RFI Mitigation Algorithm for the Eight-meter-wavelength Transient Array

Kshitija Deshpande* and S. W. Ellingson

March 10, 2009

Contents

1 Introduction 2

2 Need for RFI Excision and Limitations of Subjective Methods 2

3 Methodology 6

4 Results and Discussion 8

*Bradley Department of Electrical & Computer Engineering, 432 Durham Hall, Virginia Polytechnic & State University, Blacksburg, VA 24061 USA. E-mail: kshitija@vt.edu
1 Introduction

We present here a radio frequency interference (RFI) mitigation algorithm and illustrate its use for the Eight-meter-wavelength Transient Array (ETA). ETA is a low frequency transient radio telescope consisting of 10 dual polarized active dipoles, located in the mountainous region near Balsam Grove, NC at the Pisgah Astronomical Research Institute (PARI) (35°12’ N, 82°52’ W). It is Galactic noise-limited over 29-47 MHz, has a center frequency of 38 MHz, and a coherent acquisition of 18 MHz from each dipole. The recorded data are sampled at 7.5 mega-samples per second (MSPS) with 7-bit I and 7-bit Q. The processed bandwidth for the data is 3.75 MHz. The off-line data processing is done using the ETA toolchain. In this report, we assume that data are already calibrated to remove variations due to instrumental frequency response and power level variations with time, as described in [2].

2 Need for RFI Excision and Limitations of Subjective Methods

Man-made RFI is a major problem in low frequency radio astronomy. This interference can be classified as narrowband, broadband, or intermittent. Figure 1(a) shows a calibrated spectrogram without any RFI mitigation of an hour-long ETA acquisition. The spectrogram has several other artifacts, namely, the fringes due to multipath at the telescope site and “dropouts”. Figure 1(b) highlights them along with RFI. We remove the fringes in a later stage of data analysis (refer [2] for more details). The other issue with ETA observations is dropouts. These are high-power out-of-band RFI events which cause gain compression in the receiver. In these time intervals, the total power drops significantly. Figure 2(a) shows a time series before RFI mitigation and Figure 2(b) shows an example of a dropout from this time series.

1 http://www.ece.vt.edu/swe/eta/
Figure 1: (a): Calibrated spectrogram from one ETA antenna without RFI mitigation. Frequency resolution: 7.3242 kHz. Time resolution: 0.5 s. (b): The same spectrogram showing dropouts, fringes and RFI. The inset view zooms in to 1/8 the time-frequency span of the spectrogram in Figure 1(a). (Crab_070212 dataset, stream 2 (6Y).)
Figure 2: (a): Time domain result without RFI mitigation. (b): Close-up from (a) showing a dropout. (Crab_070212 dataset, stream 2 (6Y).)
A simple iterative procedure of time-frequency pixel blanking was used previous to the development of the algorithm presented in Section 3 of this report. This procedure was as follows:

1. Examine each pixel in the time-frequency matrix making up the spectrogram. Any pixel which is above a specified threshold is considered RFI. Find such pixels and set them equal to the mean power spectral density (PSD).

2. Obtain a time series by summing the frequency channels. Identify pixels associated with times at which the total power is greater than a specified total power threshold, and set them equal to the mean value of the total power.

3. Obtain a spectrum by taking the spectrogram resulting from Step 2 and summing it over time. Search for frequency bins with magnitude greater than a specified threshold, and set the associated pixels to the mean value of the integrated spectrum.

4. Iterate steps 1 to 3, adjusting thresholds, until a satisfactory level of RFI excision is achieved.

Figure 3 shows the spectrogram resulting from this algorithm. Figures 6(c) and 6(d) show the resulting time series and frequency spectrum, for a particular choice of threshold values. The thresholds applied were 1.4, 1.03 and 1.02 in time-frequency, time, and frequency domains respectively. Thus, in the time-frequency domain, for instance, the pixels with magnitudes greater than 1.4 times the mean PSD were replaced by the mean PSD. It should be noted that all three domains have means equal to 1. Notice that RFI is reduced but not eliminated. A more aggressive choice of thresholds could be more effective but that could also unacceptably damage underlying astronomy. Furthermore, the method is tedious and subjective; that is, different persons applying the algorithm are likely to prefer different thresholds, resulting in different and possibly inconsistent results. We therefore seek an algorithm that automates the process of excising RFI for all low frequency observations from the ETA telescope.
3 Methodology

The proposed new RFI mitigation algorithm is described below.

1. Obtain a time series by averaging the spectrograms over frequency channels. Identify dropouts by visually inspecting the time series and set them equal to the mean power. Since the receiver behaves nonlinearly when the dropouts occur, these regions must be identified and replaced by mean power \textit{a priori}.

