A review of current marine mammal detection and classification algorithms for use in automated passive acoustic monitoring.

Michael Bittle (1) and Alec Duncan (1)

(1) Centre for Marine Science and Technology, Curtin University, Bentley, WA, Australia

ABSTRACT

The detection and classification of marine mammal vocalisations is an important component in noise mitigation strategies and in the tracking of animals for research purposes. These complex vocalisations span a broad range of frequencies with differences between and within species, and with temporal and geographical variations adding further complexity. Passive Acoustic Monitoring (PAM) systems can be deployed for long periods and can collect large volumes of data, becoming impractical for human operators to manually process due to the significant effort required. Many signal processing algorithms to automate this process have been produced with mixed results. Some are focused on the identification of single species while others handle a variety. No single algorithm is ideal for detecting and classifying all species concurrently, so any automated system requires a suite of these algorithms. A number of these algorithms are summarised here as part of an initial step in the construction of a PAM system incorporating real-time detection and classification.

INTRODUCTION

Passive Acoustic Monitoring (PAM) of marine mammals provides a method of observation which can supplement or replace the visual method of monitoring which has traditionally been used. Acoustics can provide a means of monitoring animals at great distances, compared to visual methods, due to the fact that sound can propagate much further in the ocean than light can (Cato, Noad & McCauley 2005). Vocalisation is common in many marine mammals, and is used for social, navigational (Verfuß, Miller & Schnitzler 2005) and predatory purposes. Visual observations can also be hindered by environmental conditions, such as inaccessible geography, remoteness, time of day or by the effects of weather, and can be limited in their use due to the short times animals may spend at the surface. While PAM is not immune to the effects of weather, it can still provide useful information where visual observations are unsuitable.

The vocalisations of some species can travel large distances underwater, able to be detected up to tens of kilometres away for whales (Medwin & Blue 2005) and their unique characteristics provide a means to differentiate between species and even individuals. Current PAM systems can be described as those which simply record all sound for post-processing, such as noise loggers, and those which perform some real-time processing performed either on-board or by offloading to a separate computing system, such as in cabled hydrophone arrays or sonar buoys.

Basic loggers are usually filtered for frequencies of interest and recorded on a duty-cycle. Eventually the data must be retrieved and transported to the lab for post-processing by a combination of human operators and computer algorithms. Other systems which transmit data back to a base can do so potentially in real time. These data can be processed immediately but the computational effort increases with every node in an array. When not processed immediately, large scale data storage poses additional problems such as costs and facilities.

PAM systems combining detection and/or classification at the source of the recording can alleviate some of these issues. This has obvious benefits since it allows the deployment of larger arrays without the increasing requirements of processing power and data storage. This can greatly reduce the effort required by analysts, reduce the time spent on data retrieval, and can increase the time that the sensors are deployed to collect data.

Some existing PAM systems already include detection. An example is the T-POD (Timed POrpoise Detector), which was designed to detect porpoise click trains and could be used for other species (Philpott et al. 2007). Despite the ability to detect some other species this was still limited in use in more diverse cetacean communities (Thompson et al. 2010). The C-POD (Cetacean POrpoise Detector) superseded the T-POD and was designed to detect clicks from a larger variety of species, however it is unable to detect clicks from the sperm whale due to limitations in the frequency range (*C-POD Species Detection* 2013). These systems are limited to only detecting click trains with limited classification capability. Another system, the PAMBuoy, was recently released and includes a detection and classification system utilising the PAMGuard software suite (Marine Instrumentation Ltd. 2012). In a recent trial monitoring beluga whales in Alaska the PAMBuoy system was found to not perform as well as a human observer although it did reliably detect all whales approaching or entering the river where the test was performed (Gillespie 2013).

In developing a PAM system which includes detection and classification for marine mammals, it is possible to draw on a large body of research investigating and describing the various characteristics of vocalisations along with algorithms that have been created to detect them. This research has shown that there is extreme variation between species and that no one algorithm is suited to detect all species concurrently. Any system with the goal of detecting multiple species must utilise a range of algorithms. A large selection of these algorithms were collected and catalogued as part of the preliminary research for the development of such a system.

Comprehensive discussions regarding most of the techniques in these algorithms can be found in Au & Hastings (2008) and Zimmer (2011). The use of those techniques in the papers cited here, for example spectrogram correlation, types of feature extraction and neural networks shows that these are accepted and tested methods, but there are also some novel techniques discussed here which are areas of new and active research.

Discussion of these algorithms is preceded by a short overview of marine mammal sounds and a discussion of signal flow in PAM systems. The review is roughly structured into four broad categories. The first is feature extraction which describes a major component of all classification algorithms. Feature extraction seeks to define relevant features of a signal which can be extracted for detection and classification. The second discusses energy based approaches, which are arguably just another feature extraction method but in some cases are simpler and all of them focus largely on energy content. The third is a category which processes information from extracted features using statistical analysis methods such as Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs) for classification. Finally, the fourth category discusses some more novel ideas such as the use of information entropy and Eigen-clustering and using neural networks for classification. Where the algorithm has been applied to a particular species it is mentioned and many of the algorithms can be adapted to different species.

