A NEURAL NETWORK FOR CLASSIFYING CLICKS OF BLAINVILLE'S BEAKED WHALES (*MESOPLODON DENSIROSTRIS*)

David K. Mellinger

Cooperative Institute for Marine Resources Studies (CIMRS), Oregon State University
and NOAA Pacific Marine Environmental Laboratory
2030 SE Marine Science Dr., Newport, OR 97365 USA
David.Mellinger@oregonstate.edu

ABSTRACT

Beaked whales are difficult to detect visually, and researchers have thus proposed using acoustic detection and classification. Because of the large data volumes often involved in acoustic detection and classification, automatic methods are often used. Here a neural network classification method is investigated. Using backpropagation, a feedforward neural network with one hidden layer was trained to classify clicks of Blainville's beaked whales and other odontocetes recorded in the Bahamas. Training and testing data consisted of approximately 1600 Blainville's beaked whale clicks and 3100 clicks from other odontocetes. Networks with 2-10 hidden units were trained and tested, with performance curves (ROC curves) calculated at several levels of signal-to-noise ratio. Results for most networks were quite good when compared with previous classification efforts, with less than 3% errors in both the wrong-classification and missed-call categories. Future work includes testing the network on sounds recorded in different noise backgrounds and from other populations of Blainville's beaked whales, and combining it with a detector and evaluating the joint performance.

SOMMAIRE

Mésoplodons sont difficiles à voir et chercheurs ont proposé d'employer la détection et la classification acoustique pour en trouver. Face à la quantité de données produites par détection et classification acoustiques, méthodes automatisées sont souvent utilisées. Ici on present une methode de réseau neuronal pour classifier. Un réseau neuronal à rétropropagation non récurrent avec une seule couche cachée a été formé pour classifier des clics des Mésoplodon de Blainville et autres odontocètes enregistrés aux Bahamas. Les données de formation se sont composées d'environs 1600 clics de Mésoplodon de Blainville et 3100 clics d'autres odonotocètes. Reseaux avec 2-10 unités cachées ont été formés et examinés par courbes caractéristiques d'opération du récepteur (ROC curves) calculés à plusieurs niveaux du ratio signal/bruit. Résultats pour la plupart des réseaux étaient tout à fait bons en comparaison avec des efforts précédents de classification avec moins de 3% d'erreurs chez les clics incorrectement classifiés ou manqués. Travaux à suivre sont essais du réseau avec les enregistrements venant d'autres niveaus deu bruit de fond et d'autres populations de Mésoplodon de Blainville, et en combinaison avec un detecteur, une evaluation d'exécution commune.

1. INTRODUCTION

Beaked and bottlenose whales – members of the family Ziphiidae, including the genera Ziphius, Mesoplodon, Berardius, Hyperoodon, and others – are among the most cryptic and least known of all mammal species. They inhabit deep-water regions (MacLeod and Zuur 2005), which are mostly distant from land and thus relatively difficult to study. They spend much of their time submerged, making it difficult to see them (Barlow et al. 2006). Indeed, visual line-transect studies have found narrower effective strip widths and lower detection probabilities on the trackline for beaked whales than for most other cetaceans (Barlow et al. 2006).

Despite being difficult to see, beaked whales have had notable interactions with humans: they have stranded in places and at times associated with anthropogenic sound use [Frantzis 1998; NMFS 2001; Fernández 2005; Aguilar de Soto 2006], and have attracted intense interest from management agencies, conservation organizations, and the public. A basic first step in preventing harm to beaked whales is to detect when they are present in an area of concern. Because of the difficulty of detecting beaked whales visually, acoustic detection and classification methods have been suggested as a tool for aiding in mitigation of the effects of human activities. Beaked whales are known to produce both echolocation clicks (Johnson et al. 2004) and whistles (Dawson et al. 1998). However, whistles appear to be relatively rare among all species of beaked whales, making clicks a potentially more useful type of sound for an acoustic detection and classification system.

