COMPARISON OF MACHINE LEARNING TECHNIQUES FOR THE CLASSIFICATION OF ECHOLOCATION CLICKS FROM THREE SPECIES OF ODONTOCETES

Marie A. Roch[1], Melissa S. Soldevilla[2], Rhonda Hoenigman[1], Sean M. Wiggins[2], and John A. Hildebrand[2]

[1] Dept. of Computer Science, San Diego State University, 5500 Campanile Dr, San Diego, California, 92182-7720 USA

[2] Scripps Institution of Oceanography, The University of California at San Diego, La Jolla, California 92093-0205 USA

ABSTRACT

A species classifier is presented which decides whether or not short groups of clicks are produced by one or more individuals from the following species: Blainville's beaked whales, short-finned pilot whales, and Risso's dolphins. The system locates individual clicks using the Teager energy operator and then constructs feature vectors for these clicks using cepstral analysis. Two different types of detectors confirm or reject the presence of each species. Gaussian mixture models (GMMs) are used to model time series independent characteristics of the species feature vector distributions. Support vector machines (SVMs) are used to model the boundaries between each species' feature distribution and that of other species. Detection error tradeoff curves for all three species are shown with the following equal error rates: Blainville's beaked whales (GMM 3.32%/SVM 5.54%), pilot whales (GMM 16.18%/SVM 15.00%), and Risso's dolphins (GMM 0.03%/SVM 0.70%).

SOMMAIRE

Ce travail concerne la création d'un système pour identifier trois espèces d'odontocètes par les clics d'écholocation: la baleine à bec de Blainville, la baleine pilote, et le dauphin de Risso. Les clics sont identifiés par l'opérateur d'énergie Teager-Kaiser, et les vecteurs cepstraux sont construits. Dans un travail de détection, on compare les résultats obtenus avec deux modèles différents : le modèle de mélange gaussiens (MMG) et la machine à vecteurs de support (MVS). Les résultats de la détection sont exprimés par les courbes de DET, « *Detection Error Tradeoff* ». Le point sur les courbes de DET où les probabilités de fausses alarmes et manques de détection sont égales est comme suit : la baleine à bec de Blainville (MMG 3,32%/MVS 5,54%), la baleine pilote (MMG 16,18%/MVS 15,00%) et le dauphin de Risso (MMG 0,03%/MVS 0,70%).

1. INTRODUCTION

The use of acoustic information for study of marine mammals is a promising method that is complimentary to visual observations. One use of acoustics is to determine the presence of species of interest, the so called detection problem. In this work, we describe a detection system implemented for the 3rd International Workshop on the Detection and Classification of Marine Mammals Using Passive Acoustics, a conference which brought together multiple groups to work on a common data set containing calls from Blainville's beaked whales (*Mesoplodon densirostris*), short-finned pilot whales (*Globicephala macrorhynchus*) and Risso's dolphins (*Grampus griseus*). Low error-rate detections were achieved for all three species using both Gaussian mixture models (GMMs) and support vector machine algorithms.

2. BACKGROUND

Building an effective machine learning solution is a combination of determining the right set of features to use and an appropriate classifier. Features should be chosen such that they capture the essence of the problem, a statement that is easy to make and frequently difficult to achieve. Once the feature set is determined, a method of detection or classification must be selected that enables the system to effectively exploit characteristics of the feature set.

2.1 Features

Bioacousticians working on detection and identification problems for odontocetes have traditionally concentrated on extracting features from whistles. Typically, systems identify a variety of measurements of the whistle such as slope, inflection points, frequency, etc. either manually or automatically (e.g. [1, 2]). There has been little effort in the examination of echolocation clicks or burst pulses as providing information that can be used to determine species, and until recently, band limitations of most field recording systems prevented serious consideration of clicks as features for species recognition tasks.

