ELSEVIER

A measure of some time–frequency distributions concentration

Ljubiša Stanković*

Elektrotehnicki fakultet, University of Montenegro, 81000 Podgorica, Montenegro, Yugoslavia

Abstract

A criterion that can provide a measure of time–frequency distribution concentration is proposed. In contrast to the norm-based concentration measures it does not need normalization in order to behave properly when cross-terms are present, and also it does not discriminate low concentrated components with respect to the highly concentrated ones within the same distribution. This measure has been used for the automatic window selection in the spectrogram, as well as in finding the optimal distribution that can be produced in a transition from the spectrogram toward the pseudo Wigner distribution. © 2001 Elsevier Science B.V. All rights reserved.

1. Introduction

Efficient time–frequency distribution concentration measurement can provide a quantitative criteria to evaluate performances of different distributions and can be used for adaptive and automatic parameters selection in time–frequency analysis, without interference by a user. Measures for distribution concentration of monocomponent signals date back to [4,9,10,17]. For more complex signals, some quantities in the statistics were the inspiration for defining measures for time–frequency distributions concentration in the form of the ratio of distribution norms by Jones and Parks [12], and the Rényi entropy by Williams et al. [8,15,18]. Distribution energy was also used, first by Baraniuk and Jones, for optimal kernel distribu-

tions design [2,3,6,7,11]. Common for all of these measures is that they are based on the distribution norms, i.e., sums over the distribution values raised to a power greater than one. They provided good quantitative measure of the auto-terms concentration. Norms themselves failed to behave in the desired way when the cross terms appeared. Various and efficient modifications were used in order to take into account the appearance of nondesirable oscillatory zero-mean distribution values. The distribution norm has been divided by a lower-order norm in [12,15], while some strict constraints were imposed on the kernel form in [2]. However, even the normalized forms of the norm-based measures are not quite appropriate for the cases where 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, due to raising of distribution values to a high power (fourth, in [12], third, in

---

*Tel./fax: +381-81-244921.

E-mail address: l.stankovic@ieee.org (L. Stanković).

[8,15]) will favor distributions with "peaky" components [12]. It means that if one component (region) is "extremely highly" concentrated, and all the others are "very poorly" concentrated, then they will not look for a compromise, for example, when all components are "very well" concentrated. In order to deal with this kind of problem, that are present in time–frequency analysis, Jones and Parks [12] introduced local concentration measure, that locally measures concentration, and increases the calculation complexity.

Here, we will present *a simple* measure for distributions concentration, that can overcome some of the mentioned drawbacks. It behaves well with respect to the auto-terms, cross-terms and does not discriminate low concentrated components against very concentrated ones during the optimization procedure. Its application will be demonstrated for *automatic* determination of the "best window length" for the spectrogram or "the optimal number of terms" in the S-method (SM) [16]. It could be used in other similar problems in time–frequency analysis.

2. A concentration measure

Consider time–frequency representation of a signal $x(t)$ denoted by $P_x(t,\omega)$. Assume that $P_x(t,\omega)$ satisfies the unbiased energy condition

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_x(t,\omega)\, \mathrm{d}t\, \mathrm{d}\omega = E_x.$$

Let us, just for the beginning, assume that $P_x(t,\omega) \neq 0$ only for $(t,\omega) \in D_x(t,\omega)$. For a large p we have that

$$M_p \triangleq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |P_x(t,\omega)|^{1/p}\, \mathrm{d}t\, \mathrm{d}\omega \to \iint_{D_x(t,\omega)} 1\, \mathrm{d}t\, \mathrm{d}\omega$$

$$= S_x, \qquad (1)$$

where S_x is the area of $D_x(t,\omega)$.

As a criterion for the distributions-concentration measure we will assume: *Among several given unbiased energy distributions, the best concentrated is the one having the smallest S_x.*

The basic idea for form (1) comes from an obvious definition of the time-limited signals duration [19]. If a signal $x(t)$ is time limited within the interval $t \in [t_1, t_2]$, i.e., $x(t) \neq 0$ only for $t \in [t_1, t_2]$, then the duration of $x(t)$ is $d = t_2 - t_1$. It can be written as $d = \lim_{p \to \infty} \int_{-\infty}^{\infty} |x(t)|^{1/p}\, \mathrm{d}t$.

Value of M_p raised to the pth power will be referred to as the concentration measure. Its discrete form is

$$\mathscr{M}[P_x] \triangleq M_p^p = \left(\sum_{n=1}^{N} \sum_{k=1}^{N} |P_x(n,k)|^{1/p} \right)^p \qquad (2)$$

with $\sum_{k=1}^{N} \sum_{n=1}^{N} P_x(n,k) = 1$ being the normalized unbiased energy constraint, and $p > 1$.

P_1: For nonnegative $P_x(n,k)$, and any $p > 1$, the constrained maximal value of $\mathscr{M}[P_x]$ is reached for the maximally spread function $P_x(n,k) = 1/N^2$ for all n, k. The global infimum is reached for the maximally concentrated function $P_x(n,k) = \delta(n - n_0, k - k_0)$, $1 \leqslant n_0, k_0 \leqslant N$.

