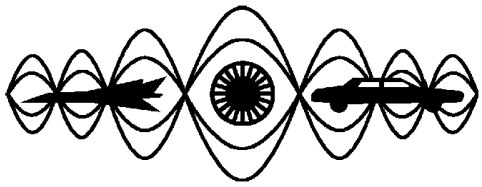
Chi-square law mixture for an unsupervised time-frequency pattern extraction

Nadine Martin, Cyril Hory, Mathieu Huchard



Tenth International Congress
on Sound and Vibration
7-10 July 2003 • Stockholm, Sweden

TENTH INTERNATIONAL CONGRESS ON SOUND AND VIBRATION, ICSV10

CHI-SQUARE LAW MIXTURE FOR AN UNSUPERVISED TIME-FREQUENCY PATTERN EXTRACTION

Nadine MARTIN, Cyril HORY* and Mathieu HUCHARD

LIS - Laboratoire des Images et des Signaux
BP 46 – F- 38402 Saint Martin d'Hères Cedex
nadine.martin@lis.inpg.fr

*Systems Division, Department of Electrical Engineering and Computer Science
The University of Michigan, 1301 Beal Avenue Ann Arbor, Mi. 48109

Abstract

Our purpose is to develop an unsupervised method of extraction and characterization of time-frequency (TF) patterns. The new idea is the notion of probability model in the time-frequency representation. Each TF point is considered as a random variable having a density distribution, which depends of its belonging class and of the TF estimator. We consider square modulus based estimates such as spectrogram and scalogram, which induces chi-square mixture densities. Density parameters are estimated by Maximum Likelihood. The classification is managed by a region-growing algorithm in a feature space built from moments evaluated in local cells. We here propose a new feature based on entropy we must consider in the continuous case. We derive the analytical expressions of the generalized Rényi entropy for a non-central chi-square distribution. We therefore can represent what we refer to as an entropy map where entropy is calculated from local histogram evaluated in gliding time-frequency cells. We present both feature space segmentation and entropy map on real signals: seismic signals (avalanches and seisms), acoustical signals (cavitation and concert halls), bioacoustical signals. Application of the proposed algorithm is also possible on other representation than time-frequency representation.

1 INTRODUCTION

This communication tackles the interpretation problem of time-frequency representations (TFR). Whatever the signal domain, TFR are 2D representations, which lead to intricate and critical interpretation. Lots of key issues arise. One of

them is why and how to decide if a time-frequency coefficient comes from noise only or has energy from signal.

Different approaches are proposed in the literature. Our approach is different in the sense that we do not want to consider *a priori* information. We think that the structure of TFR can be used to discriminate components having different behavior such as random signal from deterministic one. For this purpose, we propose a new processing, which could be applied after a TF analysis and right before a decision process. This interpretation process is general and can be of interest to all applications where a non-stationary analysis is necessary.

Analyzed signals are random due to the presence of noise. So, TF points are also random variables defined by their probability distribution. In the TFR, we are so able to define a probability model that we use for the purpose of classification. The observation set $\mathcal{L}_x$ associated to a time discrete signal $x[n]$ is here a TFR or a time-scale space. So, $\mathcal{L}_x$ is a subset of $\mathbb{R}^2$ such as for each element (n,k) of this set, k being the frequency variable, a TF coefficient $S_x[n,k]$ exists. It writes:

$$\mathcal{L}_x = \left\{(n,k) \in \mathbb{R}^2 \,/\, \exists\, S_x[n,k] \textit{ for a given } x[n]\right\} \tag{1}$$

Elements of $\mathcal{L}_x$ satisfy one of the two following assumptions:

-Hypothesis H_0: the signal is a stationary zero mean Gaussian noise $b[n]$ of unknown variance σ^2 to which matches a sub-set $\mathcal{L}_{H_0}$ defined by:

$$\mathcal{L}_{H_0} = \left\{(n,k) \in \mathcal{L}_x \;/\; x[n] = b[n]\right\} \tag{2}$$

-Hypothesis H_1: the signal is a non-stationary deterministic signal $d[n]$ in an additive noise $b[n]$ to which matches a subset $\mathcal{L}_{H_1}$, complementary of $\mathcal{L}_{H_0}$:

$$\mathcal{L}_{H_1} = \left\{(n,k) \in \mathcal{L}_x \;/\; x[n] = d[n] + b[n]\right\} \tag{3}$$

There is no restriction on $d[n]$, which can be composed of several non-connected components. These components could be wideband signals, transients, pure or modulated frequencies. This communication addresses the problem of determining the subsets $\mathcal{L}_{H_0}$ and $\mathcal{L}_{H_1}$ from the knowledge of $\mathcal{L}_x$ only. A solution is proposed without *a priori* knowledge on the nature of $d[n]$, on the number of their components and on their localization in the TFR.

