

# Detection, Classification, and Localization of Cetaceans by groups at the Scripps Institution of Oceanography and San Diego State University (2003-2013)

Marie A. Roch \*<sup>◇</sup>

Ana Širović <sup>◇</sup>

Simone Baumann-Pickering <sup>◇</sup>

\* *San Diego State University, Department of Computer Science, 5500 Campanile Drive, San Diego, California 92182-7720*

<sup>◇</sup> *Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0205*

## I. Introduction

Cetaceans produce a wide variety of sounds, each presenting different sets of challenges in detection, classification, and localization. In general, baleen whales, or mysticetes, tend to have lower frequency calls. While the calls themselves tend to be simpler than those of some other species, the low frequency environment is a challenging one. Mysticete calls typically have high source levels ( $>180$  dB re 1  $\mu$ Pa) and can travel over hundreds of kilometers. In contrast, toothed whales and dolphins, the odontocetes, tend to produce sounds that are higher in frequency and thus limited to propagation over short distances on the order of kilometers. For most odontocetes, sounds tend to be more variable than those of the majority of their mysticete counterparts.

Analysis algorithms must typically deal with a great deal of acoustic clutter, signals that may mask the detection of the sounds of interest. Examples of clutter include anthropogenic signals (e.g. sonar, ship traffic), natural phenomena such as rainfall or wind and waves, recording instrument noise (e.g. cable strumming), distortions of the signal itself due to multipath propagation, and calls from other species. Acoustic clutter is particularly difficult in lower frequency bands as many of the sources of clutter tend to be concentrated in this range. In addition, as low frequency energy is absorbed less quickly than that of high frequencies, the low frequency calls tend to travel farther and are thus susceptible to distortion as a result of long distance propagation effects.

The early focus of the whale acoustics lab from Scripps Institution of Oceanography (SIO) was on the detection of mysticete calls, primarily using energy ratio matches (Burtenshaw et al., 2004; Širović et al., 2004) or spectrogram correlation following the work of Mellinger and Clark (2000). In 2006, with the start of the collaboration between labs at SIO and San Diego State University, we began using methods inspired from speech recognition technologies. This brought new ways of thinking about feature extraction and statistical machine learning techniques. We have elected to organize the review by theme rather than project and separate

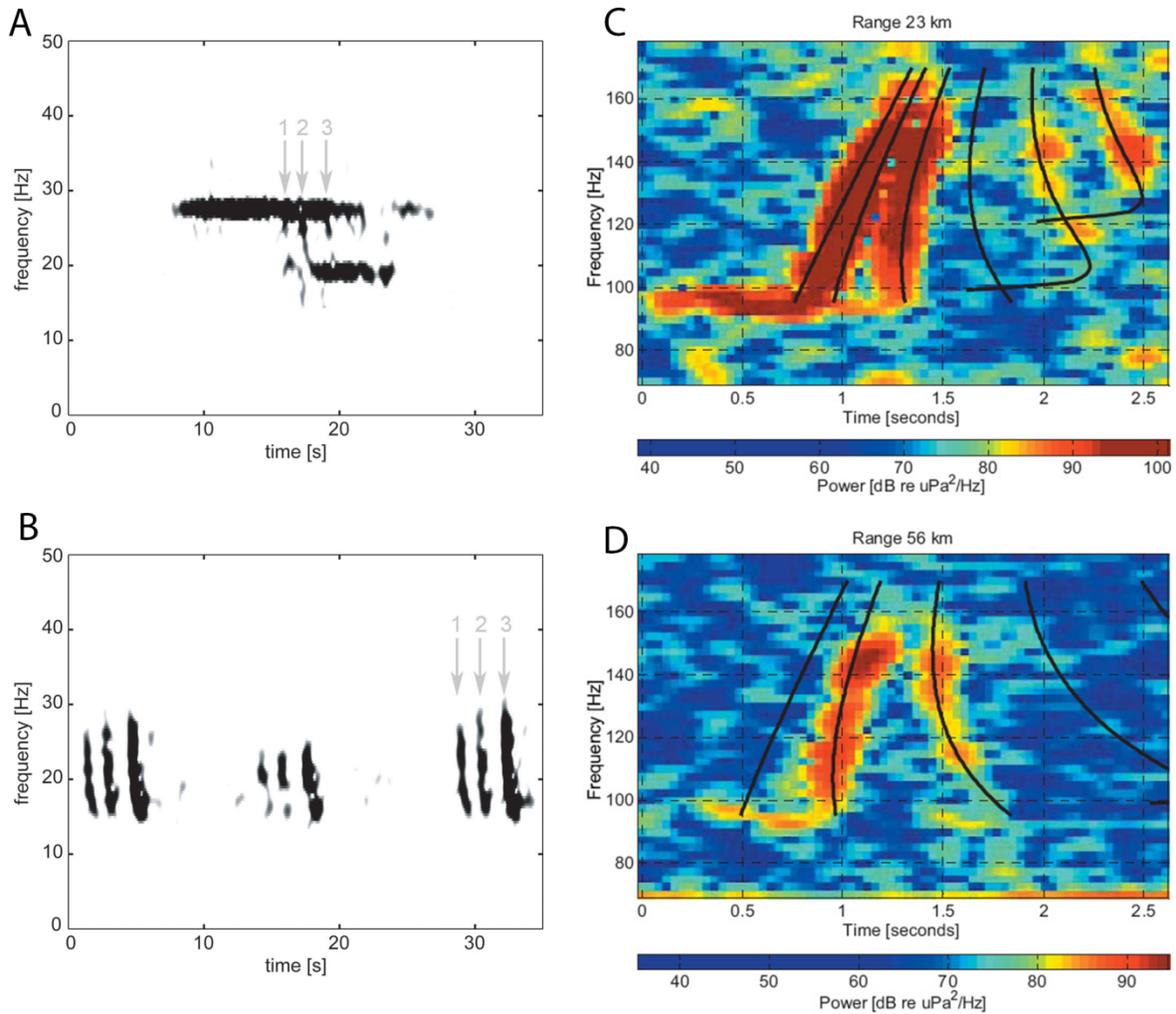
our discussion broadly into signal representation, classification, and localization. While we do cite other groups upon whose work we have built, this is a review of work by our labs rather than a general literature review.

## **II. Representation: Call properties and feature extraction**

One of the most difficult tasks in designing pattern recognition algorithms is to determine what makes a particular call distinct. For calls that are highly stereotyped, that is with little variation between productions, primary features include duration and frequency information (e.g. peak or trajectory). The received level is rarely of importance unless distance to the animal has been estimated and a reverse propagation model can infer an estimate of the source level. Calls that are variable in nature require different techniques. Some of the features can be similar, but capturing the essence of a call is much more difficult.

For any call type, one must consider what are the other potential calls or phenomena that might have similar measurements. When these occur in the same geographic region as the call of interest, they can be a source of confusion for the ensuing classification algorithm. In general, the goal is to try to select features that not only describe the call of interest, but provide contrast to any other type of signal that is likely to occur. In our labs, we have focused our attention largely on detecting distinct, stereotyped calls with minimal distortions, but when propagation distortions were present, sometimes they were used to calculate valuable information on the location of the calling animal, as described below (section II.A).

It is important to note that many times when we believe a call to be highly stereotyped, it is not necessarily so. A good example of this can be seen in the vocalizations produced by blue whales (*Balaenoptera musculus*). B calls produced by these animals in the North Pacific appear to be very consistent when examined over a short time scale. However, it has been shown by McDonald et al. (2009) that there is a worldwide trend by most blue whale populations to shift their B call frequencies lower every year. For populations in the North Pacific, where we have the longest time series, the dominant frequency of B calls declined by 31% in the period from 1963 to 2008. This makes the design of what would appear to be a very simple call to recognize significantly more difficult. As a solution, instead of a static call detection kernel, year-specific kernels have been applied for the automated detection of blue whale B calls (Oleson et al., 2007b). In addition, when working with low-frequency calls, propagation effects can severely distort the call and the arrival of different modes or paths can cause a call to be smeared temporally (Figure 1).