Any function $P_x(n,k)$ uniformly spread over $N_1 \leqslant N$ points has the measure $\mathscr{M}[P_x]$ greater than any other function over the same number of points, including any other function uniformly spread over N_2 points, if $N_1 > N_2$ and $p > 1$. The ratio of measures for uniform functions is $(N_1/N_2)^{p-1}$.

Note that $\mathscr{M}[P_x]$ *does not satisfy the "triangle inequality" condition to be a norm* of $P_x(n,k)$ for $p > 1$. It would be a norm for $0 < p \leqslant 1$, only, when the opposite conclusion would hold: Greater measure value better concentration.

P_2: If $P_x(n,k)$ additionally assumes oscillatory zero-mean values (cross-terms) within a region which does not overlap with its basic values (auto-terms) then its measure $\mathscr{M}[P_x]$ will be increased. This will mean worse concentration measure, for $p > 1$. For norms, $p < 1$, it will go in the same direction as the concentration improvement, thus the normalizations would be necessary [12,15].

P_3: If $P_x(n,k)$ is spread over two separate regions with N_1 and N_2 points having values $1/(2N_1)$ and $1/(2N_2)$, respectively (two-component signal with the same energy of components), with a very different concentration $N_1 \gg N_2$, then $\mathscr{M}[P_x] = \frac{1}{2} N_1^{p-1}[1 + (N_2/N_1)^{(p-1)/p}]^p$. For a large order norm [8,12,15], when $0 < p \ll 1$, the measure

$\mathscr{M}[P_x]$ is dominated with the component-N_1 concentration and it is almost not sensitive to the component-N_2 concentration. However, for $p > 1$, the measure proposed here, produces theoretically $(N_2/N_1)^{(p-1)/p} \to N_2/N_1$, and both components contribute with almost equal weights.

In reality, there is no sharp edge between $P_x(n,k) \neq 0$ and $P_x(n,k) = 0$, so the value of (2) could, for very large p, be sensitive to small values of $P_x(n,k)$. The robustness may be achieved by using lower-order forms, for example, with $p = 2$ or 4.

3. Applications and examples

1. Consider the spectrogram

$$\text{SPEC}_w(n,k) = \frac{1}{E}|\text{STFT}(n,k)|^2,$$

where $\text{STFT}(n,k) = \text{DFT}_m\{w(m)x(n+m)\}$ is the short-time Fourier transform; 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 proposed measure with $p = 2$, will be that which minimizes

$$w^+ = \arg \min_w \left(\sum_{n=1}^{N} \sum_{k=1}^{N} |\text{SPEC}_w(n,k)|^{1/2} \right)^2. \quad (3)$$

Let us illustrate this by an example. Consider the signal

$$x_1(t) = \cos(50\cos(\pi t) + 10\pi t^2 + 70\pi t)$$
$$+ \cos(25\pi t^2 + 180\pi t), \quad (4)$$

sampled at $T = \frac{1}{256}$. The Hanning window $w(m)$ with different lengths is used in the spectrogram calculation. The results are shown in Fig. 1. For wide lag windows, signal nonstationarity makes the spectrogram very spread in the time–frequency plane, having relatively large measure $\mathscr{M}[P_x] = M_2^2$ (Fig. 1(a) and (b)). For narrow lag window its Fourier transform is very wide, causing spread distributions and large M_2^2 (Fig. 1(d) and (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
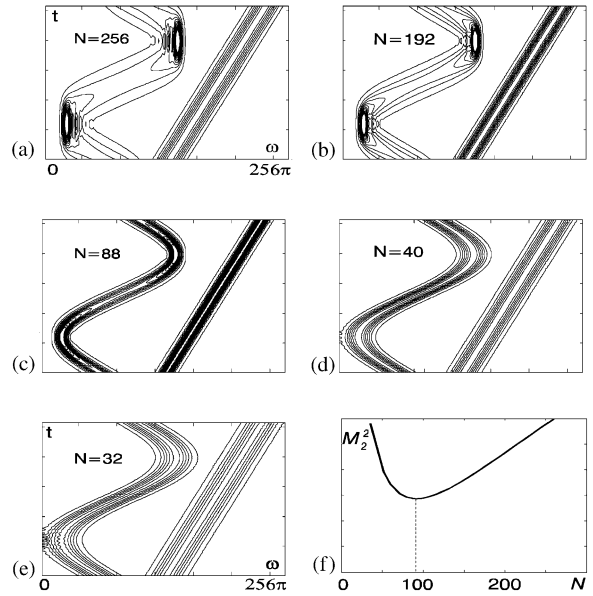

Fig. 1. Spectrogram for various window lengths and its measure M_2^2. The lowest M_2^2 is achieved for $N = 88$, being the best window length choice according to this measure.

for a set of spectrograms with $N = 32$ up to 256 window length, Fig. 1(f). The minimal 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).

