# Detection and Classification of North Atlantic Right Whales in the Bay of Fundy using Independent Component Analysis

**Brian R. La Cour and Michael A. Linford**
Applied Research Laboratories, The University of Texas at Austin
P.O. Box 8029, Austin, TX 78713, E-mail: blacour@arlut.utexas.edu

## ABSTRACT

A novel method of detection and classification for marine mammals is presented which uses techniques from independent component analysis to solve the blind source separation problem for North Atlantic right whales (*Eubalaena glacialis*). Using the fundamentally non-Gaussian nature of marine mammal vocalizations and data collected on multiple hydrophones, we are able to separate right whale source spectra, up to an unknown scale, from ambient noise. This technique assumes that the array data is a linear combination of non-Gaussian source signals but does not require specific knowledge of the array geometry. A detection algorithm which separates right whale vocalizations from ambient background using a Kolmogorov-Smirnov test statistic is presented and tested on data collected in the Bay of Fundy. The performance of the detector was found to be such that it was possible to achieve a probability of detection of about three-fourths with a false alarm probability of about one-third. Independent component analysis was found to provide little improvement over standard principle component analysis, which was used as preprocessing step.

## RÉSUMÉ

Une nouvelle méthode de détection et de classification pour les mammifères marins est présentée. Elle utilise des techniques d'analyse par composantes indépendantes pour résoudre des problèmes de séparation aveugle de sources pour des baleines franches de l'Atlantique Nord (*Eubalaena glacialis*). En se basant sur la nature non gaussienne des vocalisations des mammifères marins et sur les données recueillies par un ensemble d'hydrophones, nous avons été capables de séparer les spectres de baleines franches, jusqu'à une échelle inconnue, du bruit ambiant. Cette technique suppose que l'ensemble des données est une combinaison linéaire des signaux sources non gaussiens, mais ne requiert pas de connaissance particulière sur la géométrie de l'ensemble des hydrophones. Un algorithme de détection permettant de séparer les vocalisations de baleines franches du bruit ambiant en utilisant un test statistique Kolmogorov-Smirnov est présenté et testé sur des données recueillies dans la Baie de Fundy. La performance du détecteur était telle qu'il a été possible de réaliser une probabilité de détection d'environ trois quarts, avec une probabilité de fausse alarme d'environ un tiers. L'analyse par composantes indépendantes n'a donné que des améliorations mineures comparé á l'analyse par composantes principales standard qui a été utilisée comme étape de pré-traitement.

## 1 INTRODUCTION

Passive acoustic detection of cetaceans has become an area of great interest in recent years due in part to the need to mitigate any possible impact due to shipping and naval training exercises on local populations. Visual observations, the traditional method of detection, are limited in several ways. They are restricted to daylight observations, require human observers, are limited in detection range, and can detect only surfacing animals. Passive acoustic detection overcomes these particular limitations, but brings also a new set of challenges. Marine mammals in the observation area must vocalize to be detected, and those vocalizations must be detectable and distinguishable from the multitude of competing background sound sources. In addition, it may be important to distinguish between different types of cetaceans based on their calls. A key advantage to using passive acoustics is the potential to perform 24-hour, real-time automated detection. Since, for purposes of impact mitigation, important time and distance scales for detection and localization are on the order of hours and miles, respectively, the demands placed upon an automated system are significantly reduced from that of more traditional anti-submarine applications.

In this paper we discuss a method for performing passive acoustic automated detection of marine mammals based upon the use of independent component analysis (ICA). ICA is a statistical analysis tool used to solve the blind source separation problem, wherein simultaneous recordings of multiple sound sources are used to separate out the individual sound sources. The approach, then, is to apply ICA to segments of the recorded acoustic time series and separate out marine mammal calls from the ambient background. The separated signals may then be used to perform classification in an automated or human-directed classifier.

