Comparison of signal detectors for time domain radio SETI

Gregory Hellbourg[1], Andrew Xu[2]

(1) Berkeley SETI Research Center - University of California - Berkeley, CA, USA

(2) Department of Electrical Engineering and Computer Sciences - University of California - Berkeley, CA, USA

Abstract

The radio Search for Extra Terrestrial Intelligence (SETI) [1] aims at identifying intelligent and communicative civilizations in the Universe through the detection of engineered transmissions. In the absence of prior knowledge concerning the expected signal, SETI detection pipelines necessitate high sensitivity, versatility, and limited computational complexity to maximize the search parameter space and minimize the probability of misses. This paper addresses the SETI detection problem as a binary hypothesis testing problem, and compares four detection schemes exploiting artificial features of the data collected by a single receiver radio telescope. After a theoretical comparison, those detectors are applied to real data collected with the Green Bank Telescope in West Virginia (USA).

1 Introduction

The radio Search for Extra Terrestrial Intelligence (SETI) aims at detecting radio transmissions from intelligent and communicative civilizations in the Universe. High sensitivity radio telescopes receivers and detection schemes, associated to dedicated voltage data recorders, are necessary to maximize the searched parameter space (directions-of-arrival, epochs, frequency, signal characteristics...) and the probability of detection. The absence of prior information concerning the signal-of-interest also necessitates the exploration various signal features to minimizes the probability of misses.

This paper addresses artificial signal detection for single receiver radio telescopes. Section 2 introduces the SETI detection problem from a binary hypothesis testing standpoint. Four detectors are compared onto three data models in this study, all are described in this same section. Section 3 details the Monte-Carlo simulation ran to achieve the comparison, and draws the main conclusions. Section 4 analyzes the results of detectors applied to real telescope data. Section 5 finally concludes this study and suggests future research directions to improve the reliability and computational cost of a single receiver radio SETI experiment.

2 SETI signal detection

2.1 Detection theory

The SETI detection problem is formulated as a binary hypothesis testing problem [3], with the following two hypotheses *H0* and *H1*:

$$\mathbf{x}[n] = \begin{cases} \mathbf{x}_{\text{noise}}[n] & (H0) \\ \mathbf{x}_{\text{ET}}[n] + \mathbf{x}_{\text{noise}}[n] & (H1) \end{cases} \quad (1)$$

where $\mathbf{x}[n]$ is the basebanded, digitized, and channelized telescope output, $\mathbf{x}_{\text{noise}}[n]$ is the system noise contribution and $\mathbf{x}_{\text{ET}}[n]$ the eventual extra terrestrial (ET) transmission.

Given a detector $\theta(\mathbf{x}[n])$ and threshold γ, the detection performance is assessed through the evaluation of its Probability of Detection $P_d(\gamma)$ and Probability of False Alarm $P_{fa}(\gamma)$ [3].

2.2 Proposed detectors

Assume N samples are received. The following detectors are proposed for comparison for a single dish telescope radio SETI experiment.

- Energy detection (Runtime: $\Theta(N)$):
  The energy detector is defined as:

$$\theta_{\text{energy}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}[n]\mathbf{x}^*[n] \quad (2)$$

with $(.)^*$ the complex conjugate operator. The energy being conservative, this detector monitors the increase of energy contributed by an additional signal component.

- Fourier-based detection (Runtime: $\Theta(\tilde{N}\log\tilde{N})$):
  The averaged periodogram detector the time series $\mathbf{x}[n]$ is defined as:

$$\theta_{\text{FT}} = \max_{f \in [1..\tilde{N}]} \frac{1}{M} \sum_{m=1}^{M} \left| \mathscr{F}_{\mathbf{x}[(m-1).\tilde{N}+1:m.\tilde{N}]}(f) \right|^2$$

where $\mathscr{F}_{\mathbf{x}[m.\tilde{N}:(m+1).\tilde{N}]}(f)$ is the Fourier transform of

the mth non-overlapping block of $\tilde{N}$ samples of $\mathbf{x}[n]$ such that $N = M \times \tilde{N}$, and $|.|$ is the absolute value operator. This detector monitors spectral features embedded in the received time series.

- Time-lag-based detection (Runtime: $\Theta(N\log N)$):
  The time-lag detector is expressed as:

$$\theta_\tau = \max_{0 < \tau < N} \frac{R_{xx^*}(n, \tau)}{\theta_{\text{energy}}} \quad (3)$$

where $R_{xx^*}(n, \tau) = \frac{1}{N-\tau-1}\sum_{n=\tau+1}^{N} \mathbf{x}[n]\mathbf{x}^*[n-\tau]$ is the autocorrelation of $\mathbf{x}[n]$ at discrete time step $\tau \in \mathbb{Z}$. This detector monitors the whiteness of the received time series.