2. The same procedure will be used for the best frequency window length L determination in the S-method (SM), whose discrete form [16] can be written using recursive relation

$$\text{SM}_L(n,k)$$
$$= \text{SM}_{L-1}(n,k)$$
$$+ 2\text{Re}\{\text{STFT}(n,k+L)\text{STFT}^*(n,k-L)\} \quad (5)$$

with $\text{SM}_0(n,k) = \text{SPEC}(n,k)$. Distributions $\text{SM}_L(n,k)$ should be properly scaled in order to satisfy unbiased energy condition.

As it is known, the SM relates to the most important time–frequency representations: the spectrogram and the pseudo Wigner distribution (WD). The optimal distribution $\text{SM}^+(n,k)$, on our way from $L = 0$ (the spectrogram) toward $L = N/2$ (the WD), is the one calculated with L producing the
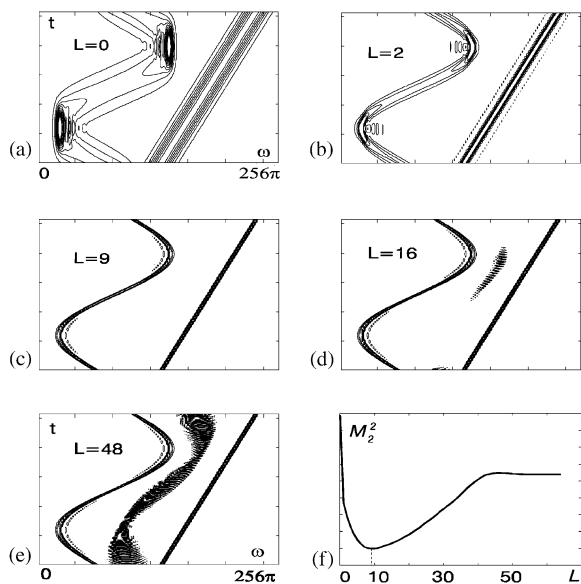
Fig. 2. The SM for various values of parameter L and its measure M_2^2. The lowest M_2^2 is achieved for $L = 9$, being the best choice according to this measure. Note that $L = 9$ corresponds to the analog cutoff frequency of $0.07\omega_m$, where ω_m is maximal frequency in the time–frequency plane.

minimal value of $\mathscr{M}[P_x]$,

$$L^+ = \arg \min_L \left(\sum_{n=1}^{N} \sum_{k=1}^{N} |SM_L(n,k)|^{1/2} \right)^2. \qquad (6)$$

Here, instead of $|SM_L(n,k)|$ a nonnegative part of $SM_L(n,k)$ could, and will be, used in examples. The same signal is used for the illustration of the SM. Since this method is a WD-based method, 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. 2(a). By increasing of L, the SM improves concentration of the spectrogram toward the WD quality, meaning lower measure M_2^2, Fig. 2(b) and (c). After L has reached the value equal to the distance between the auto-terms, cross-terms start to appear, increasing M_2^2, Fig. 2(d) and (e). Minimal M_2^2 means a trade-off between the auto-terms concentration and the cross-terms appearances, Fig. 2(f). The SM with L corresponding to minimal M_2^2 is shown in Fig. 2(c).

The proposed measure is illustrated on the signal whose components intersect:

$$x_2(t) = \cos(30\pi t^2 + 128\pi t) + \cos(-30\pi t^2 + 128\pi t). \qquad (7)$$
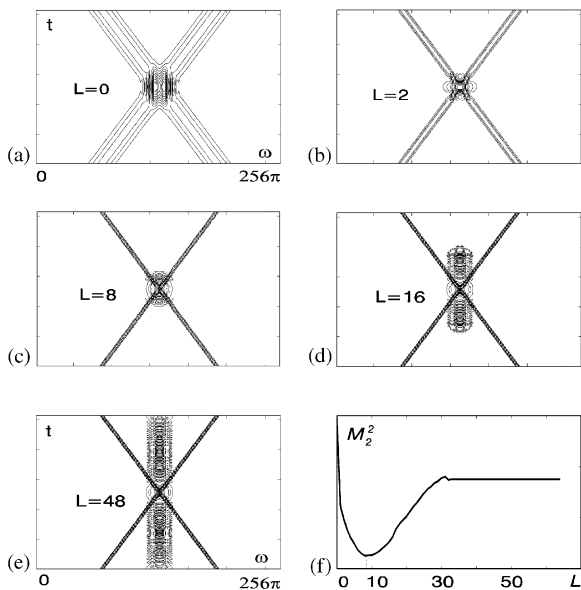

Fig. 3. The SM for various values of parameter L and its measure M_2^2. The lowest M_2^2 is achieved for $L = 8$, being the best choice according to this measure.

Here, there are not separate regions for the auto-terms concentration and the cross-terms appearance. The proposed measure has to decide when the auto-terms are concentrated enough and the cross-terms starts to dominate, Fig. 3.

Finally, the measure is illustrated on time–frequency analysis of the real data pressure signal in the BMW engine with speed 2000[rot/min], Fig. 4.

