TIME-FREQUENCY BASED DISTANCE AND DIVERGENCE MEASURES

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

[□] 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
Houston, TX 77251–1892, USA

ABSTRACT

A study of the phase and amplitude sensitivity of the recently proposed Rényi time-frequency information measure leads to the introduction of a new "Jensen-like" divergence measure. While this quantity promises to be a useful indicator of the distance between two time-frequency distributions, it is limited to the analysis of positive definite TFDs. In spite of this rather severe limitation, this measure could prove useful for time-frequency based detection. We illustrate with an example of detecting a signal in additive noise.

1. TIME-FREQUENCY INFORMATION MEASURES

Information and entropy functionals have shed new light on the question "what is a signal component" by providing quantitative measures of signal complexity in the time-frequency plane [1, 2, 3]. Their theoretical basis relies on the formal analogy between the time-frequency distributions (TFDs) of Cohen's class and bidimensional probability density functions. In particular, a large class of TFDs possesses the following marginal characteristics (for unit energy signals):

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

$$\iint C_s(t,f)\,dt\,df \;=\; 1.$$

The probabilistic interpretation of a TFD suggests the Shannon entropy

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

as a natural candidate for a time-frequency information measure. Unfortunately, however, most Cohen's class TFDs take on negative values, prohibiting its application.

Recent research has concentrated on alternative information measures, in particular the class of Rényi entropies [4]

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

parameterized by $\alpha > 0$ [1, 2, 3]. The Shannon entropy belongs to this class, in the limit as $\alpha \to 1$. Unlike the Shannon entropy, however, the Rényi measures are defined for virtually all Cohen's class TFDs for all integer $\alpha \geq 2$ [2, 3]. The primary property of the Rényi entropies studied thus far has been the *component counting property:* as two identical, overlapping components are separated in the time-frequency plane, the measure $H_\alpha(C_s)$ increases by 1 bit (for α odd). As an example of this property [1, 2, 3], consider the $H_3(W_s)$ information of the signal $s(t) = g(t)+g(t+T)$, with g a lowpass Gaussian pulse and W_s the Wigner distribution. This information is plotted in Fig. 1(a) versus the separation distance T (in units of the RMS time duration of $g(t)$). (At $T = 0$, the two pulses coincide and therefore, because of the assumed energy normalization, have the same information content as a solitary pulse.) The time-bandwidth product (TBP) of the signal is also plotted with a dashed line. 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 $H_3(W_g) = -0.208$.[1] 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.

2. PHASE SENSITIVITY

The results of Fig. 1(a) are very appealing, but also incomplete and unrealistic, because no phase differences were introduced between the two signal components. Figure 1(b) illustrates a more complete set of curves of the $H_3(W_s)$ information for the signal $g(t)\cos(\pi t/6)+g(t+T)\cos(\pi(t+T)/6+\varphi)$. Each curve corresponds to a different relative phase angle φ 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" [1] of information content occurs.

The phase sensitivity of the Rényi information measure stems from cross-terms in the Wigner distribution and can be studied by developing an expansion for $H_\alpha(W_s)$ with s a simple two-component signal. Our signal model is

$$s(t) \;=\; g(t) \;+\; (\mathcal{D}_{d,\varphi}\, g)(t),$$

where $g(t)$ is a Gaussian logon of the form

$$g(t) \;=\; \frac{1}{(\pi\sigma^2)^{\frac{1}{4}}}\, e^{-\frac{t^2}{2\sigma^2}}$$

and

$$(\mathcal{D}_{d,\varphi}\, g)(t) \;\stackrel{\text{def}}{=}\; e^{j(2\pi t\Delta f+\varphi)}\, g(t-\Delta t)$$

denotes the same logon phase-shifted by angle φ and translated in time-frequency by the normalized distance

$$d^2 \;=\; (2\pi\Delta f\,\sigma)^2 \;+\; \frac{\Delta t^2}{\sigma^2}, \tag{2}$$

*This work was supported by URA 1325 CNRS, the National Science Foundation, grant no. MIP-9457438, and the Texas Advanced Technology Program, grant no. TX-ATP 003604-002.

[1] 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.

with Δt and Δf the separation distances in time and frequency.

