

DETECTING SIGNAL STRUCTURE FROM RANDOMLY-SAMPLED DATA

Frank A. Boyle(*), Jarvis Haupt(**), Gerald L. Fudge(*), Chen-Chu A. Yeh(*)

(*) L-3 Communications, Integrated Systems, 10001 Jack Finney Blvd., Greenville TX 75402,

(**) Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison WI 53706

ABSTRACT

Recent theoretical results in Compressive Sensing (CS) show that sparse (or compressible) signals can be accurately reconstructed from a reduced set of linear measurements in the form of projections onto random vectors. The associated reconstruction consists of a nonlinear optimization that requires knowledge of the actual projection vectors. This work demonstrates that random time samples of a data stream could be used to identify certain signal features, even when no time reference is available. Since random sampling suppresses aliasing, a small (sub-Nyquist) set of samples can represent high-bandwidth signals. Simulations were carried out to explore the utility of such a procedure for detecting and classifying signals of interest.

1. INTRODUCTION

Common signal processing problems include detection and identification of features in data that is incompletely represented or reduced in some manner. One such example is wide band processing with limited sample support, in which signals must be digitally represented at well below their Nyquist rate. Several reduced-sampling techniques have been proposed, including non-constant sampling [1,2,3], adaptive sampling [4,5], and compressive sensing [e.g., 6-12]. The latter of these approaches has recently generated significant interest in the signal processing community.

While these methods all allow for significant reduction in the amount of sampled data, they require knowledge of sample times to preserve signal structures. A natural question emerges about the possibility of detecting signals in randomly collected samples *without* time references.

Certainly, signal structures can influence the probability distributions of randomly-sampled datasets, suggesting their histograms may be valuable for detecting and/or classifying signals. Histogram methods have been effectively applied for feature detection in other contexts, including Facial recognition [13], handwriting analysis [14], and the recognition of randomly rotated objects [15].

The remainder of the paper is organized as follows. Section 2 briefly discusses the influence of signal structure on probability densities. In section 3, histogram analysis techniques are applied to random samplings of simulated data sets containing various structures. Section 4 is a brief illustration of an instantaneous histogram technique. Conclusions are discussed in section 5.

2. PROBABILITY DENSITIES OF RANDOMLY-SAMPLED DATA

The probability density function (PDF) $P_s(x)$ for a data sequence consisting of a single sinusoid is illustrated in Figure 1. It can be expressed by the projection of a circle onto the x axis,

$$P_s(x) = \begin{cases} \left(\pi \sqrt{A^2 - x^2}\right)^{-1}, & x \in (-A, A) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where A is the total amplitude. The PDF in Fig. 1 has a strong peak at the signal amplitude, indicating that a histogram of measured values would convey the signal amplitude in cases where a single signal dominates.

Note that Eq. 1 is independent of frequency, with one caveat; the frequency must be such that a large (or integral) number of cycles is contained in the data stream. As long as this condition is satisfied, the PDF is not affected by frequency, which can be arbitrarily high. Signal detection is therefore not bound by the Nyquist criterion.

The Probability density for a data stream consisting of a sum of several components can be expressed as the convolution of the densities of the individual components. An example with two CW components is shown in Fig. 2. The resulting PDF peaks at the difference between the component amplitudes. An edge in the curve appears at the sum of amplitudes. Certainly, more signal components can be added in this fashion. Structural features in the histograms will be more subtle and susceptible to noise and resolution limitations as more components are added. Preliminary simulations indicate that three continuous wave signals can sometimes be resolved in quiet environments.

