First Data from ACoRNE and Signal Processing Techniques

Seán Danaher
for the ACoRNE Collaboration

School of Computing, Engineering and Information Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, United Kingdom
sean.danaher@unn.ac.uk

Abstract. This paper attempts to estimate the performance of the Rona naval acoustic hydrophone array which has been recently upgraded to act as a test bed for an EeV neutrino detector. Preliminary estimation of the array performance is presented together with some preliminary data analysis and reduction techniques.

1. Introduction
This paper describes the Rona hydrophone array, estimate the sensitivity to EeV acoustic neutrino signals and presents a number of signal processing ideas and some preliminary findings from our first attempt at data analysis.

2. The Rona Array
The Rona array is situated near the island of Rona, between the island of Skye and the Scottish mainland. The sea is approximately 230m deep in the area of the array. Eight of the hydrophones have been instrumented for high frequency continuous data acquisition. Six of these are approximately in mid-water, one (hydrophone 8) is on the sea bed and one (hydrophone 3) is approximately 30m above the sea bed. The instrumented part of the array is approximately 1.5 km in length and hence has an acoustic transit time of about 1 s.

The first data recording for acoustic UHE neutrino studies was taken in December 2005. Data were taken at a sampling frequency of 140 kHz and with 16 bit precision. There are 8 Channels yielding a data rate of 1.9x10^{11} B per day. Data was recorded in the first instance for 15 days, yielding 2.8 TB of data.

3. The Data Acquisition Chain
Underwater acoustic systems are most commonly used to monitor continuous signals and are not normally used for the detection of transients. In our application - the detection of bipolar pulses - particular care must be undertaken such that no distortion can occur. Typical hydrophone detection systems depend on a number of stages; the most important for pulse distortion being the response of the hydrophone itself and that of any filters used in the system. Provided however the phase response of the hydrophone/amplifier/filter system is linear over the region of interest then the pulse is simply

http://www.shef.ac.uk/physics/research/pppa/research/acorne.htm

© 2007 IOP Publishing Ltd
delayed. If the phase response is nonlinear, distortion will occur. A linear phase response gives a constant group delay this is analogous to group velocity in continuous systems.

3.1. The Hydrophone System

The Rona hydrophones are commercially available, wide-band, and have a resonance frequency of around 50kHz. In common with many hydrophones only the amplitude response is given by the manufacturers. It is naturally possible to determine the phase response of the hydrophone directly through measurement; however as hydrophones are electromechanical systems [1] it is possible to infer the phase response from the amplitude response. In this study the hydrophones were modelled in terms of ordinary differential equations

\[ \sum_{i=1}^{n} a_i \frac{d^i y}{dt^i} = \sum_{i=1}^{m} b_i \frac{d^i u}{dt^i} \]  

where \( u \) is the input, \( y \) the output, \( a \) and \( b \) constants determined by the physical parameters of the hydrophone and the maximum of \( n \) and \( m \) giving the order of the system; this is equivalent to using a network of inductors, resistors and capacitors [2]. The order of the system is the combined number of inductors and capacitors e.g. a 4th order model is equivalent to a L²C²R circuit. In figure 1 a) a range of fits, to the data sheet amplitude response are shown, from 2nd to 5th order; the 4th order fit is omitted from the graph to avoid clutter. The low amplitude ripples in the hydrophone response between 20 and 30 kHz are caused by manufacturing tolerances in hydrophone manufacture and vary from hydrophone to hydrophone. The corresponding phase responses are shown in figure 1 b). In figures 1 c) and d) the expected hydrophone response is shown. The bipolar pulse keeps its integrity,

![Figure 1](image-url)
independent of the order of the model, with the anticipated pulse shape, based on the simulated acoustic pulse at 1km from a $10^{16}$GeV neutrino interaction. However if the pulse were three times higher in frequency the effect of non-linear phase response becomes very apparent. Further details of this technique are given in [3].

3.2. The filtering and Amplification chain.
One of the major causes of noise in marine acoustic systems is wave noise, which is dominated by low frequencies [4]. At Rona measurements, made during the December campaign, indicated that 85% of the acoustic background energy was below 2.5kHz. As only a few percent of the anticipated neutrino signal is predicted to be in this region, it is sensible to remove these. Analogue filters can easily achieve this; however nearly all designs e.g. Butterworth, Chebychev and elliptical have very non-linear phase responses [5] and will create significant distortion of pulses well above the cut off frequency. Whereas it is possible to design analogue filters with nearly linear phase response e.g. Bessel filters [5], it was decided to record unfiltered data and use digital filters [6]. Not only can these be designed with a perfectly linear phase response but they can be implemented in software giving us much greater flexibility and the option of re-filtering our data using different strategies.

