LOCALISATION OF WHALES USING ACOUSTICS

P.R. White ISVR, University of Southampton, Highfield, Hants, UK
D.C. Finfer ISVR, University of Southampton, Highfield, Hants, UK
T.G. Leighton ISVR, University of Southampton, Highfield, Hants, UK
C. Powles ISVR, University of Southampton, Highfield, Hants, UK
O.N. Baumann ISVR, University of Southampton, Highfield, Hants, UK

1 INTRODUCTION

There is an increasing awareness of the potential environmental impact of anthropogenic noise, in particular sonar, in underwater environments. This impact is most dramatically illustrated by the mass strandings of beaked whales that have been attributed to the use of military sonars [1, 2]. In order to avoid endangering cetaceans (whales and dolphins) Navies monitor for cetaceans prior to employing high power sonars. Similarly, in many part of the world, there are legal requirements on commercial operations, such as seismic surveying, to undertake suitable monitoring for cetacean so as to avoid unnecessary disturbance.

Presently most cetacean surveying is undertaken visually. Evidently visual monitoring is of limited utility since its effectiveness is strongly dependent upon the prevailing weather conditions and much of the time cannot be conducted at all, e.g. at night and in fog. Further problems arise because some species undertake long dives, spending a high proportion of their time below the surface. Other species are particularly cryptic, the prime examples being the beaked whales; spotting these animals even when they are at the surface can only be reliably achieved in good weather conditions.

Passive acoustic monitoring offers a powerful alternative methodology. One is able to detect a vocalising animal at considerable distances, commonly at ranges that exceed those at which visual monitoring is feasible. The performance of passive acoustics also is much less dependent on the weather and can be implemented 24 hours a day. Further there is considerable potential for automating such processing. Passive acoustic monitoring is not without shortcomings; obviously passive acoustics can only detect animals when they vocalise. Some species of baleen whales are not very vocal, whilst some species of odontocetes (toothed whales) predominantly vocalise during dives, largely remaining quiet when they are near the surface. In many respects visual and acoustic monitoring complement each other and a comprehensive monitoring system may employ elements of both modalities.

The distance between an animal and the survey platform is an important piece of information in any decision regarding mitigation. In the case of visual monitoring it is relatively straightforward to estimate range once an animal has been spotted. This paper seeks to explore techniques for estimating range to animals based on passive acoustics. There has been considerable work on the use of acoustics for localisation of animals. In particular the work of Spiesberger describes methods for localisation based on hyperbolic and isodiachronic methods [3]. These methods rely upon one estimating the relative delay between sensor pairs and then inverting a propagation model to infer source location.

There are broadly two classes of passive acoustic monitoring system that are considered. The first, the short aperture systems, which are typified by towed arrays. These are convenient for deployment, but the accuracy of the localisation is severely impaired by their short aperture. Most such systems typically employ arrays with lengths <10m. The large aperture systems employ sensors positioned over distances on the order of kilometres. Such systems may consist of sensors deployed through towed arrays, sonobuoys, bottom mounted sensors or combinations of these. The obvious disadvantage of large aperture systems is the difficulty in deploying them, bearing in mind that data needs to be synchronously captured from all of the sensor sites. However the large

aperture systems provide location estimates with potentially much greater accuracy than the small aperture systems.

Passive acoustic localisation methods have been applied predominantly to the vocalisations of sperm whales (*Physeter macrocephalus*). During dives, these animals produce distinctive trains of clicks, which they use for echolocation purposes. These clicks are emitted at extremely high amplitudes (on-axis sound pressure levels 230-240 dB re 1μPa @ 1m [4,5]) and so are audible across large ranges. Most commonly the train of clicks occur at regular intervals. The precise interval between click pairs varies as a function of time, so that the resulting time series is approximately periodic. The inter-click intervals are typically of the order of 1s.

Estimating the relative delay for sperm whale vocalisations is a non-trivial problem. For large aperture systems the time delay between sensors maybe of the order of seconds and so encompasses periods that may span several inter-click intervals. This adds considerable complexity to the delay estimation process, which will be discussed in more detail later. When processing data from short aperture arrays, the expected time delays are short, normally in the millisecond regime. But the arrays are commonly towed at relatively shallow depths, around 10-20 m, and so the received signals often contain strong surface reflections and may also exhibit reflections from the sea bottom. The problem of delay estimation becomes significantly more complex when there are multiple animals vocalising simultaneously, resulting in interleaved click trains contaminated by multi-path effects.

This paper describes the analysis of sperm whale recordings collected from a bottom mounted sensor array with a long aperture. These data being made available as part of a recent workshop (2nd International Workshop on Detection and Localization using Passive Acoustics, Monaco, 2004). The data were collected at The Atlantic Undersea Test and Evaluation Center located at the tongue of the ocean, close to Andros Island, Bahamas. The problem of estimating delay information from such data is explored, examining and contrasting various approaches. The resulting delay data are then used to infer the animal location using an inversion routine.

2 DELAY ESTIMATION

The estimation of delay can be optimally computed (assuming a Gaussian model for the noise and signal) based on cross-correlation [6]. In practice delay estimators based on cross-correlation offer good performance even when the measured signals have properties that deviate from those invoked in order to claim optimality. Consequently the most common methods for computing delay are based on some form of cross-correlation. In this section we shall discuss three methods for performing delay estimation.

2.1 Correlation of the Raw Time-Series

When analysing raw time series data from a vocalising sperm whale on a large aperture array, there are several practical issues that need to be considered. One example of this arises because the cross-correlation function contains multiple peaks. The reason why this can be problematical can be seen when considering the acoustic data from a sperm whale shown in Figure 1. The correlation function for this signal is shown in Figure 2. For robust delay estimation one desires this function to contain a single peak, corresponding to the appropriate delay. However for a click train sequence the correlation function contains many peaks. It is a difficult task to determine which of these many peaks is the “correct” value that aligns the two signals (further analysis of this data suggests that the correct delay is -2.1s).

