WAVELET PACKET ANALYSIS AS A MEANS OF SEARCHING FOR WEAK NARROW BAND SIGNALS

T. Anastasia[1], G. Maenza[2], R. Polikar[3]

[1]L3 Communications Systems East, Camden, NJ 08103, Thomas.Anastasia@L-3Com.com, [2]Sony Music, Pitman, NJ 08080

[3]Electrical and Computer Engineering, Rowan University, Glassboro, NJ 08028, polikar@rowan.edu

Abstract- Potential benefits of wavelet packet analysis are explored as a way of searching over a wide frequency band for weak, narrow band signals. The specific application of the technique discussed in this paper is for SETI (Search for ExtraTerrestrial Intelligence) data where conventional Fourier techniques have been used in the past. SETI data analysis is based on searching for pulsatile signals that are buried in noise and the wavelet transforms are known to be superior to Fourier transforms in analyzing such highly corrupted nonstationary signals. The time and frequency localization characteristics of wavelet packets are used to efficiently localize weak signals with an algorithm that is potentially more sensitive and less time consuming than the FFT.

1. INTRODUCTION

Traditional Fourier based methods for searching signals have been used as part of various SETI (Search for ExtraTerrestrial Intelligence) programs over several decades to find artificial extraterrestrial signals in radio telescope data. A popular search technique that is commonly used today is based on the SETI@home program, which makes use of the volunteered computing resources of participating Internet users. The search software (the 'client' program) is installed on users' personal computer and runs as a screen saver, whenever the CPU is idle, to prevent interfering with normal tasks. Data gathered from a radio telescope is divided into small packets called "work units", which are sent to each client via the Internet for analysis [1]. When the client is finished, the results are returned and another work unit is requested. Today, there are over 3 million volunteers running the client software.

The SETI@home program, like its predecessors, uses traditional Fourier based techniques to search and analyze signals that are possibly of intelligent origin. However, since the signals searched by the program are highly nonstationary and extremely low SNR pulses, wavelet analysis appears to be a better fit for their detection and time localization. The potential for wavelet packet analysis in analyzing SETI signals are therefore explored in this study. Although the application that is targeted in this study is the SETI data, the proposed method is in fact applicable to the more general problem of searching for very weak signals over a wide band that is many times the bandwidth of the target signal.

2. METHODOLOGY

Radio based SETI is based on the assumption that an artificial signal would be narrowband since signals created by human activity are narrowband, and those created by natural sources are broadband. Not only does a narrowband signal stand out from natural sources, it also uses power more efficiently so the signal can be detected at greater distances. Early analog based efforts at SETI used narrowband filters to analyze a radio signal, whereas the current practice is using dedicated FFT processors separating signals into a billion 1 Hz wide channels that can be examined for a signal.

The drawback of examining a narrow band is that the signal is unlikely to be stable in frequency. This is because of the Doppler shift created by movement of the transmitter and receiver. The earth rotates fast enough to cause a 1 Hz wide signal to drift out of band in only 6 seconds. The transmitting end can also drift at some unknown rate. The transmitting party can compensate for this drift if the receiver's Doppler shift is known, but this cannot be assumed. Therefore, the search algorithm used by the SETI@home client spends a significant portion of its time performing Doppler shift compensation to remove potential positive and negative drift rates.

A. Radio Spectrum Search Region

The region of the spectrum used by SETI@home is a narrow band of 2.5 MHz around the mean frequency of 1420 MHz. If a signal is transmitted at any other frequency it will not be detected by SETI@home. This frequency was chosen because it lies in an area of the spectrum where the universe is naturally quiet, making it easier to detect a signal. This is also the frequency emitted by Hydrogen, the most common element in the universe, and it also is close to the frequency of another common molecule hydroxyl (OH) at 1640 MHz. Hydrogen and hydroxyl combine to form water, the fundamental component to life as we know it. Any technologically advanced civilization would know these facts, so there has been wide acceptance of this band as a search area.