Causal digital filters implement the following operation

\[ y[n] = \sum_{k=0}^{M} b_k y[n-k] + \sum_{k=0}^{N} a_k u[n-k] \]  

(2)

where \( u \) is the input sequence, \( y \) the output sequence, \( a \) and \( b \) are the filter coefficients, \( n \) the sample number and \( k \) a dummy index used in the summation. Filters come in two broad types. Finite impulse response (FIR) filters have outputs which only depend on the current and previous inputs; all the \( a \) coefficients are zero apart from \( a_0 \) which equals one. Infinite impulse response (IIR) filters use feedback, they have outputs which depend not only on the current and previous inputs but also on previous outputs. Non causal filters allow the value of \( k \) to become negative. They produce outputs whose values depend on future inputs or outputs.
Zero pulse distortion using digital filters can be obtained in two ways; either by designing a digital filter which intrinsically has a linear phase response or by filtering the data both backwards and forwards in time. The latter is not possible in real time, however it is perfectly feasible with recorded data. In figure 2 the effect of filtering an acoustic pulse comprising an idealized bipolar signal superimposed upon synthesized noise with a $1/f^2$ spectrum using a number of filters is illustrated. The linear phase response 512 tap Finite Impulse Response Filter (FIR) shows zero pulse distortion, however it has a time delay of 256 samples as is intrinsic to this design. The non-causal Infinite Impulse Response (IIR) filter is used to filter the data both forward and backwards in time. Phase shift is equivalent to a time delay, hence removes any phase distortion and produces a zero time delayed filtered pulse. The behaviour of the Butterworth Analogue filter (7.5 kHz cut off) as used by SAUND I (ref) is shown for comparison.

The decision at ACoRNE was to record unfiltered data and to digitally filter during the data analysis stage. As we are using high specification broadband linear amplifiers we are confident that a bipolar pulse in the water will produce a bipolar electric signal at the point of digitization.

4. Acoustic spectra at Rona
A typical acoustic spectrum at Rona is shown in Figure 3(a). The binning on the x axis is linear which is a common practice with the Signal Processing Community, whereas the Acoustic Community commonly use logarithmic bins; normally fractions of an octave. With linear binning, a white noise spectrum will look flat, whereas with logarithmic binning a white noise spectrum will increase linearly with frequency. Similarly a $1/f$ fall off on a logarithmic plot will become a $1/f^2$ plot on a linear plot. Standard Knusden curves [4] predict a $1/f$ fall off (log) above around 100Hz which will become a $1/f^2$ on these plots.

Figure 3(a) shows a fall off of approximately $1/f^2$ up to around 40 kHz. The feature centred around 50kHz is caused by the hydrophone resonance in Figure 1. In Figure 3(b) the effect of high pass filtering with a cut off frequency of 2.5kHz is shown. In figures 3(c) and 3(d) the effect of using a 1st and 2nd derivative on the signal. Clearly Figure 3(d) shows the flattest spectrum in the region between
5 and 40kHz, which is the region in which we expect the neutrino acoustic signal to lie. Though there is some roll-off at the lower frequencies. This spectral flatness is very important for the next stage in the process, matched filtering.

5. Matched Filtering
Matched Filtering is a technique of optimizing the signal to noise; in our case we wish to optimize the signal to noise ratio produced from a standard bipolar pulse. Firstly consider the discrete autocorrelation function

$$ R_{uu}[n] = \sum_{k=-\infty}^{\infty} u[k]u[n + k] $$

Where $n u$ and $k$ have been previously defined. When $k=0$ the autocorrelation function gives a maximum value which is proportional to the total energy of the signal. Given a sampled signal $u[n]$ and a digital filter with an impulse response (the output from a one padded by an infinite number of zeros) sequence of $h[n]$ then the output from the filter is given by the convolution sum

$$ y[n] = \sum_{k=-\infty}^{\infty} u[k]h[n - k]. $$

This is very similar to autocorrelation apart from a different sequence $h[n]$ being used and the change in sign. If however $h[n]$ is a time reversed copy of $u[n]$ the two processes are identical as far as the signal is concerned. If the noise is completely white then the anticipated noise contribution will be independent of the shape of the filter (apart from an overall gain). For white noise the maximum signal

![Matched Filtering](image.png)
to noise ratio is achieved by filtering the data with a time-reversed copy of the anticipated signal (bipolar pulse).