A wide variety of methods have been used for detection and classification of cetacean sounds. A method that has worked well for a number of species is a neural network (Ghosh et al. 1992; Potter et al. 1994; Kundu and Chen 1997; Murray et al. 1998; Deecke et al. 1999; Houser et al. 1999; Mellinger 2004). Neural networks combine a design phase, in which the structure of the network is chosen, a training phase, in which the parameters of the network's units are configured, and a testing and use phase, in which the network is operated with the parameters fixed. Here a method is presented for acoustic classification of clicks of Blainville's beaked whale (*Mesoplodon densirostris*) and other odontocetes using a neural network. Although a simple automated detector is used for finding clicks of other odontocetes to use in training and testing, the focus here is on classification of the clicks.

2. METHODS

2.1 Classification method

An input sound signal is used to compute a spectrogram, to which conditioning steps – spectrum level equalization, rectification, and normalization – are then applied. The conditioned spectrogram is then used as input to the neural network, resulting in a classification value indicating the certainty that a Blainville's beaked whale click is present.

In more detail, the spectrogram is calculated using a frame length of 0.000667 s (64 samples at a sample rate of 96 kHz), overlap of 50%, and a Hann window. This short frame size was chosen because of the brief nature of beaked whale clicks (Johnson and Tyack 2005), and indeed other known odontocete clicks (e.g., Au 1993). At this frame length, time resolution is relatively good, while the spectrogram filter bandwidth is a relatively poor 6.0 kHz. Nevertheless, the upsweeping nature of these clicks can still be resolved in these spectrograms (Fig. 1). The logarithm of each spectrogram cell is used, compressing spectrogram values to a range typically in the range of ±10.

After calculation of the spectrogram, the next step is spectrogram level equalization, rectification, and normalization. This is similar to the method described by Mellinger (2004), and will be explained only briefly here. Equalization is performed by subtracting the time-averaged spectrum (Van Trees 1968) from each spectrogram frame; that is, the spectrum for each spectrogram frame is multiplied by a small positive constant α near zero and added to the product of the long-term average spectrum and 1-α. Rectification consists of hard-limiting the minimum value in the spectrogram with a (constant) floor value to remove small and negative values. Normalization consists of subtracting the floor value from each spectrogram cell, so that the minimum spectrogram value becomes zero. In other words, the time-averaged spectrogram value is calculated for each frequency band of the spectrogram; this is subtracted from the spectrogram at each time step, a floor value is applied, and the floor constant is subtracted so that the minimum value in the resulting spectrogram is 0. The time constant used for equalization here was 0.02 s, while the floor value of the (logarithmic) spectrogram was 0.3 (this is equivalent to $e^{0.3}\approx1.35$ as a raw FFT value).

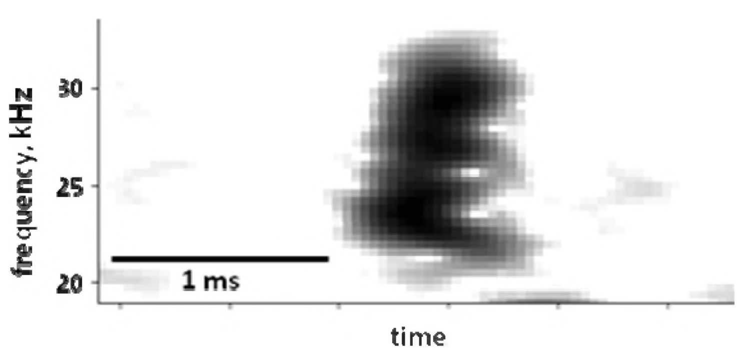

Fig. 1. An example click of a Blainville's beaked whale showing the upsweeping nature of these clicks. Spectrogram parameters: frame size 0.000667 s (64 samples), FFT size 128 samples, hop size 1/16 frame, Hann window, for a filter bandwidth of 6.0 kHz.