The algorithm presented here is applied to and tested on data gathered in Canada's Bay of Fundy that contains several calls from North Atlantic right whales (*Eubalaena glacialis*).[1] The focus on this species is motivated by its rapid decline in recent years [1], with over a third of deaths now attributable to ship collisions [2]. The data was collected in September of 2002, a time when right whales come to the bay in great numbers to feed themselves and their newborn calves, and was recorded on several bottom-mounted hydrophones.

The organization of this paper is as follows. In Sec. 2 we provide a brief introduction to Independent Component Analysis, with particular emphasis on its application to marine mammal detection. Sec. 3 contains a description of the detection algorithm used, which is based on the observed non-Gaussianity in the statistics of right whale calls. The application to right whale data from the Bay of Fundy is analyzed in Sec. 4, while a discussion of results and conclusions are given in Sec. 5.

## 2 INDEPENDENT COMPONENT ANALYSIS

### 2.1 Description

Independent component analysis (ICA) is a method for solving the blind source separation problem [3]. The basic model assumed by ICA is one in which the data, $\mathbf{x}(t) = [x_1(t), \ldots, x_M(t)]^T$, on $M$ receivers at time $t$ is a linear combination of $P$ sources, $\mathbf{s}(t) = [s_1(t), \ldots, s_P(t)]^T$, plus ambient noise, $\mathbf{n}(t)$. The sources are assumed to be realizations of mutually independent, stationary stochastic processes in which at most one of the sources has a Gaussian marginal distribution [4]. Denoting by $\mathbf{A}$ the linear mixing matrix, we therefore have the following data model.

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t). \tag{1}$$

The goal of ICA is to find a suitable unmixing matrix, $\mathbf{W}$, such that $\mathbf{y} = \mathbf{W}\mathbf{x}$ is an approximation to $\mathbf{s}$. Due to the nature of the problem and the type of solution offered by ICA, there is a fundamental ambiguity such that $\mathbf{y}$ may differ from $\mathbf{s}$ by an arbitrary permutation and scaling of its rows.

[1] This data was provided courtesy of Defense Research and Development Canada—Atlantic and Dalhousie University.

What allows ICA to perform this inversion is the assumed non-Gaussianity of the sources. This may be understood as follows. For a given choice of $\mathbf{W}$, each element of $\mathbf{y}$ will be a linear combination of the sources. In general, this will result in a $\mathbf{y}$ which appears more Gaussian, due to the central limit theorem [5]. If $\mathbf{W}$ is suitably chosen, however, then each element of $\mathbf{y}$ will correspond to only one source, up to an arbitrary scaling, and thus will appear non-Gaussian. In this manner, non-Gaussianity may be used as a metric of independence, and the independent components estimated by $\mathbf{y}$ are maximally independent linear combinations of the original data.

It may be noted that ICA resembles Principle Component Analysis (PCA), which produces a linear combination of the data that is uncorrelated. In PCA the unmixing matrix $\mathbf{V}$ is given in terms of the eigenvalue decomposition of the data covariance matrix $\mathbf{C}$. If $\mathbf{U}$ is the matrix of eigenvectors of $\mathbf{C}$ and $\mathbf{D} = \mathbf{U}^H\mathbf{C}\mathbf{U}$ is the diagonal matrix of corresponding eigenvalues, then $\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{U}^H$ is the PCA unmixing matrix. In applying ICA, PCA is often used as a preprocessing step. When this is done, the ICA algorithm is applied to the transformed variable, $\mathbf{z} = \mathbf{V}\mathbf{x}$, to yield the unmixing matrix $\mathbf{Q}$. The final unmixing matrix, applied to $\mathbf{x}$, is therefore $\mathbf{W} = \mathbf{Q}\mathbf{V}$.

### 2.2 Mutual Information Approach

Bell and Sejnowski [6] have suggested a method for performing ICA based on mutual information. More traditional ICA methods have relied on kurtosis as a measure of non-Gaussianity and, by inference, statistical independence [7]. Mutual information, on the other hand, provides a direct measure of statistical independence between a joint probability density function (PDF) and the product of its marginals.