**Figure 1 - Examples of multipath arrivals of blue (A) and fin (B) whale calls from the Antarctic marked with arrows (from Širović et al. 2007) and modes of North Pacific right whale up calls overlaid with synthetic received calls from normal-mode modeling at two distances (C and D) distances (from Wiggins et al. 2004). Note that mode four is not excited in either spectrogram, and only modes two and three are above the background noise at 56 km range (D). Permission pending.**

The calls of humpback whales (*Megaptera novaeangliae*), odontocete whistles, and echolocation clicks all fall in the category of calls that are highly variable. As an example, the received level, duration, and frequency content of echolocation clicks varies dramatically depending on the angle of the echolocation beam to the receiving hydrophone (Au, 1993; Au et al., 2012a; Au et al., 2012b; Lammers and Castellote, 2009). This presents a significant challenge for feature extraction. One must either find features that are invariant to the orientation of a free ranging odontocete and the receiving hydrophone or one must develop models that characterize the range of signals that can be received.

To date, nobody has been able to develop a set of features for echolocation clicks that are invariant to axis orientation or solve the inverse problem, which would enable us to develop

features with respect to orientation. Much of our lab's classification work with odontocetes has focused on their echolocation signals. With respect to feature extraction, our work has had two principal directions: 1) characterizing spectral signals by trends or mean spectra and 2) developing low dimensional representations that capture the shape of individual click spectra.

### **A. Properties of stereotyped mysticete calls**

To describe stereotyped, mysticete calls in a way that enabled us to distinguish them from other calls and phenomena, we have generally employed relatively simple measurements of the frequency and temporal characteristics, as well as the patterning. The basic set of frequency features that have been measured included start and stop frequency. In calls that have multiple parts with variable level of frequency modulation, start and stop frequency of each part of the call was measured. Thus for blue whale B calls from the northeastern Pacific, four values were reported (Oleson et al., 2007b). Similarly, three start and end frequencies were used to describe blue whales calls from the Antarctic (Širović et al., 2004). In contrast, fin whale (*B. physalus*) 20 Hz pulses and Pacific right whale (*Eubalaena japonica*) up-calls are simpler, and a single frequency pair is sufficient (Munger et al., 2005; Širović et al., 2004). In conjunction with frequency, start and end time of each segment is measured, allowing us to define durations of each part of the call and creating a kernel that can be used in automatic detection. These features were generally measured from spectrograms generated with appropriate frequency ( $\leq 1$  Hz) and temporal ( $\leq 0.1$  s) resolution.

Another valuable parameter to describe regularly repeated songs produced by some mysticetes (e.g. blue and fin whales) is the inter-pulse interval (IPI). This is the interval between the start of subsequent individual calls in a sequence of stereotyped calls. While the IPI is not always constant over time or space, it can be useful for identifying calling sequences from an individual animal.

### **B. Echolocation click trends and mean spectra**

When spectra of a longer echolocation sequence are sorted by peak frequency, it becomes apparent by how much the received spectra of echolocation clicks can vary. An example of this is shown below (Figure 2) for Risso's dolphins (*Grampus griseus*). In this figure, individual echolocation clicks have been detected and their spectra are displayed consecutively after ordering them by peak frequency. Some of the variation in peak frequency can be attributed to the animals themselves (variation from one animal to another or between individual clicks), but most of the variation is likely due to orientation and to a lesser extent distance.

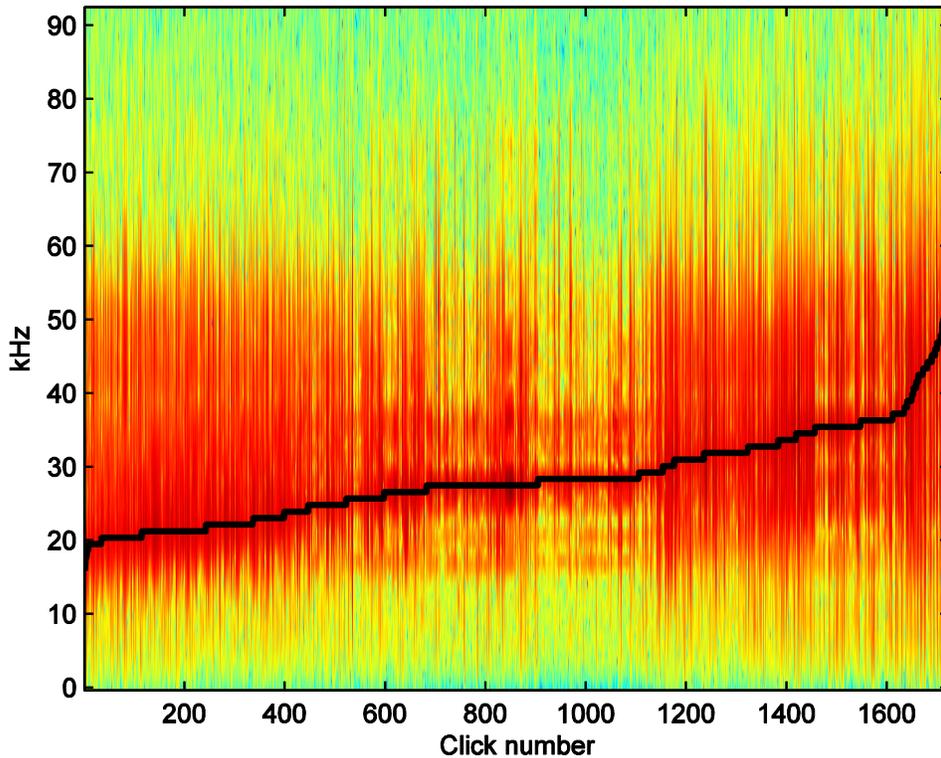


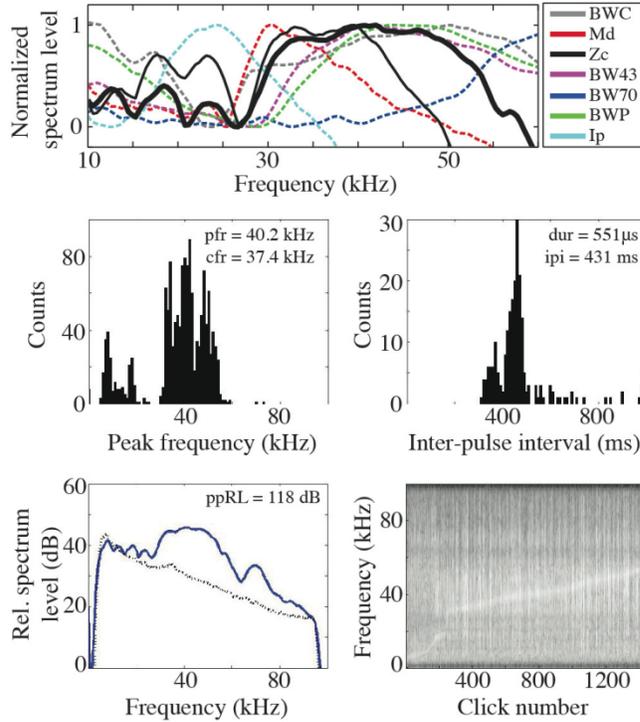
Figure 2 – Variability in the spectra of Risso's dolphin (*Grampus griseus*) echolocation clicks recorded on a HARP. Clicks are sorted by peak frequency. The solid black line shows the peak frequency for each click. Banding patterns in the sorted spectra due to spectral peaks in the click structure of some odontocetes such as Risso's dolphin become apparent in this type of visualization.

By examining these stacked spectra, it became apparent to us that some species tend to have spectral peak and notch structure in many, but not all of their echolocation clicks. The first two species we found this for were Risso's and Pacific white-sided dolphin (*Lagenorhynchus obliquidens*) clicks (Soldevilla *et al.*, 2008). While it is not known conclusively, we believe this spectral structure to be in off-axis clicks, and the aforementioned work by Lammers and Catellote (2009) as well as Au *et al.* (2012a; 2012b) lends some credence to this hypothesis.