PAM SIGNAL FLOW

Referring to Figure 1, signal flow in a PAM system can be described as follows. A hydrophone converts the acoustic energy into electrical energy which enters the system as an analogue signal. Systems which don't perform any detection or classification and which do not digitise the signal will transmit the analogue signal and processing will be done elsewhere.

Some approaches, discussed in the overview of detection and classification methods section, perform detection in the analogue domain. For example some methods use band-pass filters to monitor amplitude levels for threshold exceedences. These detections will be digitised and either transmitted back to base or stored for later retrieval.

Many systems operate in the digital domain. Some will digitally transmit back to base for later processing or store the raw digitised and unprocessed data. Those that perform detection and classification at the source will likely monitor spectral energy in specific frequency bands for detection in the same way as the analogue detection mentioned, i.e. watching for amplitude threshold exceedences in certain bands to indicate a signal of interest might be present.

Referring to Figure 2 there are broadly three stages a signal will go through when being examined for the presence of a vocalisation. The first is the initial detection, which as discussed will rely on energy content, monitored through filter banks or a digital method like Fast Fourier Transforms (FFTs). If the system seeks to classify the signal and associate it with a particular species then it will extract some features to perform an analysis. The feature extraction and classification stage is what differentiates the techniques. For some algorithms, the features extracted are simply the amount of energy in certain frequency bands (Gillespie & Chappell 2002; Ward et al. 2000) where others collect more spectral and temporal features, such as a slope in the frequency over the length of a call (Gavrilov et al. 2011; Harland & Armstrong 2004), the duration of the call and the inter-click-interval for echolocation clicks (Ma et al. 2010).

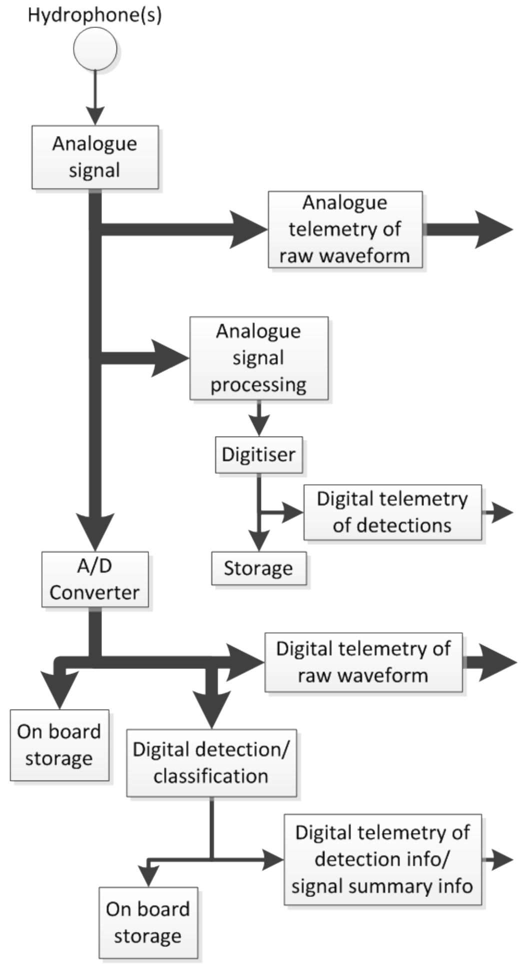

Figure 1. Overview of signal flow in a PAM system

Figure 2. The stages of detection and classification.

There are also a variety of classification methods in the algorithms discussed here. For example, a number use statistical methods to analyse the extracted features (Brown & Smaragdis 2009; Rickwood & Taylor 2008; Roch et al. 2007), others use neural networks (Deecke & Janik 2006; Dugan et al. 2010a, 2010b) and some use spectrogram correlation (Mellinger & Clark 1997, 2000; Mellinger, Stafford & Fox 2004).

OVERVIEW OF MARINE MAMMAL SOUNDS

The characteristics of marine mammal vocalisations have been well studied and documented over the past 50 years (Au & Hastings 2008; Backus & Schevill 1966; Gavrilov et al. 2011; Payne & McVay 1971; Rankin & Barlow 2005; Schevill 1964; Thompson & Friedl 1982; Watkins & Schevill 1977; Weilgart & Whitehead 1997; Weilgart & Whitehead

1988). The vocalisations can be broken down broadly into two categories, social sounds and echolocation sounds (Au & Hastings 2008). The sounds have many names, for example, moans, grunts, buzzes, clicks, pulses etc. but it is reasonable to say that the sounds are mostly "species-specific" and for the purposes of classification of species should be considered separately (Zimmer 2011), although this can be a risky assumption as some species calls sound very similar to other species, such as the humpback call (Baumgartner & Mussoline 2011). Acoustically, the sounds can be categorised as periodic or a-periodic and are a mixture of frequency and amplitude modulated signals (Zimmer 2011).

The sounds made by geographically isolated groups of individuals within the same species are known to vary (Weilgart & Whitehead 1997), though there is evidence of horizontal cultural transmission of songs over a vast geographic region (Garland et al. 2011). There is also evidence of some whales experiencing a subtle lowering of frequency over a number of years (Gavrilov et al. 2011). The humpback whale song is thought to only be produced by males travelling mostly in isolation, but whales within a population sing the same basic song, although it may undergo slight changes throughout the breeding season (Au & Hastings 2008).