Alternatively in the frequency domain, if the DFT is the discrete Fourier Transform and the IDFT is the inverse DFT then

$$y[n] = \text{IFFT}(\text{DFT}(u[n])\text{DFT}(h[n]))$$  \hspace{1cm} (5)

The output from the filter can be determined either by doing a convolution in the time domain or a multiplication in the frequency domain. In order to get the maximum signal to noise ratio out of the filter, the filter has to be matched to signal (bipolar pulse) in both relative amplitude and phase at every frequency.

In terms of amplitude for frequency regions where signal to noise ratio is low the filter is made less sensitive and for regions with a high signal to noise the sensitivity is increased. Again if the noise is white (or made white by a filtering process) the spectrum of the filter has to simply match the spectrum of the pre-whitened signal. As illustrated in Figure 3 d) taking the 2nd derivative of the noise produces a good approximation to a white noise distribution, though naturally more sophisticated (and computationally intensive processes) could be used.

The phase response of the filter has to also match that of the pre-whitened signal. Consider the DFT as a vector in $n$ dimensional space. Given two DFTs, without loss of generalization we can adjust the amplitude of the DFT inputs such that the vectors are of unit length. The process of multiplication is equivalent to a dot product. If the vectors are real (which will be the case if the inputs to the DFT are even functions; symmetrical round zero time) then the maximum product will be 1 where the two DFTs are identical. This will yield a value of $y[0]$ also equal to one. If the vectors are complex then the maximum will also be 1 but will be achieved if the vectors are complex conjugates. Taking a complex conjugate in the frequency domain is equivalent to reversal in the time domain. A good approximation of matched filtering can be achieved by passing the 2nd derivative of the data through a filter with an impulse response which is the time reversed 2nd derivative of the anticipated pulse. In Figure 4 a 20 mPa bipolar pulse, corresponding to a $2 \times 10^{10}$ GeV neutrino at 1 km, is artificially injected into typical data from Rona. (hydrophone 5, 9th December 2005). The pulse, highlighted with a continuous line is difficult to see in the original data but very easy to spot after using the matched filter.

### 6. Signal triggering

Whereas a matched filter is designed for a “steady state” noise background, i.e. one whose spectrum remains constant it is not good at filtering out transient events such as that created by bio-noise. The philosophy to be used at Rona is to investigate the additional sources of background triggering in the first instance. Once these sources have been identified and characterised the idea is to generate alternative filtering strategies to remove these transients. It was decided therefore to produce a large number of triggers based on the energy inside a window of similar duration to the expected

|   | 4s | 5s | 6s |
|---|----|----|----|
| H1| 5.1| 5.1,5.6| 5.6 |
| H2| 4.1| 5.8 |
| H3| 4.9| 4.9,5.9| 5.9 |
| H4| 5.3| 5.3 |
| H5| 3.9 |
| H6| 5.7 | 5.7 |
| H7| 5.2 | 5.2 |
| H8| 5.8 | 5.8 |

Table 1. Allowing for combinatorial triggers
pulse width (25 samples at \( f_s = 140\text{kHz} \)). Four triggers were chosen: pressure, \( p \) (after high pass filtering), \( dp/dt \), \( d^2 p/dt^2 \) and the matched filter output based on the anticipated bipolar pulse at 1km. These are designated types 1, 2, 3 and 4 respectively. The data is analysed in 10 minute segments and time stamp of the highest 20 triggers, (5 of each type), are saved together with 2001 points around the trigger; from sample -1000 to 1000 - equivalent to a \( \pm 7.1\text{ms} \) window around the trigger. As the more energetic events will fire more than one trigger, priority is given to type 4 then type 3 etc. Two triggers are assumed to be coincident if the time stamps are within 500 samples.

This reduces the data by a factor of about 50 down to approximately 50 GB, which is easily stored using current hard disk technology. The energy of the triggers naturally depends on sea state and the presence of other noise sources, but the triggers typically are in the region of \( 10^{-10} \text{ to } 10^{-11} \text{ GeV} \) at a distance of 1km.