We briefly introduce the probability model, a local mixture density (section 2). We focus on main features of this model that enable to discriminate TF patterns in TFR such as spectrogram or scalogram. The studied features are first local moments. The statistical behavior of these moments (section 3.1) is used to propose a complete segmentation process (section 3.2). As a complementary feature, we also begin to

work about the local entropy (section 4.1) and present here our first results in entropy map form (section 4.2). We illustrate both methods on real signals.

2 LOCAL MIXTURE DENSITY MODEL

$S_x[n,k]$ is the square modulus of the projection of x on a function base $\Phi_{n,k}$:

$$S_x[n,k] = \left| \left\langle x[m], \Phi_{n,k}[m] \right\rangle \right|^2 \tag{4}$$

If $\Phi_{n,k}$ is the complex exponential function centered on time n, at frequency k and smoothed by a N_w points time window w such as $\Phi_{n,k}[m] = w[m-n]e^{2i\pi\frac{mk}{N_w}}$, $S_x[n,k]$ is the spectrogram of x. If $\Phi_{n,k}$ is a wavelet computed from a mother wavelet ψ centered on n and scaled from a scale factor k_0/k such as $\Phi_{n,k}[m] = \frac{1}{k_0/k}\psi\left[\frac{m-n}{k_0/k}\right]$, $S_x[n,k]$ is the scalogram of x. In both cases, under hypothesis H_1, $S_x[n,k]$ is a non-central χ^2 distribution with r degrees of freedom, proportionality parameter α equal to σ^2/r and non-centrality parameter δ equal to $S_d[n,k]$, [HOR 02c,d], [WHA 71], [JOH 95]:

$$p_{S_x[n,k]|H_1}(s) = f_{\chi^2\left(r,\, \sigma^2/r,\, S_d[n,k]\right)}(s) \tag{5}$$

r is asymptotically equal to 2 with respect to the support of $\Phi_{n,k}$, the extreme frequency values excepted. Under hypothesis H_0, $S_x[n,k]$ does not contain deterministic signal energy. So, δ is null and has central χ^2 distribution:

$$p_{S_x[n,k]|H_0}(s) = f_{\chi^2\left(r,\, \sigma^2/r,\, 0\right)}(s) \tag{6}$$

Therefore, each point of the TFR is a mixture of central and non-central χ^2 distribution. Identifying the deterministic components turns out to estimate the non-centrality parameters δ. Identifying the noise alike turns out to estimate the proportionality parameter α.

In that communication, we consider a TF local approach [HOR 02c,d]. Let us define a cell C of N TF coefficients gliding over the whole TFR. In that case, the mixture distribution $p_C(s)$ of the parent random variable associated to C is defined from the total probability as:

$$p_c(s) = \frac{1}{N}\sum_{(n,k)\in C} f_{\chi^2\left(r,\, \sigma^2/r,\, S_d[n,k]\right)}(s) + (1-p)\, f_{\chi^2\left(r,\, \sigma^2/r,\, 0\right)}(s) \tag{7}$$

where p is the proportion of deterministic coefficients in cell C. In a local approach, r is different from 2 and is an unknown parameter and so are σ^2, $S_d[n,k]$ and p. The cell size is fixed as the time-frequency support of the reproducing kernel of the TF method [HOR 02c].

3 MOMENT FEATURE SPACE AND SEGMENTATION

3.1 Local TFR moments and Chi-2 laws

Our purpose is the estimation of the mixture density parameters. In a global approach, resolution by maximum likelihood turns out to be an incomplete data parameterization. EM algorithm is able to provide a sequence of iterates that are guaranteed to find a maximum of the likelihood function. In that case, the problem must be formulated for a density mixture involving non-central χ^2 distribution. A solution can be read in [HOR 03].