There are many resources available to researchers studying and seeking to detect and classify the vocalisations, including annotated datasets (Mellinger 2010; Mellinger & Clark 2006) and a compiled repertoire specifically for automatic detection and classification (Erbe 2004). A frequency band of 7 Hz to 180 kHz is thought to contain all frequencies within the range of marine mammal vocalisations and hearing (Barker & Lepper 2012).

OVERVIEW OF DETECTION AND CLASSIFICATION METHODS

Detection and Feature Extraction

The majority of detection and classification algorithms use a method of feature or attribute extraction. These algorithms do not all extract the same features but they seek to generate a profile of the signal of interest and compare it with some known parameters. Most detection components focus on watching a range of frequency bands where the signals of interest reside and record a detection when a number of thresholds are exceeded.

Earlier classification algorithms simply used this detection technique, looking for increased energy levels in certain frequency bands tailored to the species of interest. Ward et al. (2000) detected sperm whales using an energy detector on 6 narrowband frequencies below 12 kHz. Wavelets were also used with a moving average in a way that was equivalent to a series of filter banks. Similarly Gillespie & Chappell (2002) applied band-pass filter banks to split signals into three bands, using the relative amplitude of the signals in these bands and the shape of the pulse for the detection of harbour porpoises. The simplicity of these approaches makes them susceptible to false positives and less robust in noisy environments, but they are well suited to narrowband signals like porpoise clicks and some clicks from other species.

Harland & Armstrong (2004) split incoming acoustic signals into five processing channels for five groups of calls from odontocetes (toothed whales) and mysticetes (baleen whales). For pulses, an FFT was applied and a static amplitude threshold was used for detection, which was confirmed with a test of the spectral slope. For tonal signals however, a combination of minimum and maximum frequency, start and stop frequency, bandwidth, duration and other parameters was collected after generating a spectrogram and searching for connected components. Johansson & White (2004) used a different approach, applying an adaptive notch filter with a dynamic detection threshold for tonal signals from right whales which worked well with simultaneous sounds in low signal-to-noise ratios.

The generalised perceptual linear prediction (gPLP) model was introduced to animal sound analysis by Clemins & Johnson (2006); Clemins et al. (2006), adapting a speech processing model. It generated features in the discrete cepstral domain and incorporated experimentally acquired perceptual information to tailor the feature extraction to the species of interest. The use of Mel-Frequency Cepstral Coefficients (MFCCs) and Greenwood Function Cepstral Coefficients (GFCCs) were also discussed, which are both traditionally used to extract features from human speech. It was not applied to marine mammals in this study but the research was used by Roch et al. (2007) who used it to classify odontocetes.

Oswald, Barlow & Norris (2003); Oswald et al. (2007) extracted features and applied discriminant function analysis, and non-parametric classification using regression tree analysis to classify whistles (from spinner, striped, pantropical spotted, long-beaked common, short-beaked common, rough-toothed and bottle nosed dolphins, as well as short-finned pilot and killer whales) with mixed accuracy. In Oswald, Barlow & Norris (2003), which only used discriminant function analysis, correct classification within species was significantly greater than expected by chance alone and ranged from 29.9% for striped dolphins to 91.2% for false killer whales. In Oswald et al. (2007) this increased to 34.2% for striped dolphins and decreased to 70% for false killer whales when regression tree analysis was added. This method of detection and classification was incorporated into the PAMGUARD software suite in 2011 (Oswald et al. 2011).

Linear discriminant analysis has been applied to the supervised classification of beaked whale clicks by Parnum et al. (2011), who used supervised classification to achieve a false alarm rate of 5% and a misdetection rate of 2%. Baumgartner & Mussoline (2011) used attribute extraction which focused on pitch-tracking and was combined with a quadratic discriminant function analysis to detect and classify sei whale and North Atlantic right whale calls. Their system was judged to be similar to that of a human analyst.

Gavrilov et al. (2011) used a signal recognition algorithm that searched for transient signals in sea noise which had known time-frequency features, similar to a documented call of the pygmy blue whale. The features that were used were the signal duration, the frequency band and the slope of frequency change with time. This same algorithm was again used by Gavrilov et al. (2012) to detect common song themes and was reported to have a misdetection and false detection rate of less than 5%.

Extraction of a mixture of temporal and spectral features was used by Zaugg et al. (2010) in the detection of sperm whale clicks, as part of a larger system including impulse detection to detect impulsive shipping noise. The system used band pass filters at 1-5 kHz and 5-20 kHz, established a dynamic threshold and detected impulses. The extracted features were fed into a feed forward neural network for classification. Their model was able to reliably automatically detect and classify sperm whale clicks and differentiate them from impulses due to ship noise. This method was combined with a short tonal detector and more band pass filters between 0 and 500 Hz and a high pass filter of 20 kHz in André et al. (2011). This also used spectral and temporal features to detect whistle-like sounds from dolphins and tonal calls from baleen whale species, and is likely able to be adapted to a variety of species. It was found to be a highly efficient ultrasonic click detector which reduced the computational load on the classifier stage.