The SETI@home observations are collected at the National Astronomy and Ionospheric Center's 305-meter radio telescope in Arecibo, Puerto Rico. The project has a dedicated feed that is piggybacked with the receiver, so that data are taken almost continuously without interfering with other uses of the telescope. Since the telescope is built into a natural depression in the terrain, most observations pass through its' field of view (0.1 degrees) at the earth's rotation rate or sidereal rate allowing about 24 seconds for a stationary point in the sky to transit its' field. The data from the dedicated feed is recorded onto data tapes on a 2.5 MHz band centered at a frequency of 1420 MHz. The tapes are sent to dedicated splitter workstations that split the signal data into smaller pieces of time and frequency to make the work units that are sent to the clients.

B. Work Units

The splitter workstations divide the 2.5 MHz bandwidth into 256 subbands using a 2048-point FFT. Since the data is sampled at the Nyquist rate of 5 MHz, the FFTs are performed at a rate of 5 MSamp/sec / 2048 samples = 2441.4 FFT/sec. This corresponds to a time period of 0.4096 ms for each FFT, the minimum time resolution possible. Each of the 256 subbands in the FFT consists of 2048/256 = 8 points (this is up to the Nyquist rate), so the aliased frequencies are included in the 8 points.

The splitter then re-creates the time based signal by performing an inverse FFT with 8 points in each subband. Therefore, in the time domain each subband has 8 points for every 0.4096 ms, for a sampling rate of 19.531 KHz. This is illustrated graphically in Fig. 1, showing an expansion of one of the subbands:

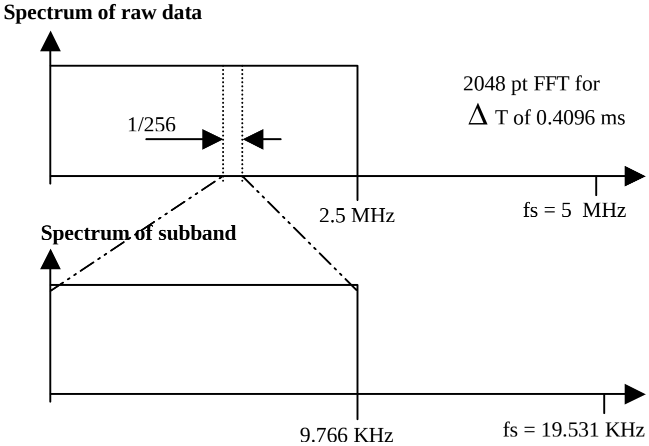

Figure 1: Breakdown of raw data into work unit subbands.

Since the subband is now sampled at 19.531 KHz, a convenient length of time is chosen that exceeds the length of time that a signal is in the beam of the radio telescope. The number of samples in the work unit must also be a power of 2 for the convenience of the FFT calculations. Two million samples are chosen, making the total length of time in a work unit 107.4 seconds. This results in a small enough file to be sent over Internet connections at various speeds. Total length of time also exceeds the 24 seconds that a signal is within the beam of the radio telescope. Information is appended onto each work unit to identify it, and sent to the client.

C. Analysis Algorithm

The client software performs a Fourier based analysis on each work unit. This can be described with a pseudo code description as in Table I.

D. Wavelet Analysis of Work Units

Breaking down the existing client analysis software into different steps in Table 1 allows choosing the parts of the algorithm that seem to have the most promise for a wavelet analysis. Step 0 splits the data into 256 bands using an FFT, then reconstruct each band in the time domain with a length of 107 seconds. Here the superior ability of wavelets to reconstruct data, especially edge information would be an asset. This could improve the time domain analysis performed later in steps 7, 8 and 9.

TABLE I: Pseudo Code for SETI@home algorithm

Step	Function
0	Server Workstation splits 2.5 MHz signal into 256 bands, 107 seconds long
1	Perform baseline smoothing (removes wideband noise fluctuations)
2	BEGIN OUTER LOOP: For Doppler drift rates of -50 Hz/s to +50 Hz/s (+/- 10 Hz/s in steps of 0.0018 Hz/s, and +/- 10 to 50 Hz steps of 0.029 Hz/s)
3	BEGIN INNER LOOP: For Bandwidths of 0.075 to 1221 Hz (step through BW's, FFT length 2^n,n=3,4..17)
4	Generate Power Spectrum (in time order)
5	Search for signal above threshold
6	If sufficient time resolution (FFT < 2^15)
7	Match signal power vs time to beam parameters (gaussian)
8	Search for pulse triplets
9	Search for faint repeating pulses
	END INNER LOOP
	END OUTER LOOP

Step 2 increments through combinations of Doppler shift, and for each Doppler increment an FFT is implemented through bandwidths of 0.075 Hz to 1221 Hz. These bandwidths are believed to be adequate to find an artificial signal. Applying wavelet packet analysis at this step generates time-frequency information in a form that can be used for signal detection as discussed in Section 3.

