Measuring Time-Frequency Distributions Concentration

LJubiša Stanković

Efficient measurement of time-frequency distributions (TFDs) concentration can provide a quantitative criterion for evaluation of various distributions performance. It can be used for adaptive and automatic parameter selection in time-frequency analysis, without supervision of a user. Measures for distribution concentration of monocomponent signals date back to [3], [8]. For more complex signals, some quantities from statistics and information theory were the inspiration for defining measures of the TFDs concentration [4]-[5]. They provided good quantitative measure of the auto-terms concentration. Various and efficient modifications are used in order to take into account the appearance of oscillatory cross-terms.

The application of concentration measures will be demonstrated on automatic determination of the "best window length" for the spectrogram or "the best number of terms" in the method that provides transition form the spectrogram toward the pseudo Wigner distribution (pseudo WD) [Article 6.2].

A. Concentration Measurement

The basic idea for measuring TFDs concentration can be explained on a simplified example motivated by the probability theory. Consider a set of N nonnegative numbers $p_1, p_2, ..., p_N \geq 0$, such that $p_1 + p_2 + ... + p_N = 1$. Form a simple test function $M(p_1, p_2, ..., p_N) = p_1^2 + p_2^2 + ... + p_N^2$. It is easy to conclude that $M(p_1, p_2, ..., p_N)$, under the constraint $p_1 + p_2 + ... + p_N = 1$, has the minimal value for $p_1 = p_2 = ... = p_N = 1/N$, i.e., for maximally spread values of $p_1, p_2, ..., p_N$. The highest value of $M(p_1, p_2, ..., p_N)$, under the same constraint, is achieved when only one p_i is different from zero, $p_i = \delta(i - i_0)$, where i_0 is an arbitrary integer $1 \leq i_0 \leq N$. This case corresponds to the maximally concentrated values of $p_1, p_2, ..., p_N$, at a single $p_{i_0} = 1$. Therefore, the function $M(p_1, p_2, ..., p_N)$ can be used as a measure of concentration of the set of numbers $p_1, p_2, ..., p_N$, under the unity sum constraint.[1] In general, the constraint can be included in the function itself by using the form $M(p_1, p_2, ..., p_N) = \left(p_1^2 + p_2^2 + ... + p_N^2\right) / \left(p_1 + p_2 + ... + p_N\right)^2$. For nonnegative $p_1, p_2, ..., p_N$ this function has the minimum for $p_1 = p_2 = ... = p_N$, and reaches its maximal value when only one p_i is different from zero.

In time-frequency analysis this idea has been used in order to measure the concentration. Several forms of the concentration measure, based on this fundamental idea, are introduced.

1. **Measure based on the ratio of norms:** For the WD of energy normalized signals, the relation $\sum_n \sum_k \rho_x^2(n,k) \equiv 1$ holds. Therefore, substitution $p_i \rightarrow \rho_x^2(n,k)$ in the basic example, gives a function that can be used for measuring the concentration of the time-frequency representation $\rho_x(n,k)$:

$$M_{JP} = \left(\frac{L_4}{L_2}\right)^4 = \frac{\sum_n \sum_k \rho_x^4(n,k)}{\left(\sum_n \sum_k \rho_x^2(n,k)\right)^2}. \quad (1)$$

This form is just the fourth power of the ratio of L_4 and L_2 norms of $\rho_x(n,k)$.[2] It has been

[1] In probability theory, the famous Shannon entropy $-\sum_i p_i \log(p_i)$ is commonly used for the same purpose. It produces the maximal value for the lowest concentration of probabilities p_i, $p_1 = p_2 = ... = p_N = 1/N$, and the minimal value for the highest concentration $p_i = \delta(i - i_0)$.

[2] In statistics, similar form (known as kurtosis) is used as a measure of the flatness or peakedness of a distribution. Kurtosis is zero for a Gaussian distribution. Values greater than zero mean that the distribu-

Time-Frequency Signal Analysis and Processing, ed. B. Boashash, Elsevier, 2003.

introduced by Jones and Parks in [4]. They have used the magnitude of the signal's short-time Fourier transform as the time-frequency representation $\rho_x(n,k)$ in (1). High values of M_{JP} indicate that the representation $\rho_x(n,k)$ is highly concentrated, and vice versa. In general, any other ratio of norms L_p and L_q, $p > q > 1$, can also be used for measuring the concentration of $\rho_x(n,k)$ [4].