Binder & Hines (2012) applied aural classification techniques, utilising a simple auditory model which used a 100-channel filter bank with gains scaled to represent the propagation of sound through the outer and middle ear. From the output of the filter bank, 46 time-frequency features and 12 purely spectral features were extracted. Discriminant analysis and principal component analysis were applied, with the goal of dimensionality reduction. The conclusion was that the aural classifier was a useful tool for classifying cetacean vocalisations, with discriminant analysis the preferred tool. It was tested successfully with five cetacean species (bowhead, humpback, North Atlantic right, minke and sperm whales) and likely could be adapted to a larger range.

Ou, Au & Oswald (2012) produced an automatic detector of minke whale ‘boing’ sounds which searches for frequency features but avoids calculating the continuous spectrogram to reduce computational time. These features are the peak frequency range, the separation between centre and side frequency bands and the duration of the sound. The technique worked well in low signal to noise ratio environments and can be adapted to any species of interest.

Zaugg et al. (2012) describe an algorithm which uses a ‘peakiness’ metric; a way to quantify the prominence of spectral peaks. The peakiness metric can be a measure of average energy via the median, the arithmetic mean or an entropy metric, which is a slightly altered measure of Shannon entropy (discussed in Information Entropy, Neural Networks, the Hilbert-Huang Transform and Eigen-Clustering). It was well suited to tonal sounds but not for echolocation clicks, and was judged particularly suitable for real-time operation.

Mellinger et al. (2011) used a spectral peak tracking technique which used frequency contour analysis to detect Risso’s dolphin whistles and minke whale boing sounds, despite the presence of interfering humpback whale songs. This technique can be adapted to a wide range of sounds and species. It also potentially has applications in the identification of individuals due to the way it tracks frequency over time.

Energy Based Algorithms

Some detection and classification algorithms focus on energy content alone but are more advanced than the early feature detection algorithms by Ward et al. (2000) and Gillespie & Chappell (2002) which also focused on spectral energy content. These energy based algorithms can be susceptible to low signal to noise ratios though, as some dynamic thresholds are based on noise levels. A method of dealing with a noisy environment is the matched filter, which is optimal when the signal is known. It was applied to bowhead whales and compared with spectrogram correlation, a traditional energy based method of analysis by Mellinger & Clark (1997). It is best suited for environments where the noise has a flat spectrum or close to.

Spectrogram correlation involves a kernel which is constructed and cross-correlated with a spectrogram of a recording, and can work fairly well in low noise environments. This produces a recognition function representing whether the sound of interest was recorded at a certain time (Mellinger & Clark 1997, 2000; Mellinger, Stafford & Fox 2004). Mellinger & Clark (2000) found it to be quite successful for detecting bowhead whale calls with an error of only 0.9%. It was recommended for detecting a call type when relatively few instances of the call type are known.

Another simple method of feature extraction is concerned only with frequency analysis and energy content. Morrissey et al. (2006) employed a click detector that purely used frequency domain energy with a time varying threshold that is associated with the noise level to detect sperm whale clicks. A 512 point FFT was calculated and a binary frequency map was generated. When the number of bins in the map that registered a threshold exceedence passed another threshold of approximately 10 bins, a click was detected. It was successfully used as part of a tracking system in real time.

Methods for detecting porpoise clicks using a band-limited energy sum technique were outlined by Gillespie & Chappell (2002) who demonstrated a sensor that performed real-time detection. Klinck & Mellinger (2011) improved on the band-limited energy sum technique by developing the Energy Ratio Mapping Algorithm (ERMA) which was tested by detecting Blainville’s beaked whale clicks while rejecting the echolocation clicks of Risso’s dolphins and pilot whales. The technique is suited for applications in low-power computing environments but suffered enough false positives that they recommended it to be the first step of a two-step detector.

Kandia & Stylianou (2006) introduced the use of the Teager-Kaiser energy operator to detect sperm whale creaks as well as regular clicks. This was compared with a rainbow click detector which uses two steps. The first stage passes a rectified input through a first order low-pass filter, then filters the signals that register as clicks with a user-defined band pass filter. If a click has energy higher than some threshold it is registered as a click. It was also used by (Ma et al. 2010) for localisation of marine mammal clicks from an unspecified species. The Teager-Kaiser energy operator is fast and efficient, only requiring 3 samples to calculate, making it near instantaneous and ideal for real-time systems.

HMMs, GMMs and Other Statistical Methods

Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs) are statistical models which have also been applied to marine mammal acoustic detection and classification algorithms. These models often use cepstral features or the MFCCs. MFCCs are a representation of the power spectrum over a short term and are often used in conjunction with HMMs and or GMMs. The Mel-frequency cepstrum differs from the regular cepstrum in the spacing of its frequency bands.

The first use of GMMs for classification of marine mammal call types was Roch et al. (2007) who extracted cepstral feature vectors from call data to train GMMs of varying orders for a selection of species (short-beaked, long-beaked common, Pacific white-sided and bottlenose dolphins). The classifier predicted the species of groups with 67%-75% accuracy. This technique is well suited to detecting clicks and was used again in Roch et al. (2008) with positive results.

Rickwood & Taylor (2008) demonstrated an energy based technique which involved the extraction of feature vectors using Hidden Markov Modelling (HMM) and used an information-theoretic approach with Minimum Message Length (MML) encoding. This allowed automatic detection and unsupervised classification of humpback whales which can be adapted to other marine mammals. It also successfully clustered similar song units together when there were significant variations in levels and interference, but was judged to require further work incorporating feedback from human operators.