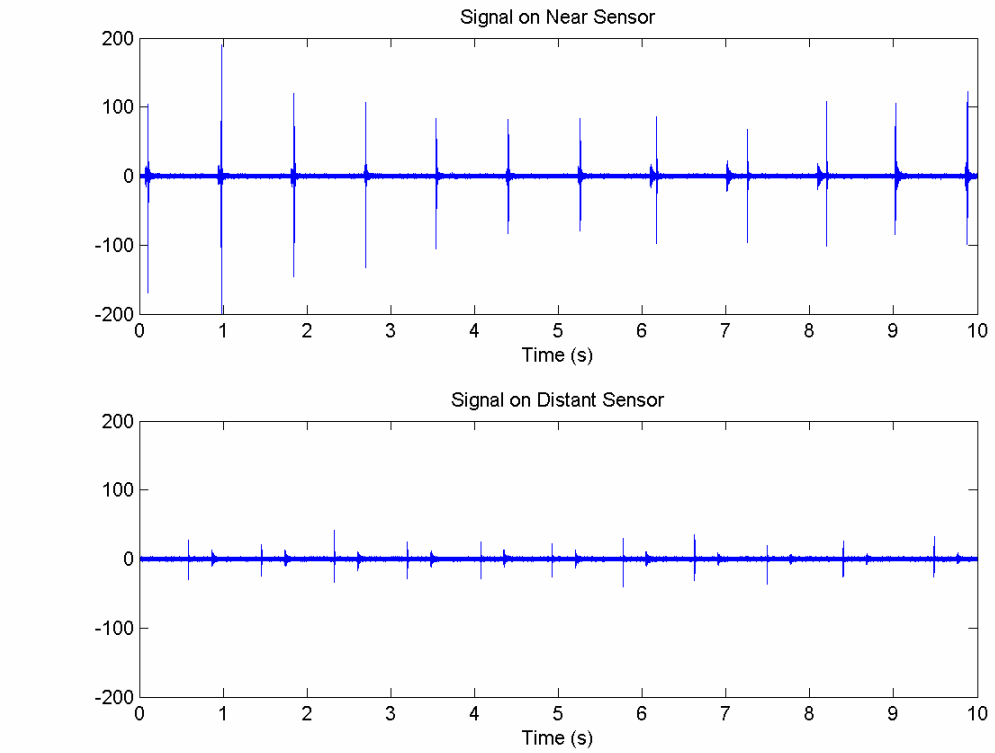

Figure 1. Example Sperm Whale Click Trains Measured on Two Bottom Mounted Sensors

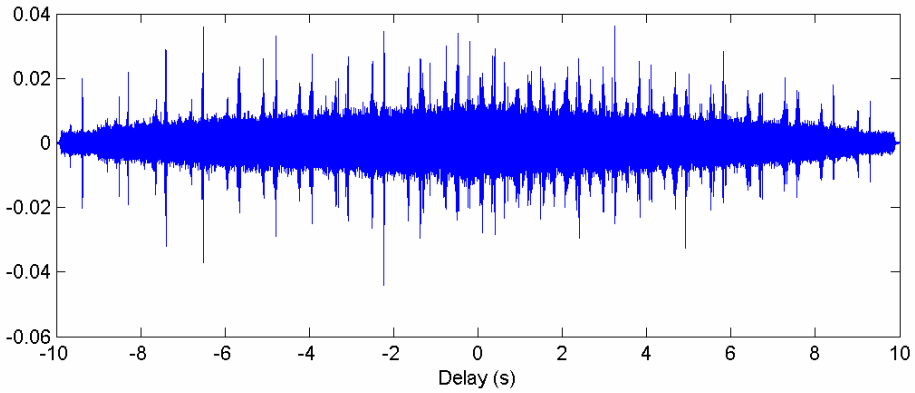

Figure 2. Cross-Correlation Function Computed for the Data Shown in Figure 1

The problem is further exacerbated by the fact that there is a lack of coherence between the signals measured at distant sensors. Figure 3 shows an expanded view of two clicks, chosen to ensure that they correspond to the same click emitted by the animal. It is evident that the detailed structure of clicks is not faithfully reproduced on the two sensors. This figure also highlights that each individual click has a complex structure, being composed of several pulses. This pulsed structure can be related to the physiology of the sperm whale [5]. The click is generated at the upper frontal region of the sperm whale. It is then reflected between a pair of air sacs before exiting the front of the whale. Additional reflections occur between these sacs. This causes each click to be made of a series of pulses. The observed character of the click depends on the orientation of the animal relative to the observer. Figure 4 depicts a series of sperm whale clicks recorded at a single sensor. It shows a series of 240 consecutive clicks aligned based on the leading edge of the initial pulse. In the plot each vertical line represents a single click, where amplitude is encoded via colour. This figure demonstrates how the internal structure of the click varies on a click-to-click basis. The

source of this variability is believed to be the varying orientation of the animal relative to the sensor as it moves through the water. This variability provides part of the reason why clicks have a different structure when observed on different sensors. Coherence is further reduced by the fact that the signals have propagated considerable distances through different paths in the ocean, and so are subject to different random perturbations of mean acoustic properties.

