A Novel Multi-class Support Vector Machine Classifier for Automated Classification of Beaked Whales and Other Small Odontocetes

Susan Jarvis [2,1] Nancy DiMarzio[1] Ronald Morrissey[1] and David Moretti[1]
1- Naval Undersea Warfare Center Division, Newport, RI, 02841
2- Worcester Polytechnic Institute, 100 Institute Rd., Worcester, MA 01609

ABSTRACT

Navy sonar has recently been implicated in several marine mammal stranding events. Beaked whales (particulary *Mesoplodon densirostris*) have been the predominant species involved in a number of these strandings. Monitoring and mitigating the effects of anthropogenic noise on marine mammals are active areas of research. Key to both monitoring and mitigation is the ability to automatically detect and classify animals, especially beaked whales. This paper presents a novel support vector machine based methodology for automated, species level classification of small odontocetes. The new classifier, called the class-specific support vector machine (CS-SVM), consists of multiple binary SVM's where each SVM discriminates between a class of interest and a common reference class. A main objective in the development of the CS-SVM was to realize a robust multi-class SVM whose implementation is simpler than existing multi-class SVM methods. A CS-SVM was trained to identify click vocalization from four species of odontocetes including *Mesoplodon densirostris*. The algorithm processes time series data in a fully automated fashion first detecting and then classifying click events. Results from the application of this automated classifier to the data sets provided by the 3rd International Workshop on Detection and Classification of Marine Mammals Using Passive Acoustics are presented.

SOMMAIRE

Le sonar a été récemment associé à un certain nombre d'événements de mammifère marin immobilisé en eau peu profond. Les Baleines a bec (en particulier le *Mesoplodon densirostris*) ont été les espèces prédominantes impliquées dans un certain nombre d'événements d'immobilisation. La surveillance et l'atténuation des effets du bruit synthétique sur les mammifères marins sont des domaines de recherche actifs. Ce qui est importante de la surveillance et la réduction des effets est la capacité automatiquement de détecter et classifier des animaux, particulièrement les baleines a bec. Cet article présente une nouvelle méthodologie basée sur une machine de support vecteur (SVM) pour automatisé le classification de niveau d'espèces de petits odontocetes. Le nouveau classificateur, appelé le "class-specific support vector machine" (CS-SVM), est composé de SVM binaire multiple où chaque SVM se distingue entre une classe d'intérêt et une classe commune de référence. Un objectif principal dans le développement du CS-SVM était de réaliser une multi-classe robuste SVM dont l'exécution est plus simple que des méthodes existantes de la multi-classe SVM. Un CS-SVM a été formé pour identifier le vocalisation de clic de quatre espèces des odontocetes incluant des *Mesoplodon densirostris*. Les données de série chronologique de processus d'algorithme sont traitées d'une mode entièrement automatisée détectant d'abord et classifiant ensuite des événements de clic. Les résultats de l'application de ce classificateur automatisé fournis par le "Troisième Atelier Internationale de Détection, Localisation, et Classification du Mammifères Marins avec les Acoutiques Passive" sont présentés.

1.   BACKGROUND

Until quite recently, little was known about the vocalizations of beaked whale. However, starting with the definitive recording of beaked whale clicks by Johnson, Tyack, et al. (using non-invasive DTAG's) [1,2] and continuing with the visually verified recording of beaked whales and other small odontocete vocalizations at AUTEC [3] there is now sufficient labeled data available to develop automated classification algorithms. To foster exchange of ideas and classification methodologies, the 3rd International Workshop on Detection and Classification of Marine Mammals Using Passive Acoustics was convened. In addition to providing a venue for scientific exchange in the topic areas, the workshop provided a data set [12] consisting of both labeled training data for 3 species of odontocetes and unlabeled test data. This paper investigates the application of a novel class-specific support vector machine classifier to the classification of vocalizations from beaked whales and other odontocetes specifically using the data set provided by

the Workshop.

At a basic level, a classification system is one that assigns the current input x membership into one of v known classes according to some set of decision metrics or functions. In general, x is a multivariate random variable where $x \sim P(x)$. For example, popular maximum likelihood classifiers assign an input data vector x membership in one of v possible class hypotheses $\{H_1, H_2, \ldots H_j \ldots H_v\}$ according to the probabilistic rule $j^* = arg\ max(p(H_j|x))$. This is equivalently written as $j^* = arg\ max(p(x|H_j)p(H_j))$ after applying Bayes rule. Theoretically, a maximum likelihood (ML) classifier is optimal in that it offers the lowest probability of error of any classifier [4]. However, in practice, it can be difficult to attain this optimal performance because the multidimensional probability density functions $p(x|H_j)$ are unknown and must be estimated from training data. The amount of training data required to estimate $p(x|H_j)$ grows exponentially with the dimension of x. This is problematic because the collection of labeled training data is usually difficult, time consuming and expensive.