When there are two or more components (or regions in time-frequency plane of a single component) of approximately equal energies (importance), whose concentrations are very different, the norm based measures will favor the distribution with a "peaky" component, due to raising of distribution values to a high power. It means that if one component (region) is "extremely highly" concentrated, and all the others are "very poorly" concentrated, then the measure will not look for a trade-off, when all components are "well" concentrated. In order to deal with this kind of problems, common in time-frequency analysis, a concentration measure could be applied to smaller, local time-frequency regions [4]:

$$M_{JPL}(n,k) = \frac{\sum_n \sum_k Q^2(m-n,l-k)\rho_x^4(m,l)}{(\sum_n \sum_k Q(m-n,l-k)\rho_x^2(m,l))^2} \quad (2)$$

The localization weighting function $Q(n,k)$ determines the region where the concentration is measured. In [4] the Gaussian form of this function is used.

2. **Rényi entropy based measures:** The second class of TFD measures is defined in analogy with the Rényi entropy. It has been introduced in time-frequency analysis by Williams et al. [5], [9], with a significant contribution of [1], [2] in establishing the properties of this measure. The Rényi entropy, applied on the TFD $\rho_x(n,k)$, has the form

$$R_\alpha = \frac{1}{1-\alpha}\log_2\Big(\sum_n \sum_k \rho_x^\alpha(n,k)\Big) \quad (3)$$

with $\alpha > 2$ being recommended for the TFD measures [2]. For $\alpha = 2$ and the WD of energy normalized signals ($\sum_n \sum_k \rho_x^2(n,k) \equiv 1$), $R_2 = 0$ for all signals. Note that the logarithm is a monotone function. Thus, the behavior of R_α is determined by the argument $\sum_n \sum_k \rho_x^\alpha(n,k)$ behavior, as explained at the beginning of this section. In contrast to the measure (1), the entropy (3) has larger values for less concentrated distributions due to a negative coefficient $1/(1-\alpha)$ for $\alpha > 2$. This will be the case for all other measures which will be presented in the sequel.

It is interesting to note that the **Shannon entropy**

$$H = -\sum_n \sum_k [\rho_x(n,k)\log_2 \rho_x(n,k)]$$

could be recovered from the Rényi entropy, from the limit case $\alpha \to 1$, [2]. The Shannon entropy could not be used for general TFDs $\rho_x(n,k)$, which can assume negative values [2].

3. **Normalized Rényi entropy measures.** In order to avoid the problem which could be caused by the fact that the Rényi entropy based measure with $\alpha = 3$ ignore the presence of oscillatory cross-terms (when the auto-terms are well separated), some kind of normalization should be done. It can be done in various ways, leading to a variety of possible measure definitions [5].

Normalization with the distribution volume is performed as:

$$RV_3 = -\frac{1}{2} \times \log_2 \sum_n \sum_k \Big[\rho_x(n,k)/\sum_n \sum_k |\rho_x(n,k)|\Big]^3. \quad (4)$$

If the distribution contains oscillatory values, then summing their absolute values means that large cross-terms will decrease the measure RV_α. This is the expected behavior of a measure, since it will seek for a balance between the cross-terms suppression and auto-terms enhancement. The volume normalized form of measure has been used for adaptive kernel design in [5].

4. The basic idea for the measure that will be presented next comes from an obvious **classical definition of the time-limited signal duration**. If a signal $x(n)$ is time-limited to

tion has more of a peak than a Gaussian distribution, while values less than zero mean flatter distributions.

the interval $n \in [n_1, n_2 - 1]$, i.e., $x(n) \neq 0$ only for $n \in [n_1, n_2 - 1]$, then the duration of $x(n)$ is $d = n_2 - n_1$. It can be written as $d = \lim_{p\to\infty} \sum_n |x(n)|^{1/p}$. The same definition applied to a two-dimensional function $\rho_x(n,k) \neq 0$ only for $(n,k) \in D_x$, gives

$$N_D = \lim_{p\to\infty} \sum\nolimits_n \sum\nolimits_k |\rho_x(n,k)|^{1/p} \qquad (5)$$

where N_D is the number of points within D_x. In reality, there is no a sharp edge between $\rho_x(n,k) \neq 0$ and $\rho_x(n,k) = 0$, so the value of (5) could, for very large p, be sensitive to small values of $\rho_x(n,k)$. The robustness may be achieved by using lower order forms, for example with $p = 2$. Therefore, the concentration can be measured with the function of the form

$$M_p^p = (\sum\nolimits_n \sum\nolimits_k |\rho_x(n,k)|^{1/p})^p, \qquad (6)$$

with $\sum_n \sum_k \rho_x(n,k) = 1$ and $p > 1$.