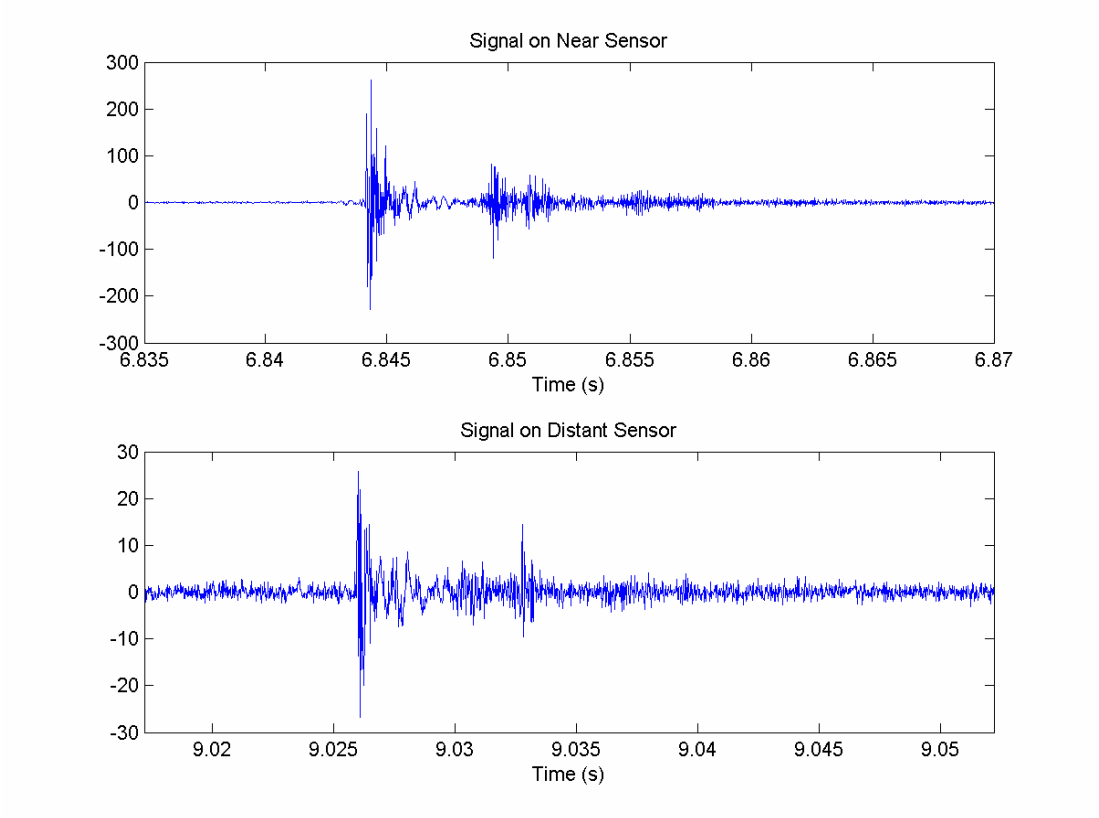

Figure 3. Example Sperm Whale Click Trains Measured on Two Bottom Mounted Sensors

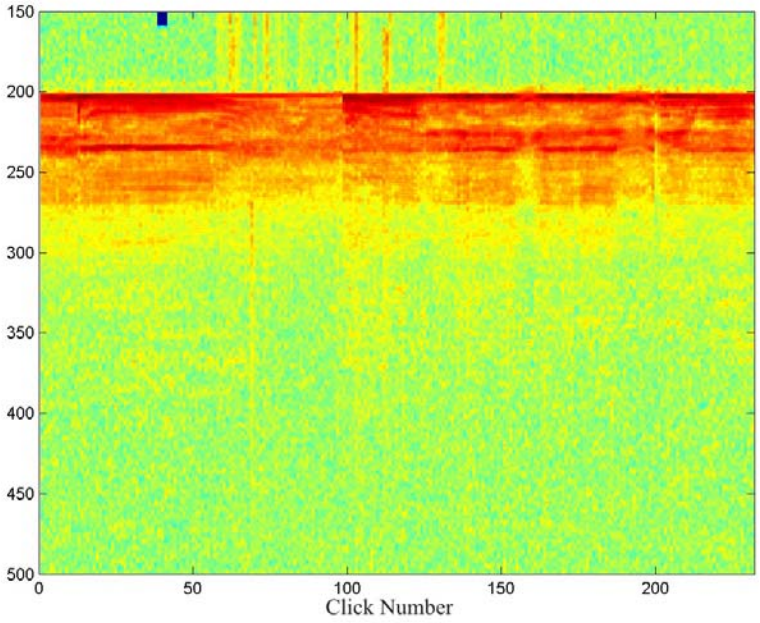

Figure 4. Plot of Clicks Aligned on the First Pulse

2.2 Correlation of the Detection Statistic

An alternative to correlating the raw time series data (coherent correlation) is to compute the correlation of the envelopes of the signals (incoherent correlation). Rather than simply analysing the envelope of the signal, better performance can be achieved if an optimised detection statistic is computed [7] that represents the envelope of the normalised whitened signal. The incoherent approach provides estimates that are more robust to subtle changes in the waveform. This robustness is attained at the cost of a reduction in potential accuracy of the method. It should be noted that the extra accuracy associated with coherent processing is only available if the signals are truly coherent and Figure 3 suggests that, for these data, this is not the case.

Figure 5 shows the detection statistic, which forms the basis of the incoherent processing, computed for the same data depicted in Figure 1. These detection statistics are smoothed temporally using a 1024 point Hanning window. The smoothing process improves the correlation function by removing fine detail in the envelope, but again affects the potential accuracy of the method. The 1024 point smoothing window has support of roughly 20 ms which corresponds to the approximate length of a click. Consequently this smoothing removes features within a click that may not be coherent between sensors. Figure 6 shows the correlation function computed based on these detection statistics. This case demonstrates the effectiveness of employing the detection statistic in preference to the raw time-series data. The correlation function in Figure 6 exhibits a clear peak close to the correct delay of -2.1s, in contrast to the correlation function in Figure 2. This correlation function still contains a large number of false peaks that can potentially generate erroneous estimates of delay, inducing large errors in the consequent estimates of location.