The narrow band search in step 3 appears to be the most interesting area to concentrate on. The narrow band search is analogous to a wavelet based radio receiver. The search is over a band of frequencies, and uses different bandwidths to which the radio operator listens. When a signal is detected above the noise, then it is examined in detail. We note that the bandwidths of the search are constant over the entire search band. This hints that wavelet packets (WP), rather than discrete wavelet transform (DWT), may work best since WP uses a constant frequency resolution, unlike the power of 2 frequency resolution (dyadic grid) in DWT.

In step 6, a potential signal is recreated in the time domain, if the FFT dataset has a fine enough resolution: if the FFT has more than 2^{15} points, then the time domain signal will be too coarse. This appears to have wavelet potential as well, since reconstructing the signal with a finer time resolution would aid the time domain analysis in steps 7, 8 and 9.

3. Narrow Band Signal Detection

Narrow band test signals were created containing multiple pulses. The bandwidth and signal to noise ratio of these signals were varied to determine whether wavelets are viable for detecting narrow band signals of extremely low SNR, and whether they can efficiently detect individual pulses.

The existing client software does a thorough search for pulse triplets at each frequency and bandwidth. Triplets are considered closer to a 'real' SETI signal than a spike, as they seem less likely to be caused by a natural source or interference. Resolving individual pulses requires better time resolution than finding an individual spike.

The triplet signal of length of 20,000 samples was created using a sampling rate of 20 KHz, resulting in 1 second of data. Signal parameters such as SNR, carrier frequency, bandwidth, sampling rate, etc. were varied. An example is shown in Fig. 2, where the signal has a bandwidth of 250 Hz (generated by filtering a string of pulses), and is on a carrier of 3 KHz, lying in the 10 KHz band that is analyzed.

The pulse triplet in Fig. 2 was then buried in noise with an SNR of -5 dB as shown in Fig. 3. This is well below (about two orders of magnitude) the threshold used in the present FFT based algorithm, which is about +13dB above mean noise power.

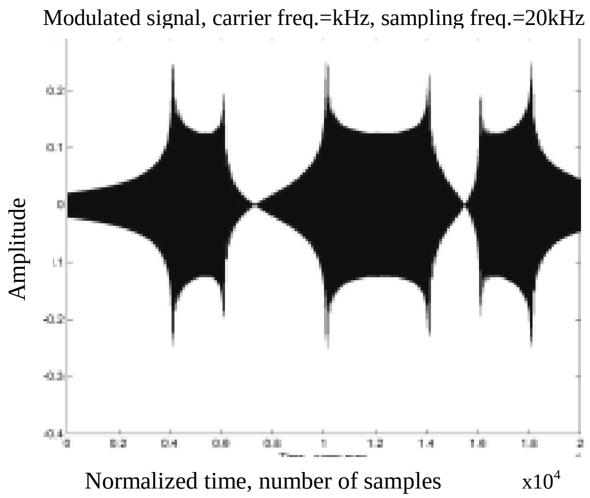

Figure 2. Pulse triplet, BW = 250 Hz, carrier freq. = 3kHz.

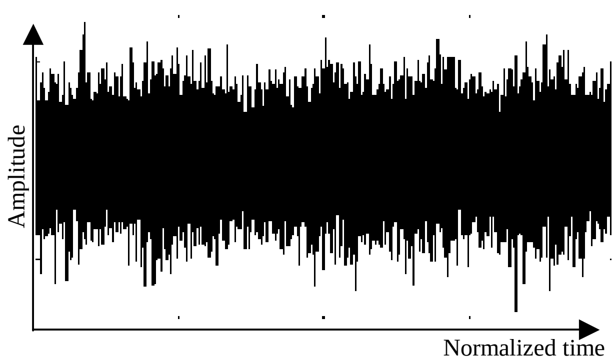

Figure 3. The same pulse triplet in noise, SNR = -5 dB.