We have noted unique spectral patterns in echolocation clicks of some species of delphinids, notably Pacific white-sided dolphins (*Lagenorhynchus obliquidens*) and Risso's dolphins [3]. Earlier work [4] on an automatic species

identification system showed good results on a species identification problem where whistles, burst-pulses, and clicks were processed in an identical manner. These results have led us to investigate the suitability of clicks as indicators of species. We see this as being a complementary task to whistle-based systems rather than a competing one. Both methods have advantages: whistles propagate farther than clicks [2], but the short duration of clicks makes call separation easier in large population groups, and some species are not known to whistle [5]. In addition, whistle production may be linked to behavioral state and we have observed species which are known to whistle producing only clicks.

A range of techniques have been used to characterize odontocete clicks [6]. In general, signal samples are squared and heuristics or distributional metrics are used to determine the beginning and ending energy. As described later, we use a technique based upon the Teager energy operator which is similar to that proposed by Kandia and Sylianou [7]. Once the click is identified, typical features include the peak frequency, 3 dB bandwidth, inter-click intervals, etc. [8]. These metrics are a very rough approximation of the spectral shape. Most of the work on echolocation has focused on on-axis clicks, but it is well known that off-axis clicks lack the coherence of on-axis ones and have significantly different spectra [9-11]. In addition to inter-species differences, click production is known to vary even in the same individual in source level, peak frequency, and bandwidth, depending upon factors such as activity and environment [8, 10]. The variation in click attributes suggests that an effective species detector needs to be able to learn a variety of click types associated with each given target species.

2.2 Classifiers and detectors

A recent discussion on applications of machine learning techniques to bioacoustics can be found in [4] and includes linear discriminant analysis, neural networks, dynamic time warping, adaptive resonance theory networks, classification and regression trees, hidden Markov models, self-organizing maps, and Gaussian mixture models (GMMs). In this study, we compare the performance of GMMs with that of support vector machines (SVMs). GMMs are well known for their ability to model arbitrary distributions whereas SVMs attempt to model the boundaries between distributions. SVMs have gained in popularity throughout the 1990s in the machine learning community and to our knowledge have only recently been considered in the bioacoustics community [12, 13].

3. METHODS

3.1 Click production of target species

The click characteristics of the three species vary greatly. Digital acoustic recording tag (DTAG) recordings of free-ranging Blainville's beaked whales have shown that they produce two types of click trains [10]. One type has been observed in prey approach, characterized by a frequency modulated (FM) sweep with inter-click intervals (ICIs) of 100 ms and a median centroid frequency of 38.3 kHz, RMS bandwidth and duration of 6.9 kHz and 271 μs, respectively. These swept clicks are presumed to be related to foraging activities. As the whales close in on their prey, they have been observed to switch to buzz clicks which have different spectral characteristics from the FM sweep clicks. The buzz clicks have greatly diminished ICIs, a higher median frequency of 51.3 kHz with wider RMS bandwidth (14.6 kHz) and an RMS duration which is about half of the FM sweep clicks (29 μs).

Analysis of clicks recorded on a ship-deployed hydrophone array [9] show that free-ranging Risso's dolphins produce clicks with ICIs generally between 40-200 ms with short click trains having ICIs of 20 ms. Centroid frequency of on-axis clicks is 75 kHz (out of band for the conference data set) with an RMS bandwidth of 25 kHz and duration of 30-50 μs. Presumed off-axis clicks from a different population of Risso's dolphins have been shown to have a spectral peak and notch structure [3].

Echolocation clicks of short-finned pilot whales recorded in the Gomera and Canary Islands have been reported [14] to produce clicks with RMS bandwidths of 27 kHz and durations of 8.4 μs. The mean centroid frequency was 68 kHz (also out of band for the conference data).

3.2 Click detection and feature extraction

Clicks are detected using a two-stage search. In the first stage, spectra are created for 20 ms frames with a 10 ms frame advance that have been windowed using a Hann window. Noise is estimated on a per frequency bin basis over a 5 s average. A frame is said to be a click candidate when frequency bins covering at least 5 kHz exceed the noise floor by 12 dB. After obtaining a set of click candidates, a second pass locates clicks with greater precision in a high pass filtered (10 kHz) signal.