The proposed measure can be generalized for the on-line recursive time–frequency distribution realizations, [1,5,16], by applying it over frequency only, for a given time instant, $m_p^p(n) = \left(\sum_{k=1}^{N} |P_x(n,k)|^{1/p} \right)^p$. The same properties as for M_p^p hold, but $m_p^p(n)$ results in time-varying "optimal" parameters, that can additionally improve time–frequency representation in highly nonstationary signal cases.

4. Optimization

Parameters optimization may be done by a straightforward computation of a distribution measure $\mathscr{M}[P_x]$ for different parameter values. The
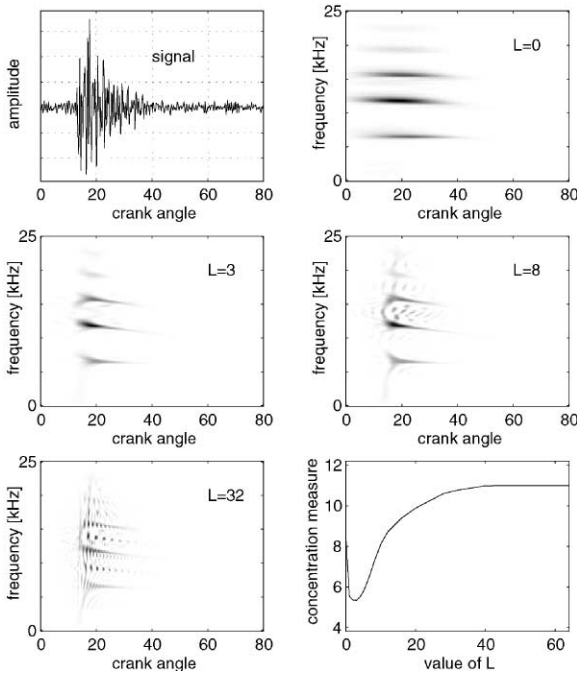
Fig. 4. Concentration measure illustration on time–frequency analysis of a car engine pressure signal. Signal, and its S-method based time–frequency representations are given. Time is rescaled into corresponding crank-angle. The best choice according to this measure was $L = 3$.

best choice according to this criterion (optimal distribution with respect to this measure) is the distribution that produces the minimal value of $\mathscr{M}[P_x]$. 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 descend approach, can be used. It will be described in this section.

The gradient of $\mathscr{M}[P_x]$ with respect to a distribution's generalized parameter, denoted by ξ, is

$$\frac{\partial \mathscr{M}[P_x]}{\partial \xi} = \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |P_x(t,\omega)|^{1/p} \right)^{p-1}$$
$$\times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |P_x(t,\omega)|^{(1-p)/p} \Psi(t,\omega) \, \mathrm{d}t \, \mathrm{d}\omega,$$
$$(8)$$

where

$$\Psi(t,\omega) = \frac{\partial |P_x(t,\omega)|}{\partial \xi} = \mathrm{Re} \left\{ \frac{P_x(t,\omega)}{|P_x(t,\omega)|} \frac{\partial P_x(t,\omega)}{\partial \xi} \right\}. \quad (9)$$

The continuous forms, rather than the discrete ones, are given since the derivatives, that will be used, are well-defined only for this kind of variables. Let us consider the factor $\Psi(t,\omega)$ in the cases of spectrogram and S-method.

In the spectrogram the optimization parameter ξ is the window width $\xi = T$. For the rectangular window of the length T, we get

$$\Psi(t,\omega) = \frac{\partial |\mathrm{STFT}_x(t,\omega)|^2}{\partial \xi}$$
$$= 2\mathrm{Re} \left\{ \mathrm{STFT}_x(t,\omega) \frac{\partial \mathrm{STFT}_x(t,\omega)}{\partial T} \right\}$$
$$= \mathrm{Re}\{\mathrm{STFT}_x(t,\omega)[x(t+T/2)\mathrm{e}^{-\mathrm{j}\omega T/2}$$
$$+ x(t-T/2)\mathrm{e}^{\mathrm{j}\omega T/2}]\}. \quad (10)$$

For the S-method, the optimization parameter is the frequency-domain window width $\xi = W_P$. Here we get

$$\Psi(t,\omega) = \frac{\partial |\mathrm{SM}_x(t,\omega)|}{\partial \xi}$$
$$= 2\mathrm{Re} \left\{ \mathrm{sign}(\mathrm{SM}_x(t,\omega)) \frac{\partial \mathrm{SM}_x(t,\omega)}{\partial W_P} \right\}$$
$$= \mathrm{sign}(\mathrm{SM}_x(t,\omega))$$
$$\times \mathrm{Re}\{\mathrm{STFT}_x(t,\omega + W_P/2)$$
$$\times \mathrm{STFT}_x(t,\omega + W_P/2)\}. \quad (11)$$

Discrete forms of these functions can easily be obtained, just by sampling the distributions and the function $\Psi(t,\omega)$. The iterations, starting from the very low-concentrated distribution toward the maximally concentrated one, i.e., toward the measure minimum, can be done according to

$$\xi_{m+1} = \xi_m - \mu \frac{\partial \mathscr{M}[P_x]}{\partial \xi}, \quad (12)$$

where μ is the step, which should be chosen in the same way as the step in the other adaptive algorithms. It should not be too small, since the convergence would be too slow, and should not be

Table 1
Window width in spectrogram during the iterations

Iteration $m =$	0	1	2	3	4	5
For $x_2(t)$, $\xi_m =$	16	20	68	82	84	86
For $x_1(t)$, $\xi_m =$	16	20	76	90		

too large to miss the minimum, or cause the divergence.