A lower threshold such as this will certainly increase the probability of a false detection, however, using such a weak signal clearly demonstrates the search capability of WP.

In standard discrete wavelet transform, each step of the procedure decomposes the approximation coefficients into two parts, resulting in a vector of approximation coefficients and a vector of detail coefficients [3]. Each successive step then repeatedly decomposes the approximation coefficients further into a coarser set of approximation coefficients and detail coefficients. This successive decomposition at each level results in the dyadic spacing of the frequency ordered coefficients, where each level has twice the frequency resolution of the previous level. Using the DWT in simulations has shown that wavelet analysis does not easily find narrow band signals, so the time-frequency property of the signal of interest is not well suited for DWT analysis.

With the WP analysis however, both the approximation and the detail coefficients are decomposed into 2 parts at each level of decomposition. This results in a complete binary tree with superior frequency localization.

Four level WP decomposition of the signal in Fig. 3 was obtained using the coiflet wavelet [2]. The resulting binary tree is shown in Fig. 4a. One of the main disadvantages of the WP analysis is that the number of coefficients from a complete decomposition can become quite large, so a more efficient decomposition is shown in Fig. 4b, obtained using the “best tree” analysis.

The best tree analysis is based on an entropy criterion at each node, allowing the search to continue down the tree if the entropy remains high. In this study the Shannon entropy [2] *E*, was used, which is defined as

$$E(s_i) = -s_i^2 \log(s_i^2) \tag{1}$$

where *s* is the signal and s_i are the coefficients of *s* so that the total entropy is

$$E(s) = -\sum_i s_i^2 \log(s_i^2) \tag{2}$$

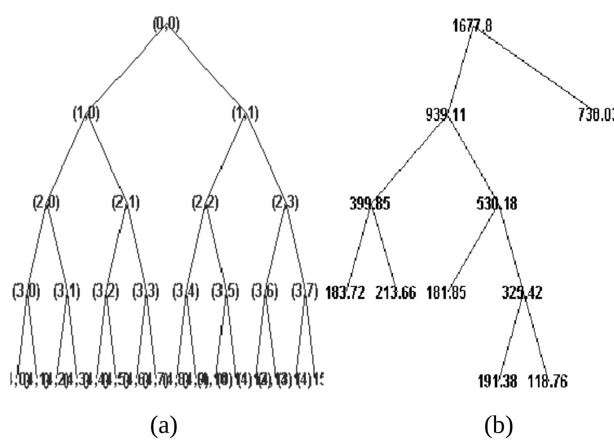

Figure 4. WP decomposition of weak triplet signal, (a) full tree, and (b) entropy based best tree

A time-frequency plot of this decomposition, using the best tree analysis in Fig. 4b, is shown in Fig. 5, illustrating the frequency localization property of wavelet packets. The signal is clearly visible around the frequency of 3kHz.

The detected signal can then be used as an input to the time based analysis routines of the client search algorithm in steps 7, 8 and 9. The reconstruction of this signal (of Fig. 3) in the time domain, shown in Figure 6, clearly indicates the presence of the pulse triplet.

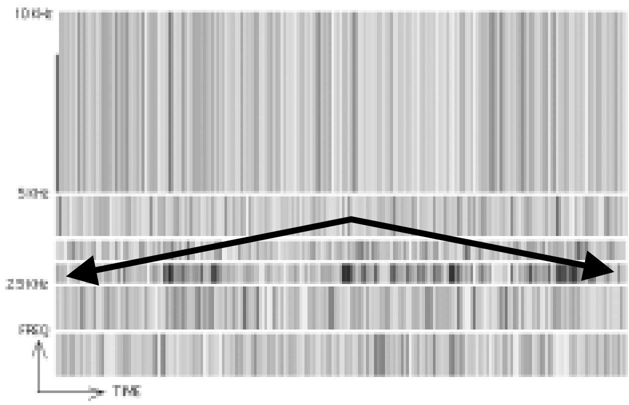

Figure 5. Time-frequency plot of WP decomposition.