Not all species have such distinct spectra, and some species such as bottlenose (*Tursiops truncatus*) and common (*Delphinus spp.*) dolphins have highly variable echolocation click spectra that are quite difficult to classify based on characteristics of individual spectra (Roch *et al.*, 2011b). One technique that has shown some promise is comparison of averaged spectra. While individual clicks vary greatly, we compared the mean spectra of three species: melon-headed whales (*Peponocephala electra*), bottlenose dolphins, and Gray's spinner dolphins (*Stenella longirostris longirostris*). Melon-headed whale echolocation clicks had the lowest peak and center frequencies, spinner dolphins had the highest frequencies and bottlenose dolphins were nested in between these two species. Feature differences were enhanced by reducing variability within a set of single clicks by calculating mean spectra for groups of clicks. A

subsequent discriminant function analysis over grouped spectra showed the ability to correctly discriminate between 93% of melon-headed whales, 75% of spinner dolphins and 54% of bottlenose dolphins. The results also showed, however, that these differences in averaged echolocation click spectra of delphinids are small and easily overwritten by minor differences introduced through imperfectly calibrated recording instrumentation.

Recent advances have shown that this technique is particularly useful in the description and discrimination of species-specific beaked whale echolocation signals. Over the past decade, research has revealed that most beaked whales use a species-specific frequency modulated (FM) upswept echolocation pulses to forage and sense their environment. Based on recordings from animal-attached, suction-cup acoustic archival tags and from towed hydrophones during concurrent visual surveys, acoustic descriptions have been made for FM pulses from Baird's (*Berardius bairdii*) (Baumann-Pickering et al., submitted; Dawson et al., 1998), Blainville's (*Mesoplodon densirostris*) (Aguilar de Soto et al., 2012; Johnson et al., 2006; Johnson et al., 2004; Madsen et al., 2005), Cuvier's (*Ziphius cavirostris*) (Zimmer et al., 2008; Zimmer et al., 2005), Gervais' (*M. europaeus*) (Gillespie et al., 2009), Longman's (*Indopacetus pacificus*) (Rankin et al., 2011), Deraniyagala's (*M. hotaula* or *M. ginkgodens hotaula*) beaked whales (Baumann-Pickering et al., 2010b), and Northern bottlenose whales (*Hyperoodon ampullatus*) (Wahlberg et al., 2011). Likewise, Stejneger's beaked whale (*M. stejnegeri*) FM pulses were recorded with bottom-moored autonomous acoustic instruments and linked to the species based on geographic location and exclusion of other species (Baumann-Pickering et al., 2012b). We developed a software tool to manually classify acoustic encounters with FM pulses in long-term data sets, likely produced by beaked whales, to certain beaked whale species (Baumann-Pickering et al., 2012a). We used histograms and median values of a number of signal parameters and an overlay of mean spectra to assign a species label to an acoustic encounter (Figure 3).



**Figure 3** – Example of classification tool used to label an acoustic encounter consisting of 1431 Cuvier’s beaked whale ( $Zc$ ) FM pulses. Top panel: Mean spectra of all automatically detected FM pulses of the example encounter denoted by black bold line. Mean spectra of templates for all other FM pulse types are denoted as thin dashed lines with the exception of  $Zc$ , which is shown as a thin solid black line to highlight the similarity with the example encounter. Middle panel: Histograms of peak frequency (left, pfr) and inter-pulse interval (IPI, right) with median values for peak, center frequency (cfr), duration (dur), and IPI. Bottom panel: mean spectra of encounter (left, solid line) and mean noise before each FM pulse (left, dashed line), with median peak-to-peak received level in dB re  $1 \mu\text{Pa}$  (ppRL) over all FM pulses in the encounter. Concatenated spectrogram of all FM pulses sorted by peak frequency showing variability (right).

### C. Representations of echolocation signals

Our examination of click spectra has focused on developing compact representations of spectra. Once a click has been identified (section III.A.3), spectra are extracted. We typically window the signal with a Hamming window and then zero-pad the click to a standard length (usually  $1200 \mu\text{s}$ ). A discrete Fourier transform (DFT) is computed and compensation for the acquisition system’s transfer function characteristics is applied. As the low end of the spectrum tends to be fairly noisy and does not appear to be a distinguishing feature in most echolocation clicks, we typically band pass the spectrum in the frequency domain between 10 and about 90 kHz. Most of our data is sampled at 192 or 200 kHz and the upper limit is simply to reduce any weak aliasing that may be left from the roll-off of our anti-alias filter.

Using the vector of spectral magnitudes across this frequency range would represent a very high dimensional feature space. At 200 kHz, there are 120 frequency bins between 0 Hz and the Nyquist rate, and our band-pass filter only discards about 20% of these. High dimensional

features are problematic for classifier systems. In a high dimensional space, it is very difficult to obtain enough data to robustly model a distribution and it becomes easy for models to be overtrained. As a consequence, one of the goals in feature extraction is to reduce the number of parameters to the minimal number needed and to ensure as much as possible that there is not redundant information in those features.

A common method for reducing the number of features is to use principal components analysis (for details, see Duda *et al.*, 2001). This method analyzes the variance-covariance matrix of a data set and finds its eigenvectors and eigenvalues. The eigenvectors can be used to form a new basis set and the data is projected onto the eigenvectors, forming a new set of coefficients. Unlike the original basis set, the bases with larger eigenvalues account for more of the variance of the data. This provides a convenient method to reduce dimensionality. One can discard the coefficients with the smallest associated eigenvalues and typically still preserve most of the variance in the dataset.

While principal components analysis is attractive, it assumes that one can estimate the covariance of the data, something that is not necessarily easy to do as there are many factors that are not necessarily linked to the production of the echolocation click that can affect how the features vary. A naïve example of this can be seen by considering the effects of using data from uncalibrated hydrophones where the signals have not been adjusted for differences in the transfer functions of the receiving systems (not recommended). Consider features derived from a single hydrophone versus data drawn from two independent deployments of hydrophones with different frequency response curves. If features are drawn from echolocation click spectra, there will be a single bias in the single hydrophone data set and two in the multiple hydrophone data. The mean of both data sets will be biased, but the second will contain a bias term that is a combination of both hydrophone biases. Consequently, one would expect the variances in the second set to be larger and principal components analysis of the second data set will reflect this change. While this particular example is controllable, there are other situations such as differences in high frequency attenuation that are not. As a consequence, while the principal component directions might reflect those of the measured dataset, they might not reflect those of the actual distribution of echolocation clicks.

As an alternative to the use of principal components, we use cepstral (pronounced  $\text{ˈ}t\text{ˌ}e\text{p}\text{ˌ}str\text{ˌ}əl$ ) analysis as a method of reducing this to a more manageable dimension. Cepstral analysis was developed independently by the two different groups (Bogert *et al.*, 1963; Oppenheim, 1964). Oppenheim and Schaffer (2004) provide a nice history of the parallel development. The name cepstral is derived from reversing the first syllable of spectral. The original goal of the Bogert *et al.* effort was to examine the effects of echoes in seismic events, and it was shown that the cepstrum was effective for identifying echo onsets. Oppenheim took a different approach and focused on the use of homomorphic transforms as a means of blind separation for signals that had been multiplied or convolved. The groups learned of one another's work and a number of

applications for the technique were developed such as the extraction of pitch from speech and the separation of the vocal tract response from a periodic glottal source. Cepstral features have been shown to be asymptotically independent (Merhav and Lee, 1993) and are widely used in speech processing systems (Huang et al., 2001; Jurafsky and Martin, 2009; Rabiner and Juang, 1993).

While there are still many open questions about how odontocetes produce echolocation clicks, there has been direct observational evidence (Cranford *et al.*, 2011) suggesting strongly that, in at least one species (*Turisiops truncatus*), pulses are generated at the phonic lips. Pressure from the intranarial passages most likely brings the lips into vibration and while each of the two pairs of lips can be controlled independently, the driving pressure appears to be coupled. While this occurs in a different anatomical structure than the vocal folds of other mammals, the process is somewhat similar and suggests that cepstral analysis may be an appropriate method of analysis for echolocation clicks.