- Karhunen-Loève Transform (Runtime: ΘM^3):
  The Karhunen-Loève Transform detector [4] extends the *Time-lag-based detection*, and is based on the eigen value decomposition of the auto-correlation matrix $\mathbf{R}$ of the received time series $\mathbf{x}[n]$:

$$\mathbf{R} = \frac{1}{N-M-1}\sum_{n=1}^{N-M} \mathbf{x}[n:n+M]\mathbf{x}^H[n:n+M] \quad (4)$$

with $(.)^H$ the Hermitian transpose operator. The *Karhunen-Loève Transform detector (KLT)* is expressed as:

$$\theta_{KLT} = \max_{0 < m < M} \lambda_m / \text{trace}(\mathbf{R}) \quad (5)$$

with λ_m the mth eigenvalue of $\mathbf{R}$, and trace(.) the matrix trace operator.

The rank of $\mathbf{R}$ quantifies the trade-off between its integration time and its dimensionality, and is the main parameter of the detector. Large dimensionality is necessary for detecting high dimensional signal subspaces [5], while long integration time ensures the estimation quality.

2.3 Data models

Three data models are considered to compare the performances of the proposed detectors.

- Noise model
  The system noise includes independent contributions of the cosmic background, stellar responses, and from the instrumentation itself. It is modeled as a white circular stationary (over short durations) centered complex, independently and identically distributed, Gaussian process with power : $\mathbf{x}_{\text{noise}}[n] \sim \mathcal{NC}(0, \sigma^2)$

- Linear chirp
  A narrow-band ET transmission undergoes a frequency drift resulting from the Doppler effect due to the Earth rotation. Frequency drifts

are assumed linear over short observation durations ($\sim$ 5 minutes), and might reach up to a few Hz/s [2]. The resulting signal model follows $\mathbf{x}_{\text{ET}}[n] = A.\exp\{i2\pi[d.n+f_0]n\}\exp\{i2\pi\phi\}$, with power $\sigma^2_{\text{chirp}} = A^2$, and $d/2 \in \mathbb{R}$ the linear chirp rate.

- Oversampled Binary Phase-Shift Keying (BPSK)
  Information-bearing transmissions are likely, although the trade-off between detectability and channel capacity remains argued in SETI. A Binary Phase-Shift Keying (BPSK) modulation scheme has been selected for this study, and is modeled following: $\mathbf{x}_{\text{ET}}[n] = \xi_m.A.\exp\{i2\pi f_0 n\}\exp\{i2\pi\phi\}$, with power $\sigma^2_{\text{BPSK}} = A^2$, and ξ_m a pseudo-random variable taking value in $\{-1,+1\}$. The baud-rate is assumed lower than the sampling frequency. The discrete message ξ_m can be convolved with a Hamming window for spectral smearing control (referred to as windowed BPSK).

3 Detector comparison on simulated data

3.1 Simulation setup

The detectors proposed in section 2.2 are compared on the data models described in section 2.3 through a Monte Carlo simulation involving 10^3 independent trials per Signal-to-Noise Ratio (SNR) defined as $\text{SNR} = A^2/\sigma^2$. Two data sets of length $N = 2^{16}$ are generated for each trial, corresponding to the two hypothesis *H0* and *H1*. The statistics of each detector are then evaluated under both hypotheses.

Figure 1 reports the results of this analysis. Each plot represents the Probability of Detection P_d associated with each detector (figures 1.(a)-(d)) for a given probability of False Alarm $P_{fa} = 0.01$ for 4 signal types (pure carrier, linear chirp, BPSK and windowed BPSK). The data models parameters[1] are randomly generated following: $f_0 \sim \mathcal{U}_{[0,1]}$, $d \sim \mathcal{U}_{[-\frac{2}{N},\frac{2}{N}]}$, $\phi \sim \mathcal{U}_{[0,1]}$, where $\mathcal{U}_{[a,b]}$ is the uniform distribution over the range $[a,b]$.