In the approach of Bell and Sejnowski, the ICA process is viewed as a neural network with input $\mathbf{x}$, output $\mathbf{y}$, and nodal weights $\mathbf{W}$. The goal, then, is to minimize the mutual information between the input and output of the network. This mutual information may be written

$$I(\mathbf{x}, \mathbf{y}) = H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}), \tag{2}$$

where $H(\mathbf{y})$ is the entropy of $\mathbf{y}$ and $H(\mathbf{y}|\mathbf{x})$ is the entropy of $\mathbf{y}$ relative to $\mathbf{x}$ [8]. Since $\mathbf{y}$ is a deterministic function of $\mathbf{x}$, the term $H(\mathbf{y}|\mathbf{x})$ is a constant (i.e., $-\infty$), independent of $\mathbf{W}$. Thus, minimization of the mutual information is equivalent to minimizing the entropy of the output.

To achieve greater sensitivity, the output is transformed via a sigmoidal function $g(\cdot)$ so that the output of the network is $\mathbf{y} = g(\mathbf{u})$, where $\mathbf{u} = \mathbf{W}\mathbf{x}$. The choice of this function is arbitrary, although the Bell and Sejnowski suggest using the logistic transfer function, $g(\mathbf{u}) = 1/(1+e^{-\mathbf{u}})$, a common choice for neural networks, and we adopt this choice in our work. Ideally, the transfer function would be formed from the cumulative distribution function (CDF) of the output $\mathbf{u}$.

Extremizing the mutual information with respect to $\mathbf{W}$ leads to the following gradient method: Starting with an initial value for $\mathbf{W}$, the matrix is updated according to the scheme $\mathbf{W} \mapsto \mathbf{W} + \Delta\mathbf{W}$, where

$$\Delta\mathbf{W} = \alpha\left[(\mathbf{W}^T)^{-1} + (1-2\mathbf{u})\mathbf{x}^T\right] \quad (3)$$

and $\alpha$ is the learning rate. In our work, the initial value of $\mathbf{W}$ was taken to be the identity matrix and $\alpha$ was taken to be $0.001/n$, where $n$ is the current iteration number.

### 2.3 Applicability to Marine Mammal Detection

To apply ICA to marine mammal detection, is it necessary to demonstrate the validity of two key model assumptions: statistical assumptions regarding the source signals and propagation assumptions under the linear mixing model.

For propagation we assume a linear medium such that the received signal is a sum of attenuated and time delayed sources. The receivers are assumed to be horizontally distributed so that time delays may exist between receivers from sources are different bearings. Thus,

$$x_m(t) = \sum_{p=1}^{P} A_{mp}s_p(t-\tau_{mp}) + n_m(t), \quad (4)$$

where $\tau_{mp}$ is the time delay from source $p$ to receiver $m$ and $n_m(t)$ is the ambient noise in receiver $m$. (For simplicity, we ignore the presence of multipaths.) We assume that a dominant source signal is present which is spatially localized in bearing so that $\tau_{mp}$ is effectively independent of $p$. (Physically, this dominant source may correspond to an individual whale or an entire pod.) Then, using a correlation technique described in Sec. 3, the received data may be transformed by the estimated relative delay time, $\hat{\tau}_m$, so that $x'_m(t) = x_m(t+\hat{\tau}_m)$ may be expected to fit the ICA linear model.