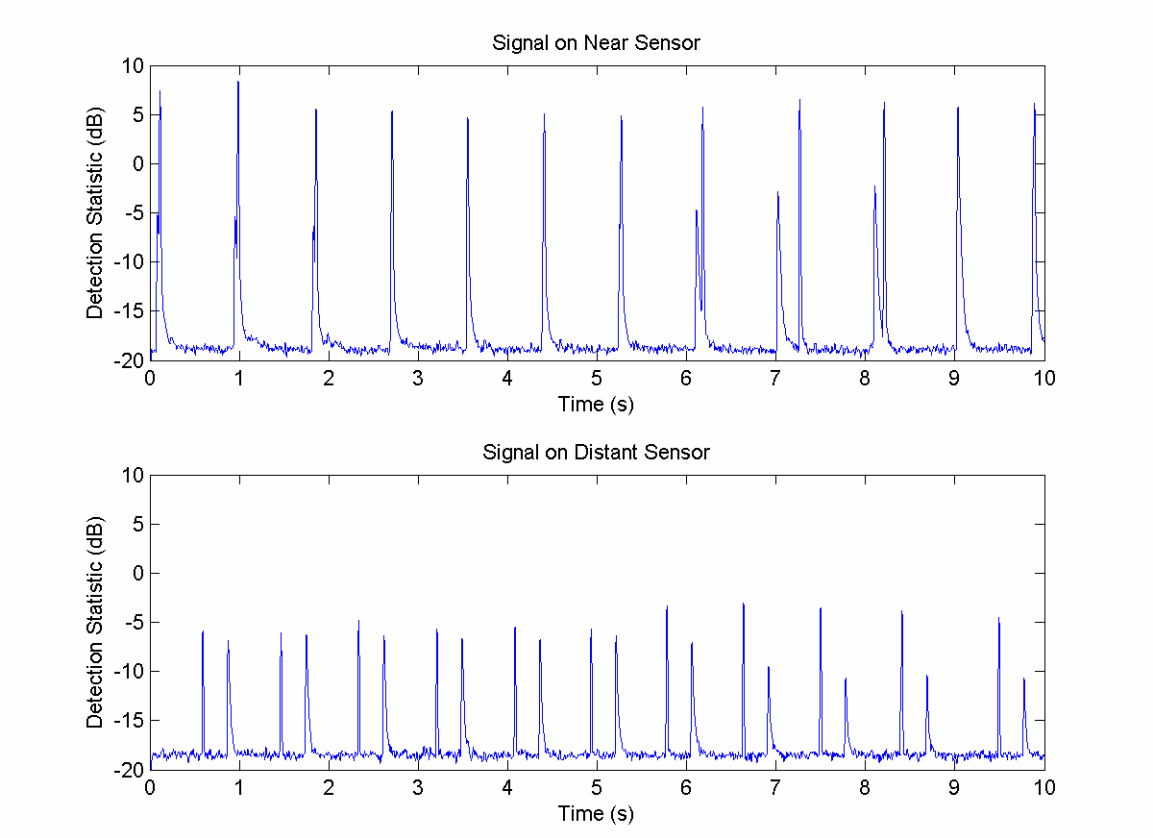

Figure 5. Smoothed Detection Statistic Plotted for Data Shown in Figure 1

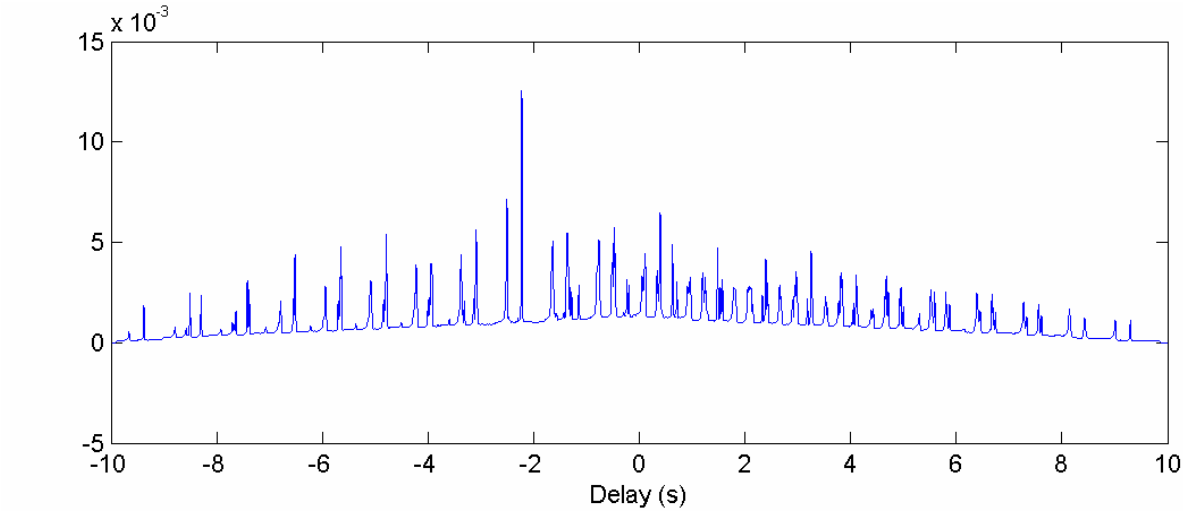

Figure 6. Correlation Function Computed Based on the Smoothed Detection Statistics

An illustration of the problems associated with this approach is shown in Figure 7. This depicts delay estimates computed using windows of 30s of data. The window is translated by 7.5s between estimates, i.e. overlapped by 75%, before the next delay estimate is computed. This process is repeated to cover a period of 5 mins. From Figure 7 we see that in this instance there are 4 periods in which the delay estimates are very inaccurate. Furthermore, close inspection of the data between 3 and 4 minutes reveals that the estimated delay in that interval is also inconsistent with the other results, being offset by roughly by roughly 1s. Hence for this data set roughly 30% of the delay estimates are seriously in error.

The continued failure of this approach results from the ambiguity caused by the fact that the sperm whale click train is close to being periodic, so that the correlation function contains a large number of spurious peaks. Correlation methods inherently introduce spurious peaks, so that in order to obtain robust delay estimates we consider an alternative philosophy.

