2008 International Symposium on Nonlinear Theory and its Applications
NOLTA'08, Budapest, Hungary, September 7-10, 2008

Complementary Aspects of Spectral and Entropic Measures of Time-series

M.R. Titchener†

†Institute of Bioengineering, University of Auckland
Uniservices House, 70 Symonds St., Auckland, New Zealand
Email: mark@tcode.auckland.ac.nz

Abstract—In times past the term *signal-analysis* was synonymous with *spectral analysis*. There is however, a growing appreciation that most natural phenomena is non-linear, capable of exhibiting variable, indeed unpredictable behaviors not able to be expressed meaningfully in harmonic terms. A relative newcomer to an evolving range of non-linear tools is the T-entropy, a computable measure of information content for strings that may be applied to suitably encoded time-series. This paper highlights firstly the complementary aspects of the FFT and T-entropy in the characterisation of time-series but also their respective and contrasting sensitivities. The simple illustrations provided here of time-series comprising a mixture of harmonic and non-stationary components, subject to sampling and coarse graining, demonstrate effects which complicate interpretation of particularly the T-entropy results.

1. Introduction

Most engineers and scientists have a well honed intuitive understanding of harmonic phenomena. The term entropy does not elicit the same degree of intuition. Many will correctly associate entropy with descriptions of *state*; the lower the entropy, the more predictable the state, the higher the entropy the more chaotic and less predictable the state. However, disorder and chaos defy easy visualisation so diverse are their possible manifestations. If presented with disorder, we seek out recognisable structures, the equivalent perhaps of performing a sort of mental FFT but on a non-stationary data set. More difficult again is the visualisation of *comparative differences* or *changes* in variability. To this end entropy, serves as an important characterisation for complex processes, albeit a largely unintuitive one.

There exist in the literature some forty or so definitions of entropy, some related, some not. Only a few of these have any practical potential. In this paper we confine ourselves to the T-entropy, a computable symbolic measure not dissimilar to Lempel and Ziv's complexity measure for finite strings [1]. For these measures the complexity of a finite string $x \in A^n$ is computed in terms of the number-of-steps required to construct it from its alphabet, A. While for the LZ76 algorithm, the constituent patterns p_i are taken to be an ordered catenation of prior patterns, identified in a linear parsing of the string, the T-entropy exploits recursively formed hierarchical-pattern structures of the form: $p_i = p_{i-1}^{m_{i-1}} \cdots p_1^{m_1} a_i$, where $a_i, p_1 \in A$. [6] proves that any $x \in A^n$ is expressible in the form : $x = p_{q-1}^{k_{q-1}} \cdots p_1^{k_1} a$, $a \in A$, subject to the aforementioned hierarchical constraints on p_i. The p_i may be systematically discovered by parsing x in a process that runs in $O(n \log n)$ time [5]. The T-complexity [6] of x, is defined by way of the production steps: $C_T = \sum_q (k_i + 1)$, and the T-entropy, denoted H_T, as a linearised rate: $H_T = \mathrm{li}^{-1}(C_T)/n$. Software implementations of the $O(n \log n)$ parsing algorithm are available for computing the T-entropy from strings of as long as 10's of millions of symbols. Dynamical systems, like the logistic map [4], have been used as generators of strings having known Kolmogorov-Sinai entropy (K-entropy), to demonstrate that the T-entropy at a practical level is a strongly indicative measure.

Whereas an FFT may be effective in deriving spectral components of a structured signal, even in the presence of noise, it is not effective in characterising the noisy or non-stationary aspects of a signal. As in the humorous description in which an *optimist* says of a glass of water that *it is half-full* and and the *pessimist*, that *it is half-empty*, the FFT and T-entropy may be seen as describing a given sequence in complementary terms. The FFT responds to structure, and the entropy to the "unstructure". When it comes to characterising chaotic, random, or non-stationary signals, an FFT presents as a blunt instrument, and yet we know now that chaotic signals may provide important indicators of state in non-linear complex phenomena. It is this that drives our interest in developing tools to sensitively quantify signal variability.