The Teager energy operator [15] is an estimate of the instantaneous energy of a signal and has been shown to be an effective method for detecting echolocation clicks [7]. It is based upon a model of the energy needed to drive a spring-mass oscillator, and measures energy with high resolution:

$$\psi_d(x[n]) = x^2[n] - x[n-1]x[n+1] . \qquad (1)$$

A noise floor is set at the 40th percentile of the Teager energy measurements across the interval detected in the previous step. Locations where the Teager energy exceeds the noise floor by a factor of 50 are assumed to be interior to the click and the click onset is found by searching for the point at which the energy dips below 1.5 times the noise floor.

Once the click has been located, cepstral features [16] are computed for a 1200 μs segment of the signal starting with

the click onset. The log magnitude of the discrete Fourier transform of the segment is computed after windowing with a Hann window. The discrete cosine transform of this result is the cepstrum. We also form an estimate of the cepstral representation of noise in the vicinity of the click and subtract the average noise. This is known as cepstral means subtraction [17] and is a method which normalizes for convolutional noise (e.g. mismatched hydrophones or filtering). Once cepstral features have been generated, they are grouped such that the first click and the last click are separated by no more than 2 s and no click is more than 1 s apart from the previous click.

3.3 Detection

Gaussian mixture models (GMMs) and support vector machines (SVMs) were both used in this study. Due to space constraints, only an outline of each technique is presented, but references to the literature where complete details can be found are provided. For both methods, our experiments are designed to answer the question: Given that we are looking for target species X, was a specific set of clicks produced by this species? This contrasts with an identification task where one attempts to determine which species produced the set of clicks.

Gaussian mixture models

For GMM classifiers, one GMM was trained for each of the three species. GMMs are frequently used to approximate arbitrary distributions as a linear combination of parametric distributions. A set of N normal distributions with separate means μ_i and diagonal covariance matrices Σ_i are scaled by a weight factor c_i such that the sum of their integral across the entire feature space is 1. The likelihood of the cepstral feature vector $\vec{x}$ which represents a click can be computed for model $M = \left[\{c_i\}, \{\mu_i\}, \{\Sigma_i\} \text{ where } 1 \le i \le N \right]$ by:

$$\Pr(\vec{x} \mid M) = \sum_{i=1}^{N} \frac{c_i}{(2\pi)^{\frac{d}{2}} \mid \Sigma_i \mid^{\frac{1}{2}}} e^{\frac{-(\vec{x} - \vec{\mu}_i)' \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i)}{2}} \quad . \quad (2)$$

The number of mixtures is typically chosen empirically. Model estimation (training) cannot be accomplished by a straightforward application of the maximum likelihood (ML) principle as the relative contribution c_i of each mixture to the total likelihood is unknown. To address this, the GMM is trained incrementally. A single mixture GMM is estimated from the sample mean and variance. This mixture is then split into two mixtures by dividing the weight in two and forming new mixtures where the means have been slightly perturbed by a small vector $\pm\vec{\delta}$. The resulting model is then refined by an application of the EM algorithm [18] where the current estimate is used to determine the expected values of the mixture weights. With the missing weights estimated, the ML estimator can be

found. This process is executed several times and the model is split again. Once the desired number of mixtures is reached, iteration is performed until a convergence threshold is reached. Convergence is guaranteed and is typically fast (5-15 iterations).

After the models have been trained, the likelihood of click groups are computed and a log likelihood ratio test is used to decide whether each group belongs to each species [19]. We make the simplifying assumption that clicks in a group are independent and compute the group likelihood as the product of the individual click likelihoods normalized for group duration by using the geometric mean. These operations are done in the log domain to prevent machine underflow. Decisions to accept or reject the hypothesis that a click group was produced by the species in question are based upon a log likelihood ratio test. Due to the small number of competing classes, we set the alternative class likelihood to be the likelihood of the highest competitor model as opposed to a background model. The system is implemented using Cambridge University's hidden Markov toolkit (HTK) [20] along with a custom set of programs written in Python and Matlab™.