In discrete implementations, the gradient $\partial \mathcal{M}[P_x]/\partial \xi$ can also very efficiently be approximated based on $\mathcal{M}[P_x]$ calculated with the considered ξ_m and $\mathcal{M}[P_x]$ calculated with the previous parameter value ξ_{m-1},

$$\xi_{m+1} = \xi_m - \mu \frac{\mathcal{M}[P_x; \xi_m] - \mathcal{M}[P_x; \xi_{m-1}]}{\xi_m - \xi_{m-1}}, \quad (13)$$

where $\mathcal{M}[P_x; \xi_m]$ means a measure of distribution P_x calculated by using parameter value ξ_m.

Example. The optimization procedure will be illustrated on the signals $x_1(t)$ and $x_2(t)$ from the previous section. The spectrogram will be used in illustrations, since in the S-method, the convergence with respect to L is already fast, and the minimum is anyway reached in few iterations. The optimal window length in spectrogram, with respect to this measure, is obtained in few iterations by using (13), starting from the very narrow window. The results for $\xi_m \equiv N$, as a function of iteration number, are given in Table 1. Note that $N = 16$ and 20 in the initial and first iteration, are assumed. The next value of $\xi_{m+1} \equiv N$ is calculated according to (13). The algorithm is stopped 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 $\mu = \frac{1}{3}$ has been used in all examples.

5. Review of the existing measures and their comparison with the proposed measure

Here we will briefly review the existing measures of time–frequency distributions concentration. For each of them, we will point out a drawback that lead us to consider and propose another form of measure for time–frequency distributions concentration.

1. *Ratio of norms based measures.* Jones and Parks proposed [12] the fourth power of the L_4 norm of time–frequency distribution divided by the L_2 norm:

$$M_{JP} = \sum_n \sum_k P_x^4(n, k) \bigg/ \left(\sum_n \sum_k P_x^2(n, k) \right)^2.$$

This norm is similar to "kurtosis" in statistics. They have also concluded that other measures based on the ratio of norms L_p/L_q, $p > q$ would behave in a similar manner. The main property of this measure is that "the fourth power in the numerator favors a peaky distribution", [12]. For multicomponent signals, this property means that if one component is extremely highly localized, with respect to other components of the same energy, the measure will not look for compromise. Jones and Parks proposed a localized application of this form, on smaller regions separately. This can improve the results and overcome this problem, but in a computationally very demanding way.

2. *Rényi entropy measures.* The second class of time–frequency distribution measures is defined in analogy to the Rényi entropy measure. It has been introduced in time–frequency analysis by Williams et al. [15,18], with a significant contribution of [8] in establishing the properties of this measure. Rényi entropy measure applied on time–frequency distribution $P_x(n, k)$ has the form

$$R_\alpha = \frac{1}{1 - \alpha} \log_2 \left(\sum_n \sum_k P_x^\alpha(n, k) \right) \quad (14)$$

with $\alpha \geqslant 2$ being values recommended for time–frequency distribution measures [8]. For the case $\alpha = 2$ (distribution energy) oscillatory cross-terms would increase the energy leading to false conclusion that the concentration improves. In combination with some constraints on the time–frequency distribution kernel, the distribution energy was used in an interesting way, by Baraniuk and Jones [3], in order to derive a signal-dependent time–frequency representation. The case with $\alpha = 3$ fails to detect the existence of oscillatory zero-mean

cross-terms (which do not overlap with auto-terms), since for odd α they do not contribute to this measure. These were the reasons for the introduction of normalized Rényi entropy measures, that will be described next. It is interesting to note that the *Shannon entropy*,

$$H = -\sum_n \sum_k P_x(n,k) \log_2(P_x(n,k)),$$

could be recovered from the Rényi entropy measure, from the limit case when $\alpha \to 1$, [8]. However, the Shannon entropy could not be used for general time–frequency distributions $P_x(n,k)$, since they can assume negative values [8].

3. *Normalized Rényi entropy measures*. In order to avoid the problem caused by the fact that the Rényi entropy measure with $\alpha = 3$ does not detect zero-mean cross-terms, as well as to reduce a distribution to the unity signal energy case, some kind of normalization must be done [15]. The normalization can be done in various ways, leading to a variety of possible measure definition. In [8] two normalization schemes of the Rényi entropy are proposed:

(a) *Normalization with the signal energy* produces

$$\mathrm{RE}_\alpha = \frac{1}{1-\alpha} \log_2 \left(\frac{\sum_n \sum_k P_x^\alpha(n,k)}{\sum_n \sum_k P_x(n,k)} \right) \quad \text{with } \alpha \geqslant 2.$$

$$(15)$$

Behavior of this measure is quite similar to the nonnormalized measure form, except in its magnitude. This kind of normalization is important for comparison of various distributions, or the same distribution when it is not energy unbiased, by definition.