### 7. Coincidence methods

In order to study all potential event classes the triggering strategy was designed to force triggers in each 10 second segment. It is believed that most of the genuine background events will effectively be from point sources and should spread isotropically. The array is approximately 1.5 km across corresponding to a 1 s window for coincidence triggers. In the first instance we looked at 5 fold plus triggers. With 8 hydrophones there is 1 possible eight-fold, 8 seven-fold, 28 six-fold and 56 five-fold combinations. An initial pass on the data yielded over \( 10^{10} \) trigger combinations.

With a 5 fold plus trigger however it is possible not only to calculate a source position using triangulation, but to determine an error based on hydrophone timing consistencies. The algorithm is based on Singular Value Decomposition (SVD) \(^7\) and determines from the hydrophone locations and a set of times the source location, which minimizes the sum of the squares of the distance errors. The error inside the algorithm is equivalent to a timing error for each hydrophone, which can be converted to an equivalent position error. For each hydrophone the consistency between the presented time and inverted times, \( dt \) can be converted to an equivalent distance error.

\[
\sum_{i=1}^{n_{hit}} |c(t_i - t_{in,i})| = \sum_{i=1}^{n_{hit}} |c(t_i - t_{in,i})| = \sum_{i=1}^{n_{hit}} (t_i - t_{in,i})
\]

Where \( n_{hit} \) is the number of hits and \( c \) the velocity of sound (\( c = 1500 \text{ ms}^{-1} \)).
A table is constructed with 8 rows, one for each hydrophone, and 600 columns corresponding to each second in the 10 minute data sample. (Table 1). The time stamp for each hydrophone trigger is placed in two columns, one column corresponding to the time in seconds at which the trigger occurs and one column above or below depending on whether the fractional second is above or below 0.5. If multiple combinations can occur then the columns are expanded up to a maximum of 1000 possible events (ordered in terms of decreasing energy) in a 1s interval. This is illustrated in table 1 where four possible combinations can occur.

Currently the four types of triggers are analysed separately as the combinatorial problem increases factorially with events and it is not unreasonable to assume the pulse shapes will remain comparatively constant over 1 km. The initial analysis was done using type 4 triggers. For each potential set of hydrophone times trigger a timing check is done based on causality (87,299 candidates). The trigger is then passed to the vertex checking code which must yield a real vertex (15,024,299 candidates), a z position in the water and a satisfy a “chi squared” cut currently set at 100m. This yields a total of 123,330 five fold, 74,839 six fold, 24,830 seven fold and 9,555 eight fold candidates; 232,554 in total.

A typical event is displayed in Figure 5. The signals in hydrophones 1 to 4 are displayed on the left and hydrophones 5 to 8 on the right. This event has therefore triggered hydrophones 1-4 and 7-8. The x-axis is in ms and the y-axis in ADC counts, each corresponding to 50 µPa. The pulse heights therefore correspond to pulses in the region of 15-30 mPa. An interesting feature of this and other pulses is that some of the hydrophones, e.g. hydrophone 4, produce pulses which are inverted with respect to the others, presumably because the pre-amp has been wired with the opposite polarity. The “chi squared” error of 37.2 corresponds to an average hydrophone position error of 37.2/6 = 6.2 m.

8. Conclusion
The ACoRNE collaboration has implemented a continuous data acquisition system at a naval hydrophone array at Rona in Scotland. The first data were acquired in December 2005 and the results of a preliminary analysis of these data are presented. To date a number of different signal processing techniques have been employed and work is underway to classify the observed background events. This will facilitate the construction of dedicated filters capable of eliminating unwanted events at Rona or in a potential future large area Mediterranean hydrophone array.

References
[1] Bentley J P 1995 Principles of Measurement Systems 3rd Edition. Longman ISBN 0 470 23445 8
[2] Levine W S 2000 Control Systems Fundamentals CRC Press ISBN 0273 30325 69
[3] Thompson L This volume
[4] Urick R J M 1975 Principles of Underwater Sound McGraw-Hill ISBN 0-070-66087 5
[5] Huelsman L P 1993 Active and Passive Analog Filter Design. McGraw-Hill. ISBN 0 07 030860
[6] Ifeachor E G, Jervis B W 1993 Digital Signal Processing (A Practical Approach) Addison Wesley. ISBN 0-201-54413 X
[7] Danaher S, Fegan D J , Hagan J, 1993 Astropart. Phys 1 357