The real-valued cepstrum (as opposed to complex) is derived by taking either the forward (Bogert *et al.*, 1963) or inverse (Oppenheim, 1964) discrete Fourier transform of the log magnitude spectrum. In practice, the discrete cosine transform (DCT) is often used which assumes symmetry and lets the practitioner use only the positive frequency coefficients. There are different forms of DCT depending upon the symmetry assumptions, with the one typically used for the cepstrum being the DCT-II which concentrates more energy in the lower index coefficients (Huang *et al.*, 2001):

$$C[k] = \sum_{n=0}^{N-1} \log(|X_w[n]|) \cos\left(\frac{\pi k(n + \frac{1}{2})}{N}\right)$$

where  $X_w[\cdot]$  is the discrete Fourier transform of the windowed signal. One way to think of the cepstrum is as a Fourier analysis of a spectral signal. Consequently, the 0<sup>th</sup> coefficient (which by tradition is not considered a cepstral coefficient) represents the sum of the energy in the log spectrum and is thus related to the energy in the original time series. When considering the Fourier analysis of the log spectrum, the lower order coefficients are responsible for the overall shape of the log spectrum and higher coefficients are responsible for quickly changing elements of the spectrum. Thus, by omitting the higher order cepstral coefficients, the remaining ones contain an intrinsic representation of the remaining spectra (Figure 4).

We have found empirically for a variety of odontocete species recorded in the Southern California Bight (Roch *et al.*, 2011b) that including more than the first fourteen cepstral coefficients did not provide any further improvements to classification error rate.

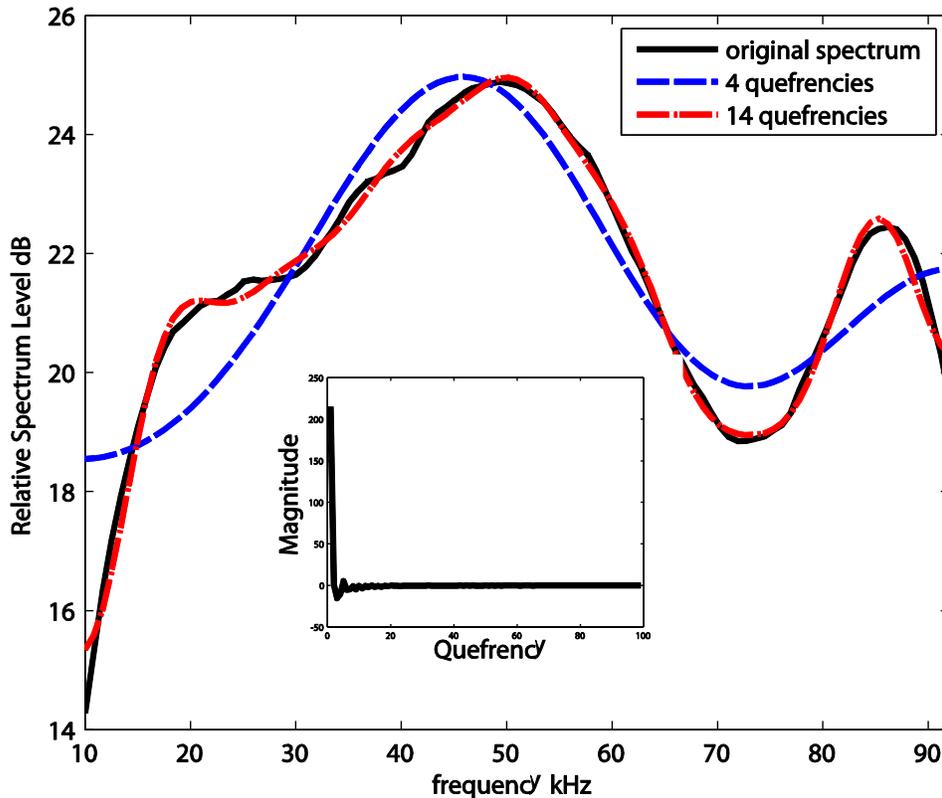


Figure 4 – Spectrum of a long-beaked common dolphin (*Delphinus capensis*) echolocation click. The cepstral representation of this click is shown as in inset where most of the energy can be seen to be concentrated in the first coefficients of the quefrendy axis. Reconstructions of the spectrum are shown using only the first four (dashed) and fourteen (dash-dot) quefrendys. (Coefficient zero was retained in each case to preserve the frequency offset.) A small number of coefficients contain most of the information about the click. From Roch et al. (2011b), permission pending.

### III. Detection and Classification

Traditionally, researchers in the bioacoustics community consider detection and classification as separate problems. A detector is a method for detecting a signal of interest, such as a call, or sequence of calls, without attempting to determine the specifics of the detected signal. In contrast, classification is the process of assigning one of a finite number of classes, a class label, to a detected signal. Class labels are highly dependent upon the question a researcher is trying to address. Valid labels for a single call produced by a blue whale might be: blue whale call, B call, song component, etc. The first of these examples might be of interest to someone considering density estimation or mitigation while the last class label might be more appropriate to people researching behavior.

A consequence of the great variety in the specificity of class labels is that the distinction between detection and classification can easily become blurred. One can see a detector as a form of classifier with coarse class labels, and while we will use the traditional term “detector,” the

difference between detection and classification is sometimes artificial. Following tradition, we will discuss detection and classification separately, but the reader is encouraged to think about the overlap between these two not so disparate problems.

## **A. Detection**

Cetaceans produce a variety of signal types. Mysticete calls are usually frequency modulated or pulsed calls. Odontocetes produce tonal calls that are referred to as whistles, and impulsive signals known as echolocation clicks. When the clicks are produced in rapid succession, they are referred to as burst pulses. In the case of beaked whales whose echolocation signals have a frequency modulated (FM) sweep, we have referred to these as echolocation FM pulses (Baumann-Pickering et al., 2010a). As the detection techniques for these different signal categories are quite different, we will describe them separately.

### **1. Mysticete calls**

Automatic detection of mysticete calls in our labs has been focused on blue whale B calls, Antarctic-type blue whale calls, fin whale 20 Hz pulses, North Pacific right whale up calls (*Eubalaena japonica*), Bryde's whale (*B. edeni*) Be4 calls, and humpback whale calls (*Megaptera novaeangliae*). Most detection of blue whale calls has been based on the spectrogram correlation method (Mellinger and Clark, 2000). Spectrogram correlation works reasonably well for stereotyped blue whale calls. In some cases such as the Western Antarctic Peninsula, where the background is relatively quiet, this method performs exceedingly well (Širović et al., 2004). In a noisier environment, interference from shipping noise can be a problem since the flat tonal nature of these calls, and their relatively long duration, make them difficult to distinguish from ships. In the Southern California Bight, fortunately, B calls have slight frequency modulation over the course of the call, which enabled development of a kernel that distinguishes the calls from ship noise (Oleson et al., 2007b; Wiggins et al., 2005). Bryde's whale Be4 calls, on the other hand, are more similar to ships lacking the frequency modulation, and thus are harder to detect automatically (Kerosky et al., 2012).

Spectrogram correlation relies on the stability of the call, and as previously mentioned, even the highly stereotyped blue whale calls change slightly over time (McDonald et al., 2009) and require annual adjustment for change in call frequency (Oleson et al., 2007b). Spectrogram correlation is also difficult when calls appear in very large numbers. Fin whale 20 Hz pulses can occur in such volume that they are not individually distinguishable, and in this case we have successfully detected these calls by comparing acoustic power in the frequency band representative of fin whale 20 Hz pulse calls to noise in adjacent bands (Širović et al., 2004). Another approach is to find peaks in a spectrogram and link them together, an approach that we used for blue whale B and D calls (Madhusudhana et al., 2008). This technique was later extended to odontocete whistles as described in the next section.

Humpback whale calls have significant variation, making them less amenable to spectral correlation techniques. In the Southern California Bight, there are relatively few competing calls

in the bandwidth of humpback whale calls. Consequently, we have been able to design a generalized power-law detector for these vocalizations (Helble *et al.*, 2012). The power-law detector, first proposed by Nutall (1994), is a statistical test that compares the hypothesis that a signal occupies a certain number of frequency bins in a spectrogram slice to a null hypothesis of no signal present. Nutall showed that raising the spectral magnitude to higher powers is more effective for many signal types, and proposed an appropriate exponent (2.5) when the number of bins is completely unknown. In Helble *et al.* we introduced steps to reduce the contribution of both near constant frequency and impulsive signals as well as a post processing step that groups detection frames and accepts or rejects them as humpback whale calls based on simple criteria such as signal length.