Brown & Smaragdis (2009) applied both HMMs and GMMs to classify marine mammal call types and concluded that they were both highly successful for automatic classification of killer whale call types, but gave special mention regarding the performance of HMMs. This technique was analysed again, on killer whales, in Brown, Smaragdis & Nousek-McGregor (2010) which came to similarly positive conclusions.

The performances of GMMs and other detection methods, including the ERMA were analysed by Yack et al. (2010) who compared a total of six beaked whale detection algorithms. It was found that all had detection rates above 60%. The conclusion, however, was that the choice of algorithm is ultimately dependant on the application (i.e. real-time or post-processing) and they suggested a qualitative and quantitative analysis should be conducted before choosing an algorithm. It also showed that the GMM had the best correct detection rate for beaked whales. Work on GMMs continues, with Roch et al. (2011), representing echolocation clicks of six species (bottlenose, short-beaked, long-beaked common, Pacific white-sided and Risso's dolphins as well as Cuvier's beaked whales) as cepstral feature vectors that are classified by GMMs. This technique was suggested as the second stage to complement the method of Klinck & Mellinger (2011).

Pace, White & Adam (2012) used HMMs and MFCCs to classify humpback whale calls, based on the concepts of subunits as building blocks. It was found that the HMM classification method potentially had a high level of performance with only a modest requirement in computational load and storage. Samaran et al. (2012) also used HMMs and MFCCs with success and suggest that in the future it should be possible to assign acoustic signatures to specific humpback whale individuals.

Using a statistical approach, Gervaise et al. (2010) applied kurtosis estimation to create a general click detection algorithm. The algorithm works on "the assumption that click trains are embedded in stochastic but Gaussian noise" and kurtosis can be applied as a "statistical test for detection". Their algorithm adapted to a varying click centre frequency. After testing on datasets containing Cuvier's beaked whale and beluga whale calls, they concluded that their method appeared to be promising for detecting click trains and isolating individual clicks, either alone or combined with additional click detectors. It also performed well with a weak signal to noise ratio, where energy detectors are less appropriate.

Taking a different approach and applying a statistical decision theory to the binary hypothesis of 'presence' or 'absence' of a call, Urazghildiiev & Clark (2006) applied a Generalised Likelihood Ratio Test (GLRT) detector to the calls of North Atlantic right whales. This was done by separating the data into chunks of 8-16s and calculating the FFT, the power spectral density and then applying a median filter. Then the inverse FFT of the product of the normalised data spectrum and the frequency response of the median filter is taken. This was compared for several 8-16s samples to generate the GLRT statistic, but GLRT showed poor performance when the transient noise rate was high. However, GLRT was improved on in Urazghildiiev & Clark (2007) and Urazghildiiev et al. (2009) by using a multistage decision-making process involving spectrogram and feature vector testing algorithms. It was made more resistant to transient noise and less computationally demanding.

Information Entropy, Neural Networks, the Hilbert-Huang Transform and Eigen-Clustering

The use of information entropy in the detection of marine mammal vocalisations is relatively new, beginning with Erbe & King (2008). A detection method using information (or Shannon) entropy was demonstrated which detected calls from a variety of marine mammals and performed considerably faster than real time. Bougher et al. (2012) used the same information entropy approach as part of their improved band-limited processing approach to detect minke whale boings. This type of detector is ideal for tonal signals or signals where the energy content is contained in a limited number of frequency bands so it is suitable for a range of call types from a range of species.

Artificial neural networks were employed by Dugan et al. (2010a, (2010b) as a component of the North Atlantic right whale CRITIC system. The CRITIC system uses several recognition methods running in parallel. These methods are the artificial neural network, a linear discriminant analysis component and a classification regression tree. In testing, it was found that although a comparison feature vector testing approach had a very low rate of false positives, the combination system had higher assignment rates.

Earlier work on artificial neural networks was done by Deecke & Janik (2006). They demonstrated the use of an adaptive resonance theory neural network to identify and automatically categorise bottlenose dolphin and killer whale bioacoustic signals which can also be adapted to other species. Bougher et al. (2012) used their band-limited processing technique to accomplish the same task on the same dataset with positive results.

There has also been studies done on the use of the Hilbert-Huang Transform (HHT) on the analysis of vocalisations by sperm and killer whales (Adam 2006a; Adam 2006b). It was found that the HHT was a viable alternative to the wavelet transform and in the case of the sperm whales it was found that only the first six modes were sufficient for regular click decomposition. It is well suited to transient, impulsive signals such as clicks. However, there does not appear to have been any more recent work on this particular analysis technique applied to marine mammal vocalisations.

Finally, there has been some interesting work done by Tao (2009) on a new technique called Eigen-clustering which may have some use in bioacoustics recognition. There does not appear to have been any study on the application of this method to marine mammal vocalisations and it appears to be an area of potential investigation.

CONCLUSION

There are a large variety of detection and classification algorithms to select from. They differ in many ways, such as performance, computational requirements, ability to cope with noise, ability to deal with multiple species and ability to be performed in real-time. This list represents the majority of techniques and sources that exist today.

Clearly the selection process in creating a robust smart, real-time PAM system with automatic detection and classification involves rigorous evaluation, and a broad knowledge of the many techniques that exist. The author's next task is to perform such an evaluation, focusing on the techniques which are most appropriate for application on low-power embedded computer systems.