The Wigner distribution of s can be expressed as sum of two auto-terms and a cross-term

$$W_s(t,f) \;=\; W_g(t,f) \;+\; W_g(t-\Delta t, f-\Delta f) \;+\; I(t,f),$$

with $I(t,f) = 2\,\mathrm{Re}\, W_{g,\mathcal{D}_{d,\varphi} g}$ corresponding to the cross-Wigner distribution between g and $\mathcal{D}_{d,\varphi} g$. Denoting the order α Rényi entropy of the Wigner distribution of $s(t)/\|s\|$ as $H_\alpha(W_g, d)$, we have

$$\begin{aligned} H_\alpha(W_g,d) &= H_\alpha(W_g,0) \;+\; 1 \;+\; \delta(d,\varphi,\alpha) \\ &\approx H_\alpha(W_g) \;+\; 1 \;+\; \frac{1}{(1-\alpha)\log_e(2)} \\ &\quad \cdot \left[c(\alpha)\,\varepsilon \;-\; \frac{1}{2} c^2(\alpha)\,\varepsilon^2 \;+\; o(\varepsilon^5) \right] \quad (3) \end{aligned}$$

where $\phi = \varphi + 2\pi\Delta f\,(t+\Delta t/2)$, $c(\alpha) = \frac{2^{(\alpha-1)H}}{\alpha} \binom{\alpha}{\frac{\alpha-1}{2}} \cos(\phi)$, $H = H_\alpha(W_g)$, and $\varepsilon = e^{-\frac{d}{4\alpha}}$. This expansion holds under the hypothesis of well-separated logons, that is, assuming that (1) there is no substantial overlap between the cross-term and the auto-terms, and (2) $\varepsilon \ll 1$

Equation (3) quantifies the effects of separation d, phase φ, and order α on the Rényi information measure:

1. As $d \to \infty$, the third term decays to zero, and we have the asymptotic, "counting" result. Clearly, then, the interesting action takes place for small d.
2. For small d, the phase difference φ, through ϕ, imposes an oscillatory structure on the entropy value. This behavior is clearly evident in Fig. 1(b).
3. The coefficient $\varepsilon = e^{-\frac{d}{4\alpha}}$ controlling the amplitude of the third term decays more slowly for larger values of α. As a consequence, larger values of α lead to larger departures from the asymptotic values of H_α. This property was to be expected; large values of α emphasize the larger values of the TFD,[2] which for the Wigner distribution occur on the cross-term.

The analysis leading to Equation (3) did not take into account the fact that for each separation and phase the Wigner distribution was normalized to unit energy — the energy $\|g + \mathcal{D}_{d,\varphi} g\|^2 = 2$ only asymptotically. Making a more exact calculation of this energy, we have

$$\begin{aligned} \log_2(\|g+\mathcal{D}_{d,\varphi} g\|^2) &= 1 \;+\; \log_2\left(1+\cos(\phi)\,e^{-\frac{d}{4}}\right) \\ &\approx 1 \;+\; \frac{\cos(\phi)\,e^{-\frac{d}{4}}}{\log_e(2)} \;+\; o(e^{-\frac{d}{4}}), \end{aligned}$$

with $o(e^{-\frac{d}{4}}) = o(\varepsilon^\alpha)$. Including the effects of normalization yields the following sharpened estimate

$$\begin{aligned} H_\alpha(W_g,d) &\approx H_\alpha \;+\; 1 \;+\; \frac{1}{(1-\alpha)\log_e(2)} \\ &\cdot \left[c(\alpha)\,\varepsilon \;-\; \frac{1}{2} c^2(\alpha)\,\varepsilon^2 \;+\; \alpha\cos(\phi)\,\varepsilon^\alpha \;+\; o(\varepsilon^5) \right]. \end{aligned}$$

Since relative phase information is carried by the cross-terms of the Wigner distribution, smoothing lessens the sensitivity of Rényi information estimates. To demonstrate the relevant variables, in Figure 2 we perform the same experiment as in Figure 1(b) for several different smoothed Wigner distributions. We employ two-dimensional Gaussian smoothing functions of varying TBPs w and consider the values $\alpha = 3, 5, 7$. Smoothing with $w = 0$ yields the Wigner distribution, plotted as the bottom curve in each plot. We can draw the following conclusions from the Figure: (1) smoothing decreases the sensitivity (excursions) of the information estimate, (2) smoothing increases the convergence rate of the information estimate, (3) larger values of α slow the convergence rate of the information estimate. The price paid for improved performance through smoothing is a signal-dependent bias of information levels compared to those estimated using the Wigner distribution. However, some amount of smoothing is crucial for accurate information estimates for complicated multicomponent signals with overlapping auto- and cross-terms.