2. Encoding of Time-Series

The application of symbolic methods to continuous dynamical systems is now a fairly standard procedure, the term *symbolic dynamics* being introduced by Morse and Hedlund in the early 1930's. Symbolic methods are of increasing interest in the analysis of complex time-series including medical, seismic, and industrial time series. Coarse grained encoding of time-series, $[y_t]$ through finite partitioning and labeling of the signal space reduces the computational effort without necessarily compromising the sensitivity of the ensuing analysis. A bipartition may be applied to encode $[y_t]$ as a binary string. The choice of partition and alphabet size etc, generally revolves around maximising the symbolic-entropy, i.e., encapsulating the full dynamics. In the absence of knowing an optimum partition [3], entropy values may be computed over a range, for example sweeping a bipartition, b_ℓ ($\ell = 1, 2, 3\ldots$) across the whole signal range to obtain an entropy profile. An adaptation of this is to further catenate the sample strings from each level to obtain a single sample string for a given window position. In this paper we use a swept bipartition resulting in a 3D depiction of an entropy surface, as in Fig 1.

Here the process of partitioning and windowing is illustrated using a uniformly distributed random series confined to the unit interval. The sample strings resulting from each partition level and window position are used to compute T-entropy values then plotted as a 3D surface, that has on one axis, time, and on the other, the partition levels. The surface profile responds to the distribution of the localised time-series, subject to the conditioning of the samples over time. Fig. 1 (middle) shows that the profile for the uniformly distributed random series closely resembles the inverted 'parabolic' function identified by Shannon in [2]. Fig. 1 (lower) compares the profile for three such distributions. In summary, the T-entropy profile offers a attractive means for visually scoping real data series as a function of time [7] [8].

3. Simulated Sources

By multiplexing scaled distributions from; (i) uniformly distributed values on the unit interval, and (ii) the logistic iterates ($y_t = r\ y_{t-1}(1 - y_{t-1}), y_0 \in [0, 1], r = 4$) and (iii) a combination of these, a test "noise"-source is simulated for a comparison of the FFT and T-entropy measure. Fig. 2 (top) shows the T-entropy surface computed for (iii), using overlapping windows, width $W = 500$ points, and increments $\Delta W = 30$ points. The resultant surface may be viewed in 3D under varying lighting conditions but here a 2D projection serves to illustrate an obvious sensitivity to changing source characteristics. Fig. 2 (bottom) shows the corresponding spectra computed using an FFT with similarly overlapping sliding windows. This yields almost no discernible spectral sensitivity to the individual contrasting sources, but a slight variation to the average energy shifts and changes in signal amplitude.

The FFT's *forte* is its ability to respond to strong spectral content. We expect the T-entropy to be much less responsive to periodic content. Our second test signal comprised a single tone, with and without added noise, for each of the two "noise" sources (i) & (ii). Fig. 3 (top left) was the result of applying the FFT by way of a sliding window to a sampled pure tone, with a frequency $0.6R/2$, R being the sampling rate. In Fig. 3 (top right) the noise amplitude (peak-to-peak) was $0.5A$ where A is the amplitude (peak-to-peak) of the tone. The surfaces were visually indistinguishable for each of the respective "noise" sources (and so we include representative examples only) confirming the FFT to be ineffective in characterising the noise component in the signal. Fig. 3 (bottom) displays the con-