In a local approach, the mixture distribution is defined by (7). In order to reduce the number of unknown parameters, one solution consists of defining only one δ for each cell. So we define it as equal to the mean of every $S_d[n,k]$ in the cell. Therefore, equation (7) writes:

$$p_c(s) = p\, f_{\chi^2\left(r,\, \sigma^2/r,\, \delta_c\right)}(s) + (1-p)\, f_{\chi^2\left(r,\, \sigma^2/r,\, 0\right)}(s) \tag{8}$$

with $\delta_c = p\, N \sum_{(n,k)\in C} S_d[n,k]$. Others approximations could be considered [HOR 02b].

The qth moment to zero μ_q of this parent variable can be expressed as [HOR 02c]:

$$\mu_q(p, snr) = q!\left(\sigma^2\right)^q \left(1 + p\sum_{i=1}^{q} \mathrm{C}_q^i \frac{snr^q}{i!}\right) \tag{9}$$

where $snr = \delta_c / \sigma^2$ is a local signal to noise ratio in the cell C. Assuming ergodicity, moments μ_q can be estimated from statistics m_q:

$$m_q = \frac{1}{N} \sum_{(n,k)\in C} S_x[n,k]^q \tag{10}$$

m_1 has a central χ^2 distribution, $\chi^2\left(r_1, \frac{\sigma^2}{r_1}, 0\right)$. The ratio r_1/r is the apparent size of the cell, lower than N, and depending from the degree of TFR coefficient correlation, so from the data weighting window and the overlapping between windows. For q greater than 1, the distribution is not so explicit. Nevertheless, the dispersion of these estimated moments can be computed [HOR02b,cd]:

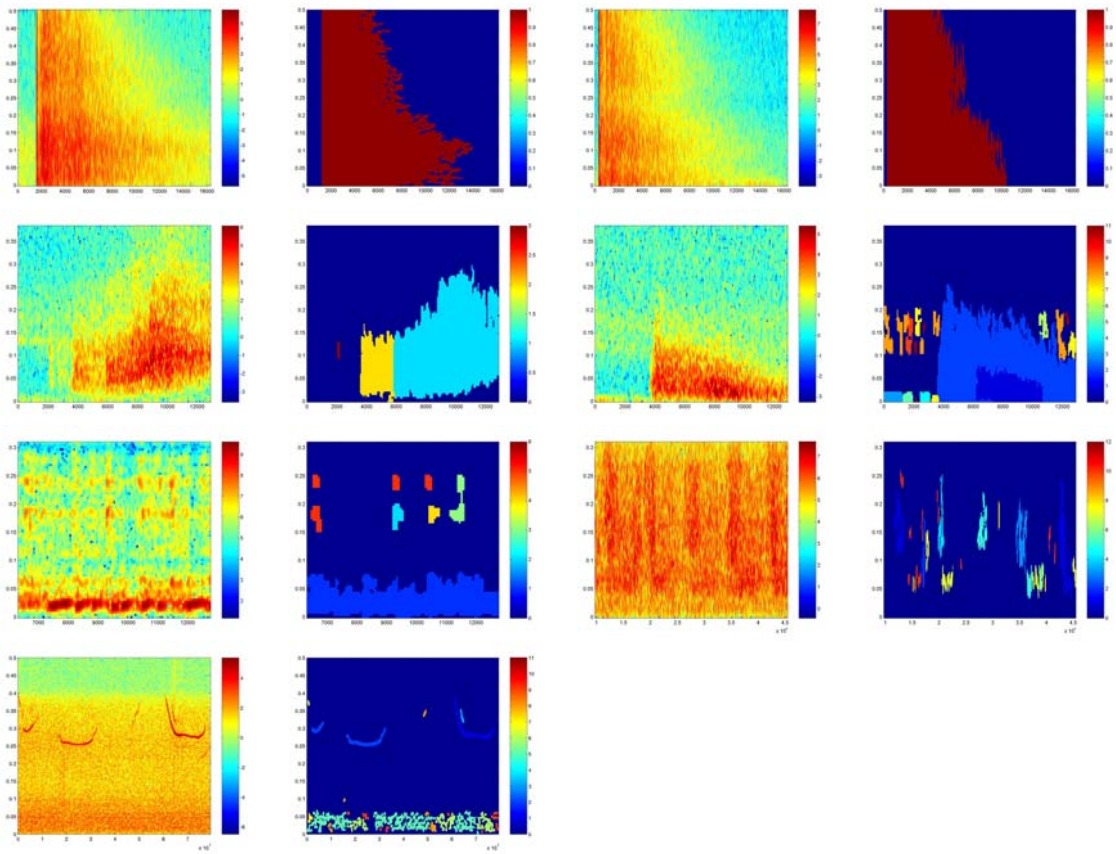


Figure 1*: TFR and FS co-segmentation. H-axis is time, V-axis is frequency. For each signal, on the left: spectrogram and on the right: segmentation result. First line: Opera "Bastille" (left) and concert hall 'Halle aux grains" (right) impulse response (sampling frequency 16 KHz), Kaiser Bessel 4 window with 63 pts, covering 50%, 256 frequency pts, cells 3×11,* $\lambda_b = 5.10^{-2}$ *. Second line: Avalanche (left) and seismic (right) signals (sampling frequency 100 Hz), Hanning window with 128 pts, covering 50%, 256 frequency pts, cells 3×7,* $\lambda_b = 10^{-2}$ *. Third line: Two types cavitation signal (sampling frequency 15 324 Hz for the 1st and 32 000 Hz for the 2nd), Kaiser Bessel 4 window with 128 pts, covering 50%, 256 frequency pts, cells 3×5,* $\lambda_b = 10^{-2}$ *. Fourth line: Bioacoustical signal (dolphin wistle), (sampling frequency 32 KHz), Hanning window with 511 pts, covering 50%, 1024 frequency pts, cells 3×7,* $\lambda_b = 10^{-2}$ *.*

$$E\left\{m_q\right\} = \mu_q\left(p, snr\right) \qquad Var\left\{m_q\right\} = \frac{r}{r_1}\left(\mu_{2q}\left(p, snr\right) - \mu_q^2\left(p, snr\right)\right)$$

$$Cov\left\{m_q, m_j\right\} = \frac{1}{N_{app}}\left(\mu_{q+j}\left(p, snr\right) - \mu_q\left(p, snr\right)\mu_j\left(p, snr\right)\right) \tag{11}$$

These local moments are features of each cell. Each cell can be transformed in a new space defined by these moments. The statistical properties given by (11) allow a description of this new space, referred to as Feature Space (FS). The FS dimension is the number of unknown parameters of the mixture model. Parameters σ^2, r and r_1

can be approximated by maximum likelihood estimators [HOR 02] and are not considered as parameters of the mixture distribution. The approximation defined in (8) leads to two unknowns, p and snr. So, in that case, FS is bi-dimensional (m_1, m_2) and presents clusters: one for the noise centered close to the origin at the point $(E\{m_1\} = \mu_1(0,0) = \sigma^2, E\{m_2\} = \mu_2(0,0) = \frac{r+2}{r}\sigma^4)$ and the other ones distributed according the values of p and snr of the analyzed signal in each cell.

3.2 Region growing algorithm and co-segmentation

The FS defined in the previous section shows off clusters that need to be segmented under the condition of belonging to connected points in the TFR. These connected points belong to the set $\mathcal{L}_{H_1}$ we want to estimate. The co-segmentation procedure based on a region growing algorithm takes into account the structure of the FS described by equations (11). A complete description of the algorithm can be read in [HOR 02c,d]. Beginning by points having maximum values for p and snr and so belonging to $\mathcal{L}_{H_1}$, the iterative process stops when the remaining points have a central χ^2 distribution, such as they belong to $\mathcal{L}_{H_0}$. The only control parameter that the user has to choose is a probability λ_b for the m_1 distribution such as its quantile defines the limit of the noise cluster in the FS. Figure 1 shows the extraction of TF patterns for real signals. Details of parameters are in the caption.

4 ENTROPY MAPS

4.1 Entropy and Chi-2 laws

As we want to extract information from TFR, it is natural to think about the notion of entropy. The entropy depends on the probability distribution only. Its main property is its behavior according to the size of the distribution support. The entropy is maximum for a uniform distribution whereas it decreases for a more localized one. The entropy can so be used for classifying distributions, which have different supports. It is the case in TFR where central and non-central χ^2 distributions discriminating the set $\mathcal{L}_{H_0}$ from the set $\mathcal{L}_{H_1}$. We again consider the local approach, which implies the distribution mixture (8).