The fundamental statistical assumption required for our work is that marine mammal calls can be distinguished from ambient background signals by their non-Gaussian statistics. Such behavior is well known in the field of human speech processing, were super-Gaussian statistics dominate.[6] In the case of right whales, this question may be answered by an examination of actual whale call recordings. In Fig. 1 we

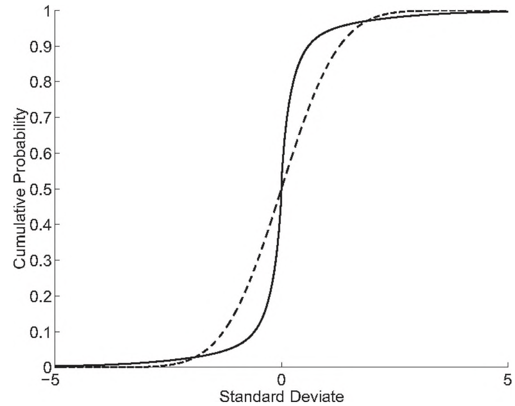


Figure 1: Plot of the cumulative distribution of right whale call statistics (solid line) versus that of a standard normal distribution (dashed line). The distribution is symmetric, but the sharp slope at the center indicates a high value of kurtosis.

have plotted the empirical CDF for a 110-sec recording of a North Atlantic right whale.[2] For comparison, the CDF for a standard normal distribution is included. It is clear from the figure that the statistics of the right whale calls exhibit a large, positive kurtosis excess (about 22.3) and thus may be considered super-Gaussian. Such statistical behavior has been found to be typical of human speech, as well as many other natural and unnatural sound sources, so this result is not at all surprising. On the other hand, and for this very reason, statistics alone cannot serve as a means of classification, though it may serve a role in detection and preprocessing.

## 3 DETECTION ALGORITHM DESCRIPTION

The detection algorithm may be described as a three step process. First, a sliding window along the time axis chooses a segment of the time series data to be analyzed. Next, independent component analysis is performed to extract the dominant non-Gaussian signal. Finally, a test statistic is computed from the dominant component and compared to a threshold to determine whether a detection is called. As a preprocessing step prior to selecting individual time segments, the time axes of the different receiver time series are aligned via cross-correlation under the assumption that all sound sources are co-located in bearing. Although not required by ICA, such an alignment is needed to ensure that a given call appears in the same time window for all hydrophones.

Time aligning the different receivers is equivalent to

[2] Data provided by Susan Parks of the Woods Hole Oceanographic Institute from recordings made by Scott Kraus of the New England Aquarium.

source localization in bearing. Since the expected source signals are transient and broadband, the usual narrowband subspace methods for bearing estimation are rejected in favor of an approach using cross-correlations. In this approach, a reference receiver, labeled $m_0$, is chosen, and cross-correlations are computed for each receiver paired with the selected reference. Rather than search for peaks in the receiver cross-correlations, a set of physically realizable time delays is computed based on the geometry of the receiver array and a hypothesized source bearing.

The delays are computed under the assumption of a constant sound speed, $c$, and direct path propagation. To do this, an asymptotic result is used. If the source range is much larger than the spatial extent of the array, then the relative time delay for receiver $m$ is given approximately by

$$\tau_m = -\left(\Delta x_m \cos\theta + \Delta y_m \sin\theta\right)/c, \qquad (5)$$

where $\Delta x_m = x_m - x_{m_0}$, $\Delta y_m = y_m - y_{m_0}$, and $\theta$ is the hypothesized source angle. At short ranges, propagation in the vertical direction will cause this approximation to underestimate the magnitude of the time delays.

Each value of $\theta$ gives a different set of delays. If we consider the cross-correlation $\rho_m(\tau)$ between receiver $m$ and the reference, then a measure of goodness of fit would be the sum of $|\rho_m(\tau_m)|^2$ over $m$; i.e., the beam intensity in the $\theta$ direction. For perfect alignment of data that differ only by a translation in time, this quantity will have a peak value of $M$. Thus, maximizing over $\theta$ gives an estimate of the source bearing and, with it, the delay estimates, $\hat{\tau}_m$, needed to perform time alignment.