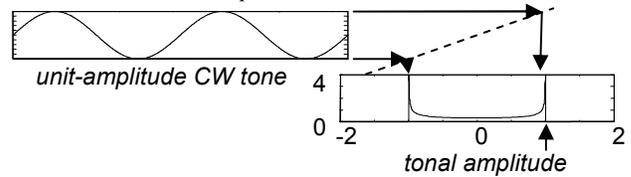


Figure 1: Probability Density for a single CW tone

Two CWs; amplitudes 1 and 0.25

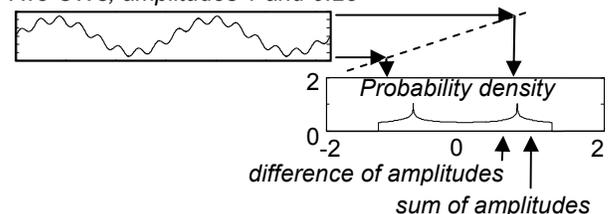


Figure 2: Probability Density - Two CW components

Noise can be included as another component. For zero-mean white Gaussian noise, the PDF is given by

$$P_n(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, \quad x \in (-\infty, \infty) \quad (2)$$

where σ is the standard deviation. Fig. 3 shows the two-component PDF of Figure 2 convolved with $P_n(x)$. Note that the edges and peaks that indicate the component signal amplitudes still appear, but are significantly obscured, even though the signal-to-noise ratio was small ($\sigma=0.1$). Preliminary simulations indicate that the signals are resolvable when the noise standard deviation σ is less than the smallest component signal amplitude.

3. SIMULATIONS

3.1. PDF estimation from histograms

To investigate PDFs of randomly-sampled data sequences, a set of simulations was carried out. A random selection of time instances t was generated. A vector $S(t)$ of simulated measurements was made by adding a Gaussian random noise vector $N(t)$ to a signal describing the continuous wave (CW) components:

$$S(t) = N(t) + \sum_n A_n \cos(2\pi f_n t + \phi_n) \quad (3)$$

where f_n , A_n , and ϕ_n are the frequency, amplitude, and phase of the n^{th} CW component, respectively.

PDF estimates were constructed by generating fine-resolution histograms of the simulated measurements. The histograms were then smoothed by convolution with a normalized smoothing kernel, as illustrated in Fig. 4.

Fig. 4a depicts the result of a simulation with two CW components, with amplitudes of 1 and 0.25. Fig. 4b is the same simulation, with a small amount Gaussian noise added, so that the CW components SNR's become 20 and 8 dB. The consistency between the simulations in Fig. 4 and the analytical results in Figs 2 and 3 supports the validity of this technique for estimating PDFs.

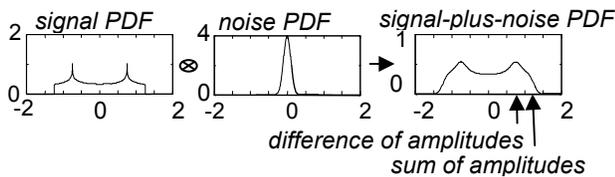


Figure 3: PDF; Two CWs in white Gaussian noise (SNRs of 20 and 8 dB).

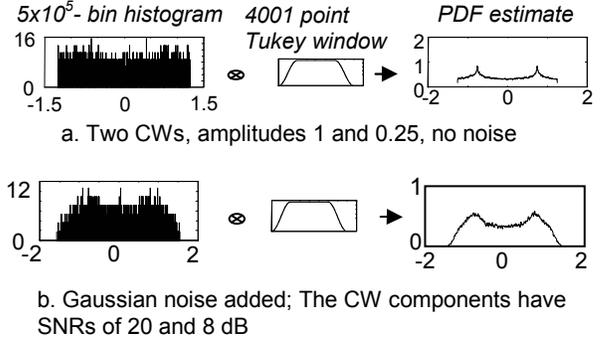


Figure 4: Estimating PDFs from fine-resolution histograms

3.2. Performance characterization

Given PDFs for noise and signal-plus-noise, a receiver operating characteristic (ROC) curve can be constructed to quantify the performance of a detector. The ROC curve conveys the balance between detection and false alarm probabilities when an amplitude threshold α is used to define signal detection for each sample. Since the noise and signal-plus-noise PDFs are symmetric about the origin, one-sided PDF curves can be used to compute the ROC value versus amplitude threshold:

$$ROC(\alpha) = \frac{PD_1(\alpha)}{PFA_1(\alpha)} = \frac{\int_{\alpha}^{\infty} PDF_{noise}(\alpha)}{\int_{\alpha}^{\infty} PDF_{signal}(\alpha)} \quad (4)$$

where PD_1 is the probability of detection and PFA_1 is the probability of false alarm for a single sample. Fig. 5 illustrates its construction for a CW white Gaussian noise with an SNR of 0 dB..

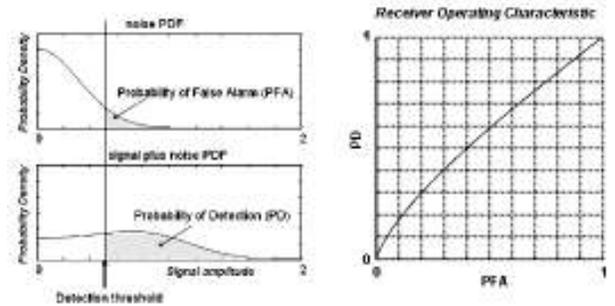


Fig. 5: Construction of a single-sample ROC curve from PDFs; CW in white Gaussian noise; SNR=0 dB.

Generally, the presence of a signal will not be assessed from a single sample, but rather from a set of successive samples. There are multiple ways of scoring such sets. One approach is to histogram the set and compare the result with expected PDFs by cross correlation or other means. A simpler approach is to assess each individual sample in the set and combine the results. A detection is determined when more than $m=n/2$ samples exceed the

detection threshold. Expressions for the probabilities of detection and false alarm can be written for such a detector [16]:

$$PD_n = \sum_{k=m}^n \binom{n}{k} PD_1^k (1 - PD_1)^{n-k} \quad (5)$$

$$PFA_n = \sum_{k=m}^n \binom{n}{k} PFA_1^k (1 - PFA_1)^{n-k} \quad (6)$$

Figure 6 is a plot of the ROC curves for various sample set lengths between 1 and 100 samples. As expected, the balance between detection and false alarm rates improves significantly with increasing n .

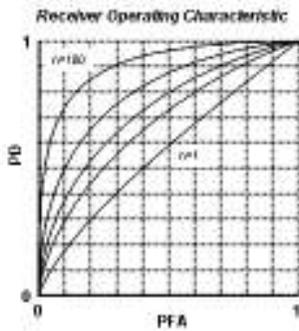


Figure 6: ROC curves for n-sample set for $n = 1, 10, 20, 40, 100$ samples; SNR=0 dB

3.3. Modulated signals

Fig. 7 illustrates the effect of modulation on the estimated PDF. Fig. 7a is an unmodulated unit-amplitude CW, with a PDF that peaks at ± 1 . In Fig. 7b the frequency is swept in time, producing no noticeable difference in the estimated PDF. This is expected, since Eq. 1 has no frequency dependence. Similarly, Fig. 7c shows that phase modulations are not expected to affect the PDF. Amplitude modulations, however, had strong influence, as shown in Figs. 7d and 7e.