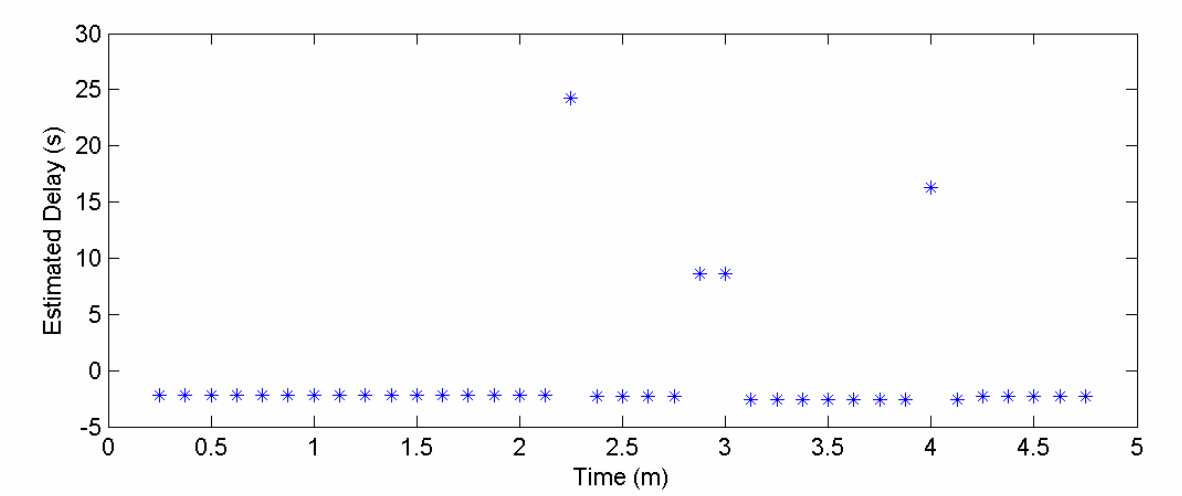

Figure 7. Delay Estimates Computed for a Sliding Window

2.3 Time of Arrival Differences (TOADs)

The final method is a histogram-based technique. The initial requirement is to detect all of the clicks in the time series. The detection statistics, plotted in Figure 5, provide an optimal basis for this detection. The fact that the detection statistic is pre-normalised means that fixed absolute thresholds can be used. A double thresholding scheme [8] is employed to estimate the extent of each click, and the peak in the detection statistic is considered as marking its location in time (this time is denoted $t_{p,n}$ for the n^{th} click on the p^{th} sensor). The fact that the clicks are not periodically spaced in time means that the time differences $d_{m,n}=t_{1,m}-t_{2,n}$ have a statistical distribution which exhibits a strong peak at the true time delay between sensors. Thus by estimating the probability density function (pdf) (here based on histogram methods), and looking for peaks in the pdf of $d_{m,n}$, one can estimate the delay.

In a multi-path environment, reflections in received time series generate a secondary peak in the distribution. If all the reflections are detected, then the height of this secondary peak matches that of the peak due to the direct path and may lead to an erroneous delay estimate. To make the method robust one needs to be able to discriminate between reflected returns and the direct paths. This discrimination only needs to detect a significant proportion of the reflections, so that the secondary peak is reduced in amplitude. There are several features of the waveforms that can be used as the basis of this discrimination. The most reliable of the ones explored for this paper was to estimate the rate of decay of the detection statistic. The reflections, having come from a rough surface (that of the sea), have a larger decay time than the direct path clicks. Hence by removing the detected peaks in the signal, corresponding to clicks with a large decay time, one removes reflections. The decision to remove a click can be based on thresholding the decay rates, or by rejecting a fixed percentage of clicks with the slowest decay rates. Mistakenly rejecting a few clicks which correspond to direct paths reduces the height of the primary peak. However it does not seriously affect performance, so long a significant majority of rejected peaks correspond to reflections.

Figure 8 displays the result of such processing. The direct paths for the detection statistic, shown in the lower frame of Figure 5, have been highlighted. Inspection of Figure 8 also illustrates how the detection statistics corresponding to the reflected paths have a significantly longer decay than the direct paths, a feature even more evident in Figure 9.

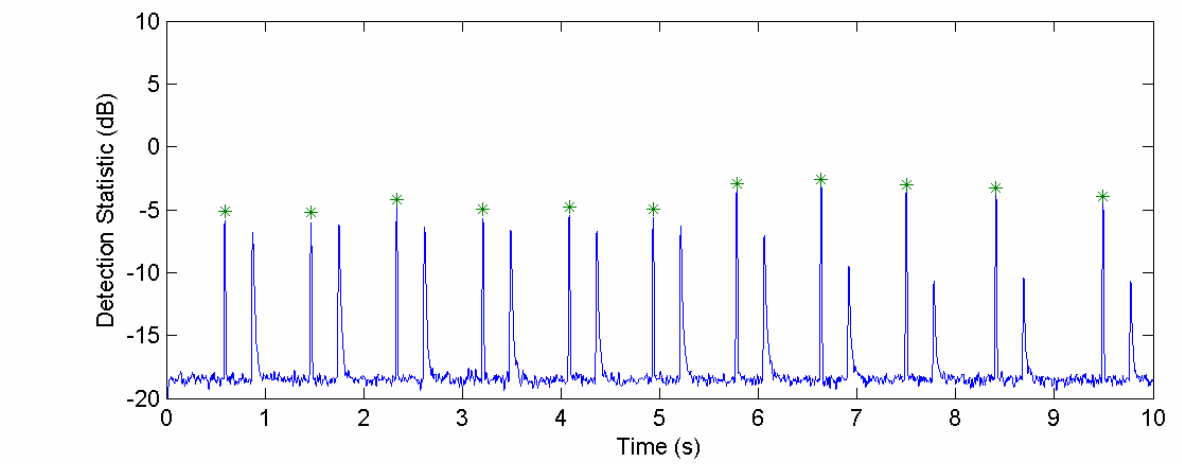

Figure 8. Smoothed Detection Statistic, as Lower Frame of Figure 5, Direct Paths Indicated by *

The localisation scheme adopted in this paper also requires one to estimate the additional delay associated with the reflected path in each channel. To achieve this, the peaks associated with each click are aligned. A new reduced detection statistic is then computed for each channel by taking the median of the aligned detection statistics. A median operator is preferred to the mean because of its inherent robustness, which means that the results are not adversely affected by a small number

of errors, such as false identification of direct or reflected path. An example of this median detection statistic is shown in Figure 9. These results are computed over an extended time axis, which covers a sufficiently long period so as to include the following click in the sequence. Because the clicks are emitted at intervals which vary, then peaks associated with the subsequent click are not reinforced by the median operator and thus the resulting amplitude of this peak is attenuated. However the echo follows the direct path by an amount which almost constant over the analysis window, so it is not significantly attenuated by the median operator.