After we have presented several possible forms for measuring the concentration of TFDs, we can summarize **a procedure for constructing a TFD measure** based on one-dimensional classical signal analysis definitions, or definitions from either probability, quantum mechanics, or information theory:

i) In the classical signal analysis definitions, consider the signal power $|x(t)|^2$ (spectral energy density $|X(f)|^2$) as the probability density function in time (frequency)). This idea comes from quantum mechanics, where the absolute square of the wave function is the position's probability density function.

ii) Assume that the TFD $\rho_x(t,f)$ can be treated as a joint two-dimensional probability density function.

iii) According to these assumptions, reintroduce one-dimensional definition into joint two-dimensional time-frequency domain.

iv) Additional modifications, interpretations, constraints, and normalizations are needed in order to get forms that can be used in time-frequency analysis. For example, several possible forms of the Rényi entropy measure in time-frequency domain have been proposed and used in various problems.

Example: Consider the classic **Leipnik entropy measure** [8]**,** and **Zakai's entropy parameter** $\delta_t = -\int_{-\infty}^{\infty} |x(t)|^2 \ln |x(t)|^2 \, dt$ of signal $x(t)$ [8]. According to the procedure for constructing a time-frequency form, based on a classical signal processing relation, we get

$$\delta_t = -\int_{-\infty}^{\infty} |x(t)|^2 \ln |x(t)|^2 \, dt$$

$$\rightarrow \quad -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \rho_x(t,f) \ln \rho_x(t,f) dt df \qquad (7)$$

This is exactly the well known **Shannon entropy**. It has already been discussed in [2] with respect to its (non)applicability in time-frequency problems. In a similar way, a logarithm of the general Zakai's signal duration (uncertainty)

$$Z_\alpha = \log_2 T_{2\alpha} = \frac{1}{1-\alpha} \log_2 \frac{\int_{-\infty}^{\infty} \left(|x(t)|^2\right)^\alpha dt}{(\int_{-\infty}^{\infty} |x(t)|^2 \, dt)^\alpha},$$

according to the proposed procedure, transforms into the Rényi entropy measure,

$$Z_\alpha \rightarrow \frac{1}{1-\alpha} \log_2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \rho_x^\alpha(t,f) dt df = R_\alpha$$

where $|x(t)|^2$ has been replaced by $\rho_x(t,f)$, and the unit signal energy is assumed.

Remark: In the probability theory all results are derived for the probability values p_i, assuming that $\sum_i p_i = 1$ and $p_i \geq 0$. The same assumptions are made in classical signal analysis for the signal power. Since a general TFD commonly does not satisfy both $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \rho_x(t,f) = 1$ and $\rho_x(t,f) \geq 0$, the obtained measures of TFD concentration may just *formally look like the original* entropies or classical signal analysis forms, while they can have different behavior and properties.[3]

[3]Quantum mechanics forms can also be used for the definition of highly concentrated signal representations. One of them is the "pseudo quantum" signal representation [6] in the form of $SD_x(t,\wp) = \int_{-\infty}^{\infty} x^{[L]}(t+\tau/(2L)) x^{*[L]}(t-\tau/(2L)) e^{-j\wp\tau} d\tau$, with $x^{[L]}(t) = A(t)\exp(jL\phi(t))$ for $x(t) = A(t)\exp(j\phi(t))$. For example, for $x(t) = A\exp(-at^2/2 + jbt^2/2 + jct)$ we get $SD_x(t,\wp) = A^2\exp(-at^2)\sqrt{4\pi/(a/L^2)}\exp(-(\wp -$

B. Numerical Examples

Consider the spectrogram

$$S_x^w(n,k) = \left|F_x^w(n,k)\right|^2 / E$$

where $F_x^w(n,k) = DFT_{m\to k}\{w(m)x(n+m)\}$ is the short-time Fourier transform (STFT); E is the energy of the lag window $w(m)$. Among several spectrograms, calculated with different window lengths or forms, the best one according to the one of proposed concentration measures, denoted by $\mathcal{M}[\rho_x(n,k)]$, will be that which minimizes (or maximizes, depending on the used measure form):

$$w^+ = \arg\min_w \left\{\mathcal{M}\left[S_x^w(n,k)\right]\right\}. \qquad (8)$$

Let us illustrate this by an example. Consider the signal

$$x(t) = \cos(50\cos(\pi t) + 10\pi t^2 + 70\pi t) + \cos(25\pi t^2 + 180\pi t) \qquad (9)$$

sampled at $\Delta t = 1/256$, within $-1 \le t < 1$. The Hanning window $w(m)$ with different lengths is used in the spectrogram calculation. Here, we have used the measure (6) with $p = 2$, although for this signal all presented measures would produce similar results [7]. Note that the presented measures would significantly differ if, for example, the second component were pure sinusoid $\cos(180\pi t)$ instead of $\cos(25\pi t^2 + 180\pi t)$.