Rényi entropy is a generalization of Shannon one [REN 61]. Given that the TFR distribution is continuous, we define Rényi entropy as:

$$H_\alpha(p_c) = \frac{1}{1-\alpha}\log_2 \int_0^1 p_c(s)^\alpha ds \qquad (12)$$

When α tends to 1, equation (12) tends to Shannon entropy. For the set $\mathcal{L}_{H_0}$ defined by distribution (6) and for $r = 2$, Rényi entropy writes:

$$H_\alpha\left(f_{\chi^2\left(2,\sigma^2/2,0\right)}\right)=\frac{\log_2\alpha}{\alpha-1}+\log_2\left(\sigma^2\right) \tag{13}$$

Rényi entropy increases with the noise variance, which is coherent since the distribution spreads out with the noise variance. For the set $\mathcal{L}_{H_1}$ defined by a non-central χ^2 distribution (5), the analytical expression is very complex, which also implies a complex analytical expression for the mixture distribution (8) [HUC 02]. For this reason, we here only comment the behavior of Rényi entropy and more particularly the choice of α. In figure 2, more α increases, more the difference between the minimum and the maximum of the entropy increases. This property can be useful for better classifying distributions. More α increases, more the concavity of the curve increases. This property can be of interest to initialize the segmentation method with cells of highest p, that is to say cells which are entirely inside a TF pattern. For α equal to 2, Rényi entropy varies linearly with p We also notice that all curves have a maximum for values of p close to 0.9. This property is useful to enhance the contours of a pattern right before detection.

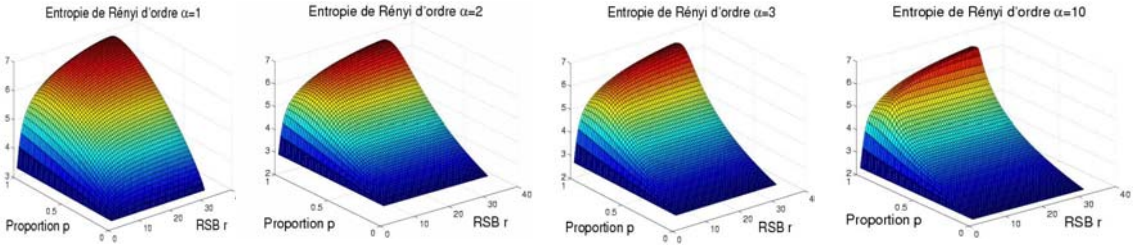


Figure 2: Rényi entropy versus the mixture proportion p and the snr for 4 values of α.

4.2 TFR Histograms with adapted bounds and entropy maps

The measure of entropy needs the measure of a distribution [SMI 01]. Equation (12) is get from a continuous density that we have to estimate from a discrete TFR. In each cell, the distribution p_c is estimated by the simple method of histogram, which provides finite values P_c. Let $\left[a_c,b_c\right]$ be the amplitude support and n_{bin} the interval number in $\left[a_c,b_c\right]$. Applying Riemann theorem and given that the considered functions are continuous, we can connect $H_\alpha\left(p_c\right)$ in (12) to $H_\alpha\left[P_c\right]$:

$$H_\alpha\left[P_c\right]=\frac{1}{1-\alpha}\log_2\sum_{i=1}^{n_{binc}}P_c[i]+\log_2\left(\frac{b_c-a_c}{n_{bin}}\right)\xrightarrow[n_{bin}\to+\infty]{}H_\alpha\left(p_c\right)=\frac{1}{1-\alpha}\log_2\int_0^1 p_c\left(s\right)^\alpha ds \tag{14}$$

We *a priori* choose to fix n_{bin} equal to $\sqrt{N}$. The choice of $\left[a_c,b_c\right]$ is more critical. If these amplitudes are chosen according to the extreme values of the global TFR, it has the advantage to compute distributions according the same bounds. But, in the case of a TFR, it leads to insignificant estimates of the distribution in cells where all the

values are small relative to the maximum of the representation. We so propose to compute histograms with adapted bounds in each cell such as $a_c = \min_c S_x[k,f]$ and $b_c = \max_c S_x[k,f]$. In this way, we get $\sqrt{N}$ values of $P_c[i]$ for each cell C. Afterwards, we can calculate the entropy associated to each cell from (14). The obtained 2D representation is referred to as entropy map. Figure 3 shows entropy maps for real signals. At this stage of the study, one important result is the improving of the entropy histogram related to the RTF histogram due to the contrast of entropy maps. We now are able to fix a threshold for computing a pattern contour. Works are in progress to introduce entropy as a feature for a co-segmentation.