3. AMPLITUDE SENSITIVITY

A similar analysis yields insight on the behavior of the Rényi measures for signal components of different amplitudes. For the signal

$$\tilde{s}(t) \;=\; g(t) \;+\; k\,(\mathcal{D}_{d,\varphi}\, g)(t),$$

with k a scaling factor, it is simple to show that in the limit as $d \to \infty$

$$H_\alpha(W_{\tilde{s}}) \;=\; H_\alpha(W_g) \;+\; H'_\alpha(p_1,p_2).$$

Here, $p_1 = \frac{1}{1+k^2}$ and $p_2 = \frac{k^2}{1+k^2}$ represent the relative energy levels between g and $\mathcal{D}_{d,\varphi}\, g$, and $H'_\alpha(p_1,p_2)$ is given by

$$H'_\alpha(p_1,p_2) \;=\; \frac{1}{1-\alpha}\,\log_2(p_1^\alpha + p_2^\alpha).$$

$H'_\alpha(p_1,p_2)$ corresponds to the order α Rényi entropy of the binary probability distribution $\{p_1,p_2\}$ [4].

4. DISTANCE AND DIVERGENCE MEASURES

The sensitivity of the Rényi entropy to phase and amplitude differences between components could limit its utility as a time-frequency analysis tool. As a better tempered alternative we now propose two time-frequency divergence functions that indicate the distance between different TFDs. These measures could prove useful, for example, as time-frequency detection statistics in applications comparing reference and data distributions.

A familiar way of deriving distance measures from entropy functionals is to form the *Jensen difference* [5]. The Jensen difference of two probability densities p and q is defined by

$$H\left(\frac{p+q}{2}\right) \;-\; \frac{H(p)+H(q)}{2}. \qquad (4)$$

Positivity of this quantity relies on the concavity of the entropy function H. Since the Rényi entropy H_α is neither concave nor convex for $\alpha \neq 1$, (4) makes sense for only the Shannon entropy. In light of this (somewhat disappointing) fact, we restrict ourselves to positive TFDs[3] for the remainder of the paper.

As noted in [2, 3], the Rényi entropy can be derived from the same set of axioms as the Shannon entropy, the only difference being the employment of a more general exponential mean rather than arithmetic mean [4]. This realization inspires the modification of (4) from an arithmetic to a geometric mean, and we have the following quantity (for two

[2]In fact, we have that $\lim_{\alpha\to\infty} H_\alpha(W_s) = \max_{(t,f)} W_s(t,f)$!

[3]Smoothed spectrograms, unless we walk out on Cohen's class.

smoothed spectrograms C_1 and C_2 and $\alpha \neq 1$):

$$J_1(C_1, C_2) \;=\; H_\alpha\big(\sqrt{C_1\,C_2}\,\big) \;-\; \frac{H_\alpha(C_1) + H_\alpha(C_2)}{2}, \qquad (5)$$

where $(\sqrt{C_1\,C_2}\,)(t,f) \stackrel{\text{def}}{=} \sqrt{C_1(t,f)\,C_2(t,f)}$. This quantity is obviously null when $C_1 = C_2$; we now show that J_1 is positive definite. The Cauchy-Schwartz inequality gives

$$\left| \iint [C_1(t,f)\,C_2(t,f)]^{\alpha/2}\,dt\,df \right|^2 \leq \iint C_1^\alpha(t,f)\,dt\,df \iint C_2^\alpha(t,f)\,dt\,df,$$

and, since the log function is monotonically increasing, we have for $\alpha > 1$

$$\frac{1}{1-\alpha}\log_2 \left| \iint [C_1(t,f)\,C_2(t,f)]^{\alpha/2}\,dt\,df \right|^2 \;\geq$$
$$\frac{1}{1-\alpha}\log_2 \iint C_1^\alpha(t,f)\,dt\,df + \frac{1}{1-\alpha}\log_2 \iint C_2^\alpha(t,f)\,dt\,df.$$

Thus, using (1) yields

$$H_\alpha\big(\sqrt{C_1\,C_2}\,\big) \;\geq\; \frac{H_\alpha(C_1) + H_\alpha(C_2)}{2},$$

and the result follows.

The J_1 divergence has the property that it diverges for disjoint TFDs such that $C_1(t,f)C_2(t,f) = 0\ \forall t,f$. To prevent this divergence of the divergence, we can consider also the modified form

$$J_2(C_1, C_2) \;=\; J_1\left(\frac{C_1 + C_2}{2}, C_1\right). \qquad (6)$$

The definite positiveness of J_2 follows directly from the same property for J_1, and $J_2 = 0$ if and only if $C_1 = C_2$. Unfortunately, however, J_2 is not a symmetric function of C_1 and C_2.