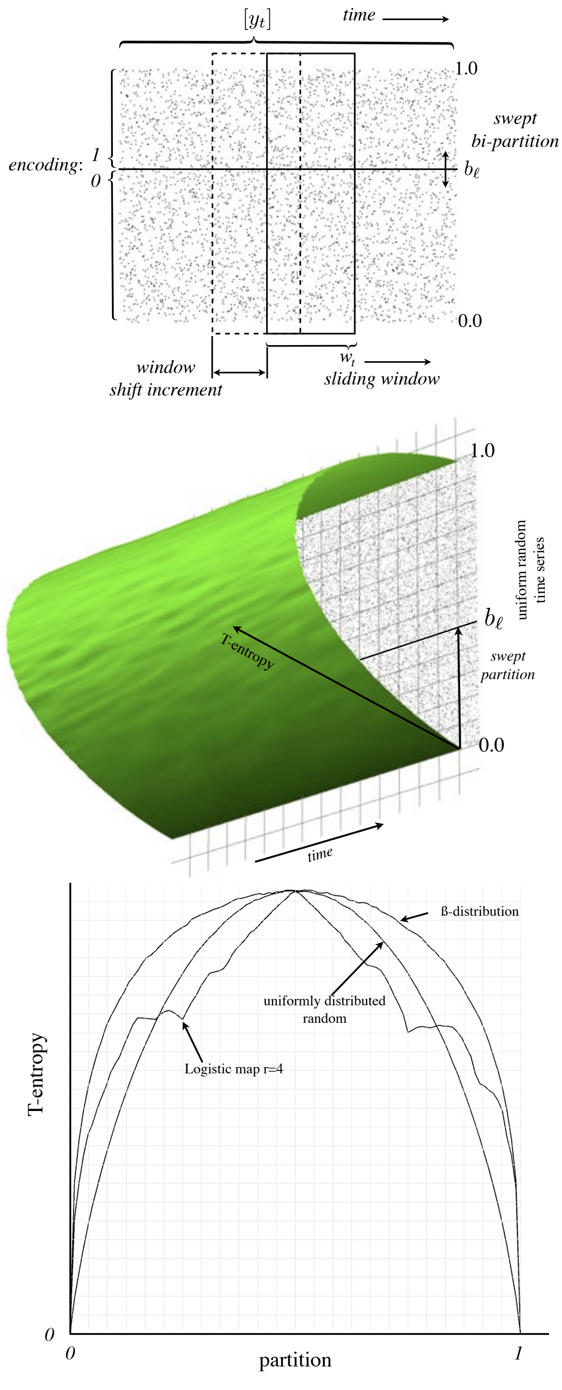

Figure 1: *Top*: depicting a random time-series uniformly distributed on the unit interval. The series is encoded by way of a swept bipartition and sliding window. *Middle*: The T-entropy is computed for each sample string and resultant array plotted as a surface. The height of the surface is the T-entropy at the position of the window and partition respectively. *Bottom*: A time-series results in its own distinctive profile: i) a uniform distribution, ii) iterates from the logistic map ($r = 4.0$), and iii) a random β-distributed series. Thus the T-entropy provides a practical way for visually scoping time-series.

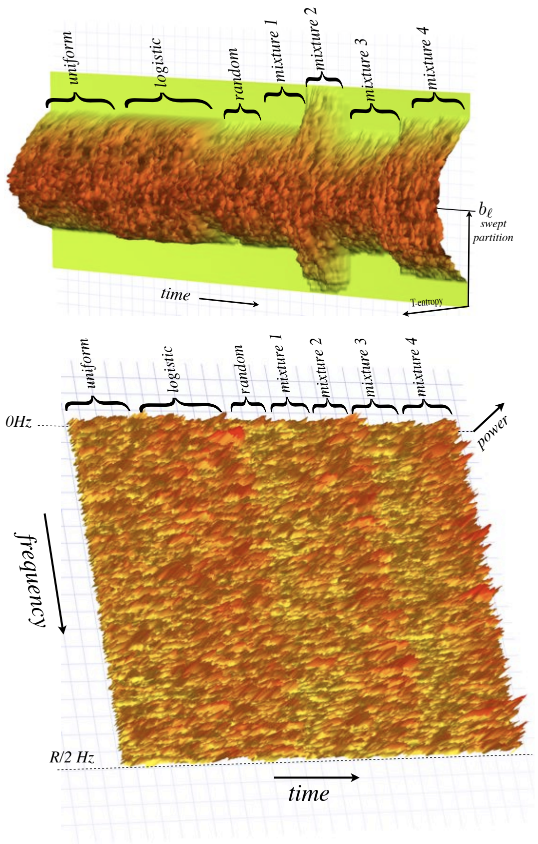

Figure 2: Multiplexed noise sources: (top) entropy surface, (bottom) FFT frequency-time series.