Statistical learning theory [5,6] represents a different paradigm for learning than the classical ML methods presented above. Statistical learning theory advocates solving specific problems directly vice solving more general problems as an intermediate step [5]. That is, if there are limited data available to train a classifier then the best course of action is to estimate a decision boundary directly from the data. This is in contrast to classical ML inference where the data are used to estimate the parameters of density functions and then the PDFs are used to form decision boundaries.

2.   DISCUSSION

One of the corner stones of statistical learning theory is the principle of structured risk minimization (SRM). Using the SRM principle, Vapnik developed a bound on the risk of classification error for a given decision function f given the empirical risk (training error) $R_{emp}(f)$ associated with the function, the training set size m, and the capacity h of the hypothesis space in which the decision function resides [6]. This bound (1) is often referred to as the guaranteed risk, and is independent of the underlying distribution of the data. According to the SRM principle, the smallest bound on classification error is achieved by minimizing training error while using the function hypothesis space of the smallest capacity [5,6].

$$R[f] \le R_{emp}[f] + \sqrt{\left(\frac{1}{m}\left(h\left(\ln\frac{2m}{h}+1\right)+\ln\frac{4}{\delta}\right)\right)} \quad [6]\ (1)$$

Support vector methods (or support vector machines, SVM) are a rich family of learning algorithms based on statistical learning theory. SVM's were originally developed to solve binary classification problems of the following type: Given a set of training data $\{(x_1, y_1), \ldots, (x_m, y_m)\}$ where each (multidimensional) input example x_i drawn from X is associated with classification label $y_i = \pm 1$, determine the decision function that maps any new x drawn from X to $y = \pm 1$ that minimizes risk of misclassification [5]. In short,

SVMs implement the SRM principle.

SVM's use the existence of a unique optimal hyperplane which separates the two classes in some feature space (figure 1). The SVM that implements the optimal hyperplane while maximizing the separation (margin) between the two classes will have the lowest risk of test error [5]. This optimal separating hyperplane is realized as

$$f(\mathbf{x}) = \sum_{k=1}^{m} \alpha_k\, y_k\, G(\mathbf{x}, \mathbf{x}_k) + b \qquad (2)$$

where G is a kernel mapping and b is an offset. The weights α_k for a "soft" margin SVM classifier [6] are found through the optimization (3)

$$\text{maximize } W(\boldsymbol{\alpha}) = \sum_{k=1}^{m} \alpha_k - \frac{1}{2}\sum_{k,l=1}^{m} \alpha_k \alpha_l\, y_k\, y_l\, G(x_k, x_l)$$

$$\text{subject to } \boldsymbol{\alpha} \in \Re^m,\ 0 \le \alpha_k \le C/m \text{ and } \sum_{k=1}^{m} \alpha_k y_k = 0.$$

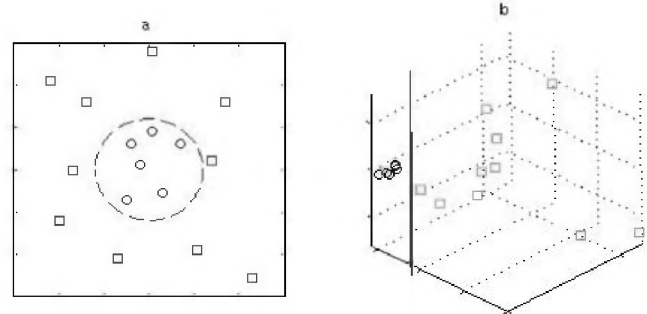**Figure 1**: A notional view of a SVM [6]. a) Training data drawn from x shows two classes. b) A transformation T(x) maps the training data to a higher dimensional space where the optimal separating hyperplane is found. The hyperplane in the higher dimensional space corresponds to a nonlinear decision boundary in the input space.

The constant C controls the degree of "slack" in the hyperplane optimization. Large C corresponds to more rigid separation of the classes and less tolerance for class overlap in the training data. Smaller C allows for moreclass overlap in the training data [7]. Equation (3) can be solved using quadradic programming techniques [6].

While SVM's were originally formulated for binary classification, many real world problems involve more than two classes. As a result, a number of methods have been developed for applying SVM's to multi-class problems. These methods tend to follow one of three basic approaches. The first approach is to form v binary "one-against-the-rest" classifiers (where v is the number of class labels) and choose the class whose decision function is maximized [5]. The second approach is to form all $v(v-1)/2$ pairwise binary classifiers and choose the class whose set of pairwise decision functions are in some way maximized [7]. The third approach is to reformulate the objective function of the SVM for the multi-class case such that the decision boundaries for all classes are optimized jointly [8].