Most mysticete calls discussed thus far are relatively common. When this is the case, it is possible to minimize the ratio of missed detections and false alarms, without the fear of losing important calls. On the other hand, when dealing with rare species, such as the North Pacific right whale (*Eubalaena japonica*), and their very rare calls, the ideal detector would not miss any calls. In the case of North Pacific right whale up calls, the most useful spectrogram correlation yielded high numbers of false alarms and low numbers of missed calls, but generally it was useful for finding time periods with these calls (Munger *et al.*, 2005; Munger *et al.*, 2008). Subsequent manual scanning of data around times with confirmed positive detections was used for detection of calls. So while the automatic detector was not a good way to reliably detect calls, it was a good tool for identifying times around which more manual effort needed to be focused, thus leading to overall decrease in the analysis effort.

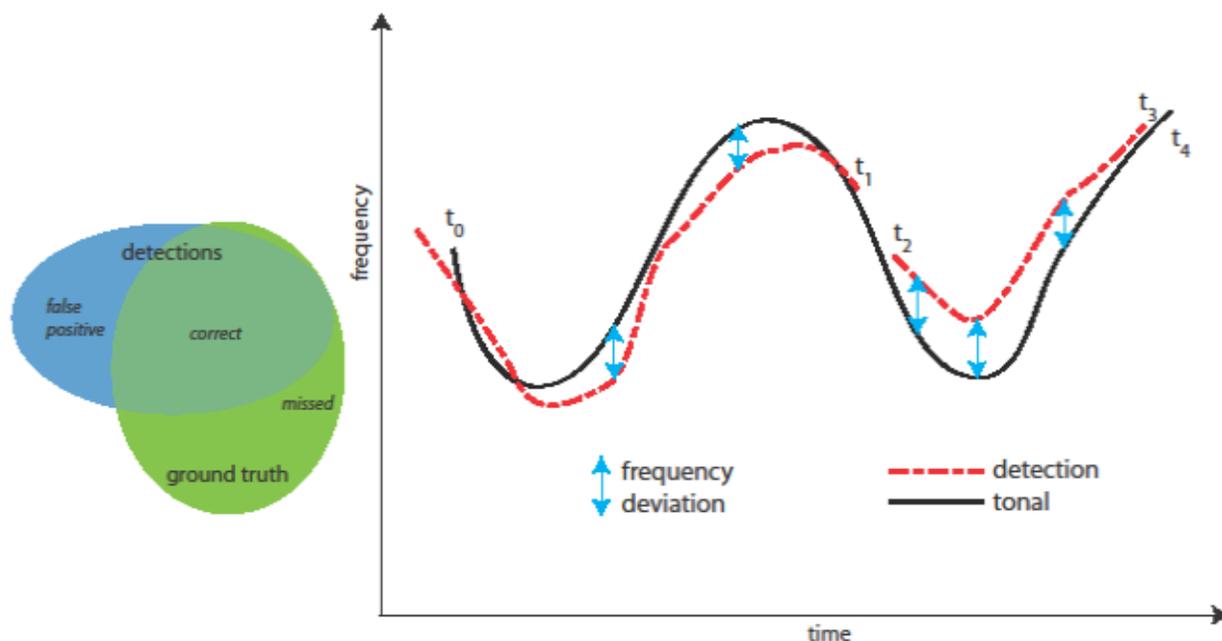
If the goal of the study is not to document presence of calling animals on a very short time scale (minutes to hours), or count the number of calls, more generalized spectral methods can be used for detection of calling baleen whales. Burtenshaw *et al.* (2004) had acoustic data that were already spectrally averaged and thus could not be used to extract individual calls and their features from the data. Instead, they used energy in a band representative of blue whale B calls (48 Hz) as a proxy for the presence of calling blue whales. Similarly, monthly spectra compared well to individual detections when used for longer time-scale estimation of the presence of calling blue and fin whales around Antarctica (Širović *et al.*, 2009).

## **2. Odontocete whistles**

The whistles of odontocetes are complex and frequently only partially observable. The missing portions of a whistle can be attributed to masking events (e.g. ship noise for low frequency whistles), high-frequency attenuation, and the high source level of echolocation clicks (e.g. Au and Snyder, 1980, observed a peak to peak source level of 220 dB re 1  $\mu$ Pa in bottlenose dolphins) which can saturate columns of the spectrogram. In addition, the whistles sometimes display nonlinearities such as sudden changes in fundamental frequency (often referred to as steps).

We developed two whistle detection algorithms that attempted to address some of these issues (Roch *et al.*, 2011a). One of our goals in this work was to provide a common test bed that other researchers could use. Many researchers do not have access to large datasets and consequently have published results on a handful of examples. We wished to provide a resource that would enable groups to have access to extensive annotated test sets in the hopes that it would enable the development of more robust algorithms. To this end, we developed an annotation tool to permit analysts to annotate dolphin whistles by specifying knots in cubic splines. We used this tool to annotate over 30,000 whistles from common dolphins, melon-headed whales, bottlenose dolphins (*Tursiops truncatus*), and spinner dolphins (*Stenella longirostris longirostris*). These whistles, and their annotations have been donated to the bioacoustics community and are available from the Moby Sound archive (<http://www.mobysound.org>, Mellinger and Clark, 2006), and were used for the whistle detection task in the 2011 Detection, Classification, Localization, and Density Estimation of Marine Mammals Using Passive Acoustics workshop.

We also defined a number of metrics to gauge the performance of whistle detectors. We used the concepts of precision and recall (Duda *et al.*, 2001), precision being the rate of detections that are correct (false positive rate = 1 - precision) and recall being their rate at which valid detections are made as compared to the number of expected detections. In addition to these basic metrics, several qualitative ones were defined: coverage, fragmentation, and deviation (Figure 5). Whistle detectors may not detect the entire whistle or may detect it as fragments, and the first two metrics are designed to measure these phenomena. The last metric provides a measurement of how closely the detector whistle contour follows the analyst detected one by measuring the average difference in frequency and provided Matlab-based software (available on Moby Sound along with the workshop data) to compute these metrics.



**Figure 5 - Metrics used to characterize detections.** The Venn diagram on the left shows the overlap between the detected tonals and ground truth (analyst annotated) data. Recall computes the rate of correct detections relative to the ground truth while precision is the rate of detections that were correct. The exaggerated caricatures of a call and associated detections on the right illustrate the quality metrics. Average deviation is the mean frequency deviation between the tonal call and detection(s). As systems may detect a call in multiple pieces, or fragments, the number of fragments per call is recorded. Coverage is an indication of the percentage of the tonal that was detected. From Roch et al. (2011a), permission pending.

Both whistle detectors used a common signal processing chain for spectrogram generation, noise compensation, and peak selection. The first detector was based on particle filters, a form of Bayesian filtering. The basic idea of Bayesian filtering in this context is that given a set of sample time×frequency peaks from a distribution that is a local estimate of the whistle contour, they can be used to estimate the posterior distribution of the next peak. The posterior distribution estimate is then used as a prior distribution for the following peak. In particle filters, each of these peaks has an importance weight attached to it that determines its contribution to the posterior estimate. An update function adjusts the weights and gradually throws out samples as they become less important, attaching higher weight to new samples. We also updated samples based upon a motion model once the contour was of sufficient length.

The second detector was based upon a graph representation of the contours. When multiple animals vocalize simultaneously, their whistles may intersect. To our knowledge, all of the whistle tracking algorithms prior to our graph detector handle whistle crossings based upon the evidence prior to the intersection. In our work, when whistles cross, we simply consider this to be a node in a graph. As new time×frequency peaks are found, they are added to paths in existing graphs or if they are too far away, new graphs are formed (Figure 6).