For wide lag windows, signal nonstationarity makes the spectrogram very spread in the time-frequency plane, having relatively large measure $\mathcal{M}\left[S_x^w(n,k)\right] = M_2^2$, Figs.1(a), 1(b). For narrow lag windows its Fourier transform is very wide, causing spread distributions and large M_2^2, Figs.1(d), 1(e). Obviously, between these two extreme situations there is a window that produces an acceptable trade-off between the signal nonstationarity and small window length effects. The measure M_2^2 is calculated for a set of spectrograms with $N = 32$ up to $N = 256$ window length, Fig.1(f). The minimal measure value, meaning the best concentrated spectrogram according to this measure, is achieved for $N = 88$, Fig.1(f). The spectrogram with $N = 88$ is shown in Fig.1(c).

The same procedure will be used for determination of the optimal number of terms L, in a transition from the spectrogram to the pseudo WD, according to the recursive form of the S-method (SM) [Article 6.2]:

$$SM_x(n,k;L) = SM_x(n,k;L-1) + 2\Re\{F_x(n,k+L)F_x^*(n,k-L)\} \qquad (10)$$

where $SM_x(n,k;0) = \left|F_x^w(n,k)\right|^2$, and both $k+L$ and $k-L$ are within the basic frequency period. Note that $SM(n,k;N/2)$ is equal to the pseudo WD. The optimal distribution $SM_x^+(n,k;L)$, on our way from $L = 0$ (the spectrogram) toward $L = N/2$ (the pseudo WD), is the one calculated with L producing the minimal value of $\mathcal{M}\left[SM_x(n,k;L)\right]$,

$$L^+ = \arg\min_L \left\{\mathcal{M}\left[SM_x(n,k;L)\right]\right\}. \qquad (11)$$

Here, instead of $\left|SM_x(n,k;L)\right|$, a nonnegative part of $SM_x(n,k;L)$ will be used. Distributions $SM_x(n,k;L)$ should be properly scaled in order to satisfy unbiased energy condition. The same signal is used for the illustration of the SM. Since this method is based on the WD, the best results will be achieved with a wide lag window in the STFT calculation, $N = 256$. The spectrogram ($L = 0$) is shown in Fig.1(g). By increasing L the SM improves concentration of the spectrogram toward the pseudo WD quality, meaning lower measure $\{\mathcal{M}\left[SM_x(n,k;L)\right]\} = M_2^2$, Fig.1(h), 1(i). After L has reached the value equal to the distance between the auto-terms, cross-terms start to appear, increasing M_2^2, Fig.1(j),1(k). Minimal M_2^2 means a trade-off between the auto-terms concentration and the cross-terms appearances, Fig.1(k). The SM with L corresponding to minimal M_2^2 is shown in Fig.1(l).

The concentration measure is illustrated on time-frequency analysis of a pressure signal in the BMW engine with speed 2000 [rev/min], Fig.2, [Article 15.2].

$bt-c)^2/(a/L^2)$). For $a/L^2 \to 0$ it results in $SD(t,\wp) = 2\pi A^2\exp(-at^2)\delta(\wp - bt - c)$, what is just an ideally concentrated distribution along the instantaneous frequency. For a large a, if L^2 is large enough so that $a/L^2 \to 0$, we get the distribution highly concentrated in a very small region around the point $(t,\wp) = (0,c)$.

