TIME-FREQUENCY COMPLEXITY AND INFORMATION

Patrick Flandrin,[□] *Richard G. Baraniuk,*[◇] *and Olivier Michel*[□] *

[□] Laboratoire de Physique (URA 1325 CNRS)
Ecole Normale Supérieure de Lyon
46 allée d'Italie, 69364 Lyon Cedex 07, France

[◇] Department of Electrical and Computer Engineering
Rice University
P.O. Box 1982, Houston, TX 77251–1892, USA

ABSTRACT

Many functions have been proposed for estimating signal information content and complexity on the time-frequency plane, including moment-based measures such as the time-bandwidth product and the Shannon and Rényi entropies. When applied to a time-frequency representation from Cohen's class, the Rényi entropy conforms closely to the visually based notion of complexity that we use when inspecting time-frequency images. A detailed discussion reveals many of the desirable properties of the Rényi information measure for both deterministic and random signals.

1. INTRODUCTION

The term *component* is ubiquitous in the literature on joint time-frequency representations (TFRs) [1]. For example, one talks of suppression of Wigner distribution (WD) cross-components, concentration and resolution of auto-components, and the property that TFRs separate signal components such as parallel chirps that overlap in both time and frequency. Very often the quality of a particular TFR is judged based on subjective criteria related to the components of the signal being analyzed. Intuitively, a component is a concentration of energy in the time-frequency plane, but this notion is difficult to translate into a quantitative concept. In fact, the concept of a signal component has never been — and may never be — clearly defined.

In this paper, rather than address the question "what is a component?" directly, we will investigate several quantitative measures of signal *complexity* and *information* on the time-frequency plane. While they do not yield direct answers regarding the locations and shapes of components, these measures are intimately related to the concept of a signal component, the connection being the intuitively reasonable assumption that signals of high complexity (and therefore high information content) must be constructed from large numbers of elementary components. Viable measures of complexity and information include the time-bandwidth product and other moments on the time-frequency plane, measures of information appropriated from probability theory by Williams *et al.* [2], and parametric techniques based on decompositions into elementary building blocks as introduced by Orr *et al.* [3].

After defining Cohen's class of TFRs in the next section, we introduce and discuss these measures in Section 3. Section 4 focuses on the many attractive properties of a measure particularly suited to time-frequency analysis, the Rényi information. Extensions, as well as several possible applications of these measures, are sketched in the Conclusions.

*This work was supported by URA 1325 CNRS and an NSERC-NATO postdoctoral fellowship.

2. TIME-FREQUENCY REPRESENTATIONS

We will study the complexity and information content of signals indirectly via Cohen's class of TFRs, a set of quadratic distributions that attempt to indicate the energy content of a signal s as a function of both time t and frequency f. A TFR $C_s(t,f)$ from Cohen's class can be expressed as[1] [1]

$$C_s(t,f) = \iiint s^*\left(u-\frac{\tau}{2}\right) s\left(u+\frac{\tau}{2}\right) \Phi(\theta,\tau) \cdot e^{j2\pi(\theta u - \theta t - \tau f)}\, du\, d\theta\, d\tau, \qquad (1)$$

where the function $\Phi(\theta,\tau)$ is called the *kernel* of the TFR. Examples of Cohen's class TFRs include the WD ($\Phi(\theta,\tau) = 1$), the spectrogram (Φ = ambiguity function of the time-reversed window function), and the exponential distribution ($\Phi(\theta,\tau) = e^{-\theta^2\tau^2/\beta}$) [1].

The kernel completely determines the properties of its corresponding TFR. For example, a fixed-kernel TFR possesses the energy preservation property

$$\iint C_s(t,f)\, dt\, df = \int |s(t)|^2\, dt \qquad (2)$$

and the marginal properties

$$\int C_s(t,f)\, df = |s(t)|^2, \qquad \int C_s(t,f)\, dt = |S(f)|^2 \qquad (3)$$

provided $\Phi(\theta,0) = \Phi(0,\tau) = 1\ \forall\, \theta,\tau$. (The function $S(f)$ denotes the Fourier transform of the signal $s(t)$.) We will assume throughout this paper that the signal energy is normalized to one, that is, $\int |s(t)|^2 dt = 1$.