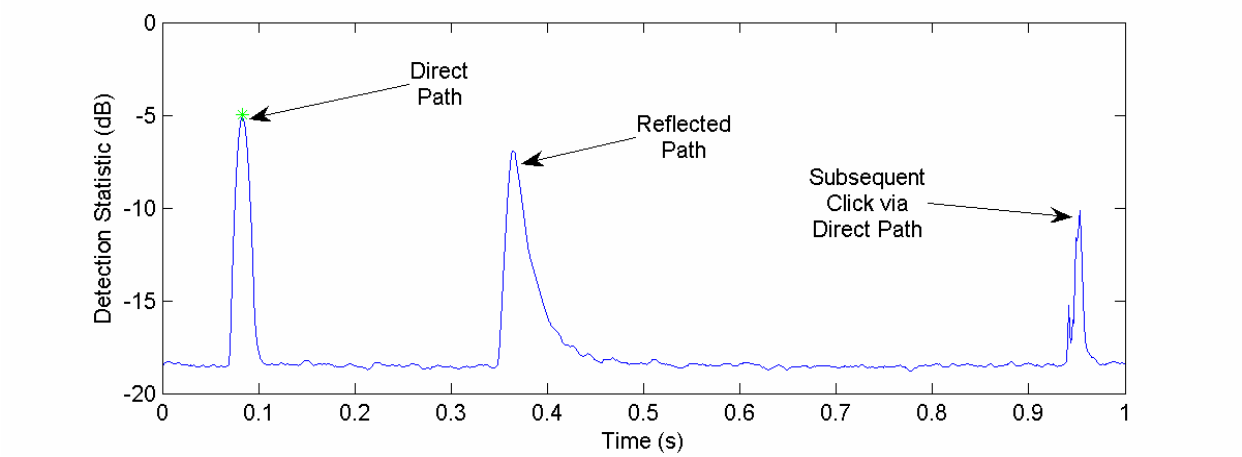

Figure 9. Median of the Aligned Detection Statistic

3 LOCALISATION

Section 2 describes how one can compute the delays between hydrophone data recording sequences of sperm whale clicks and, furthermore, how one can compute the additional delay associated with the surface reflected path. This section discusses the method used to infer source location based on the measured time delays.

There are a variety of methods that can be used to achieve this localisation [9,10,11,12] based on propagation models of varying complexity. We seek a scheme in which the computational demand is small, which leads us to consider methods based on simplified propagation models. The technique employed in this work exploits the simplest propagation scenario, namely an isospeed medium – implying rectilinear propagation. In practice the medium will certainly not be isospeed but we accepted degradation associated with this assumption as the penalty to be paid to achieve the requisite algorithmic simplification.

The algorithm described is a hyperbolic localisation scheme that is a variant of the method attributed to [11] in [10]. However we are dealing with an over-determined system of equations: six sensor measurements are used, whereas in fact four would be sufficient for this application. The location we compute minimises the cost function

$$\Psi(\mathbf{r}_s) = c\sum_{k=2}^{N} \left(\tau_{1,k} - \hat{\tau}_{1,k}(\mathbf{r}_s)\right)^2 \quad (1)$$

$$\hat{\tau}_{1,k}(\mathbf{r}_s) = \left(d_k(\mathbf{r}_s) - d_k(\mathbf{r}_s)\right)/c \quad (2)$$

$$d_k(\mathbf{r}) = \|\mathbf{r}_k - \mathbf{r}_s\| \quad (3)$$

where $\mathbf{r}_s$ is the assumed position vector of the source, $\mathbf{r}_k$ is the position vector of the k^{th} sensor, $d_k(\mathbf{r})$, is the distance from the source to the k^{th} sensor, c is the speed of sound in the medium, $\tau_{1,k}$ is the measured delay between the first sensor and the k^{th} sensor and $\hat{\tau}_{1,k}(\mathbf{r}_s)$ is the modelled delay between the two sensors assuming the source location $\mathbf{r}_s$.

The above cost function is expressed in terms of path length differences rather than delays, as witnessed through the inclusion of the factor c in (1). This factor ensures that the cost function and its parameters are dimensionally matched. Because the speed of sound is a constant, this choice makes no difference to the optimum source location. It does, however, slightly simplify the optimisation procedure, mitigating the numerical problems that can occur because of the very different scales of the cost function (which temporally is measured in seconds) and the source location (which is measured in metres).

Our localisation algorithm utilises the surface reflected paths [12] (see Figure 10). The surface reflections offer measurements effectively made on a virtual sensors, doubling the number of elements in the array and dramatically increasing the vertical aperture, and hence increasing the accuracy with which one can estimate depth.