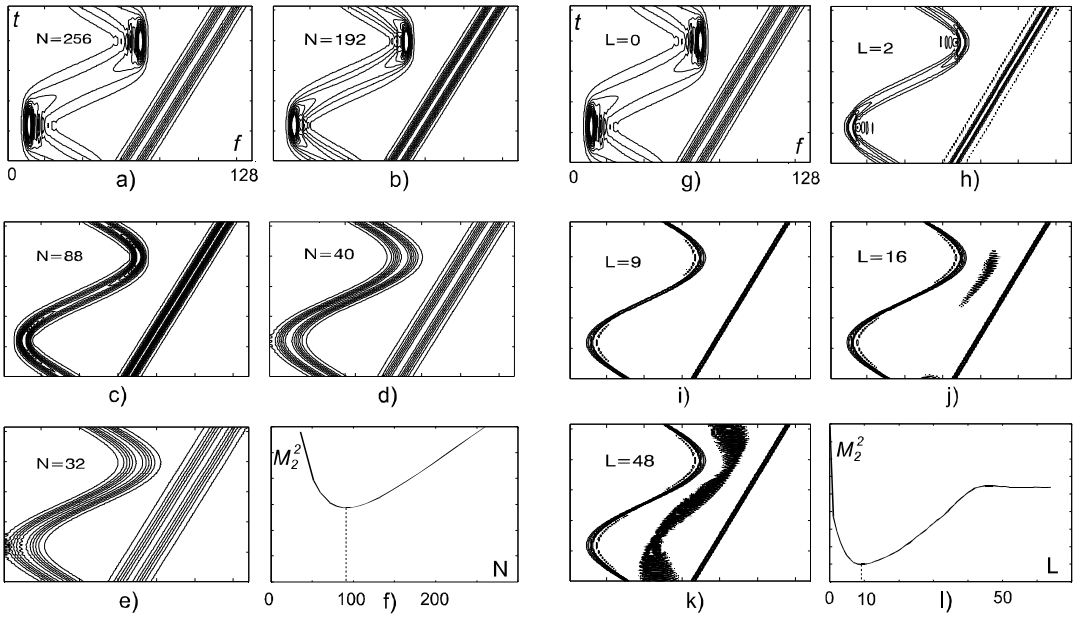

Fig. 1. a)-e) Spectrogram for various window lengths, and f) its measure $\mathcal{M}\left[S_x^w(n,k)\right] = M_2^2$. The lowest M_2^2 is achieved for $N = 88$, being the best window length choice according to this measure. g)-k) The S-method for various values of parameter L, and l) its measure $\mathcal{M}\left[SM_x(n,k;L)\right] = M_2^2$. The lowest M_2^2 is obtained for $L = 9$.

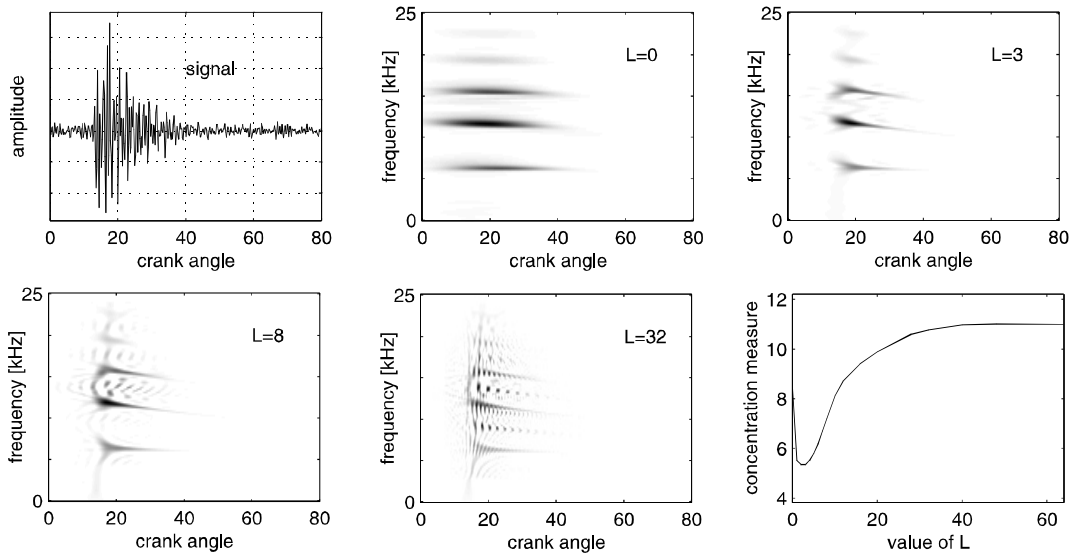

Fig. 2. Concentration measure illustration on time-frequency analysis of a car engine pressure signal. Signal, and its S-method based time-frequency representations are given. Time is rescaled into corresponding crank-angle. The best choice according to this measure was $L = 3$.