The formulas (2) and (3) evoke an analogy between a TFR and the probability density function (PDF) of a two-dimensional random variable. This parallel has been exploited with much success in the past [1,2]. In fact, as we will see, most measures on the time-frequency plane have been borrowed directly from probability theory. However, there are two key points at which this analogy breaks down. First, because of the freedom of choice of kernel function, the TFR of a given signal is nonunique. Second, most Cohen's class TFRs are nonpositive and, therefore, cannot be interpreted strictly as densities of signal energy.[2] Nevertheless, concepts from probability theory still have considerable merit in time-frequency analysis, provided caution is exercised in their interpretation.

[1] All integrals run from $-\infty$ to ∞.

[2] While there do exist classes of positive TFRs that satisfy (2) and (3) [4], we will consider only quadratic Cohen's class TFRs in this paper.

3. INFORMATION AND COMPLEXITY MEASURES

Classical moment-based measures of signal complexity include the time-bandwidth product (TBP) and its generalization, the second-order moment in the time-frequency plane [1]. While the the physical significance of these quantities cannot be denied, they do not truly measure signal complexity or information content [2,3]. To demonstrate, consider the signal $s_1(t) + s_2(t+T)$ constructed from two components of compact support, and note that each of the above measures increases without bound with the separation distance T. However, once s_1 and s_2 are disjoint in time, there is clearly no increase in signal complexity or information content when they are separated further.

More promising as measures of time-frequency complexity are measures of information borrowed from probability theory, as proposed by Williams *et al.* [2]. The celebrated Shannon entropy

$$H(C_s) \;=\; -\iint C_s(t,f)\,\log_2 C_s(t,f)\,dt\,df, \qquad (4)$$

applied here to a normalized TFR $C_s(t,f)$, belongs to the class of Rényi entropy measures [5]

$$R_\alpha(C_s) \;=\; \frac{1}{1-\alpha}\log_2 \iint C_s^\alpha(t,f)\,dt\,df, \qquad (5)$$

parameterized by $\alpha > 0$. (The Shannon entropy is recovered as the limit of R_α as $\alpha \to 1$.) As the passage from Shannon to Rényi entropy involves only the relaxation of the mean value property of entropy from an arithmetic to an exponential mean [5], R_α behaves much like H. In particular, these functionals can be interpreted as inverse measures of concentration or "peakiness," since by analogy to probability theory, the outcomes of random experiments governed by concentrated PDFs are relatively certain and, hence, yield little information. When applied to a TFR, we will refer to R_α as the Rényi time-frequency information measure. We will study this measure in greater detail in Section 4.

A different approach to signal complexity estimation has been proposed by Orr. In [3], the complexity of a signal with respect to a given discrete basis is defined as essentially the Shannon entropy of the basis expansion coefficients. Since the resulting complexity value varies with the choice of basis, it is necessary to carry out a minimization over all "nice" bases to obtain the true estimate of signal complexity. Use of the TFRs of Cohen's class rather than the Gabor expansions considered in [3] circumvents to some extent the problem of optimizing over different bases, but there remains the choice of particular TFR. Fortunately, we will see in the next section that the properties of the Rényi entropy of a TFR appear relatively insensitive to the particular choice of representation.

4. RÉNYI INFORMATION

Choice of order α. While the Shannon entropy can be employed as a time-frequency plane information measure, it is appropriate only for positive Cohen's class TFRs (spectrograms and smoothed spectrograms), all of which have large bias. The Shannon entropy of other TFRs is not defined, in general, since the locally negative values of these representations play havoc with the logarithm in (4).

This is not the case with the Rényi entropy for integer orders $\alpha \geq 2$, all of which are real valued (noninteger orders yield complex-valued results). While for even α this fact is obvious, for odd α we need the following new result:

Theorem 1 *Let $C_s(t,f)$ be a TFR with kernel Φ such that $\Phi \in L_\infty$ and $\Phi(0,0) > 0$. Then for all integers $\alpha \geq 1$, we have*[3]

$$\iint C_s^\alpha(t,f)\,dt\,df \;\geq\; 0. \qquad (6)$$

For the purpose of measuring information content, the first of the values $\alpha \geq 2$ (that is, $\alpha = 2$) is easily ruled out, since for the important WD we have the property that $\iint W_s^2(t,f)\,dt\,df = 1$ [6], and thus $R_2(W_s) = 0$ for all signals. The next possible choice, $\alpha = 3$, is the first to yield a well-defined, useful information measure [2]. Along with all Rényi entropies of odd orders, it possesses the following asymptotic invariance to cross-components [1,6] in the time-frequency plane [9].