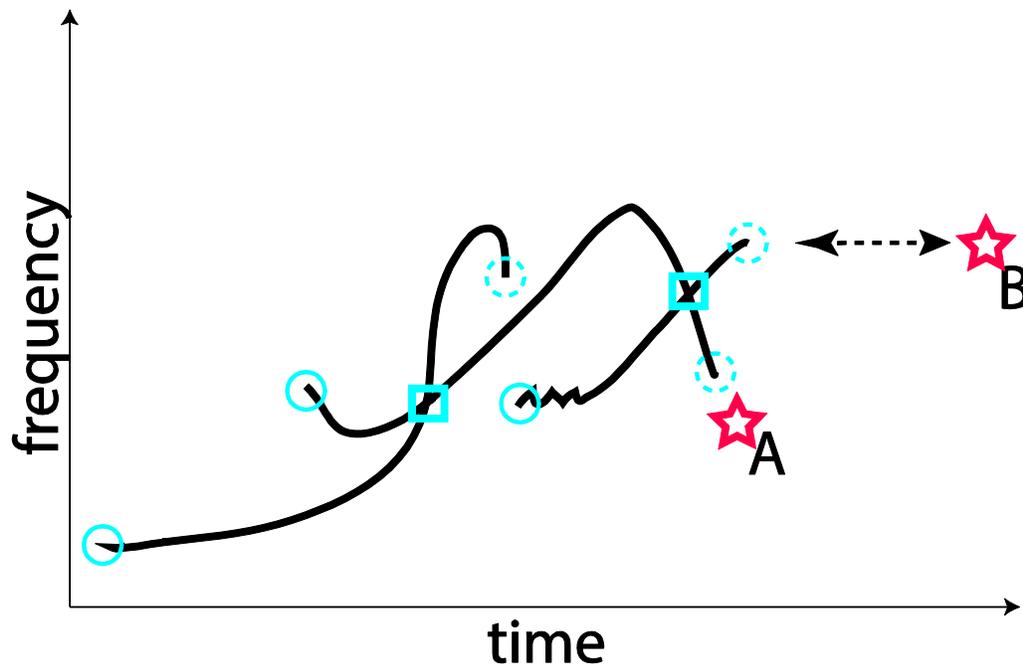


Figure 6 – Graph representation of whistles. Stylized common dolphin contours are shown above with two whistle crossings (squares). Circles represent beginnings (solid) or ends (dashed) of graphs. When a new node is discovered the graph can be extended from one of the end points if it is consistent with the local history of the whistle (e.g. star A). When processing advances to the point that it is no longer reasonable to add a peak (e.g. star B), the interior nodes are examined to determine how the crossings should be resolved.

The common signal processing chain permitted a direct comparison between the two algorithms. On a subset of over 3000 whistles (those annotated at the time that the experiments were conducted), both algorithms showed good recall (70% particle filter and 80% graph) and reasonable precision (61% particle filter and 80% graph) with good quality metrics. The vast majority of the false positives were for very short detections, although other transient signals such as echo sounders proved problematic.

The annotation tool, graph detection algorithm, and scoring metrics are part of the *silbido*<sup>1</sup> detection suite that is publicly available at <http://roch.sdsu.edu>.

### 3. Odontocete echolocation clicks

Our echolocation click detector is based upon the Teager energy operator (Kaiser, 1990). The Teager energy operator, sometimes called the Teager-Kaiser, energy operator, is based on a model of the energy to drive a mass spring oscillator. Each sample estimate of Teager energy only requires the current acoustic sample and its previous and subsequent neighbor. Thus, it responds rapidly to changes in energy such as those produced by echolocation clicks. Some of the assumptions for the underlying model are not met for echolocation clicks, but the Teager

<sup>1</sup> *Silbido* is Spanish for whistle.

energy operator has been shown empirically to be an effective basis for echolocation click detectors and is widely used.

Our implementation was developed at about the same time as the Teager energy click detector proposed by Kandia and Stylianou (2006) who were the first to publish results. While our click detector uses the same energy operator, it has several differences from the one proposed by Kandia and Stylianou. We use a two stage detection model, where the spectral content of longer frames (10 ms) are examined for sufficient energy in the click bandwidth. When regions with possible echolocation clicks are identified, the time domain signal is high-pass filtered and the Teager energy operator is used to determine possible click regions. Each of these regions is grown using an energy growth algorithm similar to that described by Au (1993). When clicks are too close together, we only retain the first one to prevent the selection of echoes. Due to Kandia and Stylianou's publication, the algorithm was never published separately, but is described more fully in some of our other papers (Roch et al., 2011a; Soldevilla, 2008).

## **B. Classification**

Due to the ambiguity between the terms detection and classification, many of the detectors previously discussed can be thought of as classifiers. For example, the spectrogram correlation methods are making the classification decision that a recording segment is either a specific call of interest or not.

Our two-stage classifier work consists of detector, which one can think of as a classifier with very broad class labels (e.g. echolocation click or something else), followed by a classifier designed to learn more about the specifics. Our first foray into this type of classification system was a paper that asked the simple question of whether or not there was a gestalt to the auditory scene associated with vocalizing odontocetes that would permit one to determine the species producing the calls (Roch *et al.*, 2007). Rather than attempt to derive features from individual clicks or whistles, which has been the focus of our later work, we simply applied an energy-based signal detector and extracted cepstra from relatively long 21 ms windows. A Gaussian mixture model (GMM) classifier exhibited reasonable performance, with an error rate ranging from .25 to .33 depending upon testing conditions for a three species task.

Later work focused on individual calls, specifically echolocation clicks. Our first effort in this area used the aforementioned Teager click detector and then examined the performance of support vector machines (SVM) and GMM classifiers (Roch *et al.*, 2008) on data from the 2007 Detection, Classification, and Localization of Marine Mammal Using Passive Acoustic Monitoring workshop. Both classifiers were comparable although the GMM exhibited a slight edge. The system was named the best performing system of the conference (Moretti *et al.*, 2008). Subsequent unpublished work examined a variety of classifiers, including the well-known adaptive boosting (Freund and Schapire, 1999), although our experiments showed that most of these classifiers did not perform significantly better than any of the others that we had already used.

As a consequence, we turned our attention to feature extraction and published a study focusing on small odontocetes (Roch *et al.*, 2011b). In this study, we significantly reduced the dimensionality of the features we used and began reporting the results based on randomization of training and test data, stressing the importance of ensuring that data from a single acoustic encounter was never split across the training and test set.

#### **IV. Localization**

After a cetacean call has been detected and classified, it can often be useful to know where the calling animal is located in space. Localization is necessary when estimating the source level of a calling animal, but it can also be useful for answering a variety of ecological questions, such as those related to population estimation or habitat modeling. The basic method for resolving the location of a sound source uses an array of recorders and the difference in the time of arrival of the signal to multiple elements of the array. Usually, the signal needs to be received on at least three elements in the array if the exact location is to be calculated. If some elements have directional capabilities, that number can be reduced. For example, only two DIFAR (DIrectional Fixing And Ranging) sonobuoys can be sufficient to localize a source. Alternatively, we were also able to use special propagation characteristics in the environment for determining the range to the calling whales using calls received at a single instrument.

Time difference of arrival of a signal to multiple receivers has been used for localization of blue whales and sei whales (*B. borealis*) in the Antarctic (McDonald *et al.*, 2005; Širović *et al.*, 2007), although in the case of sei whales, DIFAR sonobuoys were also used for localization. DIFAR sonobuoys have also been used in the Bering Sea to find right whales (McDonald and Moore, 2002) and in the Southern California Bight to assist in finding blue whales (Oleson *et al.*, 2007a).

In shallow water, well mixed environments, waveguide propagation creates range-dependent waveform dispersion. When frequency modulated sounds travel in a waveguide, the difference in travel speeds at different frequencies, with lower frequencies traveling slower, results in different arrival times of individual modes. By modeling propagation in such an environment, it is possible to determine the distance to the source from the characteristics of the received signal for low frequency signals. One such shallow water environment with waveguide propagation is the Bering Sea. Wiggins *et al.* (2004) used this characteristic of the propagating environment in the Bering Sea to determine ranges to North Pacific right whales producing up calls using calls recorded at a single instrument. While they were not able to localize on individual whales, the range information was sufficient to calculate source level of the up calls (Munger *et al.*, 2011).

In deep water, we were able to determine the range to a calling animal using multipath arrivals of a call also to a single instrument. When a call travels from the source to the receiver, it can travel in the direct path, but it can also bounce from the ocean bottom and the surface. The latter

paths are longer than the direct path and thus take longer time, resulting in multiple arrivals of the same call to one instrument. By measuring the difference in the time of arrival of the different paths to the receiver, it is possible to determine the range to the source. This method has been applied to fin whale 20 Hz pulses in the North Pacific and the Antarctic (McDonald et al., 1995; Širović et al., 2007), as well as on the down swept portion of blue whale calls in the Antarctic (Širović et al., 2007), to determine the range to the calling animals over 10s of km, establish a pattern in calling by a number of individuals, and estimate call source levels.