Having aligned the receiver data, we next run a window of fixed width $w$ and offset $t_n = n\Delta t$ such that $0 \leq t_n \leq T - w$, where $T$ is the maximum time recorded. In general, $\Delta t$ may be taken to be smaller than $w$. In our work we found that $w = 2$ sec, which adequately bounds the duration of a typical right whale call, and $\Delta t = w/2$ appear to work quite well.

The ICA algorithm described in Sec. 2 is applied to the data after preprocessing through PCA. The number of independent components to be extracted is variable (up to the number of receivers), but we have found that for detection purposes it is best to extract only two independent components (ICs). These are ranked according to their absolute kurtosis value, and the component with the largest such value is taken to be the one that would contain the source signal.

Detection is based on the non-Gaussianity of the first independent component extracted. For a detection statistic, we chose the Kolmogorov-Smirnov (KS) statistic, which measures the largest difference between

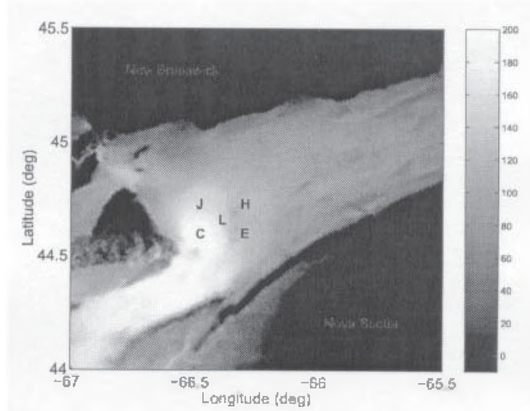


Figure 2: Bathymetry for the Bay of Fundy at 0.1-minute resolution. New Brunswick is at the top of the figure; Nova Scotia appears on the bottom right. The depth is in meters. The positions of the five OBHs are indicated in the figure, at the left and center of each letter. (Bathymetric data courtesy of the Naval Oceanographic Office)

the empirical CDF of the first IC, suitably standardized, and a standard normal distribution [5]. The KS statistic has the desirable property that, for large samples, its distribution is independent of that of the underlying data. This allows one to set a detection threshold for a desired probability of false alarm (PFA). For a standard PFA of 5%, the critical value of the KS statistic is about 1.36.

## 4 APPLICATION TO RIGHT WHALE DATA

### 4.1 Data Description

The detection algorithm was applied to data collected from the Bay of Fundy in September of 2002 on five Ocean Bottom Hydrophones (OBHs). The OBHs (OAS model E-2SD) are omnidirectional hydrophones that are moored about 0.9 m from the seafloor when normally deployed. The positions of the OBHs are shown in Fig. 2 in relation to the local bathymetry.

The locally measured sound speed profile (SSP) is shown in Fig. 3. The strongly downward refracting profile together with a shallow bottom implies that sound propagation will tend to be limited in range due to multiple bottom interactions. From reciprocity we may also observe that very little sound from the surface propagates to the bottom via a direct path.

The data set itself consists of 15 segments, each about 30 sec in length. All contain vocalizations from one or more right whales in the area. Of the 15 segments, the first five are so-called "gunshot" calls, which appear as broadband impulsive transients. The remaining segments contain low (segments 6–13) and mid-

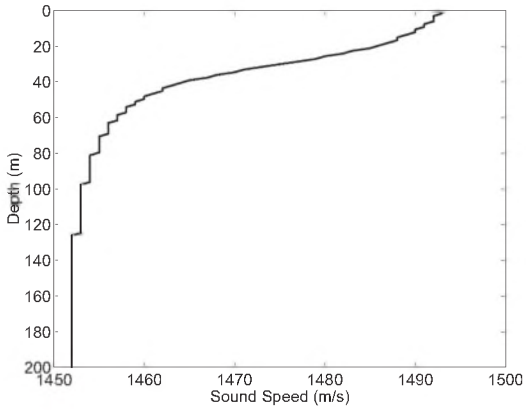


Figure 3: Plot of the sound speed profile in the Bay of Fundy during September 2002, as determined from local Expendable Bathythermograph (XBT) and Conductivity-Temperature-Depth (CTD) measurements . A mixed layer of some 40 m is evident.