C. Parameter Optimization

Parameter optimization may be done by a straightforward computation of a distribution measure $\mathcal{M}[\rho_x(n,k)]$, for various parameter values. The best choice according to this criterion (optimal distribution with respect to this measure) is the distribution which produces the minimal value of $\mathcal{M}[\rho_x(n,k)]$. In the cases when one has to consider a wide region of possible parameter values for the distribution calculation (like for example window lengths in spectrogram), this approach can be numerically inefficient. Then, some more sophisticated optimization procedures, like the one using the steepest descent approach described in [5], can be used. Its simplified version will be presented here [7].

The gradient of a measure $\mathcal{M}[\rho_x(n,k)]$, with respect to a distribution's generalized optimization parameter denoted by ξ, is

$$\frac{\partial \mathcal{M}[\rho_x(n,k)]}{\partial \xi} = \frac{\partial \mathcal{M}[\rho_x(n,k)]}{\partial \rho_x(n,k)} \frac{\partial \rho_x(n,k)}{\partial \xi}.$$

Iterations, starting from a very low concentrated distribution toward the maximally concentrated one, i.e., toward the measure minimum, can be done according to

$$\xi_{m+1} = \xi_m - \mu \partial \mathcal{M}[\rho_x(n,k)]/\partial \xi \qquad (12)$$

where μ is the step, which should be chosen in the same way as the step in the other adaptive algorithms. The step should not be too small (since the convergence would be too slow), and not too large (to miss the minimum, or cause the divergence).

In discrete implementations, the gradient $\partial \mathcal{M}[\rho_x(n,k)]/\partial \xi$ can be approximated based on $\mathcal{M}[\rho_x(n,k;\xi_m)]$ calculated with ξ_m and its previous value ξ_{m-1}

$$\xi_{m+1} = \xi_m - \mu \frac{\mathcal{M}[\rho_x(n,k;\xi_m)] - \mathcal{M}[\rho_x(n,k;\xi_{m-1})]}{\xi_m - \xi_{m-1}}. \qquad (13)$$

Example: The optimization procedure will be illustrated on the signal $x(t)$, its spectrogram, and the measure form Example 2. The optimal window length is obtained in few iterations by using (13), starting from the very narrow window. Values of $\xi_0 = N = 16$ and $\xi_1 = N = 20$ in the initial and first iteration, are assumed. The next value of $\xi_{m+1} \equiv N$ is calculated according to (13). During the iterations we get $\xi_m = 16, 20, 76,$ and 90. The algorithm is stopped at $\xi_m = 90$, when $|\xi_{m+1} - \xi_m| < 2$, since even number of samples are used in the realization. Note that the obtained optimal value is within ± 2 of the value obtained by direct calculation. The value of parameter $\mu = 1/3$ has been used in all examples.

D. Summary

Measurement of time-frequency distributions concentration, with application to an automatic optimization of distribution parameters, is presented. It is based on the forms borrowed from the classical signal analysis, probability, or information theory, with appropriate interpretations and adjustments.

References

[1] R.G. Baraniuk, P. Flandrin, A.J.E.M. Janssen and O.J.J. Michel, "Measuring time-frequency information content using the Rényi Entropies," *IEEE Trans. IT,* vol.47, no.4, pp.1391-1409, May 2001.

[2] P. Flandrin, R.G. Baraniuk and O. Michel, "Time-frequency complexity and information," *in Proc. ICASSP,* vol.3, 1994, pp.329-332.

[3] D. Gabor, "Theory of communications," *J. Inst. EE,* vol.93, pp.429-441, 1946.

[4] D.L. Jones and T.W. Parks, "A high resolution data-adaptive time-frequency representation," *IEEE Trans. SP*, vol.38, pp.2127-2135, Dec.1990.

[5] T.H. Sang and W.J. Williams, "Rényi information and signal dependent optimal kernel design," *in Proc. ICASSP,* vol.2, pp.997-1000, 1995.

[6] LJ. Stanković, "Highly Concentrated Time-Frequency Distributions; Pseudo Quantum Signal Representation," *IEEE Trans. SP*, vol.45, pp.543-551, March 1997.

[7] LJ. Stanković, "A measure of some time-frequency distributions concentration," *Signal Processing*, vol.81, no.3, pp.621-631, March 2001.

[8] D.E. Vakman, "*Sophisticated Signals and the Uncertainty Principle in radar,*" Springer-Verlag, 1968.

[9] W.J. Williams, M.L. Brown and A.O. Hero, "Uncertainty, information and time-frequency distributions," *SPIE Adv., SPAA, Imp.,* vol.1566, pp.144-156, 1991.