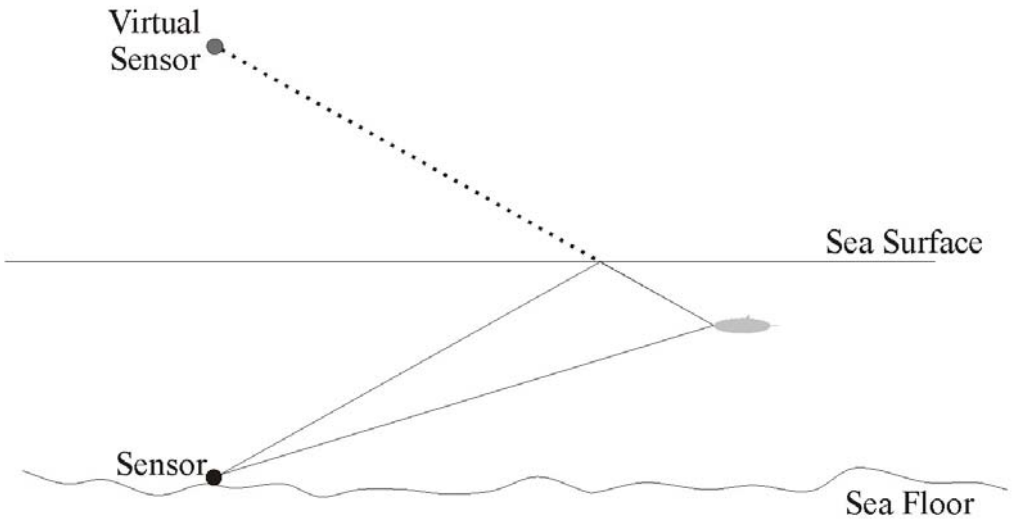

Figure 10. Geometry of Virtual Sensors via the Surface Reflections

We choose to locate the minimum of the cost function (1) using the Nelder-Mead simplex algorithm. In an attempt to reduce the probability of convergence to a local minimum, the algorithm is initialised using several different starting conditions, and the minimum yielding the lowest overall value for the cost function is selected.

4 RESULTS

The results in this section describe the above methods applied to data collected on a series of bottom mounted sensors. The sensor locations for the particular subset of elements used in this study are shown in Table 1. It should be noted that the aperture of the array (excluding virtual sensors) in the horizontal plane is roughly 3 km by 5.6 km, whereas the vertical aperture is only 0.2 km. This asymmetry in the distribution of sensors provides the motivation for employing the surface reflections. By exploiting the virtual sensors the vertical aperture increases to over 3 km, providing

a greater balance in the sensor geometry, increasing the robustness to the inherent shortcomings of the rectilinear propagation assumption.

Sensor Number	x-Coordinate (m)	y-Coordinate (m)	z-Coordinate (m)
1	3661	-1236	-1554
4	-1985	-3111	-1362
5	1349	-4285	-1523

Table 1. Sensor Locations

Figures 11 to 13 summarises results obtained from a total of 5 minutes of acoustic data, from three hydrophones, the sampling rate being 48 kHz. Each estimated source location is computed on the basis of a 30 s window, which is translated through the data using a 75% overlap, providing 37 location estimates covering the 5 minute period. In Figures 11-13 each position estimate is indicated by a "*".

Figure 11 shows a plan view of the results over an extended spatial scale. This allows one to see the estimated source locations and the sensor locations in the same plot. On this scale the detail of the track associated with the source motion are obscured. In order to present these details, Figure 12 shows the same results on an expanded scale (on this plot the direction of motion is indicated). Figure 13 shows a similar plot in the vertical plane, so that one can observe the estimated depth of the animal and how it moves through the water column.