A neural network (Hagan et al. 1996) was designed with 192 input elements, a variable number of hidden units, and 1 output unit. The network was strictly feedforward, i.e., without any backward loops. Each hidden unit consisted of a weighted sum with bias followed by an arc-tangent nonlinearity. The output unit was linear, consisting of just a weighted sum. The number of hidden units was varied between 2 and 10 to estimate what the optimal number would be. The network was trained using the data set described below; batch training in each epoch was used to remove any bias in order of presentation. The training method was backpropagation (Rumelhart et al. 1987), so that network weights were adjusted according to a back-propagated error function, and a momentum term was used to prevent the network from getting 'stuck' in local maxima.

2.2 Data set

The data set consisted of recordings made at the Atlantic Undersea Test and Evaluation Center (AUTEC) in the Bahamas that contained clicks of Blainville's beaked whales. The whales were visually identified in the field by trained observers; the visual sightings coincided with the acoustically localized positions of the clicks (Moretti et al. 2006). Recordings were made at a sample rate of 96 kHz.

The recordings were manually scanned to detect clicks of Blainville's beaked whales. Manual scanning was used to remove the possibility of bias in detection of clicks; automated methods were not used to detect sounds for use in training and testing, as the methods themselves may introduce bias. The recordings were annotated to indicate the time and frequency bounds of each Blainville's beaked whale click. A total of 1595 Blainville's beaked whale clicks were found, and are henceforth called the BBW clicks.

The AUTEC recordings were also scanned to find clicks, presumably echolocation clicks, of other odontocetes. Known species on these recordings included Risso's dolphins (*Grampus griseus*) and long-finned pilot whales (*Globicephala macrorhynchus*). This scanning was done automatically, using a simple detector that found energy in the 20-38 kHz range of the Blainville's beaked whale clicks (Moretti et al. 2006): a ratio of the long-term to short-term averages was calculated, and when this ratio exceeded a threshold, a click was registered. These clicks were annotated similarly to the beaked whale clicks, with a total of 3096 clicks found. These clicks were named the 'other' clicks.

2.3 Training and testing

Conditioned spectrograms of the annotated clicks, both BBW and 'other', were calculated and used for training and testing the neural network. Only a portion of the spectrogram was used, namely the portion from 15 kHz to 38 kHz, as this frequency band contained most of the energy of beaked whale clicks present in these recordings (Moretti et al. 2006). Also, it was important to exclude frequencies below 14 kHz, as some of the recordings were filtered with a high-pass cutoff at this frequency. Using the entire bandwidth of such recordings would provide an unrealistic cue to the neural network for distinguishing BBW and other clicks. This frequency range contains 16 bands of the spectrogram.

For each click, a conditioned spectrogram centered on the click and lasting 0.004 seconds was used. For BBW clicks, the center was defined as the midpoint of the annotated sound; for 'other' clicks, the center was defined as the peak of the summed energy in the 20-38 kHz range. The 0.004-second spectrogram comprised 12 spectrogram frames, for a total of $12 \times 16 = 192$ cells. It was these cells for each click, arranged into a 192-element vector, that were used as input to the neural network.

The click data set was randomly divided into training and testing data. The BBW data were divided such that 9/10 of the clicks were used for training, with the remaining 1/10 used for testing; the 'other' clicks were divided similarly. The network was trained using the two datasets, with target outputs of +0.5 and -0.5 for the BBW and 'other' clicks, respectively.

2.4 Performance evaluation

Performance was measured using the one-tenth of the BBW and 'other' clicks reserved for testing. Testing was done by calculating the output of the network for the two sets of test data – typical output values were between -1 and 1, though other values occurred too – and applying a set of thresholds. For each threshold, the fraction of wrong classifications (false positives) and missed clicks (false negatives) was calculated; as the threshold was increased, there were fewer false classifications but more missed clicks. Varying the threshold and calculating the fractions of wrong classifications and missed clicks for each threshold yielded a parametric curve, the Receiver Operating Characteristic curve (Fawcett 2006). Training and testing was done five times, and the ROC curve calculated five times, for each number of hidden units in the network, and the five ROCs were averaged to produce the final results.