We now demonstrate the usefulness of these new divergences using a simple time-frequency detection experiment (see Figs. 3 and 4). Our goal is detecting a Gaussian logon embedded in additive white Gaussian noise. As a reference TFD, we employ the spectrogram of the logon at high SNR (30dB), computed using the logon itself as the window (matched-filter spectrogram). In the Figures, we compare the distances between this reference TFD and matched-filter spectrograms of the same logon submerged in varying levels of noise (SNR range: -30dB to 30dB). As distance measures, we took J_1, J_2, $(\alpha = 3)$ and the following Kullback divergences [5]:

$$K_{\rm d}(C_1, C_2) \;=\; \iint C_1(t,f)\,\log_2 \frac{C_1(t,f)}{C_2(t,f)}\,dt\,df,$$
$$K_{\rm i}(C_1, C_2) \;=\; -\iint C_2(t,f)\,\log_2 \frac{C_1(t,f)}{C_2(t,f)}\,dt\,df,$$
$$]K_{\rm div}(C_1, C_2) \;=\; K_{\rm d}(C_1, C_2) \;+\; K_{\rm i}(C_1, C_2).$$

All curves are plotted normalized by their maximum value, for increased readability. The new J divergences appear to yield the most efficient statistics for detecting the logon embedded in noise; in fact, their values stay quite small down to negative SNRs.

5. CONCLUSIONS

While the Rényi entropy has great potential for time-frequency applications, its phase and amplitude sensitivity must be taken into consideration. Our attempts at designing an inter-TFD distance measure avoiding these problems have met with some degree of success; however, our results hold only for the positive TFDs of Cohen's class (smoothed spectrograms). Extending our results to all of Cohen's class will require either a different approach for combining Rényi entropies or a completely different entropy base function.

REFERENCES

[1] W. J. Williams, M. L. Brown, and A. O. Hero, "Uncertainty, information, and time-frequency distributions," in *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 1566, pp. 144–156, 1991.

[2] R. G. Baraniuk, P. Flandrin, and O. Michel, "Information and complexity on the time-frequency plane," in *14ème Colloque GRETSI*, (Juan Les Pins, France), pp. 359–362, 1993.

[3] P. Flandrin, R. G. Baraniuk, and O. Michel, "Time-Frequency complexity and information," in *Proc. ICASSP '94*, vol. III, pp. 329–332, 1994.

[4] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Stat. and Prob.*, vol. 1, pp. 547–561, 1961.

[5] C. R. Rao and T. N. Nayak, "Cross-entropy dissimilarity measures and characterizations of quadratic entropy," *IEEE Trans. Inform. Theory*, vol. 31, pp. 589–593, Sept. 1985.

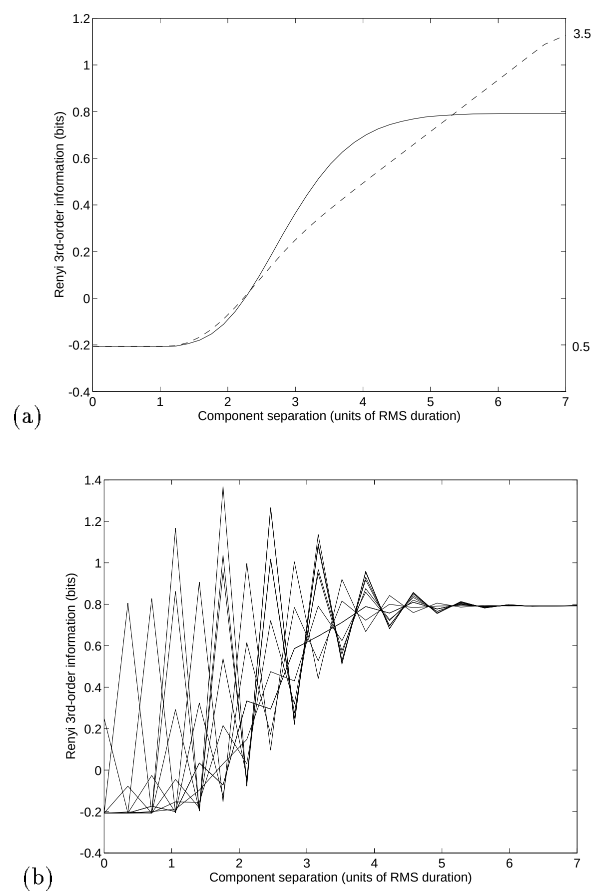
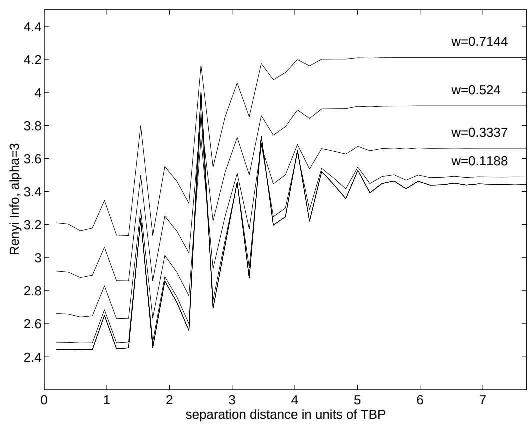
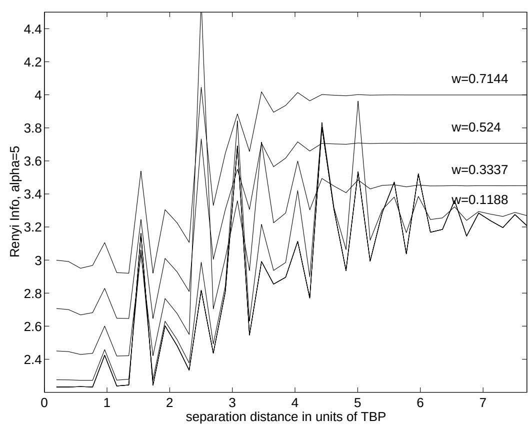