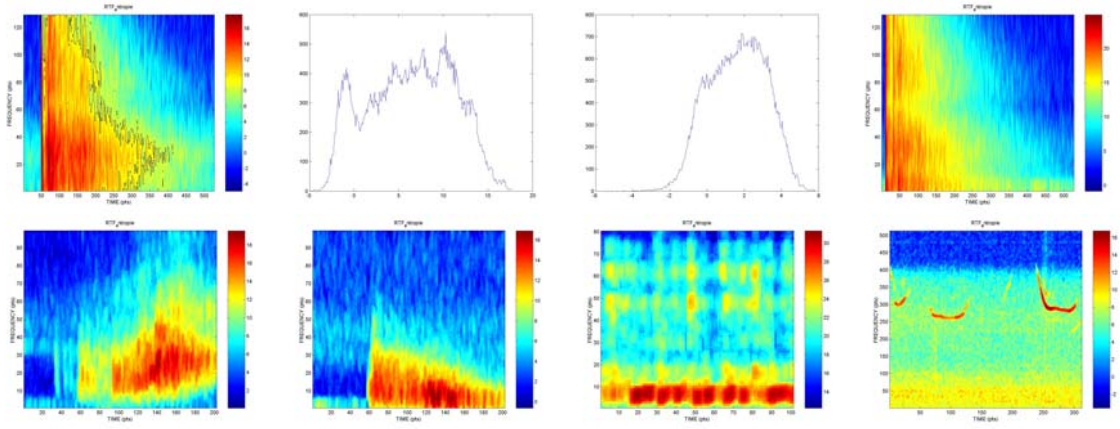


Figure 3: Entropy maps (α=10) for the signals of the figure 1. First line, entropy map of "Bastille" TFR, followed by its histogram and the RTF histogram,then the entropy map of "Halle aux grains". Second line, entropy maps of avalanche, seismic, 1st type cavitation and bioacoustical signals.

REFERENCES

[HOR 02a] C. Hory, N. Martin, Maximum Likelihood estimation for Spectrogram segmentation, ICASSP 02, Orlando, Florida, Vol. II, pp. 1581-1584, May 13-17 (2002)

[HOR 02b] C. Hory, N. Martin, Mixture Densities Formulation of a Spectrogram Segmentation Task, EUropean SIgnal Processing COnference, EUSIPCO 20002, Toulouse, France, September 3-6 (2002)

[HOR 02c] C. Hory, N. Martin, A. Chéhikian, Spectrogram segmentation by Means of Statistical Features for Non-Stationary Signal Interpretation, IEEE Trans. on SP, Vol. **50**, n°12,2915-2925 (2002)

[HOR 02d] C. Hory, Mélanges de distributions du chi-deux pour l'interprétation d'une représentation temps-fréquence, PhD Thesis, INPG, December 13 (2002)

[HOR 03] C. Hory, Formulation of the EM Algorithm for a Mixture of Central and Non-Central Chi2 Distributions, Submitted to Statistic and Probability Letters (2002)

[HUC 02] M. Huchard, Etude de l'entropie d'un mélange de lois du chi-deux. Application à la caractérisation du spectrogramme d'un signal bruité. Internal Report, DEA graduation, INPG (2002)

[JOH 95] N. L. Johnson, S. Kotz, N.Balakrishnan, *Continuous univariate Distributions*. (Wiley Series in Probability and Mathematical Statistics, Vol. 1 & 2, 1994)

[JOH 99] P.E. Johnson, D.G Long, The probability Density of Spectral Estimates Based on Modified Periodogram Averages, IEEE Transactions on Signal Processing, Vol. **47**, N°5 (1999)

[KOO 74] L. H. Koopman, *The spectral analysis of time series.* (Academic press, 1974)

[REN 61] A. Renyi, On measures of entropy and information, Hungarian Academy of Science (1961)

[SMI 01] J. D. H. Smith, Some observations on the concepts of information theoretic entropy and randomness, Entropy, Vol. **3**, pp. 1-11 (2001)

[WHA 71] A.D. Whalen, *Detection of Signals and Noise*. (Academic Press, 1971)