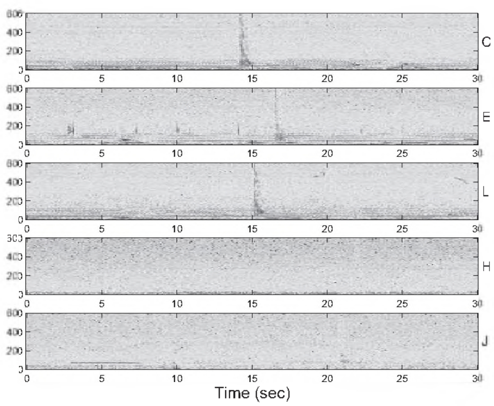


Figure 4: Plot of the spectrograms for the first data segment on all five OBHs. The gunshot is visible at about 15 sec on OBH L. OBH H and J did not register the gunshot well. The frequency, in Hz, is plotted along the vertical axis. The magnitude of the spectrogram is given in arbitrary units on a logarithmic scale.

frequency (segments 14 & 15) upsweeps. Fig. 4 shows the spectrograms for the first time segment, illustrating a gunshot call. Note that the data on OBHs H and J is very poor.[3]

### 4.2 Detection Results

Due to the nature of the data collected, ground truth is not available for a true assessment of detector performance. In lieu of this, detections based on human observations were used as truth to baseline the performance of the detector. For each of the 15 data segments, the start and stop times of each right whale call were determined by aural clues and visual inspec-

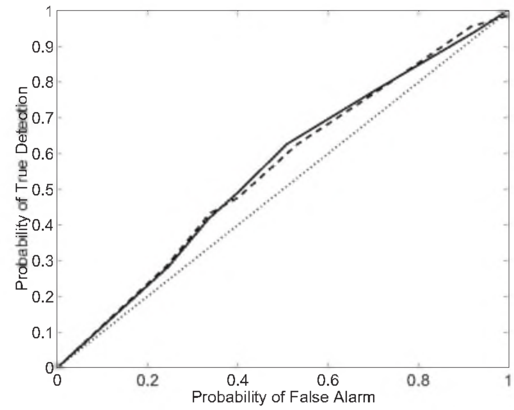


Figure 5: ROC curve comparing detection performance of PCA (solid curve) versus ICA (dashed curve). The dotted diagonal line represents random chance.

tion of the spectrograms. All such times were based relative to OBH L, which lay in the center of the array.

A called detection was considered correct (true positive) if the time window $[t_n, t_n + w]$ in which the KS statistic exceeded the threshold intersected with the truthed call interval described above. The probability of detection (PD) was then defined as the ratio of the number of true positives to the total number of true positives and true negatives. Similarly, the probability of false alarm (PFA) was defined as the ratio of the number of false positives to the total number of false positives and false negatives. By sweeping through a range of threshold values, a receiver operating characteristics (ROC) curve of PD versus PFA could then be generated. It is this ROC curve that we use as our metric of performance.

In Fig. 5 we compare the detection performance of ICA, with PCA as a preprocessing step, versus PCA alone. The results suggest that there is very little improvement gained by ICA over PCA, with the two ROC curves being almost identical. Clearly, the bulk of the work done in separating independent signals is done simply by linearly transforming the data so that the phones are uncorrelated. Since the PCA algorithm is much faster (about 100 times faster than ICA), this suggests that it may be the better choice for a real-time system.

Detection is performed by computing the KS statistic on the first independent (or principle) component estimated. If we compare the discriminating power of this first component to the second, we see from Fig. 6 that detection performance is severely degraded by using the latter. This suggests that the signal contained in the first component really does have information content useful for detection.

We have developed a method for time aligning data from spatially separated receivers. As shown in Fig.