REFERENCES

Adam, O. 2006a, 'Advantages of the Hilbert Huang transform for marine mammals signals analysis', *The Journal of the Acoustical Society of America*, vol. 120, p. 2965.

Adam, O. 2006b, 'The use of the Hilbert-Huang transform to analyze transient signals emitted by sperm whales', *Applied Acoustics*, vol. 67, no. 11, pp. 1134-43.

André, M., Van Der Schaar, M., Zaugg, S., Houégnigan, L., Sánchez, A. & Castell, J. 2011, 'Listening to the Deep: Live monitoring of ocean noise and cetacean acoustic signals', *Marine pollution bulletin*, vol. 63, no. 1, pp. 18-26.

Au, W.W.L. & Hastings, M.C. 2008, *Principles of marine bioacoustics*, Springer.

Backus, R. & Schevill, W. 1966, 'Physeter clicks', *Whales, dolphins, and porpoises*, pp. 510-27.

Barker, P. & Lepper, P. 2012, 'Development of a versatile platform for long-term underwater acoustic monitoring', *11th European Conference on Underwater Acoustics*, Edinburgh, UK, pp. 627-33.

Baumgartner, M.F. & Mussoline, S.E. 2011, 'A generalized baleen whale call detection and classification system', *The Journal of the Acoustical Society of America*, vol. 129, p. 2889.

Binder, C.M. & Hines, P.C. 2012, 'Applying automatic aural classification to cetacean vocalizations', *11th European Conference on Underwater Acoustics*, Edinburgh, UK, pp. 323-30.

Bougher, B.B., Hood, J.D., Theriault, J.A. & Moors, H. 2012, 'Generalized marine mammal detection based on improved band-limited processing', *11th European Conference on Underwater Acoustics*, Edinburgh, UK, pp. 835-42.

Brown, J.C. & Smaragdis, P. 2009, 'Hidden Markov and Gaussian mixture models for automatic call classification'.

Brown, J.C., Smaragdis, P. & Nousek-McGregor, A. 2010, 'Automatic identification of individual killer whales', *The Journal of the Acoustical Society of America*, vol. 128, no. 3, pp. EL93-EL8.

C-POD Species Detection 2013, viewed 24th August 2013, <http://www.chelonia.co.uk/species_detection.htm>.

Cato, D.H., Noad, M.J. & McCauley, R.D. 2005, 'Passive acoustics as a key to the study of marine animals', *Sounds in the Sea*, Cambridge University Press, NY, USA, pp. 411-29.

Clemins, P.J. & Johnson, M.T. 2006, 'Generalized perceptual linear prediction features for animal vocalization analysis', *The Journal of the Acoustical Society of America*, vol. 120, p. 527.

Clemins, P.J., Trawicki, M.B., Adi, K., Tao, J. & Johnson, M.T. 2006, 'Generalized perceptual features for vocalization analysis across multiple species', vol. 1, IEEE, pp. I-I.

Deecke, V.B. & Janik, V.M. 2006, 'Automated categorization of bioacoustic signals: Avoiding perceptual pitfalls', *The Journal of the Acoustical Society of America*, vol. 119, no. 1, pp. 645-53.

Dugan, P.J., Rice, A.N., Urazghildiiev, I.R. & Clark, C.W. 2010a, 'North Atlantic right whale acoustic signal processing: Part I. Comparison of machine learning recognition algorithms', *Applications and Technology Conference (LISAT), 2010 Long Island Systems*, IEEE, pp. 1-6.

Dugan, P.J., Rice, A.N., Urazghildiiev, I.R. & Clark, C.W. 2010b, 'North Atlantic right whale acoustic signal processing: Part II. Improved decision architecture for auto-detection using multi-classifier combination methodology', *Applications and Technology Conference (LISAT), 2010 Long Island Systems*, IEEE, pp. 1-6.

Erbe, C. 2004, *The Acoustic Repertoire of Odontecetes as a Basis for Developing Automatic Detectors and Classifiers*, DTIC Document.

Erbe, C. & King, A.R. 2008, 'Automatic detection of marine mammals using information entropy', *The Journal of the Acoustical Society of America*, vol. 124, no. 5, pp. 2833-40.

Garland, E.C., Goldizen, A.W., Rekdahl, M.L., Constantine, R., Garrigue, C., Hauser, N.D., Poole, M.M., Robbins, J. & Noad, M.J. 2011, 'Dynamic horizontal cultural transmission of humpback whale song at the ocean basin scale', *Current biology*, vol. 21, no. 8, pp. 687-91.

Gavrilov, A.N., McCauley, R.D., Pattiaratchi, C. & Bondarenko, O. 2012, 'The use of passive acoustics to observe the presence and movement of pygmy blue whales (Balaenoptera musculus brevicauda) in the Perth canyon, WA', *11th European Conference on Underwater Acoustics*, Edinburgh, UK, pp. 1802-9.

Gavrilov, A.N., McCauley, R.D., Salgado-Kent, C., Tripovich, J. & Burton, C. 2011, 'Vocal characteristics of pygmy blue whales and their change over time', *The Journal of the Acoustical Society of America*, vol. 130, no. 6, pp. 3651-60.