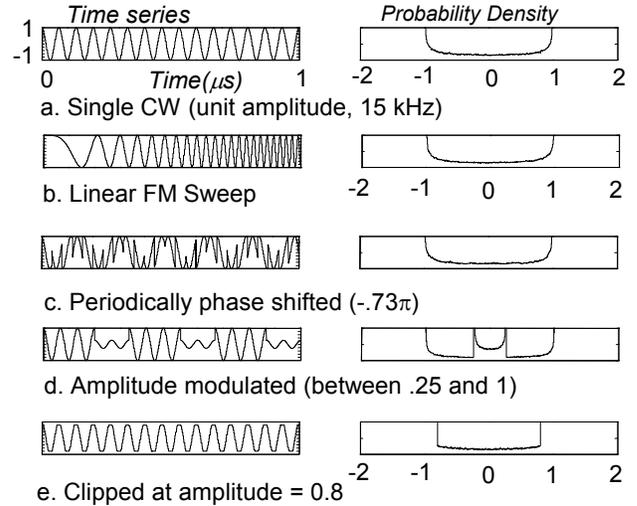


Figure 7: Modulated waveforms and their PDF's

3.4. Harmonically related signals

Given a small set of CW components, the PDFs can be influenced by harmonic relationships. Fig. 8 is an illustration of this effect. PDF estimates are constructed for a simulation like that of Fig. 4b, except that the two component tones are given a harmonic relationship; the frequency one tone is exactly twice that of the other. In this case, the two components line up in phase once per cycle of the low-frequency component. Fig. 8 shows that the PDFs depend sensitively on the phase angle at which alignment occurs.

The contrast in the shapes of the histograms between Fig. 4b and Fig. 8 reveal a strong sensitivity to the harmonic relationships. This implies that, while absolute frequency information is not conveyed, the histograms may allow for the detection of harmonic structure. The observance of an asymmetric histogram, for example, may imply that some of the signal components are harmonically related, and originate from the same source.

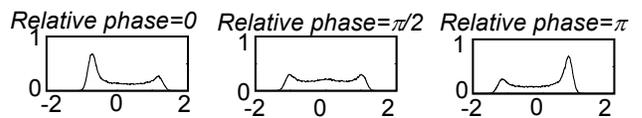


Figure 8: Estimated PDF's ; 2 harmonically-related CW tones.

3.5. Histogram detection of time-dependent features

In a real system, a continuous randomly-sampled data stream would likely be processed in blocks, with successive histograms carrying time-dependent information about the data stream. Events of interest might manifest themselves as detectable changes in the evolution of the histogram.

Data stream time duration:	1 sec
Total number of random samples:	3×10^4
Number of time segments:	100
RF background:	Gaussian Noise
Signal (pulsed):	
Frequency:	354 kHz
Level:	0 dB RE noise background
Pulsed:	4 cycles; duty cycle=0.3

A simulated data stream was constructed as described in Table 1. The data stream included a pulsed CW in 0 dB of Gaussian noise. The data, originally sampled at 1 kHz, was randomly subsampled by a factor of 10. The subsampled datastream was divided into time segments, each of which was histogrammed as illustrated in Fig. 9, to produce a time-evolving PDF estimate.

The inserted pulsed signal is clearly observable in statistics obtained from the PDF. The variance V and Kurtosis K are given by:

$$V = \int_{-\infty}^{\infty} PDF(x)x dx \quad (7)$$

$$K = \int_{-\infty}^{\infty} PDF(x)x^3 dx \quad (8)$$

In Fig. 9, the signal's presence is clear in the variance and kurtosis plots. In this example, there's very little difference in character between the variance and kurtosis, suggesting that the variance is adequate for detecting CWs. Fig. 10 depicts an example in which the kurtosis outperformed the variance. In this case the RF background is non-Gaussian. In general, the effectiveness of higher order statistics like kurtosis will depend on the character of the signal and noise.

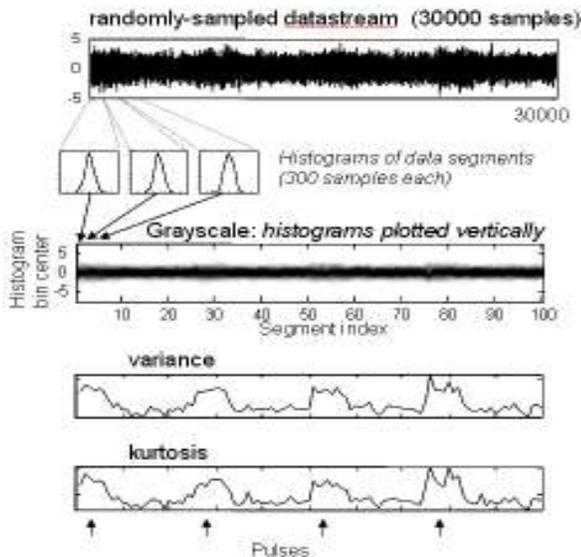


Fig. 9: Time-evolution of PDF estimate for simulation in Table 1.