[3]OBH J was know to suffer from hardware problems. The low amplitude in OBH H may be due to its location relative to the right whale pod, which was known to be south of the array.

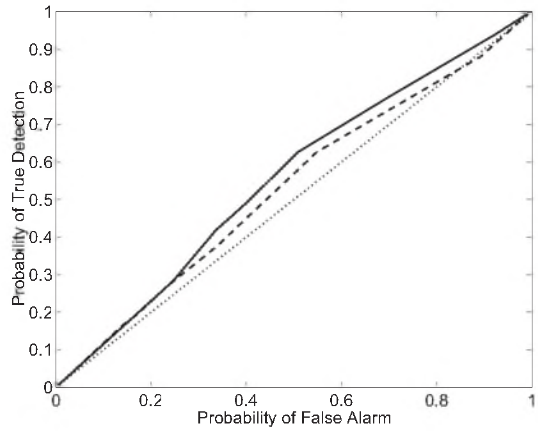


Figure 6: ROC curve comparing detection performance using the first (solid curve) and second (dashed curve) independent components.

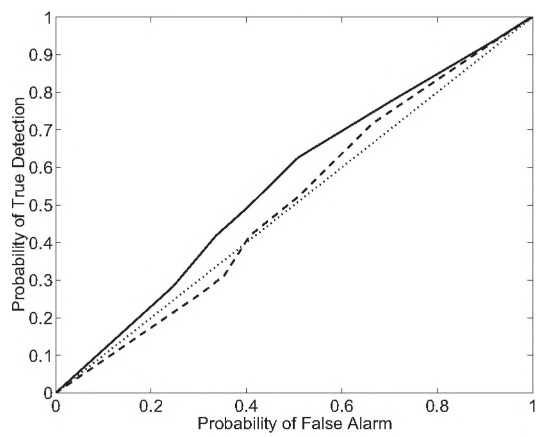


Figure 7: ROC curve comparing detection performance without time alignment (solid curve) and with time alignment (dashed curve).

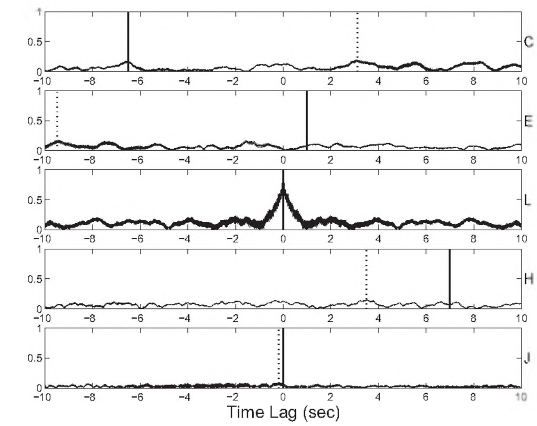


Figure 8: Magnitude of normalized cross-correlations for the five OBHs relative to OBH L. The solid vertical line indicates the estimated time delay; the dotted line indicates the peak value.

7, however, this procedure has been found to actually decrease overall detector performance. The reason for this appears to be that the cross-phone correlations contain very little information with which to discriminate different lag times. This can be seen in Fig. 8, where we have plotted the cross-correlations for each OBH relative to OBH L. The cross-correlations are flat and noisy, making estimation of time delays difficult, and small errors in estimating the peak can translate into huge errors in the delays.

Finally, we considered the effect of removing the two problematic OBHs, H and J, from the analysis. Interestingly, though the two phones were very noisy, there is a significant drop in the performance of the detector. Furthermore, if time aligning is performed with these two OBHs removed, the results are comparable to that when no time aligning is performed.