Occasionally, it is possible to use multiple methods for localization or ranging to calling animals, which can serve as verification of the methodologies. In the Antarctic, we were able to calculate distances to calling blue whales using time difference of arrivals and multipath propagation and found that the two results were different by not more than 7% (Širović et al., 2007). In the Bering Sea, the comparison of time difference of arrivals to normal mode propagation ranging yielded results, which were on average within 12% of each other (Munger et al., 2011).

## V. Summary

The last ten years have shown tremendous growth in the capabilities of detection, classification, and localization algorithms, both in our labs and the field in general. Many algorithms have moved from the field of research into the hands of trained acoustic analysts, packaged with graphical user interfaces (Figueroa and Robbins, 2007; Gillespie et al., 2008; Mellinger, 2001; Wiggins et al., 2007). Yet there remain many unanswered questions and the error of many state of the art algorithms still leaves much to be desired. The next ten years are likely to bring continued improvements to these algorithms, common use of passive acoustic monitoring in density estimation, and the study of calls in context as our ability to observe detailed movements continues to improve.

## References

- Aguilar de Soto, N., Madsen, P. T., Tyack, P., Arranz, P., Marrero, J., Fais, A., Revelli, E. and Johnson, M.** (2012). No shallow talk: Cryptic strategy in the vocal communication of Blainville's beaked whales. *Mar. Mamm. Sci.* **28**, E75-E92.
- Au, W. W. L.** (1993). *The Sonar of Dolphins*. New York: Springer-Verlag.
- Au, W. W. L., Branstetter, B., Moore, P. W. and Finneran, J. J.** (2012a). The biosonar field around an Atlantic bottlenose dolphin (*Tursiops truncatus*). *J. Acous. Soc. Am.* **131**, 569-576.
- Au, W. W. L., Branstetter, B., Moore, P. W. and Finneran, J. J.** (2012b). Dolphin biosonar signals measured at extreme off-axis angles: Insights to sound propagation in the head. *J. Acous. Soc. Am.* **132**, 1199-1206.
- Au, W. W. L. and Snyder, K. J.** (1980). Long-range target detection in open waters by an echolocating Atlantic Bottlenose dolphin (*Tursiops truncatus*). *J. Acous. Soc. Am.* **68**, 1077-1084.
- Baumann-Pickering, S., Simonis, A. E., Roch, M. A., McDonald, M. A., Solsona-Berga, A., Oleson, E. M., Wiggins, S. M., Brownell Jr, R. L. and Hildebrand, J. A.** (2012a). Spatio-temporal patterns of beaked whale echolocation signals in the North Pacific, pp. 16. Panamá, Panamá: Intl. Whaling Commission.

- Baumann-Pickering, S., Simonis, A. E., Wiggins, S. M., Brownell, R. L. and Hildebrand, J. A.** (2012b). Aleutian Islands beaked whale echolocation signals. *Marine Mammal Science*.
- Baumann-Pickering, S., Wiggins, S. M., Roth, E. H., Roch, M. A., Schnitzler, H.-U. and Hildebrand, J. A.** (2010a). Echolocation signals of a beaked whale at Palmyra Atoll. *J. Acous. Soc. Am.* **127**, 3790-3799.
- Baumann-Pickering, S., Wiggins, S. M., Roth, E. H., Roch, M. A., Schnitzler, H. U. and Hildebrand, J. A.** (2010b). Echolocation signals of a beaked whale at Palmyra Atoll. *J. Acous. Soc. Am.* **127**, 3790-3799.
- Baumann-Pickering, S., Yack, T. M., Wiggins, S. M., Hildebrand, J. A. and Barlow, J.** (submitted). Description of Baird's beaked whale echolocation signals. *J. Acous. Soc. Am.*.
- Bogert, B. P., Healy, M. J. R. and Tukey, J. W.** (1963). The quefrency analysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking. In *Time Series Analysis*, (ed. M. Rosenblatt), pp. 209-243.
- Burtenshaw, J. C., Oleson, E. M., Hildebrand, J. A., McDonald, M. A., Andrew, R. K., Howe, B. M. and Mercer, J. A.** (2004). Acoustic and satellite remote sensing of blue whale seasonality and habitat in the Northeast Pacific. *Deep Sea Res. Pt. II* **51**, 967-986.
- Cranford, T. W., Elsberry, W. R., Van Bonn, W. G., Jeffress, J. A., Chaplin, M. S., Blackwood, D. J., Carder, D. A., Kamolnick, T., Todd, M. A. and Ridgway, S. H.** (2011). Observation and analysis of sonar signal generation in the bottlenose dolphin (*Tursiops truncatus*): Evidence for two sonar sources. *J. Exp. Mar. Biol. Ecol.* **407**, 81-96.
- Dawson, S., Barlow, J. and Ljungblad, D.** (1998). Sounds recorded from Baird's beaked whale, *Berardius bairdii*. *Mar. Mamm. Sci.* **14**, 335-344.
- Duda, R. O., Hart, P. E. and Stork, D. G.** (2001). *Pattern Classification*. New York, NY: Wiley-Interscience.
- Figueroa, H. and Robbins, M.** (2007). XBAT: An Open-Source Extensible Platform for Bioacoustic Research and Monitoring. In *Intl. Academy for Nature Conservation*, eds. K.-H. Frommolt R. Bardeli and M. Calusen), pp. 143-155. Isle of Vilm, Germany: BfN Skripten.
- Freund, Y. and Schapire, R. E.** (1999). A Short Introduction to Boosting (Japanese, English version available at <http://cseweb.ucsd.edu/~yfreund>). *J. Japanese Soc. Artif. Int.* **14**, 771-780.
- Gillespie, D., Dunn, C., Gordon, J., Claridge, D., Embling, C. and Boyd, I.** (2009). Field recordings of Gervais' beaked whales *Mesoplodon europaeus* from the Bahamas. *J. Acous. Soc. Am.* **125**, 3428-3433.
- Gillespie, D., Gordon, J., McHugh, R., McLaren, D., Mellinger, D. K., Redmond, P., Thode, A., Trinder, P. and Deng, X.-Y.** (2008). PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localisation of cetaceans. In *Proc. Inst. Acoustics*, vol. 30, pp. 9.
- Helble, T., Ierley, G. R., D'Spain, G. L., Roch, M. A. and Hildebrand, J. A.** (2012). A generalized power-law detection algorithm for humpback whale vocalizations. *J. Acous. Soc. Am.* **131**, 2682-2699.
- Huang, X., Acero, A. and Hon, H. W.** (2001). *Spoken Language Processing*. Upper Saddle River, NJ: Prentice Hall PTR.
- Johnson, M., Madsen, P. T., Zimmer, W. M. X., Aguilar de Soto, N. and Tyack, P.** (2006). Foraging Blainville's beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation. *Journal of Experimental Biology* **209**, 5038-5050.
- Johnson, M., Madsen, P. T., Zimmer, W. M. X., de Soto, N. A. and Tyack, P. L.** (2004). Beaked whales echolocate on prey. *Proceedings of the Royal Society B: Biological Sciences* **271**, S383-S386.
- Jurafsky, D. and Martin, J. H.** (2009). *Speech and Language Processing*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Kaiser, J. F.** (1990). On a simple algorithm to calculate the "energy" of a signal. In *Intl. Conf. Acoust., Speech, and Signal Proces.*, vol. 1, pp. 381-384. Albuquerque, NM: IEEE.