Gervaise, C., Barazzutti, A., Busson, S., Simard, Y. & Roy, N. 2010, 'Automatic detection of bioacoustics impulses based on kurtosis under weak signal to noise ratio', *Applied Acoustics*, vol. 71, no. 11, pp. 1020-6.

Gillespie, D. 2013, *Real-time monitoring for beluga whales (Delphinapterus leucas) in the Eagle River, Alaska using a PAMBuoy® detection system*, SMRU, LLC and Marine Instruments Ltd.

Gillespie, D. & Chappell, O. 2002, 'An automatic system for detecting and classifying the vocalisations of harbour porpoises', *Bioacoustics*, vol. 13, no. 1, pp. 37-61.

Harland, E. & Armstrong, M. 2004, 'The real-time detection of the calls of cetacean species', *Canadian Acoustics*, vol. 32, no. 2, pp. 76-82.

Johansson, A. & White, P. 2004, 'Detection and characterization of marine mammal calls by parametric modelling', *Canadian Acoustics*, vol. 32, no. 2, pp. 83-92.

Kandia, V. & Stylianou, Y. 2006, 'Detection of sperm whale clicks based on the Teager–Kaiser energy operator', *Applied Acoustics*, vol. 67, no. 11, pp. 1144-63.

Klinck, H. & Mellinger, D.K. 2011, 'The energy ratio mapping algorithm: A tool to improve the energy-based detection of odontocete echolocation clicks', *The Journal of the Acoustical Society of America*, vol. 129, p. 1807.

Ma, T., Collison, N., Theriault, J., Hood, J., Pecknold, S. & Bougher, B. 2010, 'Detection of Precise Time Events for Marine Mammal Clicks', *Canadian Acoustics*, vol. 38, no. 3, pp. 30-1.

Marine Instrumentation Ltd. 2012, *PAMBuoy*, viewed 9th November 2012, <http://www.pambuoy.co.uk/>.

Medwin, H. & Blue, J.E. 2005, 'Sounds in the sea: From ocean acoustics to acoustical oceanography'Cambridge University Press.

Mellinger, D.K. 2010, *Datasets of odontocete sounds annotated for developing automatic detection methods*, DTIC Document.

Mellinger, D.K. & Clark, C.W. 1997, 'Methods for automatic detection of mysticete sounds', *Marine & Freshwater Behaviour & Phy*, vol. 29, no. 1-4, pp. 163-81.

Mellinger, D.K. & Clark, C.W. 2000, 'Recognizing transient low-frequency whale sounds by spectrogram correlation', *The Journal of the Acoustical Society of America*, vol. 107, p. 3518.

Mellinger, D.K. & Clark, C.W. 2006, 'MobySound: A reference archive for studying automatic recognition of marine mammal sounds', *Applied Acoustics*, vol. 67, no. 11, pp. 1226-42.

Mellinger, D.K., Martin, S.W., Morrissey, R.P., Thomas, L. & Yosco, J.J. 2011, 'A method for detecting whistles, moans, and other frequency contour sounds', *The Journal of the Acoustical Society of America*, vol. 129, p. 4055.

Mellinger, D.K., Stafford, K.M. & Fox, C.G. 2004, 'Seasonal occurrence of sperm whale (Physeter macrocephalus) sounds in the Gulf of Alaska, 1999–2001', *Marine Mammal Science*, vol. 20, no. 1, pp. 48-62.

Morrissey, R., Ward, J., DiMarzio, N., Jarvis, S. & Moretti, D. 2006, 'Passive acoustic detection and localization of sperm whales (Physeter macrocephalus) in the tongue of the ocean', *Applied Acoustics*, vol. 67, no. 11, pp. 1091-105.

Oswald, J.N., Barlow, J. & Norris, T.F. 2003, 'Acoustic identification of nine delphinid species in the eastern tropical Pacific Ocean', *Marine Mammal Science*, vol. 19, no. 1, pp. 20-037.

Oswald, J.N., Rankin, S., Barlow, J. & Lammers, M.O. 2007, 'A tool for real-time acoustic species identification of delphinid whistles', *The Journal of the Acoustical Society of America*, vol. 122, p. 587.

Oswald, M., Oswald, J.N., Lammers, M.O., Rankin, S. & Au, W.W.L. 2011, 'Integration of real-time odontocete call classification algorithm into PAMGUARD signal processing software', *The Journal of the Acoustical Society of America*, vol. 129, no. 4, pp. 2639-.

Ou, H., Au, W.W.L. & Oswald, J.N. 2012, 'A non-spectrogram-correlation method of automatically detecting minke whale boings'.

Pace, F., White, P.R. & Adam, O. 2012, 'Hidden Markov modelling for humpback whale (megaptera novaeangliae) call classification', *11th European Conference on Underwater Acoustics*, Edinburgh, UK, pp. 874-81.

Parnum, I., McCauley, R., Cato, D., Thomas, F., Duncan, A. & Johnson, M. 2011, 'Detection of beaked whale clicks in underwater noise recordings', *Proceedings, Acoustics 2011, Surfers Paradise, Queensland, 3-4 Nov*.

Payne, R.S. & McVay, S. 1971, 'Songs of Humpback Whales'.