Support Vector Machines

Support vector machines do not model the distribution of classes, but rather their separation [21]. SVMs find the separating hyperplane that minimizes the risk of a classifier under a 0-1 loss rule. Let $f_{\vec{\theta}}(\cdot)$ be a function parameterized by θ that maps examples to negative and positive class labels $y \in \{-1, 1\}$. As we almost never have access to the actual risk, we can attempt to minimize the empirical risk:

$$R_{emp}(\vec{\theta}) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} \mid y_i - f_{\vec{\theta}}(\vec{x}_i) \mid . \quad (3)$$

Thus, optimizing the parameter vector $\vec{\theta}$ is likely to result in lowering the misclassification rate. For a given family of classifiers, it can be shown that there exists an upper bound on the actual risk with any desired level of certainty [21, 22]. For SVMs, each $f_{\vec{\theta} = [\vec{w}, b]}(\cdot)$ specifies a hyperplane $\vec{w}\vec{x} + b = 0$ which separates the two classes of linearly separable training data (nonseparable data is discussed later). The hyperplane normal vector $\vec{w}$ and bias b are scaled such that $\vec{w}\vec{x} + b = \pm 1$ holds for the closest positive and negative training example, resulting in an empirical risk of 0. The separating line for a two dimensional synthetic data set and the parallel lines that occur at $\vec{w}\vec{x} + b = \pm 1$ are shown in **Figure 1**. As points on the hyperplane satisfy $\vec{w}\vec{x} + b = 0$, the distance between the closest point of each class and the hyperplane is $1/\|\vec{w}\|$.
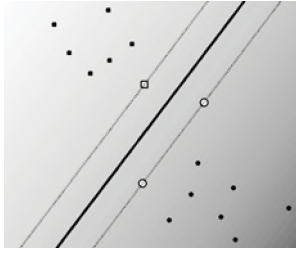
Figure 1 - Separating hyperplane (solid line) between squares and circles that maximizes the distance between the closest vectors (margin). Support vectors lying on $\vec{w}\vec{x} + b = \pm 1$ are outlined.

Figure 2 – Squares and circles that are not linearly separable. Hyperplane with dot product kernel (left) versus Gaussian kernel (right).

Consequently, the separation between the two closest points and the hyperplane is $\frac{2}{\|\vec{w}\|}$. This quantity is referred to as the margin and we can learn the appropriate parameters for the SVM by maximizing the margin subject to the constraints of the closest vectors. This is done by minimizing $\|\vec{w}\|$ or equivalently $\|\vec{w}\|^2$ subject to constraints:

$$\left| \vec{w}\vec{x}_i + b \right| \geq 1. \tag{4}$$

This is a constrained convex optimization problem, which can be solved by optimizing the dual of the Lagrange multiplier representation [21]. The Lagrange multipliers $\alpha_{1 \leq i \leq N}$ will only be nonzero for training examples which satisfy equality in (4). These vectors are called support vectors. The SVM normal vector $\vec{w}$ can be constructed from the dual solution: $\vec{w} = \sum_i \alpha_i y_i \vec{x}_i$, and b is a more complicated function of the support vectors which we omit. We decide the class of test vector $\vec{t}$ by examining the sign of $\vec{w}\vec{t} + b$, or equivalently in the dual representation:

$$f_{\vec{\theta}}(\vec{t}) = \begin{cases} 1 & \sum_i \alpha_i y_i \vec{x}_i \vec{t} + b \geq 0 \\ -1 & \sum_i \alpha_i y_i \vec{x}_i \vec{t} + b < 0 \end{cases}. \tag{5}$$

The above discussion is for sets that are linearly separable, and can be extended in two ways. The first is to introduce a slack variable $\xi_i \geq 0$ for each training vector which permits support vectors to be on the wrong side of the hyperplane:

$$\begin{aligned} \vec{w}\vec{x}_i + b &\geq 1 - \xi_i & y_i = 1 \\ \vec{w}\vec{x}_i + b &\leq -1 + \xi_i & y_i = -1. \end{aligned} \tag{6}$$