(b) *Normalization with the distribution volume*,

$$\mathrm{RV}_3 = -\frac{1}{2} \log_2 \left(\frac{\sum_n \sum_k P_x^3(n,k)}{\sum_n \sum_k |P_x(n,k)|} \right). \qquad (16)$$

If the distribution contains oscillatory values, then summing them in absolute value means that large cross-terms will decrease measure RV_α. This indicates smaller concentration, due to cross-terms appearance. The volume-normalized form of measure has been used for adaptive kernel design in [15]. Note that the term within logarithm is just the ratio of norms L_3 and L_1, while the logarithm is

a monotone function. Thus, measure (16) can be considered as L_3/L_1, reducing to the general case in 1. Therefore, it will have the same problem with components having very different concentration.

4. *Uncertainty principle form in time–frequency analysis*. A direct extension of some classical signal analysis definitions to joint time–frequency domain can be done by using marginal properties. They read, in analog domain,

$$|x(t)|^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} P_x(t,\omega)\, \mathrm{d}\omega,$$

$$|X(\omega)|^2 = \int_{-\infty}^{\infty} P_x(t,\omega)\, \mathrm{d}t. \qquad (17)$$

Thus, the classical signal analysis results, which are linear in $|x(t)|^2$ or $|X(\omega)|^2$, can easily be extended to time–frequency analysis. For example, the classical signal analysis uncertainty principle reads

$$\left(\int_{-\infty}^{\infty} t^2 |x(t)|^2\, \mathrm{d}t \right) \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^2 |X(\omega)|^2\, \mathrm{d}t \right) \geqslant 1/2.$$

By a simple substitution of $|x(t)|^2$ and $|X(\omega)|^2$, according to (17), it becomes

$$\left(\frac{1}{2\pi} \int_{-\infty}^{\infty} (t - t_{g|\omega})^2 P_x(t,\omega)\, \mathrm{d}\omega\, \mathrm{d}t \right)$$

$$\times \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} (\omega - \omega_{i|t})^2 |X(\omega)|^2\, \mathrm{d}t \right) \geqslant 1/2. \qquad (18)$$

It has been assumed that the signals are of energy unity, while $t_{g|\omega}$ and $\omega_{i|t}$ are the group delay and instantaneous frequency. The first integral represents the signal local duration, while the second integral defines local (instantaneous) bandwidth $\sigma_{\omega|t}^2$. These forms have been considered in detail by Cohen [4,5].

Cohen has used the definitions of instantaneous bandwidth and local time duration in order to define the optimal window width in the spectrogram [5]. He showed that the optimal width is inversely proportional to the instantaneous frequency derivative [5]. However, this result is not of a significant practical importance, since it can be used for monocomponent signals (when the instantaneous frequency is well defined), as well as when the derivative of the instantaneous frequency is

a priori known. Another more practical approach for window width optimization in the spectrogram of monocomponent noisy signals, which does not require that the instantaneous frequency is known in advance, is presented in [13].

In the same way, like in the case of uncertainty principle, we can try to generalize other one-dimensional classical signal analysis definitions, or other definitions from the information theory.

The procedure that can be used for generalization is:

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

(ii) Assume that time–frequency distribution $P_x(t,\omega)$ 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 Rényi entropy measures, in time–frequency, have been proposed and used in various problems.

Example. Let us consider the classic *Leipnik entropy measure* [14]. By using *Zakai's entropy parameter* $\delta_t = -\int_{-\infty}^{\infty} |x(t)|^2 \ln|x(t)|^2 \, dt$ of signal $x(t)$ [19], and its frequency-domain equivalent δ_f, Leipnik proved that for the entropy measures of uncertainty hold: $\delta_t + \delta_f \geqslant -\ln(1/2)$. According to the presented procedure for constructing a time–frequency form, based on a classical signal processing relation, we get

$$
\delta_t = -\int_{-\infty}^{\infty} |s(t)|^2 \ln|s(t)|^2 \, dt
$$
$$
\rightarrow -\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_x(t,\omega) \ln P_x(t,\omega) \, dt \, d\omega.
$$
(19)

This is exactly the well-known *Shannon entropy*. It has already been discussed in [8] with respect to its (non) applicability in time–frequency problems. In a similar way, a logarithm of the general signal duration (uncertainty) form, proposed by Zakai [19]

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

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

$$
\rightarrow \frac{1}{1-\alpha} \log_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_x^\alpha(t,\omega) \, dt \frac{d\omega}{2\pi} \equiv R_\alpha,
$$

where $|s(t)|^2$ has been replaced with $P_x(t,\omega)$, and the unity energy of signals $(1/2\pi)\int_{-\infty}^{\infty} P_x(t,\omega) \, dt \, d\omega = 1$ is assumed.