Theorem 2 *If the auto-components of a TFR $C_s(t,f)$ are separated in the time-frequency plane such that they do not overlap with any cross-components, then as the auto-component separation distance increases, we have*

$$\iint_{\mathcal{X}} C_s^\alpha(t,f)\,dt\,df \;\to\; 0 \qquad (7)$$

for all odd $\alpha \geq 1$, where $\mathcal{X}$ denotes the region containing the cross-components.[4]

Counting property. For signals satisfying the separation conditions of Theorem 2, the TFR $C_s(t,f)$ is "quasi-linear," and therefore each auto-component contributes separately to the overall $R_\alpha(C_s)$ information value. In this case, the similarity to composite PDFs of statistically independent events suggests that we should expect an additive or counting behavior from $R_\alpha(C_s)$.

Theorem 3 *Let $s(t)$ and $z(t) = s(t) + s(t+T)$ be signals with T chosen such that the conditions of Theorem 2 hold. Then for odd $\alpha \geq 3$, we have $R_\alpha(C_z) = R_\alpha(C_s) + 1$.*[5]

As an example of this property [2], consider the $R_3(W_s)$ information of the signal $g(t) + g(t+T)$, with g a lowpass Gaussian pulse. This information is plotted in Fig. 1 versus the separation distance T in units of Δ_t (the RMS time duration) for $g(t)$. (At $T = 0$, the two pulses coincide and therefore, because of the assumed energy renormalization, have

[3] *Proof:* We prove only the WD case ($\Phi = 1$); other cases are straightforward generalizations. Assume that α is an odd integer (the proof for even α is immediate). Then certainly we have $\iint W(t,f)\,W^\alpha(t,f)\,dt\,df \geq 0$, and so W^α is a Wigner weight symbol [7]. Therefore, $\iint W^\alpha(t,f)\,dt\,df \geq 0$ if $W^\alpha \in L_1 \cap L_2$ [7], which is easily shown to be the case if the signal $s \in L_2$ [8].

[4] *Sketch of proof:* For simplicity, assume that the signal consists of just two components with compact supports T_1 and T_2 separated by a distance T (the generalization to bounds on the decay of (7) for arbitrary signals in L_2 is straightforward). Let $A'(\theta,\tau)$ denote the ambiguity function (the 2-d Fourier transform of the WD [1]) of the cross-components of $C_s(t,f)$. It is also compactly supported along the τ (delay) axis in the ranges $\tau \in \pm[|T|, |T|+T_1+T_2]$. Note that the integral (7) equals the value of the α-fold 2-d convolution $[(A'\Phi) * (A'\Phi) * \ldots * (A'\Phi)](0,0)$, where Φ is the kernel of $C_s(t,f)$. Performing this convolution yields the constraint $T > \frac{\alpha-1}{2}(T_1+T_2)$ on the separation distance that must be satisfied for (7) to hold. Note that this requirement grows linearly with the Renyi order parameter, justifying a preference for 3rd-order Renyi information over other, larger orders.

[5] The result follows directly from Theorem 2 and the quasi-linearity of $C_s(t,f)$ for well-separated auto-components.

the same information content as a solitary pulse.) The TBP of the signal is also plotted. It is clear from the figure that, unlike the TBP, which grows without bound with T, the information measure saturates exactly one bit above the value $R_3(W_g) = -0.208$.[6] Similar results hold for three separated copies of $g(t)$ ($\log_2 3$ bits information gain), four copies (2 bits information gain), and so on.

This counting property of the Rényi information suggests a concept of *Rényi dimension* D_α for the WD, defined as

$$\log_2 D_\alpha(W_s) = R_\alpha(W_s) - R_\alpha(W_g). \tag{8}$$

This dimension indicates — relative to a "basis" of Gaussian building blocks — the number of blocks required to "cover" the WD of a given signal. Similar measures can be derived for other TFRs and other elementary building blocks.[7]

Phase sensitivity. The results of Fig. 1 are very appealing, but are also incomplete and unrealistic, because no modulation or phase differences were introduced between the two signal components. Figure 2 illustrates a more complete set of curves of the $R_3(W_s)$ information for the signal $g(t)\cos(\pi t/6) + g(t+T)\cos(\pi(t+T)/6 + \psi)$. Each curve corresponds to a different relative phase angle ψ between 0 and π rad. It is apparent from the curves that while phase changes do not affect the saturation levels of the information measure, they allow many possible trajectories between the two levels, including even trajectories where an "overestimation" [2] of information content occurs.

To interpret these results, note that as we decrease T, the auto- and cross-components of the signal begin to overlap in the time-frequency plane so that Theorem 2 no longer holds. At this point, relative phase plays a key role in determining information content.