When minimizing the risk, a cost factor C is introduced which scales the sum of the slack variables, with high values of C resulting in higher penalties for crossing the margin. Like the linearly separable case, this can also be solved as a constrained optimization problem. The complexity of solving these problems results in selecting strategies such as the sequential minimal optimization algorithm [22] to provide solutions within a reasonable time frame.
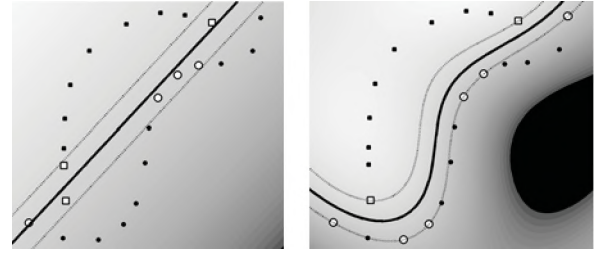
Typically, the normal vector $\vec{w}$ is not actually constructed, but left as a linear combination of the Lagrange multipliers α_i and their associated training data $\vec{x}_i$ and class y_i: $\vec{w}\vec{t} = \sum_i \alpha_i y_i \vec{x}_i \vec{t}$. A second key element to address nonlinearly separable data is to use a kernel function $K(\cdot, \cdot)$ to transform the data into a different space where linear separation is possible. The examples that we have seen so far use what is known as the dot product kernel $K(\vec{x}, \vec{t}) = \vec{x}\vec{t}$. While numerous kernels have been proposed [22], we will restrict ourselves to nonlinear Gaussian kernels

$$K(\vec{x}, \vec{t}) = e^{-\frac{\|\vec{x} - \vec{t}\|^2}{2\sigma^2}} \tag{7}$$

where σ is a tunable parameter. **Figure 2** shows an example of separating hyperplanes for nonlinearly separable data.

When multiple test vectors are classified as a group, the decision to accept a hypothesis that the clicks are produced by a specific species is based upon the threshold of a statistic of the group's click scores. We use as our statistic the percentage of clicks for which $f_{\vec{\theta}}(\cdot) \geq 0$. The system is implemented using the Torch machine learning library [23] and custom C++, Matlab™, and Python code.

For both types of classifiers, we used all available training data for the final classifier. During development, training data was jackknifed by recording date so that the system could be evaluated with test data separate from the evaluation test reported in the results section.

3.4   Evaluating results

Results are plotted using the detection error tradeoff (DET) curve [24]. DET curves are similar to receiver operator curves (ROC) except that in the former error rate normal deviates are plotted on both axes, whereas in the latter the correct detection and false alarm probabilities are plotted. When the false alarm and missed detection probabilities are normally distributed, the result is a straight line in DET
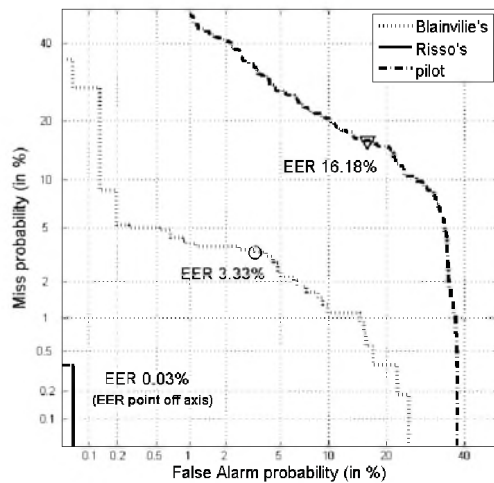
Figure 3 – Detection error tradeoff curves for GMM detector on evaluation data.
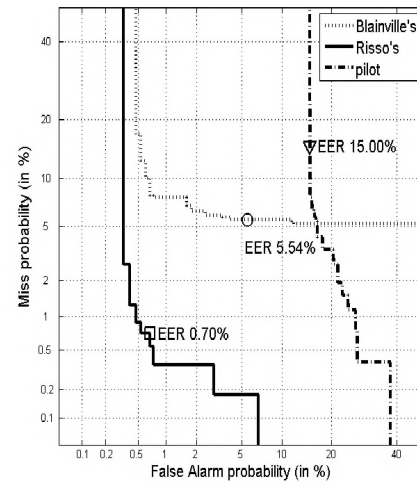