Note: The Zakai's signal duration has been defined in signal analysis *earlier* than the Rényi entropy measure in information theory. The Zakai entropy parameter is contained in the general Zakai duration, as a limit for $\alpha \rightarrow 1$. This is the expected result, since the same relation holds between the Rényi and the Shannon entropy.

Important remark: Since a general time–frequency distribution cannot satisfy both $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_x(t,\omega) = 1$ and $P_x(t,\omega) \geqslant 0$ the obtained measures of time–frequency distribution concentration will just *formally look like the original* entropies or classical signal analysis forms. However, they do not have the same properties nor the same conclusions and results derived in classical signal analysis, information theory, or thermodynamics, hold in time–frequency any more. Even calling the obtained forms "entropies" can mislead a reader. Note that other possible transformations could be based on substitution of the probability density function by the distribution's absolute value $|P_x(t,\omega)|$ or by its nonnegative part. The positivity will be preserved here, but not the unity energy condition. Finally, we could combine the previous transformation schemes, within the same relation (like, for example, in the volume normalized "Rényi entropy measures"), in order to get a usable measure for time–frequency distributions concentration.

The full analogy and transformation of classical results into possible corresponding joint time–frequency forms is beyond the scope of this, or any

other single paper. Here, the aim is to provide one more very simple measure for time–frequency distributions concentration, that can produce better results than *the already existing measures of time–frequency distributions concentration*, in some practically important cases.

Numerical comparison: We will now demonstrate how the proposed measure can overcome some of the drawbacks of other measures in time–frequency analysis, pointed out in the short review within this section.

Various measures of concentration are illustrated on two simplified examples, Fig. 5. The first example, presented in Fig. 5 (rows 1–4), corresponds to a time–frequency representation of a two-component linear frequency modulated signal. Measures based on the normalized norms, Rényi information, normalized Rényi information, and the one proposed in this paper, are given. From this table, we can conclude that all, except the third-order Rényi information based measure, would select as the best, the distribution corresponding to the Wigner distribution without cross-terms (row 2).

The same measures are used in the second simplified example, Fig. 5 (rows 5–7). This case corresponds to a two-component signal when the distribution may be adjusted to one of the components (for example, spectrogram applied on a signal whose one component has a constant frequency and the other component is linear frequency modulated). All measures, except the one proposed in this paper, selected the distribution when one component is extremely highly concentrated, while the concentration of the other component is very low (row 5). The measure proposed in this paper selected the distribution when both components are well concentrated (row 6).

Very similar conclusions can be made from the second example, Fig. 6. The normalized measures for signal $x_1(t)$, given by (4), are presented in Fig. 6a. All measures behave in a similar manner, indicating that the spectrogram optimal window width is around $N = 88$. However, when the second component is a pure sinusoid, then the other norms do not look for a compromise. All of them indicate the widest possible window as the optimal one, Fig. 6e. It produces high concentration of sinusoidal com-

Fig. 5. Simplified distributions and their measures. *Columns*: I – Simplified distributions; II – Signal energy; III – Squared norm two; IV – Normalized norm four (Jones–Parks); V – Rényi information with $\alpha = 3$; VI – Volume Normalized Rényi information $\alpha = 3$; VII – Proposed measure. Large-shaded numbers are for selected distributions by using corresponding measure. *Rows*: 1 – Wigner distribution; 2 – Wigner distribution without cross-terms; 3 – Spectrogram of two linear FM signals; 4 – A reduced interference distribution (RID); 5 – Spectrogram of a sum of sinusoid and linear FM signal, using a very wide window; 6 – Spectrogram of a sum of sinusoid and linear FM signal, using a compromised window producing "good" concentration of both components; 7 – A distribution adjusted to the second signal component. The following notations are used: $L_\alpha = \sum_n \sum_k P_x^\alpha(n,k)$, $A_\alpha = \sum_n \sum_k P_x(n,k)|^\alpha$.

ponent, ignoring the concentration of the other component, Fig. 6(g). The proposed measure chooses a compromised window width for both components, Figs. 6(e) and (f).