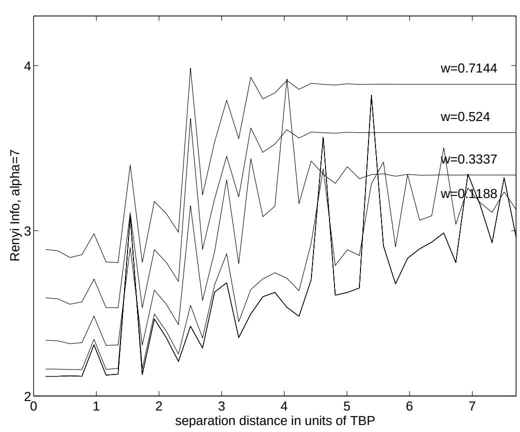
Figure 1: (a) $H_3(W_s)$ information of the Wigner distribution and TBP (dashed) vs. component separation. (b) $H_3(W_s)$ information vs. component separation, various relative phases.



($\alpha = 3$)



($\alpha = 5$)



($\alpha = 7$)

Figure 2: Rényi information of Gaussian-smoothed Wigner distributions of a signal composed of two Gabor logons. The spread of the smoothing kernel is expressed in TBP units w, normalized by the TBP of the logons composing the signal. The lowest curve gives the $H_\alpha(W_s)$ measure.

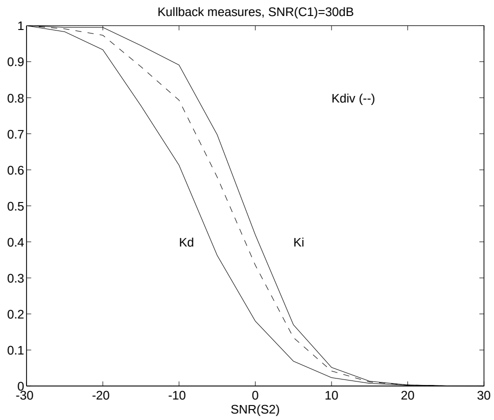


Figure 3: Kullback measures between a reference spectrogram of one logon (SNR=30dB) and the spectrogram of the same signal, for different SNRs.

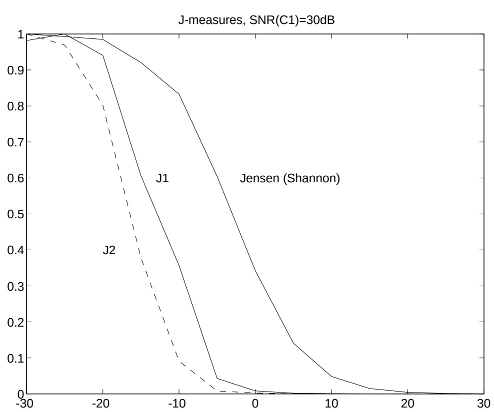


Figure 4: J-Measures (5) and (6) and Jensen distance (for Shannon entropy, $\alpha = 1$) between a reference spectrogram of one logon (SNR=30dB) and the spectrogram of the same signal for different SNRs.