, A measure of some time-frequency distributions concentration, Signal Processing, Vol.81, No.3, Mar.2001, pp.621-631.

[13] P. Borgnat, and P. Flandrin, Time-frequency localization from sparsity constraints, in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Las Vegas, NV, USA, 2008, doi:10.1109/ICASSP.2008.4518477.

[14] P. Flandrin, and P. Borgnat, Time-frequency energy distributions meet compressed sensing, in IEEE Transactions on Signal Processing, vol.58, no.6, 2010, pp. 2974-2982, doi:10.1109/TSP.2010.2044839.

[15] Z. Deprem, and A. E. Çetin, Cross-term-free time-frequency distribution reconstruction via lifted projections, in IEEE Transactions on Aerospace and Electronic Systems, vol.51, no.1, 2015, pp. 479-491, doi:10.1109/TAES.2014.140080.

[16] V. Katkovnik, A new method for varying adaptive bandwidth selection, in IEEE Transaction on Signal Processing, vol.47, no.9, 1999, pp. 2567-2571, doi:10.1109/78.782208.

[17] J. Lerga, M. Vrankić, and V. Sucic, A signal denoising method based on the improved ICI rule, in IEEE Signal Processing Letters, vol.15, 2008, pp. 601-604, doi:10.1109/LSP.2008.2001817.

[18] A. W. Rihaczek, Signal energy distribution in time and frequency, in IEEE Transactions on Information Theory, vol.14, no.3, 1968, pp. 369-374, doi:10.1109/TIT.1968.1054157.

[19] C. H. Page, Instantaneous power spectra, in Journal of Applied Physics, vol.23, no.103, 1952, pp. 103-106. doi:http://dx.doi.org/10.1063/1.1701949.

[20] M. J. Levin, Instantaneous spectra and ambiguity functions, in IEEE Transactions Information Theory, vol.10, no.1, 1964, pp. 95-97, doi:10.1109/TIT.1964.1053638.

[21] H.-I. Choi, and W. J. Williams, Improved time-frequency representation of multi-component signals using exponential kernels, in IEEE Transactions Acoustics, Speech and Signal Processing, vol.37, no.6, 1989, pp. 862-871, doi:10.1109/ASSP.1989.28057.

[22] Y. Zhao, L. E. Atlas, and R. J. Marks, The use of cone-shaped kernels for generalized time-frequency representations of non-stationary signals, in IEEE Transactions Acoustics, Speech and Signal Processing, vol.38, no.7, 1990, pp. 1084-1091, doi:10.1109/29.57537.

[23] Z. M. Hussain, and B. Boashash, Adaptive instantaneous frequency estimation of multicomponent FM signals using quadratic time-frequency distributions, in IEEE Transaction on Signal Processing, vol.50, no.8, 2002, pp. 1866-1876, doi: 10.1109/TSP.2002.800406.

[24] H.-I. Choi, and W. J. Willams, Improved time-frequency representation of multicomponent signals using exponential kenels, in IEEE Transactions on Acoustics Speech and Signal Processing, vol.37, no.6, 1989, pp. 862-871, doi:10.1109/ASSP.1989.28057.

[25] W. J. Williams, and J. Jeong, Reduced interference time-frequency distributions, in B. Boashash (Ed.), *Time-frequency Signal Analysis: Methods and Applications* Longman-Cheshire/Wiley, Melbourne/New York, 1992, pp. 74-97.