Figure 4 – Detection error tradeoff curves for SVM detector on evaluation data.

space. DET plots are more effective at highlighting differences between similar systems than ROC curves.

4. RESULTS

Mean normalized cepstral features were extracted for all files of the dataset. Tests on the jackknifed training data were used to tune the parameters of each classifier. For the GMMs, 2, 4, 8, 16, 32, and 64 mixture models were created, with 16 mixture models outperforming other parameters. For SVMs, a grid search on the penalty and standard deviation was performed ($C \in \{100, 200, ..., 600\}$, $\sigma \in \{100, 200, ..., 1000\}$). Equal error rates (EERs), the point at which a decision threshold results in the same percentage of false alarms (false positives) and missed detections are summarized in **Table 1**. Tests on the last day's training data performed poorly for SVMs, leading to the high overall EERs.

The best performing models from the development data were then used to classify click groups from the nine evaluation files whose content is summarized in **Table 2**. The evaluation dataset contained calls from the three aforementioned species plus an additional two: Atlantic spotted dolphins (*Stenella frontalis*) and sperm whales (*Physeter macrocephalus*).

File 1 had mixed Blainville's and pilot whale clicks. We manually established "correct" labels for each click group in the file based upon known characteristics of the species and our observations of the calls in the development data. A total of 2040 click groups with a mean of 10.1 clicks per group (min=1, max=103, std dev.=7.2) were classified. DET curves and EERs for all three target species are produced for the GMM and SVM detectors in Figures 3 and 4. The curves show the tradeoff between false alarms and missed detections for various detection thresholds. Note that the thresholds themselves would add a third dimension to the plot and are not reported.

5. DISCUSSION

For both classifiers, the detector performance on Risso's dolphins appears to be nearly perfect in the evaluation data, but the Risso's calls in the conference data were filtered, leading us to suspect that part of the accuracy is due to environment detection as opposed to species detection. It is also worth noting that much of the error on the SVM development set for Risso's dolphins comes from one particular split where the data from August 19th 2006 was used as test data. This was the one day for which the Risso's dolphin data contained clicks with spectra above 40 kHz. The GMM classifier dealt better with this situation, recognizing other similarities in the data. The pilot whale

EER %	GMM	SVM
Blainville's	2.8	21.4
pilot	3.7	21.1
Risso's	2.3	14.7

Table 1– Equal error rates for jackknifed development data with 16 mixture GMMs and $C = 100, \sigma = 200$ SVMs for the best parameter set across all jackknife splits.

Species producing calls in the test files						
1	Blainville's + some pilot	4	spotted	7	Risso's
2	Blainville's	5	Risso's	8	pilot
3	spotted	6	Blainville's	9	sperm

Table 2 – Contents of evaluation files 1–9.

detectors had the worst performance on the evaluation data, with the majority of errors being in the 661 out-of-set (species not seen in training) click groups from the spotted dolphins and sperm whales. Using the EER threshold, 42.97% (GMM) and 39.79% (SVM) of the out-of-set click groups were incorrectly identified as pilot whales, indicating that rejection of out-of-set clicks is an area for future work.

For any out-of-set test, the impostor click will most closely fit one of the three distributions, making its GMM likelihood higher than the others. The likelihood ratio between the two highest ranked models may be large, and it is not unexpected that a greater number of errors will occur in this situation. When examining the likelihoods produced by the pilot whale model without the normalizing alternative hypothesis, there is significant overlap. Consequently, setting a threshold based upon the pilot whale model alone would not have improved the results. Adding enough species to the alternative hypothesis to better represent the variability of clicks across species may improve out-of-set rejection. For SVMs, the lack of a distributional approach means that even if a click is far from the target species' distribution, if it lies on the target side of the hyperplane, it will be considered a target, making the need for additional data critical.