Data stream time duration:	1 sec
Total number of random samples:	3×10^4
Number of time segments:	100
RF background:	Gaussian Noise + 20 dB SNR CW interferor @ 676 kHz
Signal (pulsed):	
Frequency:	354 kHz
Level:	20 dB RE Gaussian noise background

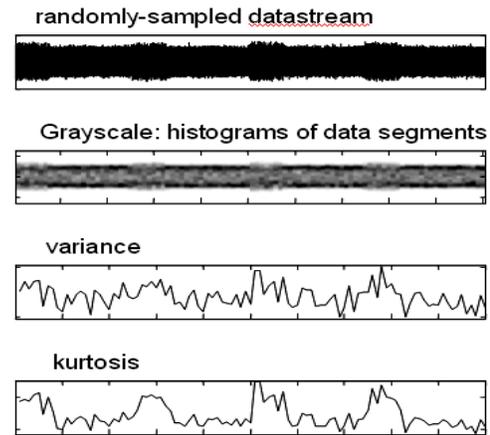


Figure 10: Time-evolution of PDF estimate for simulation in Table 2; A pulsed tone in a nongaussian RF background.

4. INSTANTANEOUS FREQUENCY HISTOGRAMS

In section 3, histogram analysis was applied to random samples from a real-valued data stream. The histograms, under the right conditions, allowed for detection of signals and estimation of their amplitudes, but provided no information about carrier frequency or bandwidth. A question that emerges is whether other information can be obtained by sampling different representations of the RF data. A representation of current interest is the instantaneous frequency, which conveys spectral information. One way of measuring the instantaneous frequency to separate the RF signal into two paths, time-delay one, and compare the phases of corresponding quadrature samples [17]. Collections of random instances of such measurements can be histogrammed to extract information on spectral content.

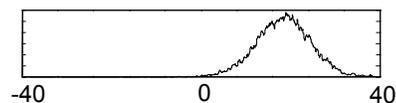


Figure 11: Instantaneous-frequency histogram for a 20 KHz CW tone in Gaussian noise (SNR=10 dB). The peak location correctly indicates the signal's frequency.

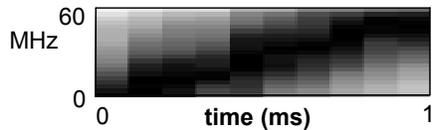


Figure 12: Instantaneous-frequency histogram vs. time for a 0-55 MHz chirp white Gaussian noise (SNR=0 dB). The mean sample rate was 12MHz.

Figure 11 is a smoothed histogram of 10000 random realizations of the instantaneous frequency for a signal containing a 20 kHz tone in white Gaussian Noise with an SNR of 10 dB. The peak in the histogram indicates the tonal frequency. Fig. 12 shows a histogram-versus-time plot for a frequency-swept signal with an SNR of 0 dB. The particular example is a base-banded 55 MHz-bandwidth chirp, as is generated by Phillips Research Laboratory's Pilot Mk3 Radar [18]. The frequency sweep of the signal is clear. These preliminary observations suggest that signals can be identified in instantaneous frequency histograms when significant amounts of noise are present. The instantaneous-frequency samples were collected at 12 MHz. Since each sample involves a comparison of two IQ samples, the implied Shannon-Nyquist limit is 24 MHz. This is well below the signal's 55 MHz bandwidth, implying a sub-Nyquist representation.