This paper presents a new type of multi-class support vector classifier called the class-specific SVM (CS-SVM). The new classifier consist of v binary SVM's where each

SVM discriminates between one of v classes of interest and a common reference class. The class whose decision function is maximized with respects to the reference class is selected. The CS-SVM extends the concept of exploiting class-specific features as proposed by other researchers for maximum likelihood classifiers [4] and neural networks [9] to the multi-class SVM problem.

Many applications involve the classification of signals which are set in additive noise. In such cases, the problem is not to differentiate between two or more of v signals present at the same time but to differentiate between one of v signals and noise. The input vectors for such problems are actually of the form $x_j = s_j + n$, for $j = 1, 2, ... v$. Currently, SVM's are designed assuming the classification problem is to distinguish $x_i = s_i$ from $x_j = s_j$. Any noise in x is assumed to be accommodated by allowing the "slack variables" in the hyperplane optimization [6].

The CS-SVM expressly acknowledges the presence of the noise by treating it as a common reference class. For a single class, the classification problem reduces to a detection problem, a decision as to whether signal s is present or not. That is, $y = sgn(f(x)) = +1$ when $x = s + n$ and $y = sgn(f(x)) = -1$ when $x = n$. In the multi-class case, x is assigned membership in the class whose decision function $f_i(x)$ against the reference is maximum, or to the noise-only class when all $f_i(x) < 0$. Note that in acknowledging the presence of a common reference class no assumptions are made about that class. Although it is intuitive to think of the reference class as Gaussian noise, the reference class could be of any arbitrary distribution. This means that the CS-SVM can actually act as a signal detector for signals of unknown distribution set in noise of unknown distribution.

Figure 2 is an notional illustration of the CS-SVM concept for two dimensional data. Optimal separating hyperplanes for each class versus the noise-only reference class are found. Since the optimal hyperplane separating any two classes is unique [5], the optimal hyperplane for class i versus n will be different from the optimal hyperplane for class j versus n. However, both hyperplanes are optimized against a common reference class. The decision function $f_i(x)$ for either signal-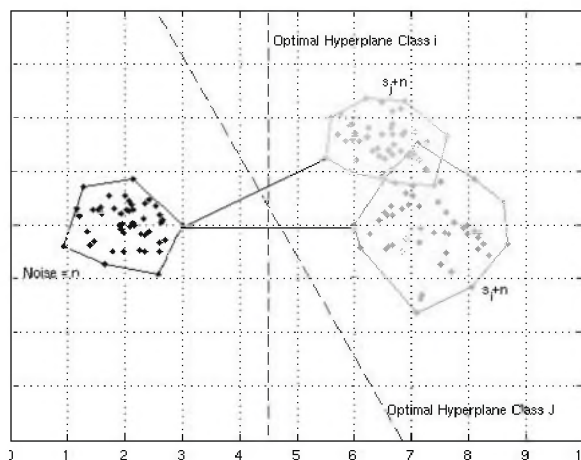present class should reject the noise-only reference case. Further, it is hypothesized that $f_i(x)$ will be greater than $f_j(x)$ whenever x is draw from class i since $f_i(x)$ is optimal for class i and $f_j(x)$ is not.

3. EXPERIMENTAL RESULTS

In the past several years there has been much interest and progress in acoustic monitoring, localization and tracking of marine mammals [3,10]. Acoustic monitoring has a number of benefits over visual monitoring. Chief among them are increased area of coverage and the ability to operate over wider weather conditions and at night. A major drawback of acoustic monitoring is associating species information with the received vocalizations. However, recent field tests combining visual verification and digital recording tags with acoustic monitoring and localization have resulted in sets of labeled acoustic data [3]. One such data set was provided as part of the 3rd International Workshop on Detection and Classification of Marine Mammals Using Passive Acoustics [12]. This section presents the development of a CS-SVM classifier using the Workshop data set.

The data set provided for the Workshop consisted of labeled training data for 3 species as well as unlabeled test data. Training data was supplied for *Mesoplodon desirostris* (Blainville's beaked whale), *Globicephala macrorhynchus* (short-fin pilot whale) and *Grampus griseus* (Risso's dolphin). The training data for each species consisted of five or more .wav files with each file containing 2 to 3 minutes of 16-bit audio data sampled at 96KHz. The test data consisted of nine longer .wav files (each 10+ minutes) also sampled at 96KHz. Within the 9 test files there were examples of the species alone, examples containing a mix of species as well as examples containing none of the 3 species given in the training data. Prior to analysis or processing, all data were passed through a 12 KHz high-pass filter.