The algorithms parameterizations are as follow:

- *time_lag*:
  Time-lagged-based detection evaluated over $N-1$ time-lags.
- *perio_{1,8,64}*:
  Fourier-based detection evaluated over $N/\{1,8,64\}$ frequency bins and averaged over $\{1,8,64\}$ non-overlapping windows.
- *perio_ham_{1,8,64}*:
  Same as above, applying an Hamming window prior to the Fourier Transform.
- *max_KLT*:
  Auto-correlation matrix of dimension $M = 64$, averaged of $N-M$ overlapping windows.

---

[1] except the amplitude term A defining the SNR of the model for $\sigma^2 = 1$
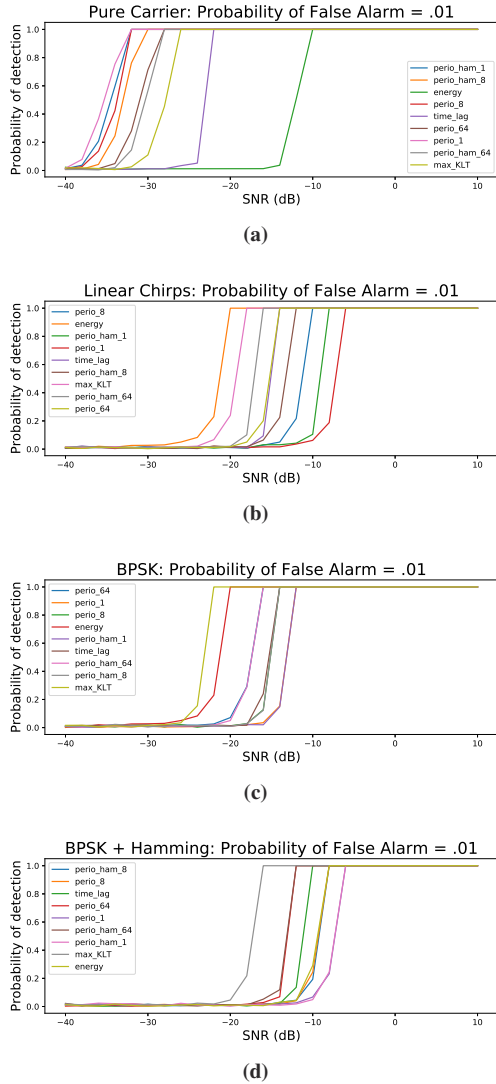
Figure 1. Probability of Detection as a function of the SNR for the considered detectors, for a given probability of False Alarm $P_{fa} = 0.01$. (a) Pure carrier model. (b) Linear chirp model. (c) BPSK - baud rate = 1/8 sampling frequency. (d) Windowed BPSK - baud rate = 1/8 sampling frequency.

3.2 Simulation results

- All detectors behave efficiently for SNR > -10 dB.
- The *Fourier-based detectors* outperform all detectors when applied to pure carrier signals, as they are assimilated to near-optimum matched filters in this context.
- Linear chirps do not respond similarly with the *Fourier-based detectors* as their spectra undergo frequency smearing after data integration. Shorter integration durations ($N/8$ or $N/64$) present lower spectral smearing and higher detection performances.
- The *Time-lag-based detector* presents average detection performances on all data models, failing to detect auto-correlation at low SNR.
- The *KLT detector* dominates the other detectors on wide-band transmissions.

4 Real Data processing

To validate the theoretical results shown in the previous section, the detectors are tested on real data collected with the Green Bank telescope in West Virginia (USA). This experiment compares the recovery of Radio Frequency Interference (RFI) on $\approx$ 3 MHz-wide channelized baseband data. A software fine channelization splits the the data into 3 / 128 MHz $\approx$ 23 kHz channels.

Figures 2.(a) and 3.(a) show the estimated Power Spectral Density of the two data sets integrated over 2.5 minutes. In red has been highlighted the "eyeballed" empty parts of the spectra used to estimate a detection threshold corresponding to a Probability of False Alarm $P_{fa} = 0.01$. The "roll-off" sections on both ends of both spectra correspond to the bandpass of the 3-MHz filter bank.

The data set #1 on figure 2 is composed of a single narrow band transmission. The data set #2 on figure 3 is more complex and composed of various wide-band signals.

The following binarized spectrograms {2,3}.(b-c-d) are the results detection results of the *energy detector*, the *KLT detector*, and the *Fourier-based detector* with 8 non-overlapping integrations. Blue pixels correspond to a non-detection, yellow pixels correspond to a detection.