In fact, the sensitivity of the $R_3(W_s)$ measure to phase is quite reasonable, given the sensitivity of closely spaced signals to relative phase. For example, Fig. 3 shows the composite signals and their respective WDs for a fixed offset T and relative phases $\psi = 0$ and $\psi = \frac{\pi}{2}$ rad. The difference in appearance is striking — clearly the components in the signal and WD at top are more separated than those on the bottom. Accordingly, the $R_3(W_s)$ informations for the two signals are very different: 1.31 and 0.31, respectively.

Effects of smoothing. Since relative phase information is carried by the cross-components of the WD, it seems reasonable that smoothing the WD (choosing kernels other than $\Phi = 1$) would lessen the effect of relative phase on information estimates. Figure 4 shows this to be the case, by repeating the same experiment as in Fig. 2, but with a matched window spectrogram TFR rather than the WD. While the spectrogram R_3 information estimate remains phase sensitive, it climbs more swiftly to the saturation level and with a reduced overshoot than $R_3(W_s)$. The price paid for this improved performance is a signal-dependent bias of information levels compared to those estimated using the Wigner distribution.

It is important to note that some smoothing is crucial for accurate information estimates for complicated multicomponent signals with overlapping auto- and cross-components.

[6] Readers should not be alarmed by negative Rényi entropy values. Even the Shannon entropy takes on negative values for certain distributions in the continuous-variable case.

[7] Although, it can be shown using the results of [8] that Gaussian signals minimize R_α, and so are the most basic signals in terms of this complexity measure.

Random signals. Time-frequency information estimates are also very useful for random signals; however, care must be taken not to confuse the Rényi information of the *TFR* of a random signal with the Rényi information of the *PDF* of the signal.

The $R_3(W_s)$ time-frequency information estimate provides an alternative to the signal-to-noise ratio (SNR) for signals embedded in additive noise. For example, Fig. 5 illustrates the relationship between the two for a single Gaussian pulse in white Gaussian noise. Interestingly, the sigmoidal characteristic of the information measure behaves more like our eyes and ears than the SNR: for high SNRs, it indicates that there is virtually only signal present, whereas for low (negative) SNRs, it indicates that there is virtually only noise present. Furthermore, the 0 dB SNR point (the point of equal signal and noise energies) occurs roughly midway between the two information extremes.

5. CONCLUSIONS

Unlike the TBP and other moment-based measures, the Rényi entropy measure shows great promise for estimating the complexity of signals via the time-frequency plane. Possible applications include adaptive transforms that minimize the complexity of the TFR, contrast measures for signals in noise, and information-theoretic distance/dissimilarity measures between different TFRs.

This paper has not addressed the question of choosing the appropriate TFR from Cohen's class for the information calculation. In fact, other classes of TFRs, such as the positive distributions [4] (which would allow the unrestricted use of the Shannon entropy) and the affine class of time-scale distributions [6] (which contains the squared magnitude of the continuous wavelet transform) could prove more appropriate for certain classes of signals.

REFERENCES

1. L. Cohen, "Time-Frequency Distributions — A Review," *Proc. IEEE*, Vol. 77, No. 7, 1989.
2. W. Williams, M. Brown, and A. Hero, "Uncertainty, Information, and Time-Frequency Distributions," *Proc. SPIE 1566*, 1991.
3. R. Orr, "Dimensionality of Signal Sets," *Proc. SPIE 1565*, 1991; J. Sweeney *et al.*, "Experiments in Dimensionally-Optimum Gabor Representations," *Proc. ICASSP 93.*
4. L. Cohen and T. Posch, "Positive Time-Frequency Distribution Functions," *IEEE Trans. Acoust., Speech, Sig. Proc.*, Vol. 33, pp. 31–38, 1985.
5. A. Rényi , "On Measures of Entropy and Information," *4th Berkeley Symp. Math., Stat., Prob.*, Vol. 1.
6. P. Flandrin, *Temps-Fréquence*, Hermès, Paris, 1993.
7. A.J.E.M. Janssen, "Positivity Properties of Phase-Plane Distribution Functions," *J. Math. Phys.*, **25** (7), pp. 2240–2252, 1984.
8. E. Lieb, "Integral Bounds for Radar Ambiguity Functions and Wigner Distributions," *J. Math. Phys.*, **31** (3), pp. 594–599, March 1990.
9. R. Baraniuk, P. Flandrin, and O. Michel, "Information and Complexity on the Time-Frequency Plane," *Quatorzième Colloque GRETSI*, Juan-Les-Pins, France, Sept. 1993, pp. 359–362.