The signal-to-noise ratio (SNR) of any sound, including a click, is a key parameter in evaluating performance. Nearly all methods work well when the SNR is high, while only some work well at low SNR. Thus it is important to distinguish differing levels of SNR in describing performance of a classification method. Here SNR is measured by calculating the energy ratio of the signal in the 20-38 kHz band in a time period ±0.01 s around each click; that is, the average band-limited energy of the click is measured and is divided by the average band-limited energy of the background noise in this time period. Separate ROC curves were calculated in 5-dB increments of SNR level, i.e., SNRs of less than 10 dB, 10-15 dB, 15-20 dB, and more than 20 dB.

3. RESULTS

ROC curves for the neural network are shown in Fig. 2. Because of the large range of values, the curve was plotted on a logarithmic scale. The left plot shows the ROC curves for varying numbers of hidden units, with the number of units indicated next to each curve. The right plot shows performance for various SNRs for the best network.

A single-point measure of performance was assessed as well: at the 1% wrong-classification rate, a total of 0.6% of all BBW clicks are missed.

4. DISCUSSION

Some of the better ROC curves for the neural network are entirely less than the 3% error bounds in both dimensions, wrong classifications and missed clicks. This performance is very good compared with previous detection methods, including neural networks, that were applied to baleen whale vocalizations (Mellinger 2004, Mellinger et al. 2004). Part of the reason for the good performance is that the training and testing data were drawn from the same recordings of presumably the same whales, so the signals to be detected were probably very similar between the training and testing data sets. However, this was also true for the data sets in the baleen whale detection studies. Another reason may be that the clicks studied here are more stereotyped than the moans of baleen whales, so that a network trained to detect clicks in the training data works well on other, adjacent clicks in the testing data. In addition, Blainville's beaked whale clicks do not travel very far (Moretti et al. 2006), and so must have been produced closer to the hydrophone than the baleen whale vocalizations. They would therefore have been affected less by variability in ocean acoustic propagation. However, successive baleen whale vocalizations – some used for training, some for testing – should have been affected by essentially the same propagation conditions, so if they were produced in a highly stereotyped manner, they should have arrived at the hydrophone with very similar structure, and should have been detected equally well. It is also possible that the reason is timing: adjacent beaked whale clicks are closer to each other in time – they are typically less than a second apart – while adjacent baleen whale sounds are tens to hundreds of seconds apart, so that the propagation conditions varied more between adjacent baleen whale vocalizations than they did between adjacent beaked whale clicks.