- Comments on Figure 2
  The *Energy detector* globally fails at recovering any feature due to the low SNR of the signal. The *KLT detector* properly retrieves the narrow band signal and additional noise outliers, and is not affected by the band-pass response of the data thanks to its power normalization. The two variations of the *Fourier-based detector* also recover the signal and noise outliers over the whole data span.
- Comments on Figure 3
  All detectors detect a broadband feature at $t \approx 80$s), possibly an instrumental artifact. The *Energy detector* recovers the various wide band features with only a few noise outliers. The *KLT detector* fails at recovering any signal due to weak auto-correlated features. Both variations of the *Fourier-based detector* show performances comparable to the *Energy detector* with higher low-SNR noise outlier detections.

5 Conclusion

Radio SETI aims at identifying intelligent and communicative civilizations in the Universe through the detection of their artificial transmissions. High sensitivity instrumentation and detection schemes are necessary in the absence of prior information concerning these transmissions. This paper compares four detection schemes for a given probability of false alarm based on the signal energy, frequency distribution and auto-correlation. The comparison is theoretically conducted on three data models before running them on real telescope data.
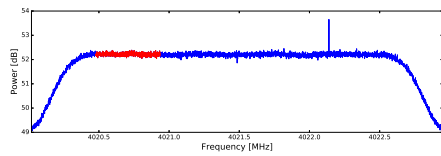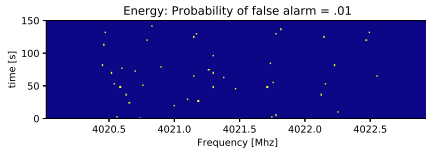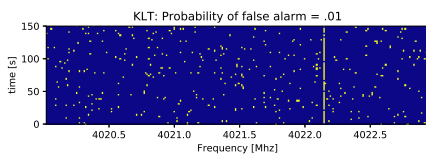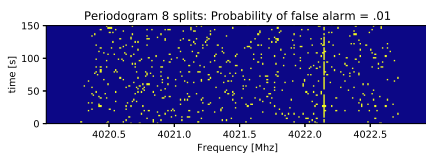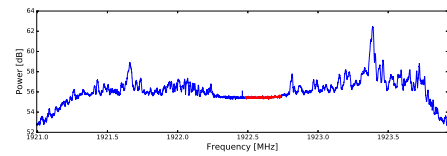
Figure 2. GBT data featuring a narrow band signal. (a) Power Spectral Density evaluated over 5 min. Following binary spectrograms resulting from the *Energy detector* (b), the *KLT detector* (c), and the *Fourier-based detector* with 8 non-overlapping spectra average (d).
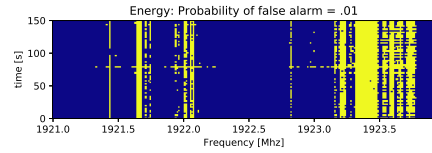

Figure 3. GBT data featuring wide band signals. (a) Power Spectral Density evaluated over 5 min. Following binary spectrograms resulting from the *Energy detector* (b), the *KLT detector* (c), and the *Fourier-based detector* with 8 non-overlapping spectra average (d).

The analysis conducted shows a high dependence of a detector performance to the data model and the computational complexity. Detection performance can be improved through careful algorithm parametrization, and require further simulations and read data processing.
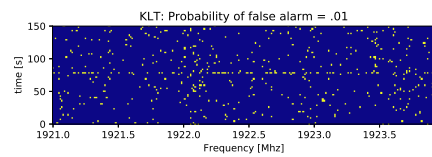
References

[1] Tarter, J. C. "SETI 2020: A roadmap for future SETI observing projects". Proc. SPIE. Vol. 4273. 2001.

[2] Isaacson, Howard, et al. "The Breakthrough Listen Search for Intelligent Life: Target Selection of Nearby Stars and Galaxies". Publications of the Astronomical Society of the Pacific 129.975 (2017): 054501.

[3] Van Trees, Harry L. "Detection, estimation, and modulation theory". John Wiley & Sons, 2004.

[4] Maccone, Claudio. "Advantages of Karhunen Loève transform over fast Fourier transform for planetary radar and space debris detection". Acta Astronautica 60.8 (2007): 775-779.

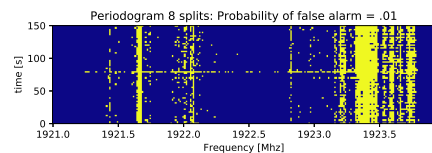[5] Wax, Mati, and Thomas Kailath. "Detection of signals by information theoretic criteria". IEEE Transactions on Acoustics, Speech, and Signal Processing 33.2 (1985): 387-392.