6. Conclusion

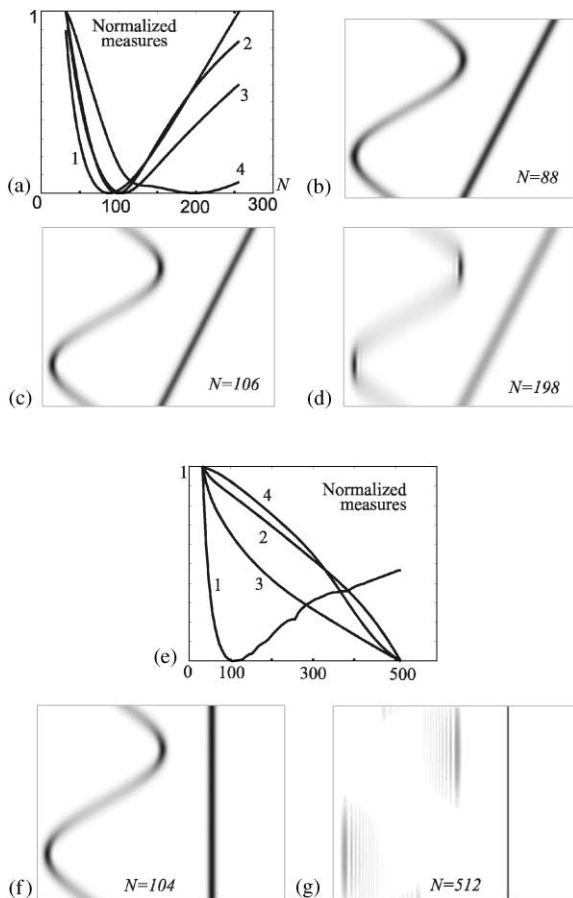A very simple criterion that can provide an objective measure for time–frequency distributions

Fig. 6. Comparison of the time–frequency distribution measures. (a) Normalized measure values: 1 – Proposed measure, 2 – Norm two L_2^2, 3 – Rényi measure with $\alpha = 3$, 4 – Park–Jones ratio of norms L_4/L_2^2. (b) Spectrogram with $N = 88$ being optimal with respect to the proposed measure, (c) Spectrogram with $N = 106$ being optimal with respect to the normalized Rényi measure, (d) Spectrogram with $N = 198$ being optimal with respect to the Parks–Jones measure, (e) Normalized measures as in (a) for the second component being pure sinusoid, (f) Spectrogram with $N = 104$ being optimal with respect to the proposed measure, (g) Spectrogram with $N = 512$ being optimal with respect to the all other measures.

concentration is presented. It has been used in automatic determination of some time–frequency distributions parameters. A review of the existing measures of time–frequency distributions concentration, and their comparison with the proposed measure is given.

References

[1] M.G. Amin, Recursive kernels for time–frequency signal representations, IEEE Signal Process. Lett. 3 (1) (January 1996) 16–18.

[2] R.G. Baraniuk, D.L. Jones, A signal dependent time–frequency representation: Optimal kernel design, IEEE Trans. Signal Process. 41 (4) (April 1993) 1589–1602.

[3] R.G. Baraniuk, D.L. Jones, Signal-dependent time–frequency analysis using radially-Gaussian kernel, IEEE Trans. Signal Process. 41 (3) (1993) 263–284.

[4] L. Cohen, Distributions concentrated along the instantaneous frequency, SPIE, Adv. Signal Process. Algebra Arch. Imp. 1348 (1990) 149–157.

[5] L. Cohen, Time–frequency Signal Analysis, McGraw Hill, New York, 1995.

[6] R.N. Czerwinski, D.L. Jones, Adaptive cone kernel time frequency analysis, IEEE Trans. Signal Process. 43 (7) (July 1995) 1715–1719.

[7] R.N. Czerwinski, D.L. Jones, Adaptive short-time Fourier analysis, IEEE Signal Process. Lett. 4 (2) (February 1997) 42–45.

[8] P. Flandrin, R.G. Baraniuk, O. Michel, Time-frequency complexity and information, Proceedings of the ICASSP, vol. 3, 1994, pp. 329–332.

[9] D. Gabor, Theory of communications, J. Inst. Electri. Eng. 93 (1946) 429–441.

[10] A.J.E.M. Janssen, On the locus and spread of pseudo functions in the time–frequency plane, Philips J. Res. 37 (3) (1982) 79–110.

[11] D.L. Jones, R.G. Baraniuk, An adaptive optimal-kernel time–frequency representation, IEEE Trans. Signal Process. 43 (10) (October 1995) 2361–2372.

[12] D.L. Jones, T.W. Parks, A high resolution data-adaptive time–frequency representation, IEEE Trans. Signal Process. 38 (12) (December 1990) 2127–2135.

[13] V. Katkovnik, L.J. Stanković, Periodogram with varying and data-driven window length, Signal Processing 67 (3) (1998) 345–358.

[14] R. Leipnik, Entropy and uncertainty, Part II, Inform. and Control 3 (1960) 18–25.

[15] T.H. Sang, W.J. Williams, Rényi information and signal dependent optimal kernel design. Proceedings of the ICASSP, vol. 2, 1995, pp. 997–1000.

[16] L.J. Stanković, A method for time–frequency analysis, IEEE Trans. Signal Process. 42 (January 1994) 225–229.

[17] D. Vakman, Optimum signals which minimize partial volume under an ambiguity surface, Radio Eng. Electron. Phys. 23 (August 1968) 1260–1268.

[18] W.J. Williams, M.L. Brown, A.O. Hero, Uncertainity, information and time–frequency distributions, SPIE Adv. Signal Process. Algebra Arch. Imp. 1566 (1991) 144–156.

[19] M. Zakai, A class of definitions of duration (or uncertainty) and the associated uncertainty relations, Inform. and Control 3 (2) (1960) 101–115.