It is worth noting that the DET curve for Blainville's beaked whales has a relatively flat slope over much of its length for both detectors. This means that the threshold is not overly sensitive, and we can reduce either the miss or false alarm probabilities significantly with a low impact on the other metric. As an example with GMMs, it is possible to have a very low false detection rate ($< 0.2\%$) and miss no more than 5% of the click groups. While the Risso's dolphin curve has a steep slope, its location in the lower left corner makes this less critical. The shape of the pilot whale curves is more problematic, with small differences in threshold having more significant impact.

When examining what appeared to be off-axis clicks, Johnson et al. [10] were able to distinguish individual pulses by cross correlation with on-axis clicks. They noted that the spectra of the off-axis clicks were "highly featured," lacking the smoothness of presumed on-axis clicks. The spectral irregularities were attributed to possible interference between pulses. We believe this to be a reasonable hypothesis, and one of the major reasons that echolocation based species detection works well. Measurements of the melon taken from CT scans of a deceased Risso's dolphin show a 30 cm length from dorsal bursae to probable signal exit and a 20 cm width at the widest section. While exact propagation paths are beyond the scope of this work, the 1200 μs window used in this study is adequately long to permit multiple paths to have interfered in constructive and/or destructive manners (assumed sound speed of 1500 m/s), even for the larger species. It is interesting to note that when we used windows smaller than 1100 μs, detection performance degraded significantly.

6. Conclusions

We have shown that cepstral feature vectors extracted from spectra over a 1200 μs window starting at the beginning of an echolocation click can be used as the basis for automated species detectors. These detectors are competitive with other state-of-the-art systems for the detection of echolocating marine mammals. It is of particular interest that the system performed well even though the echolocation clicks extended beyond the bandwidth supported by the recording equipment. EERs for this dataset ranged between 0.03% and 16.8% for GMMs and 0.70% and 15.0% for SVMs. Further work is needed on rejecting out-of-set species whose clicks bear a stronger resemblance to the target species than to any of the species used to build the impostor set.

While other explanations may exist, we also believe that the observed degradation of performance when the analysis window was shortened is a strong indicator that interference patterns may play a role in the spectral patterns. Further experiments may help to confirm or reject this hypothesis.

REFERENCES

[1]   L. E. Rendell, J. N. Matthews, A. Gill, J. C. D. Gordon, and D. W. Macdonald, "Quantitative analysis of tonal calls from five odontocete species, examining interspecific and intraspecific variation," *J. Zool., Lond.,* vol. 249, pp. 403-410, 1999.

[2]   J. N. Oswald, S. Rankin, J. Barlow, and M. O. Lammers, "A tool for real-time acoustic species identification of delphinid whistles," *J. Acous. Soc. Am.,* vol. 122, no. 1, pp. 587-595, 2007.

[3]   M. S. Soldevilla, M. A. Roch, S. M. Wiggins, J. Calambokidis, and J. A. Hildebrand, "The use of delphinid echolocation click spectral properties for species classification," *J. Acous. Soc. Am.,* submitted.

[4]   M. A. Roch, M. S. Soldevilla, J. C. Burtenshaw, E. E. Henderson, and J. A. Hildebrand, "Gaussian mixture model classification of odontocetes in the

Southern California Bight and The Gulf of California," *J. Acous. Soc. Am.,* vol. 121, no. 3, pp. 1737-1748, 2007.

[5] T. Morisaka, and R. C. Connor, "Predation by killer whales (*Orcinus orca*) and the evolution of whistle loss and narrow-band high frequency clicks in odontocetes," *J. Evol. Biol.,* vol. 20, no. 4, pp. 1439-1458, 2007.