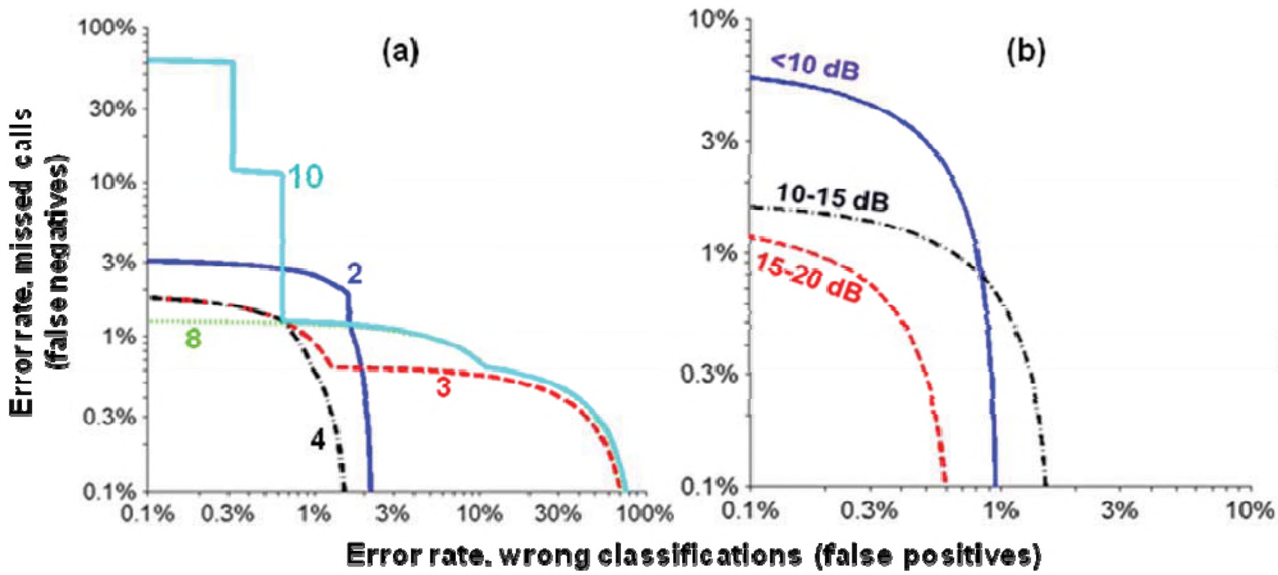

Fig. 2. Receiver Operating Characteristic curves for the neural network detector applied to the training data. Values near the lower left corner, representing smaller numbers of false detections and missed calls, are better. (a) ROC curve for different numbers of hidden units in the neural network. [The 8-hidden-unit curve is hidden by the 10-hidden-unit curve to the right of the 1% false positive point.] (b) ROC curves for the 4-hidden-unit network, with the curve for each 5-dB SNR bin plotted separately. The bin with SNR greater than 20 dB had no missed calls, and so could not be plotted on a logarithmic scale.

It appears that the curve for the 4-hidden-unit network performed best over much of the range, with the 8-hidden-unit network best over the remainder (Fig. 2). However, re-running the training procedure on another 4-hidden-unit network resulted in a performance curve somewhat worse than this one, closer to the 10-hidden-unit curve shown here to the right of the 1% false positive mark. So the superior performance of this network cannot be attributed solely to the number of hidden units.

It is possible that the performance shown here is due to over-fitting. For instance, the network with 4 hidden units has 192*4+4=772 weights, which were trained using a data set of 90% of the whole – i.e., 1436 BBW clicks and 2786 'other' clicks, or 4222 data points in total. This is about 5.5 training clicks per weight, which could be insufficient. The networks with fewer hidden units are less likely to have suffered over-fitting, with e.g. 11 data points per weight for the 2-hidden-unit network, and vice versa – the network with 10 hidden units had only 2.2 data points per weight.

Future work includes testing this network on sounds recorded from Blainville's beaked whale populations elsewhere in the world and in different noise conditions.

One might expect a neural network to perform poorly when confronted with different background noise. However, there is some hope that this one will do well, as the spectrogram conditioning steps reduce the influence of stationary or slowly-varying noise – indeed, of any noise source that is stationary on roughly the time scale at which the spectrum equalization occurs, 0.02 s.

Also, this classifier needs to be combined with a detector and the two evaluated together so that they can be useful for detection of beaked whales in the field, and can be used to mitigate the effects of human activities, including anthropogenic noise, upon these cryptic and little-understood animals.

ACKNOWLEDGEMENTS

Thanks to Dave Moretti, Nancy DiMarzio, Ron Morrissey, and Jessica Ward for providing the data used here. Thanks also to Sara Heimlich and Sharon Nieukirk for the many hours spent annotating beaked whale clicks; you hung in there. Thanks to Kate Stafford for the French translation. The work was supported by Navy awards N00014-03-1-0099 from the Office of Naval Research (thanks to Bob Gisiner), and by N00014-03-1-0735, N00244-06-P-1870, and N00244-07-1-0005 from the Office of the Chief of Naval Operations and the Naval Postgraduate School (thanks to Frank Stone, Ernie Young, and Curt Collins). This is PMEL contribution #3147.

REFERENCES

Aguilar de Soto, N., M. Johnson, P.T. Madsen, P.L. Tyack, A. Bocconcelli, and J.F. Borsani. Does intense ship noise disrupt foraging in deep-diving Cuvier's beaked whales (*Ziphius cavirostris*)? Mar. Mamm. Sci. 22:690-699.

Au, W.W.L. 1993. *The Sonar of Dolphins*. Springer-Verlag: New York.

Barlow, J., M.C. Ferguson, W.F. Perrin, L. Balance, T. Gerrodette, G. Joyce, C.D. Macleod, K. Mullin, D.L. Palka, and G. Waring. 2006. Abundance and densities of beaked and bottlenose whales (family Ziphiidae). J. Cetacean Res. Manage. 7:263-270.