Philpott, E., Englund, A., Ingram, S. & Rogan, E. 2007, 'Using T-PODs to investigate the echolocation of coastal bottlenose dolphins', *Journal of the Marine Biological Association of the United Kingdom*, vol. 87, no. 1, p. 11.

Rankin, S. & Barlow, J. 2005, 'Source of the North Pacific "boing" sound attributed to minke whales', *The Journal of the Acoustical Society of America*, vol. 118, p. 3346.

Rickwood, P. & Taylor, A. 2008, 'Methods for automatically analyzing humpback song units', *The Journal of the Acoustical Society of America*, vol. 123, no. 3, pp. 1763-72.

Roch, M.A., Klinck, H., Baumann-Pickering, S., Mellinger, D.K., Qui, S., Soldevilla, M.S. & Hildebrand, J.A. 2011, 'Classification of echolocation clicks from odontocetes in the Southern California Bight', *The Journal of the Acoustical Society of America*, vol. 129, no. 1, pp. 467-75.

Roch, M.A., Soldevilla, M.S., Burtenshaw, J.C., Henderson, E.E. & Hildebrand, J.A. 2007, 'Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California', *The Journal of the Acoustical Society of America*, vol. 121, p. 1737.

Roch, M.A., Soldevilla, M.S., Hoenigman, R., Wiggins, S.M. & Hildebrand, J.A. 2008, 'Comparison of machine learning techniques for the classification of echolocation clicks from three species of odontocetes'.

Samaran, F., Gandilhon, N., Gonzalez, R.P., Pace, F., Kennedy, A. & Adam, O. 2012, 'Passive hydro-acoustics for cetacean census and localisation', *Sensors for ecology*, pp. 63-82.

Schevill, W.E. 1964, 'Underwater sounds of cetaceans', *Marine bio-acoustics*, vol. 1, pp. 307-16.

Tao, J. 2009, 'Acoustic model adaptation for automatic speech recognition and animal vocalization classification', Faculty of the Graduate School, Marquette University.

Thompson, P.M., Lusseau, D., Barton, T., Simmons, D., Rusin, J. & Bailey, H. 2010, 'Assessing the responses of coastal cetaceans to the construction of offshore wind turbines', *Marine pollution bulletin*, vol. 60, no. 8, pp. 1200-8.

Thompson, P.O. & Friedl, W.A. 1982, 'A long term study of low frequency sound from several species of whales off Oahu, Hawaii', *Cetology*, vol. 45, pp. 1-19.

Urazghildiiev, I.R. & Clark, C.W. 2006, 'Acoustic detection of North Atlantic right whale contact calls using the generalized likelihood ratio test', *The Journal of the Acoustical Society of America*, vol. 120, p. 1956.

Urazghildiiev, I.R. & Clark, C.W. 2007, 'Acoustic detection of North Atlantic right whale contact calls using spectrogram-based statistics', *The Journal of the Acoustical Society of America*, vol. 122, p. 769.

Urazghildiiev, I.R., Clark, C.W., Krein, T.P. & Parks, S.E. 2009, 'Detection and recognition of North Atlantic right whale contact calls in the presence of ambient noise', *Oceanic Engineering, IEEE Journal of*, vol. 34, no. 3, pp. 358-68.

Verfuß, U.K., Miller, L.A. & Schnitzler, H.-U. 2005, 'Spatial orientation in echolocating harbour porpoises (Phocoena phocoena)', *Journal of Experimental Biology*, vol. 208, no. 17, pp. 3385-94.

Ward, J., Fitzpatrick, M., DiMarzio, N., Moretti, D. & Morrissey, R. 2000, 'New algorithms for open ocean marine mammal monitoring', *OCEANS 2000 MTS/IEEE Conference and Exhibition*, vol. 3, IEEE, pp. 1749-52.

Watkins, W.A. & Schevill, W.E. 1977, 'Sperm whale codas', *The Journal of the Acoustical Society of America*, vol. 62, p. 1485.

Weilgart, L. & Whitehead, H. 1997, 'Group-specific dialects and geographical variation in coda repertoire in South Pacific sperm whales', *Behavioral Ecology and Sociobiology*, vol. 40, no. 5, pp. 277-85.

Weilgart, L.S. & Whitehead, H. 1988, 'Distinctive vocalizations from mature male sperm whales (Physeter macrocephalus)', *Canadian Journal of Zoology*, vol. 66, no. 9, pp. 1931-7.

Yack, T.M., Barlow, J., Roch, M.A., Klinck, H., Martin, S., Mellinger, D.K. & Gillespie, D. 2010, 'Comparison of beaked whale detection algorithms', *Applied Acoustics*, vol. 71, no. 11, pp. 1043-9.

Zaugg, S., Schaar, M.v.d., Houégnigan, L. & André, M. 2012, 'A framework for the automated real-time detection of short tonal sounds from ocean observatories', *Applied Acoustics*, vol. 73, no. 3, pp. 281-90.

Zaugg, S., Van Der Schaar, M., Houégnigan, L., Gervaise, C. & André, M. 2010, 'Real-time acoustic classification of sperm whale clicks and shipping impulses from deep-sea observatories', *Applied Acoustics*, vol. 71, no. 11, pp. 1011-9.

Zimmer, W. 2011, *Passive Acoustic Monitoring of Cetaceans*, Cambridge Univ Pr.