2. Replace the largest magnitude pixels in the time-frequency domain data obtained from the previous step, by the mean PSD. Let this fraction of pixels replaced be \( n \). An appropriate value of \( n \) is proposed later. This step removes very strong RFI, which otherwise creates problems in subsequent steps.
3. Compute the standard deviations in time-frequency ($\sigma_{tf}$), time ($\sigma_t$), and frequency ($\sigma_f$) domains for one relatively RFI-free interval. A convenient interval length for ETA data is one file, corresponding to 17.897 s.

4. Examine all time-frequency pixels and replace any pixel with magnitude greater than $m\sigma_{tf}$ by the mean value of PSD. An appropriate value of $m$ is proposed below.

5. Obtain a time series by averaging over frequency channels. Set all pixels associated with times having power larger than $m\sigma_t$ equal to the mean power value.

6. Integrate over spectra and then set all pixels associated with frequencies having values larger than $m\sigma_f$ equal to the mean integrated spectrum value.

The above algorithm is similar to the original algorithm described in Section 2, but it has a better control over the amount of astronomical data loss. We now consider reasonable values for $m$ and $n$.

The criterion for choosing $n$ is that it should not be so large that it significantly affects RFI-free portions of the data. A nominal value of $n$ was found by experimenting with $n = 0.1\%$, $1\%$ and $10\%$. It was noticed that $n = 0.1\%$ was leaving behind unacceptable levels of RFI, whereas $n = 10\%$ was discarding an excessive amount of data in addition to RFI. Thus, our experiments suggest $n = 1\%$ as a reasonable choice.

We now consider how $m$ should be set. Figure 4 shows how the choice of $m$ affects the fraction of data removed. $m = 1$ and $m = 2$ result in an excessive data loss (more than 4.5\%). Values of $m$ from 3 to 5 seem to work fine in the subsequent RFI mitigation steps. Comparing the noise only and noise with RFI results shown in Figure 4 for these values, it is seen that significant amounts of RFI are being removed. Further experiments with datasets of different degrees of RFI contamination showed that $m = 4$ and $m = 5$ leave unacceptable amounts of residual RFI. It can also be seen in Figure 4 that $m \geq 5$ yields diminishing returns. Thus, $m = 3$ is a reasonable choice.
Figure 4: Fraction of data blanked for various values of $m$ with $n = 1\%$ for Crab_070212 dataset, stream 2 (6Y) (solid) and simulated dataset, noise only (dashed).

4 Results and Discussion

Figure 5 shows the spectrogram with the proposed RFI mitigation algorithm applied to the stream 2 of Crab_070212 dataset. Since the dropouts were not removed before applying the old algorithm, comparison of spectrograms in Figures 3 and 5 reveals that the old RFI mitigation algorithm gives poorer results than the algorithm proposed in Section 3. The time domain and frequency domain results of RFI mitigation using algorithms from Section 2 and Section 3 are shown in Figure 6. The time domain result obtained from the new proposed algorithm looks “cleaner” than the one obtained using the old algorithm. The frequency domain results are similar for the two algorithms.

Any algorithm which removes data has the potential to degrade the astronomical signal
as well; we refer to this as “toxicity”. In order to determine the toxicity of the proposed algorithm, it was applied to a simulated dataset with noise containing a dispersed pulse. The pulse has signal to noise ratio of $\sim 10$ at 8.738 ms time resolution and is dispersed at Crab dispersion measure (DM) of 56.791 pc cm$^{-3}$. Figure 7 shows the result. The pulse degrades from $9.9\sigma$ to $8.9\sigma$, which means that RFI mitigation is achieved at the price of some degradation in detection performance. It should be noted that the control parameters of new proposed algorithm can be changed to less aggressive values for datasets with less RFI.

Finally, it should be emphasized that although this algorithm mitigates most of the RFI, it obviously cannot clean the files completely. This is due to the fact that in the operating frequency range of ETA, data are corrupted with some kinds of RFI which are not well-modeled as being localized strictly in the time domain or the frequency domain.
Figure 6: Results without and with RFI mitigation. Thresholds used for mitigation with the old algorithm (described in Section 2) are 1.4, 1.03 and 1.02. For RFI mitigation using the proposed new algorithm from Section 3, values of $n$ and $m$ are taken to be 1 and 3 respectively. (Dataset Crab_070212, stream 2 (6Y).)
Figure 7: Simulated dataset with pulse and noise. (a) and (b) show the dedispersed time series without and with RFI mitigation, respectively. Time resolution: 0.01 s.
References