5. CONCLUSIONS

Preliminary simulations indicate that simple histograms of randomly-sampled data can provide important information about signal structure. This could lead to the development of practical tools for detection and classification when sample support is limited.

Simulations were carried out to explore the influence of signal structures on measured histograms. The histograms were clearly influenced by amplitude variations, clipping, and the presence of multiple CW components. Coherence between components introduced an observable asymmetry that depended on the phase relationship of the components. Pulsed CW components had measurable effect on the second moment histograms, which exhibited identifiable features even in the presence of significant noise. Histograms of other signal attributes may reveal more information about signal structure; for example, instantaneous phase histograms carry information about spectral content.

In the simulations, histograms were constructed from random samplings, with effective sample rates well below the Nyquist limit. A relevant question is how much effective compression can be realized, while preserving structures of interest.

Future work will focus on histogram estimation techniques, and will examine other signal attributes such as instantaneous phase and higher derivatives of instantaneous frequency. We also plan to quantify the effects of noise and interference to determine the practicality and limitations of the proposed methods.

6. REFERENCES

[1] Belinskis, I. and Mikelsons, A., *Randomized Signal Processing*, Prentice Hall, London, 1992.

[2] Marvasti, F., Ed., *Nonuniform Sampling, Theory and Practice*, Kluwer, New York, 2001.

[3] Pace, P.E., *Advanced Techniques for Digital Receivers*, Artech House, Boston, 2000.

[4] Steven K. Thompson and George A. F. Seber, *Adaptive Sampling*, Wiley, Hoboken, NJ, 1996.

[5] Rigau, j., Feixas, M., Sbert, M., "Entropy based adaptive sampling," *Graphics Interface 2003 (Online)*

[6] David Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, Vol. 52, (4), 1289 – 1306, April 2006

[7] Emmanuel Candès, Justin Romberg and Terence Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information," *IEEE Trans. Information Theory*, 52(2), 489 – 509, Feb. 2006

[8] Emmanuel Candès and Terence Tao, "Near Optimal Signal Recovery From Random Projections: Universal Encoding Strategies

[9] J. Haupt, R. Nowak, "Signal Reconstruction from Noisy Random Projections," *Submitted to IEEE transactions on Information Theory*, March 2005.

[10] Emmanuel Candès and Justin Romberg, "Practical Signal Recovery from Random Projections," (preprint, 2004)

[11] Joel Tropp and Anna Gilbert, "Signal Recovery From Partial Information Via Orthogonal Matching Pursuit," (preprint, 2005)

[12] Joel Tropp, Michael Wakin, Marco Duarte, Dror Baron and Richard Baraniuk, "Random Filters for Compressive Sampling and Reconstruction," *Proc. IEEE 2006 International Conference on Acoustics, Speech, and Signal Processing*, May 2006

[13] Crowley, J.L.; Berard, F, "Multi-Modal Tracking of Faces for Video Communications," *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp.640-645, 1997

[14] Wilkinson, T. S., Goodman, Joseph W., "Slope histogram detection of forged handwritten signatures" *Proc. SPIE* Vol. 1384, pp. 293-304, 1991

[15] Schneiderman, H., Kanade, T., "A histogram-based method for detection of faces and cars," *Proc. International Conference on Image Processing*, Vol. 3, 504-507, (2000)

[16] Cooper, C.R., McGillem, C. D., *Probabilistic Methods of Signal and System Analysis*, (3rd ed.), Oxford University Press, New York, 1999, p. 36.

[17] Gruchala, H.; Czyzewski, M. "The Instantaneous frequency measurement receiver in the complex electromagnetic environment," 15th International Conf. on Microwaves, Radar, and wireless communication Vol. 1, 155-158, (2004)

[18] Pace, P.E., *Detecting and Classifying Low Probability of Intercept Radar*, Artech House, Boston, 2004, p. 30.