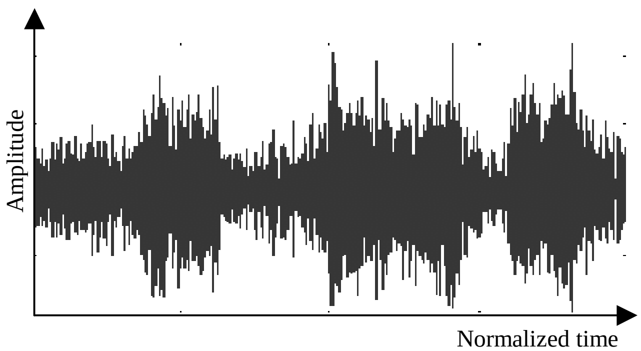

Figure 6. Time domain reconstruction of the -5 dB signal.

The same wavelet packet decomposition technique was then applied to a Doppler shifted signal that was swept from 3 kHz to 7 kHz with additional wideband noise. The time-frequency plot is shown in Fig. 7 where the signal of interest is visible with a positive slope due to the Doppler shift. Removing the slope has the effect of removing the Doppler shift, which then allows reconstructing the unshifted signal.

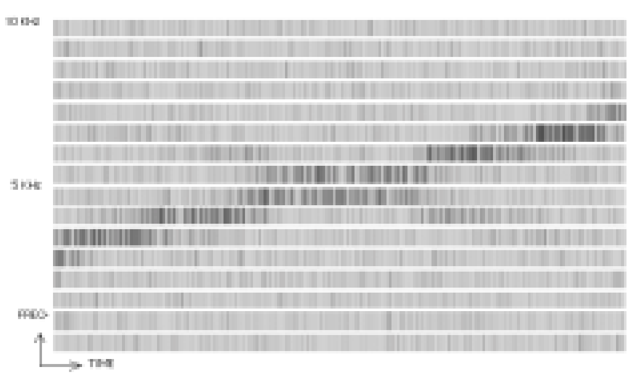

Figure 7. Time-frequency plot of a Doppler shifted signal.

4. Discussion

Wavelet packets appear to be superior to the continuous wavelet transform or the discrete wavelet transform in SETI signal detection due to the nature of the signals that are examined. When used in compression or de-noising DWT allows removing certain frequency bands over a wide range of frequencies, and a dyadic spacing of the frequencies serves well for most types of signals. However, in this application, where a search is performed in a narrow frequency band, wavelet packets appear to perform better at frequency localization than the DWT. Narrow band signals consitute an excellent match for this capability. The time domain analysis of the signal could also be potentially aided because of the superior ability of wavelet packets to accurately reconstruct the signal.

A concern with wavelet packets is the potentially larger computational load. If a signal is present, an entropy-based decomposition efficiently removes the frequency bands that do not contain any significant information, so the number of nodes in the tree remains minimal. Using the above-described approach, a very weak signal was found with only four levels of decomposition, which requires fewer computations then the $N \log N$ calculations FFT requires. More study is needed to estimate the computation time required when a signal is *not* present (the more common case), and to estimate the number of levels needed to reach such a decision.

5. Conclusion

Wavelet packets show significant potential as a means of searching for weak narrow band signals. They compare favorably to the Fourier transform, and the discrete and continuous wavelet transforms. If the time domain analysis of the detected signal is needed, wavelet packets also show a superior capability of reconstructing the signal, preserving time localization information.

Future work will include comparing wavelet packets to the present Fourier techniques to compare detection sensitivity and time domain resolution. An increase in sensitivity that maintains the same probability of false detection could be significant in that it expands the number of sources of the signal. Future work will also examine the computational efficiency of this technique with optimized code.

6. References

[1] E. Korpela, D. Werthimer, D. Anderson, J. Cobb, M. Lebofsky, “SETI @ Home - Massively Distributed Computing for SETI,” *Computing in Science & Engineering*, vol. 3, no. 1, January/February 2001.

[2] M. Misiti, Y. Misiti, G. Oppenheim, J. Poggi, *The Wavelet Toolbox*,” The MathWorks, Inc., pp.6-125 – 6-144, Natick, MA, 1999.

[3] C. Burrus, R. Gopinath, H. Guo, *Wavelets and Wavelet Transforms, A Primer*,” pp. 196-218, Prentice Hall, Upper Saddle River, NJ, 1998.