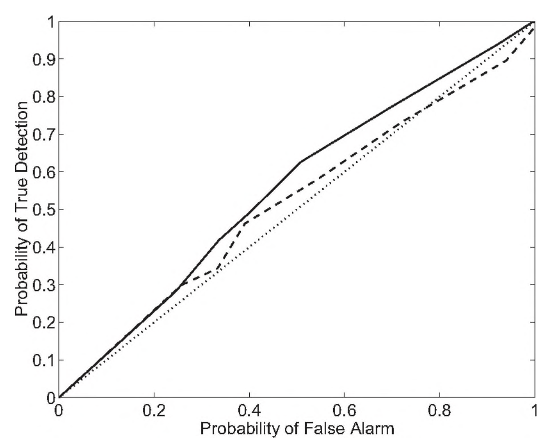


Figure 9: ROC curve comparing detection performance using all five OBHs (solid curve) versus that when OBHs H and J are removed (dashed curve).

## 5 DISCUSSION OF RESULTS

In this paper we have described an algorithm for passive acoustic detection of marine mammals using independent component analysis. This algorithm was implemented and tested on data collected in the Bay of Fundy and containing a variety of right whale calls. The performance of the detector was such that it was possible to find an operating point on the ROC curve such that about three-fourths of whales are detected with about a third of all calls being false alarms.

In comparing ICA against PCA, it was found that ICA provides little improvement over PCA when the latter is used as a preprocessing step. This suggests that decorrelating the data goes a long way towards achieving statistical independence. PCA is much faster than ICA, at least in the implementation used based on mutual information, and is therefore recommended for use in place of ICA for detection purposes. Using PCA alone, the detection algorithm is able to run in well under real time.

Procedures for estimating relative delays and time aligning the receiver data were frustrated by hardware issues and noisy data. A consequence of poor delay estimation was that detector performance actually worsened when these corrective techniques were applied. Since the PCA component of the algorithm is sufficiently fast to allow for additional processing, it may be that inclusion of a better localization algorithm, such as matched field processing, is possible within a real-time system.

For marine mammal classification, the ICA algorithm presented may be suitable as a preprocessing step to feed into a large classifier. Non-Gaussianity as a metric for detection is useful for separating out ambient background sounds, but it alone is not suitable for classification. ICA may be useful in this regard by providing an estimate of the extracted source signal, which may then be used to classify the source into more specific categories.

## REFERENCES

[1] S. L. Perry, D. P. Demaster, and G. K. Silber. Special issue: The great whales: History and status of six species listed as endangered under the U.S. Endangered Species Act of 1973. *Marine Fisheries Review*, 61:1–74, 1999.

[2] D. W. Laist, A. R. Knowlton, J. G. Mead, A. S. Colliet, and M. Podesta. Collisions between ships and whales. *Marine Mammal Science*, 17:35–75,

[3] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*. J. Wiley, New York, 2001.

[4] Sheldon M. Ross. *Stochastic Processes*. John Wiley & Sons, New York, 1983.

[5] Alexander M. Mood, Franklin A. Graybill, and Duane C. Boes. *Introduction to the Theory of Statistics*. McGraw-Hill, New York, 3rd edition, 1974.

[6] Anthony J. Bell and Terrence J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1004–1034, 1995.

[7] Juha Karhunen. Neural approaches to independent component analysis and source separation. In *Proc. 4th European Symposium on Artificial Neural Networks (ESANN'96)*, pages 249–266, Bruges, Belgium, 24–26 April 1996.

[8] Athanasios Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Boston, 4th edition, 2002.

## 6 ACKNOWLEDGMENTS

The authors would like to thank Ms. Francine Desharnais and her colleagues at Defense R&D Canada—Atlantic for providing the right whale data used in the development and testing of this algorithm. They would also like to thank Dr. Susan Parks of the Woods Hole Oceanographic Office for providing the set of right whale recordings made by Scott Kraus of the New England Aquarium. Dr. La Cour would like to acknowledge Applied Research Laboratories of the University of Texas at Austin (ARL:UT) for funding this work under Independent Research and Development Grant No. 946-9. Mr. Linford's work was funded by ARL:UT's Honors Scholars program.