Dawson, S., J. Barlow, and D. Ljungblad. 1998. Sounds recorded from Baird's beaked whale, *Berardius bairdii*. Mar. Mamm. Sci. 14:335-344.

Fawcett, T. 2006. An introduction to ROC analysis. Patt. Recogn. Lett. 27:861-874.

Fernández, A., J.F. Edwards, F. Rodríguez, A.E.de los Monteros, P. Herráez, P. Castro, J.R. Jaber, V. Martin, and M. Arbelo. 2005. "Gas and Fat Embolic Syndrome" involving a mass stranding of beaked whales (family Ziphiidae) exposed to anthropogenic sonar signals. Veterinary Pathology 42:446-457.

Frantzis, A. 1998. Does acoustic testing strand whales? Nature 392:29.

Hagan, M., H. Demuth, and M. Beale. 1996. *Neural Network Design*. Brooks/Cole: Pacific Grove.

Johnson, M., P.T. Madsen, W.M.X. Zimmer, N. Aguilar de Soto, and P.L. Tyack. 2004. Beaked whales echolocate on prey. Proc. Royal Soc. London B Supplement 6, Biology Letters:S383-S386 (DOI 10.1098/rsbl.2004.0208).

Johnson, M. and P. Tyack. 2005. Measuring the behavior and response to sound of beaked whales using recording tags. National Oceanographic Partnership Program Report: Award Number OCE-0427577.

MacLeod, C.D., and A.F. Zuur. 2005. Habitat utilization by Blainville's beaked whales off Great Abaco, northern Bahamas, in relation to seabed topography. Mar. Biol. 147:1-11.

NMFS. 2001. Bahamas marine mammal stranding event of 15-16 March 2000. Joint Interim Report, National Mar. Fish. Serv., Washington, DC. 66 pp.

Deecke, V.B., J.K.B. Ford, and P. Spong. 1999. Quantifying complex patterns of bioacoustic variation: Use of a neural network to compare killer whale (*Orcinus orca*) dialects. J. Acoust. Soc. Am. 105:2499-2507.

Ghosh, J., L.M. Deuser, and S.D. Beck. 1992. A neural network-based hybrid system for detection, characterization, and classification of short-duration oceanic signals. IEEE J. Oceanic Engr. 17:351-363.

Houser, D.S., D.A. Helweg, and P.W. Moore. 1999. Classification of dolphin echolocation clicks by energy and frequency distributions. J. Acoust. Soc. Am. 106:1579-1585.

Kundu, A., and G.C. Chen. 1997. An integrated hybrid neural network and hidden Markov model classifier for sonar signals. IEEE Trans. Sig. Process. 45:2566-2570.

Mellinger, D.K. 2004. A comparison of methods for detecting right whale calls. Can. Acoust. 32:55-65.

Mellinger, D.K., S. Heimlich, and S. Nieukirk. 2004. A comparison of optimized methods for detecting blue whale calls. J. Acoust. Soc. Am. 116:2587(A).

Moretti, D., N. DiMarzio, R. Morrissey, J. Ward, and S. Jarvis. 2006. Estimating the density of Blainville's beaked whale (*Mesoplodon densirostris*) in the Tongue of the Ocean (TOTO) using passive acoustics. Proc. IEEE Oceans '06. 5 pp.

Murray, S.O., E. Mercado, and H.L. Roitblat. 1998. The neural network classification of false killer whale (*Pseudorca crassidens*) vocalizations. J. Acoust. Soc. Am. 104:3626-3633.

Potter, J.R., D.K. Mellinger, and C.W. Clark. 1994. Marine mammal call discrimination using artificial neural networks. J. Acoust. Soc. Am. 96:1255-1262.

Rumelhart, D.E., J.L. McClelland, and the PDP Research Group. 1987. *Parallel Distributed Processing*. MIT: Cambridge.

Van Trees, H.L. 1968. *Detection, Estimation, and Modulation Theory*, Vol. I. John Wiley Sons: New York.