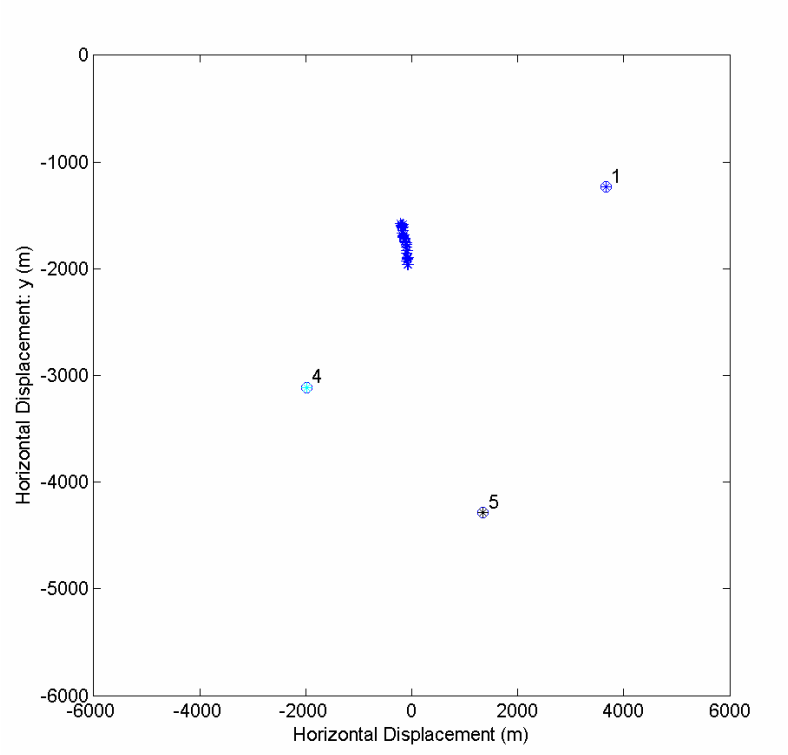

Figure 11. The results of localisation, shown in plan view on a small scale

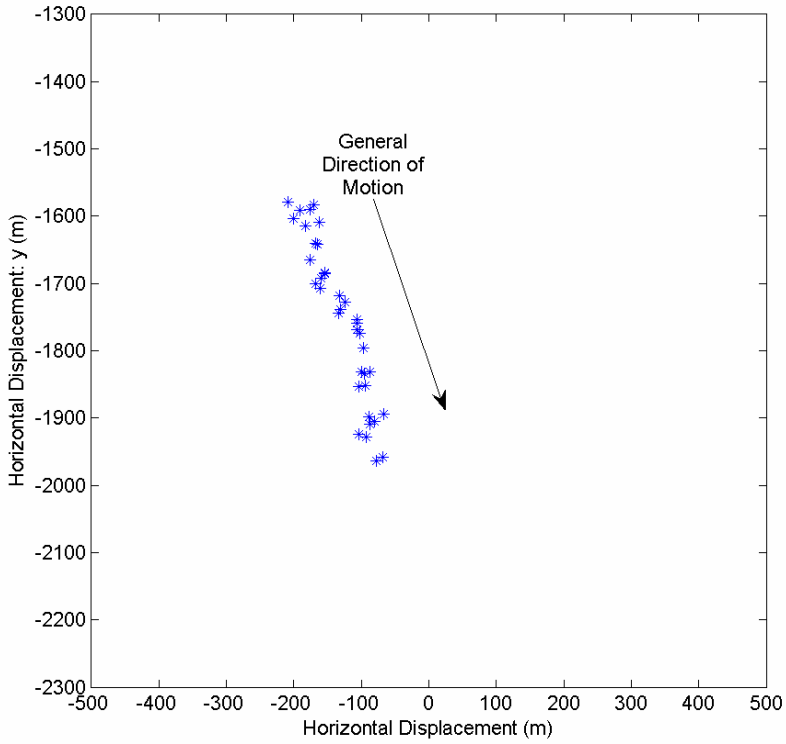

Figure 12. The results of localisation, shown in an expanded plan view compared to Figure 11.

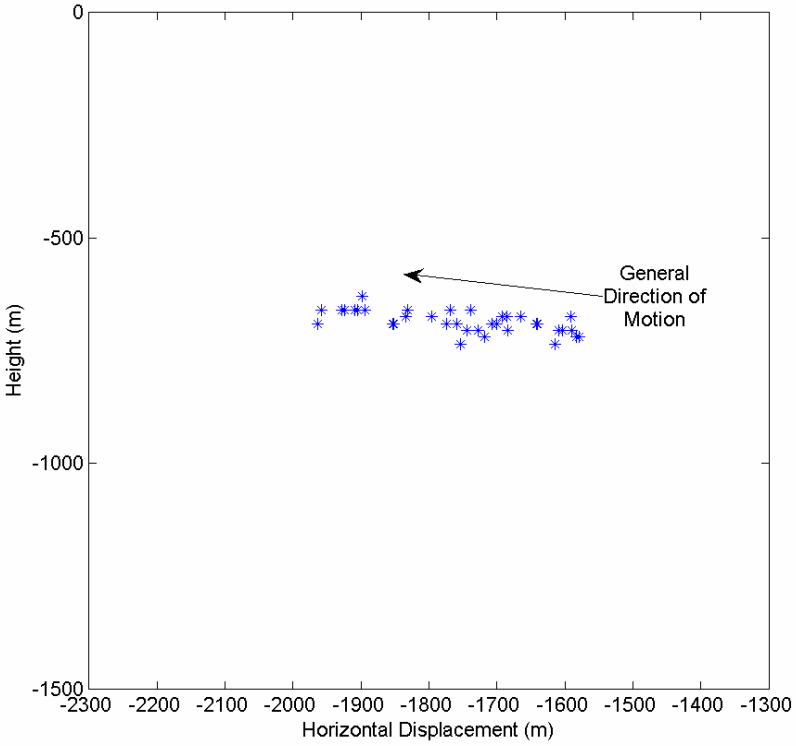

Figure 13. The results of localisation, shown in an expanded side view compared to Figure 11

5 CONCLUSIONS

The effectiveness of tracking algorithms for passive acoustic monitoring has been explored. The problem of estimating the delay between trains of sperm whale clicks has been considered and the problems associated with the use of correlation based methods have been highlighted. A simple source location method that exploits surface reflections has been implemented and shown to yield realistic predictions of the motion of an animal over a 5 minute period.

6 ACKNOWLEDGEMENTS

The authors would like to thank the organisers of the 2nd International Workshop on Detection and Localization using Passive Acoustics, Monaco, 2004 for allowing access to the data used in this work.

7 REFERENCES

1. A. Frantzis, 'Does acoustic testing strand whales?', Nature, 392(6671), 29, (March 1998).
2. D.L. Evans, G.R. England (Eds) Joint Interim Report Bahamas Marine Mammal Stranding Event of 14-16 March 2000. US Department of Commerce (NOAA)/US Navy. 61pp (2001).
3. J.L. Spiesberger, M. Wahlberg, 'Probability density functions for hyperbolic and isodiachronic locations', J. Acoust. Soc. Am., 112(6), 3046-3052, (December 2002)
4. B. Mohl et al, 'Sperm whale clicks: Directionality and source level revisited', J. Acoust. Soc. Am., 107(1), 638-648, (January 1998).
5. B. Mohl et al, 'The monopulsed nature of sperm whale clicks', J. Acoust. Soc. Am., 114(2), 1143-1154, (August 2003).
6. G.C. Carter, 'Coherence and Time-Delay Estimation', Proc. of the IEEE, 75(2), 236-255, (February 1987).
7. P.R. White, Detection Algorithms for Underwater Acoustic Transients. In Control and Dynamic Systems, Advances in Theory and Applications, C.T.Leondes Ed., Pub. Academic Press, 193-224 (1996).
8. L.R. Rabiner, R.W. Schafer, Digital Processing of Speech Signals, Pub. Prentice Hall, (1978).
9. A. Thode, Three-dimensional passive acoustic tracking of sperm whales (Physeter macrocephalus) in ray-refracting environments, J. Acoust. Soc. Am., 118(6), 3575-3584, (December 2005).
10. J.L. Spiesberger, M. Wahlberg, Probability density functions for hyperbolic and isodiachronic locations, J. Acoust. Soc. Am., 112(6), 3046-3052, (December 2002).
11. W. A. Watkins, W. E. Schevill, Four hydrophone array for acoustic three-dimensional location, Woods Hole Oceanographic Technical Report 71–60 (1971).
12. A. Thode, Tracking sperm whale (Physeter macrocephalus) dive profiles using a towed passive acoustic array, J. Acoust. Soc. Am., 116(1), 245-253, (July 2004).