[6] W. R. Elsberry, "Interrelationships between intranarial pressure and biosonar clicks in bottlenose dolphins (*Tursiops truncatus*)," Wildlife and Fisheries Sciences, Texas A&M, College Station, TX, 2003.

[7] V. Kandia, and Y. Stylianou, "Detection of sperm whale clicks based on the Teager-Kaiser energy operator," *Appl. Acous.,* vol. 67, no. 11-12, pp. 1144-1163, 2006.

[8] W. W. L. Au, *The sonar of dolphins*, New York: Springer-Verlag, 1993.

[9] P. T. Madsen, I. Kerr, and R. Payne, "Echolocation clicks of two free-ranging, oceanic delphinids with different food preferences: false killer whales (*Pseudorca crassidens)* and Risso's dolphins (*Grampus griseus)*," *J. Exp. Biol.,* vol. 207, no. 11, pp. 1811-1823, 2004.

[10] M. Johnson, P. T. Madsen, W. M. X. Zimmer, N. A. de Soto, and P. L. Tyack, "Foraging Blainville's beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation," *J. Exp. Biol.,* vol. 209, no. 24, pp. 5038-5050, 2006.

[11] W. M. X. Zimmer, P. T. Madsen, V. Teloni, M. P. Johnson, and P. L. Tyack, "Off-axis effects on the multipulse structure of sperm whale usual clicks with implications for sound production," *J. Acous. Soc. Am.,* vol. 118, no. 5, pp. 3337-3345, 2005.

[12] S. Fagerlund, "Bird species recognition using support vector machines," *EURASIP J. Adv. Sig. Proc. (online journal)*, doi:10.1155/2007/38637, [Accessed 2007].

[13] S. M. Jarvis, N. A. DiMarzio, R. P. Morrissey, and D. J. Moretti., "A novel multi-class support vector machine classifier for automated classification of beaked whales and other small odontocetes," *Canadian Acoust.*, submitted.

[14] T. Götz, A. Boonman, U. Verfuss, and H.-U. Schnitzler, *Echolocation Behaviour of Five Sympatric Dolphin Species off la Gomera/Canary Islands [abstract A-14]*, 2005. [http://www.univ-lr.fr/labo/lbem/ecs2005/Abstract%20book.pdf, Accessed December 2007].

[15] T. F. Quatieri, *Discrete-time speech processing: principles and practice*, Upper Saddle River, NJ: Prentice Hall PTR, 2002.

[16] J. W. Picone, "Signal modeling techniques in speech recognition," *Proc. IEEE,* vol. 81, no. 9, pp. 1215-1247, 1993.

[17] H. Hermansky, "Exploring temporal domain for robustness in speech recognition," in *The 15th Intl. Congress on Acoustics,* Trondheim, Norway, 1995. pp. 61-64.

[18] X. Huang, A. Acero, and H. W. Hon, *Spoken language processing*, Upper Saddle River, NJ: Prentice Hall PTR, 2001.

[19] F. Bimbot, J. F. Bonastre, F. C, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska-Delacretaz, and D. A. Reynolds, "A tutorial on text-independent speaker verification," *EURASIP J. Appl. Sig. Proc.,* vol. 2004, no. 4, pp. 430-451, 2004.

[20] S. Young, G. Evermann, T. Hain, D. Kershaw, X. A. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, *The HTK book, version 3.4*, 2006. [http://htk.eng.cam.ac.uk, Accessed March, 2007].

[21] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Disc.,* vol. 2, no. 2, pp. 121-167, 1998.

[22] N. Cristianini, and J. Shawe-Taylor, *Support vector machines and other kernel-based learning methods*, Cambridge, UK: Cambridge University Press, 2000.

[23] R. Collobert, S. Bengio, and J. Mariéthoz, *Torch: A modular machine learning software library*, IDIAP-RR 02-46, IDIAP, Valais, Switzerland, 2002.

[24] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET curve in assessment of detection task performance," in *Eurospeech*, Rhodes, Greece, 1997. pp. 1895-1898.