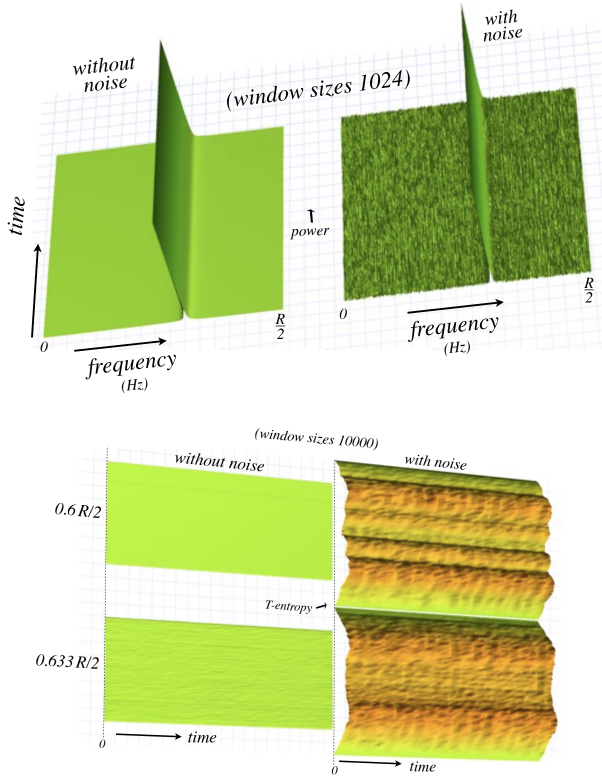

Figure 3: Applying the FFT (top) and the Tentropy (bottom) to a source comprising a singe tone with/without added noise

trasting T-entropy surfaces for the same pair of tests. The surfaces (left/right) are strongly indicative of the noise, but also we see a new source of variation. In shifting the frequency of the test tone from $0.6R/2$ or $0.633R/2$ we create a time varying relationship between the sample times and phase of the tone. This is already marginally apparent as surface texture (Fig. 3, bottom-left) but is strongly amplified with the addition of the noise. The dramatic increase in sensitivity to the spectral shift in the raised surfaces to the right appears as a manifestation of a stochastic resonance. Whereas the profiles for the noise distributions in Fig. 1 lent themselves to ready interpretation, here the addition of a single structured tone yield profiles exhibiting spectral sensitivity that seem to defy easy interpretation.

Our next experiment involves two time-multiplexed tones. There are no surprises in the spectral surfaces in Fig. 4 (top). The T-entropy surfaces are computed using two contrasting sized windows. The shorter of the two window sizes gives a more responsive resolution of the transient changes as the tones are switched on/off, but show artifacts from the mixing of the discretely sampled pair of frequencies. The dimensions of the resultant surface features are primarily a reflection of the choice of window size. With noise added, the surface takes on a more complicated profile, again strongly influenced by the structured component. How should this be interpreted? Again, these are seen as sensitive to the frequency (e.g.$0.6R/2$ versus $0.633R/2$), compounding their interpretation. Extending the window size allows one to average the T-entropy yielding profiles that bear a striking similarity to those derived from experimentally recorded EEG [7]. Thus EEG may well turn out to comprises a superposition of relatively stable harmonic components together with the more complex additive "noise" from non-linear sources.

A further experiment involves a simple swept-frequency tone. Fig. 5 (top) illustrates effects when computing the FFT, arising in relation to contrasting window sizes. Top left, a relatively small window (256 points) results in the modulation of the peak amplitude, sensitivity to the local phasing of tone and the FFT window. The averaging implicit from a larger window results in a smearing/broadening of the spectral peak. The choice of window width may be critical when it comes to scoping data with the T-entropy measure as well. For the T-entropy a swept tone represents a non-stationary signal, with its variability further underscored by relative changes in the phasing of the sampling and tone cycle. A modulated tone is precisely the way we exchange information over a communications channel. We expect a raised entropy value. The choice of the window width determines time sensitivity to local en-

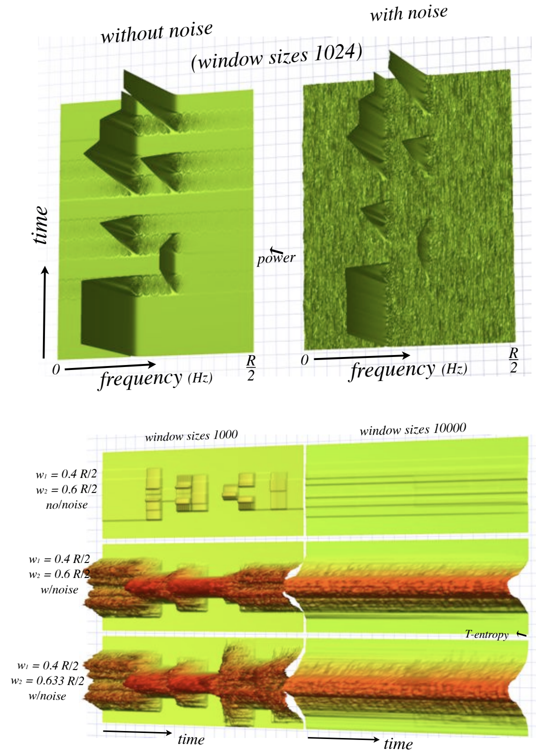

Figure 4: Modulated pair of tones, with/without added noise. At top: spectral surfaces, and at bottom: entropy surfaces.