- Kandia, V. and Stylianou, Y.** (2006). Detection of sperm whale clicks based on the Teager-Kaiser energy operator. *Appl. Acous.* **67**, 1144-1163.
- Kerosky, S. M., Sirovic, A., Roche, L. K., Baumann-Pickering, S., Wiggins, S. M. and Hildebrand, J. A.** (2012). Bryde's whale seasonal range expansion and increasing presence in the Southern California Bight from 2000 to 2010. *Deep-Sea Research Part I-Oceanographic Research Papers* **65**, 125-132.
- Lammers, M. O. and Castellote, M.** (2009). The beluga whale produces two pulses to form its sonar signal. *Biol. Lett.* **5**, 297-301.
- Madhusudhana, S. K., Oleson, E. M., Soldevilla, M. S., Roch, M. A. and Hildebrand, J. A.** (2008). Frequency based algorithm for robust contour extraction of blue whale B and D calls. In *OCEANS'08*, pp. 8. Kobe, Japan.
- Madsen, P. T., Johnson, M., de Soto, N. A., Zimmer, W. M. X. and Tyack, P.** (2005). Biosonar performance of foraging beaked whales (*Mesoplodon densirostris*). *Journal of Experimental Biology* **208**, 181-194.
- McDonald, M. A., Hildebrand, J. A. and Sarah, M.** (2009). Worldwide decline in tonal frequencies of blue whale songs. *Endang. Species Res.* **9**, 13-21.
- McDonald, M. A., Hildebrand, J. A. and Webb, S. C.** (1995). Blue and Fin Whales Observed on a Sea-Floor Array in the Northeast Pacific. *J. Acous. Soc. Am.* **98**, 712-721.
- McDonald, M. A., Hildebrand, J. A., Wiggins, S. M., Thiele, D., Glasgow, D. and Moore, S. E.** (2005). Sei whale sounds recorded in the Antarctic. *J. Acous. Soc. Am.* **118**, 3934-3940.
- McDonald, M. A. and Moore, S. E.** (2002). Calls recorded from North Pacific right whales (*Eubalaena japonica*) in the eastern Bering Sea. *J. Cetacean Res. Mgt.* **4**, 261-266.
- Mellinger, D. and Clark, C.** (2006). MobySound: A reference archive for studying automatic recognition of marine mammal sounds. *Appl. Acous.* **67**, 1226-1242.
- Mellinger, D. K.** (2001). Ishmael 1.0 User's Guide, pp. 30. Seattle: NOAA PMEL.
- Mellinger, D. K. and Clark, C. W.** (2000). Recognizing transient low-frequency whale sounds by spectrogram correlation. *J. Acous. Soc. Am.* **107**, 3518-3529.
- Merhav, N. and Lee, C.-H.** (1993). On the asymptotic statistical behavior of empirical cepstral coefficients. *IEEE Trans. Signal Processing* **41**, 1990-1993.
- Moretti, D., DiMarzio, N., Morrissey, R., Mellinger, D. K., Heimlich, S. and Pettis, H.** (2008). Overview of the 3<sup>rd</sup> international workshop on the detection and classification of marine mammals using passive acoustics. *Canadian Acoust.* **36**, 7-11.
- Munger, L. M., Mellinger, D. K., Wiggins, S. M., Moore, S. E. and Hildebrand, J. A.** (2005). Performance of Spectrogram Cross-Correlation in Detecting Right Whale Calls in Long-Term Recordings from the Bering Sea. *Canadian Acoust.* **33**, 25-34.
- Munger, L. M., Wiggins, S. M. and Hildebrand, J. A.** (2011). North Pacific right whale up-call source levels and propagation distance on the southeastern Bering Sea shelf. *J. Acous. Soc. Am.* **129**, 4047.
- Munger, L. M., Wiggins, S. M., Moore, S. E. and Hildebrand, J. A.** (2008). North Pacific right whale (*Eubalaena japonica*) seasonal and diel calling patterns from long-term acoustic recordings in the southeastern Bering Sea, 2000-2006. *Mar. Mammal Sci.* **24**, 795-814.
- Nuttall, A. H.** (1994). Detection Performance of Power-Law Processors for Random Signals of Unknown Location, Structure, Extent, and Length. Newport, RI: Naval Underwater Warfare Center (NUWC)-NPT.
- Oleson, E. M., Calambokidis, J., Barlow, J. and Hildebrand, J. A.** (2007a). Blue whale visual and acoustic encounter rates in the southern California bight. *Mar. Mammal Sci.* **23**, 574-597.
- Oleson, E. M., Wiggins, S. M. and Hildebrand, J. A.** (2007b). Temporal separation of blue whale call types on a southern California feeding ground. *Anim. Behav.* **74**, 881-894.
- Oppenheim, A. V.** (1964). Superposition in a class of nonlinear systems. PhD thesis, Massachusetts Inst. Tech.
- Oppenheim, A. V. and Schaffer, R. W.** (2004). From frequency to quefrequency: a history of the cepstrum. *IEEE Signal Proc. Mag.* **21**, 95-99,106.

- Rabiner, L. R. and Juang, B.-H.** (1993). Fundamentals of speech recognition. Englewood Cliffs, NJ 07632: Prentice-Hall.
- Rankin, S., Baumann-Pickering, S., Yack, T. and Barlow, J.** (2011). Description of sounds recorded from Longman's beaked whale, *Indopacetus pacificus*. *J. Acous. Soc. Am.* **130**, EL339-EL344.
- Roch, M. A., Brandes, T. S., Patel, B., Barkley, Y., Baumann-Pickering, S. and Soldevilla, M. S.** (2011a). Automated extraction of odontocete whistle contours. *J. Acous. Soc. Am.* **130**, 2212-2223.
- Roch, M. A., Klinck, H., Baumann-Pickering, S., Mellinger, D. K., Qui, S., Soldevilla, M. S. and Hildebrand, J. A.** (2011b). Classification of echolocation clicks from odontocetes in the Southern California Bight. *J. Acous. Soc. Am.* **129**, 467-475.
- Roch, M. A., Soldevilla, M. S., Burtenshaw, J. C., Henderson, E. E. and Hildebrand, J. A.** (2007). Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California. *J. Acous. Soc. Am.* **121**, 1737-1748.
- Roch, M. A., Soldevilla, M. S., Hoenigman, R., Wiggins, S. M. and Hildebrand, J. A.** (2008). Comparison of machine learning techniques for the classification of echolocation clicks from three species of odontocetes. *Canadian Acoust.* **36**, 41-47.
- Širović, A., Hildebrand, J. A. and Wiggins, S. M.** (2007). Blue and fin whale call source levels and propagation range in the Southern Ocean. *J. Acous. Soc. Am.* **122**, 1208-1215.
- Širović, A., Hildebrand, J. A., Wiggins, S. M., McDonald, M. A., Moore, S. E. and Thiele, D.** (2004). Seasonality of blue and fin whale calls and the influence of sea lee in the Western Antarctic Peninsula. *Deep-Sea Res. Pt. II* **51**, 2327-2344.
- Širović, A., Hildebrand, J. A., Wiggins, S. M. and Thiele, D.** (2009). Blue and fin whale acoustic presence around Antarctica during 2003 and 2004. *Mar. Mammal Sci.* **25**, 125-136.
- Soldevilla, M. S.** (2008). Risso's and Pacific white-sided dolphins in the Southern California Bight using echolocation clicks to study dolphin ecology. PhD thesis, University of California at San Diego.
- Soldevilla, M. S., Henderson, E. E., Campbell, G. S., Wiggins, S. M., Hildebrand, J. A. and Roch, M. A.** (2008). Classification of Risso's and Pacific white-sided dolphins using spectral properties of echolocation clicks. *J. Acous. Soc. Am.* **124**, 609-624.
- Wahlberg, M., Beedholm, K., Heerfordt, A. and Mohl, B.** (2011). Characteristics of biosonar signals from the northern bottlenose whale, *Hyperoodon ampullatus*. *J. Acous. Soc. Am.* **130**, 3077-3084.
- Wiggins, S. M., McDonald, M. A., Munger, L. M., Moore, S. E. and Hildebrand, J. A.** (2004). Waveguide Propagation Allows Range Estimates for North Pacific Right Whales in the Bering Sea. *Canadian Acoust.* **32**, 146-154.
- Wiggins, S. M., Oleson, E. M., McDonald, M. A. and Hildebrand, J. A.** (2005). Blue whale (*Balaenoptera musculus*) diel call patterns offshore of Southern California. *Aquat. Mamm.* **31**, 161-168.
- Wiggins, S. M., Soldevilla, M. S., Roch, M. A. and Hoenigman, R.** (2007). Triton User Manual: Scripps Institution of Oceanography.
- Zimmer, W. M. X., Harwood, J., Tyack, P. L., Johnson, M. P. and Madsen, P. T.** (2008). Passive acoustic detection of deep-diving beaked whales. *J. Acous. Soc. Am.* **124**, 2823-2832.
- Zimmer, W. M. X., Madsen, P. T., Teloni, V., Johnson, M. P. and Tyack, P. L.** (2005). Off-axis effects on the multipulse structure of sperm whale usual clicks with implications for sound production. *J. Acous. Soc. Am.* **118**, 3337-3345.