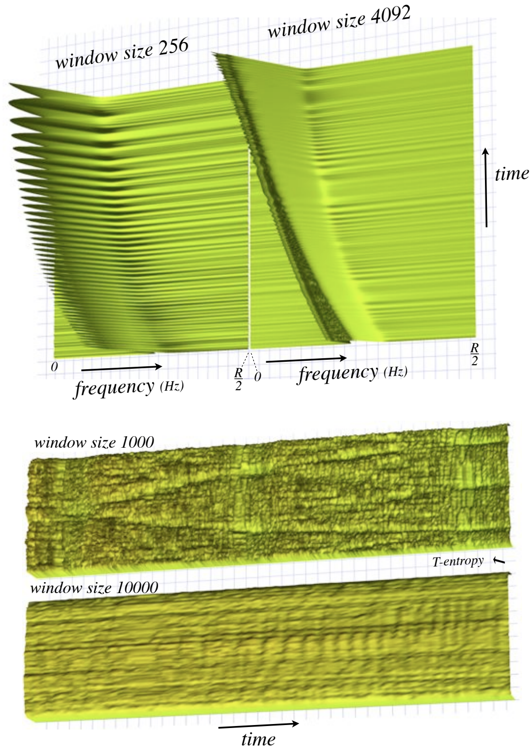

Figure 5: With non-stationary data, other effects may be observed. Here a swept tone starting at $0.6R/2$, with spectral (top) and T-entropy (bottom) surfaces displaying contrasting artifacts dependent on the averaging within the window width W.

tropy rate changes, as in EEG signals.

4. Concluding Remarks

This paper briefly compared the FFT and T-entropy with a range of contrived "test" signals. While at the outset we suggested the FFT and T-entropy might be viewed as providing complimentary descriptions of times series, the reality is that signals comprising a substantive mix of harmonic and chaotic components present considerable challenges when it comes to interpreting them. This seems especially so for the T-entropy results. It is simply not sufficient to assume that a T-entropy surface profile will be indicative of the random noise aspects of a series alone. For this reason we must develop sampling and encoding strategies, which help to separate out the effects of mixing spectral and structured components from the chaotic components. These simple examples lead us to conclude that it may be advantageous to apply subtractive smoothing filters to eliminate harmonic effects, prior to encoding and application of the T-entropy computation.

References

[1] A. Lempel, J. Ziv, "On the complexity of finite sequences," *IEEE Trans. on IT.* , vol.IT-22, No. 1, pp. 75–81, 1976.

[2] C. E. Shannon, W. Weaver "The Mathematical Theory of Communication," University of Illinois Press, Urbana, IL. (1949)

[3] R. Steuer, L. Molgedey, W. Ebeling and M.A. Jimenez-Montano : "Entropy and optimal Partition for Data Analysis", European Physics Journal B 19 , 2001, pp 265-269

[4] W. Ebeling, R. Steuer, M. R. Titchener, "Partition-based entropies of deterministic and stochastic maps," *Stochastics and Dynamics*, vol.1, No. 1, pp. 45–61, 2001.

[5] Jia Yang, Ulrich Speidel: "A T-Decomposition Algorithm with $O(n \log n)$ Time and Space Complexity," Proceedings of the IEEE International Symposium on Information Theory, 4-9 September 2005, Adelaide, pp. 23–27.

[6] M. R. Titchener, R. Nicolescu, L. Staiger, A. Gulliver, U. Speidel,"Deterministic Complexity and Entropy," *Fundamenta Informaticae*, IOS Press, vol.64, pp.443–461, 2005.

[7] M. R. Titchener,"T-Entropy of EEG/EOG Sensitive to Sleep State," *NOLTA 2006*, Bologna Italy,, pp.859–862, 2006.

[8] A. J. Hughes, M. R. Titchener, J. J. J. Chen, M. P. Taylor, "Pseudoresistance entropy as an approach to diagnostics and control in aluminium production", Asia Pacific Journal of Chemical Engineering, Aug. 2007