3.1 Training Data and CS-SVM Feature Selection

The first challenge in working with the Workshop data was deciding which events and signal features the classifier should be trained to recognize. Design of a classification algorithm generally requires selection a set of distinguishing features q_i to represent the raw data such that the input vector to the classifier is $x = [q_1 \ q_2 ... q_n]^T$. Ideally the the feature set should be a sufficient statistic for the raw data but it also must be of reasonably low dimension as the amount of training data required grows with the dimension of x.

One goal for the CS-SVM classifier presented here is for it to become incorporated into the acoustic marine mammal monitoring system, M3R [3,10]. This means that the classifier would have to be fully automated and run in real-time. In turn, that required selection of features that can be readily extracted "on the fly". Thus, it was decided that the classifier should classify individual click events rather than attempting to analyze click trains.

In previous experiments [11], the times between consecutive zero crossings were successfully used as features for classifying odontocete clicks. A zero crossing detector is easy to implement and the periods between

Figure 2: A geometric view of the optimal separating hyperplanes for two CS-SVMs for class i and class j, respectively, versus a common reference class in a 2-D decision space.

crossings capture the time-frequency structure of the signals. Additionally, the envelope shape of the clicks can be captured by using the normalized peak values between crossings (figure 3).
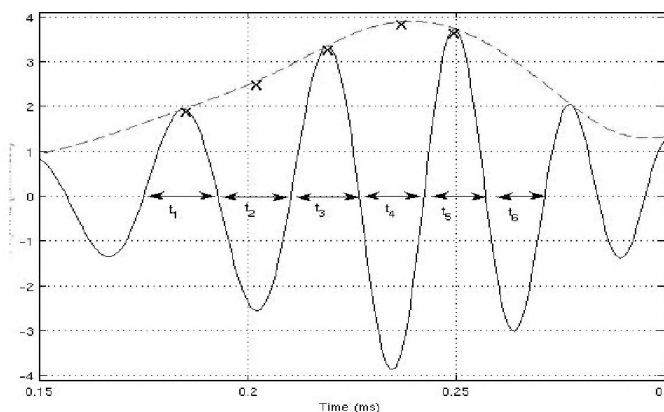

Figure 3: **The times between zero-crossings and normalized envelope amplitudes were selected as features for classifying individual clicks.**

The next step in the training process was to analyze the training data from each species to select the set of click events to be used in training the CS-SVM. The idea was to select several hundred representative clicks from each species, then to extract the zero-crossing and amplitude envelope features from them. Time-frequency analysis of the training data for *Mesoplodon desirostris* identified two distinct click waveforms, foraging clicks and buzz clicks (figure 4). A large majority of the clicks were the stereotypical foraging clicks, but several buzzes were also detected [13]. Since the foraging and buzz click waveforms are distinct, separate CS-SVM's were trained for the two click types.

(a)

(b)
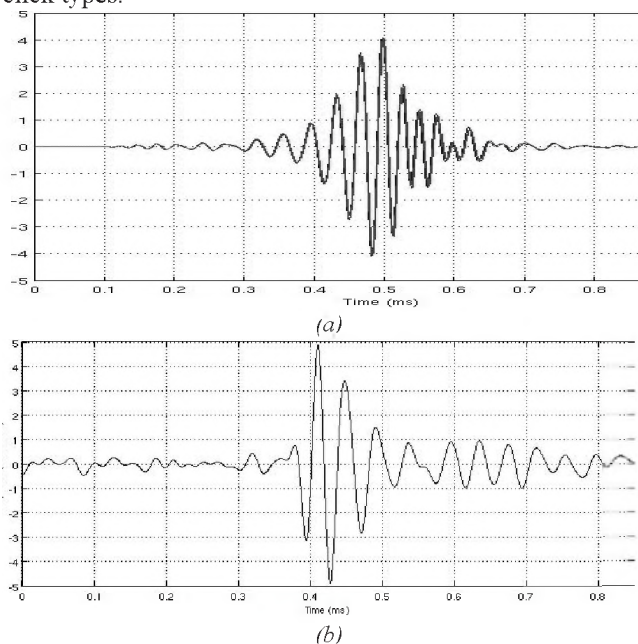
Figure 4: **(a) Overlay of time series data for 3 *Mesoplodon desirostris* foraging clicks. (b) Time series of a single buzz click.**

Time-frequency analysis of the training data from *Globicephala macrorhynchus* and *Grampus griseus* showed the clicks contained in those files to be highly variable. In contrast to the *Mesoplodon* data where the regularity of click waveforms is almost uncanny, it was difficult to identify representative click waveforms in the pilot whale and Risso's dolphin data (figure 5). After cross correlating all the clicks extracted from the training files, the click from each species most highly correlated with most of the other clicks was selected as a replica. Then, the training clicks which were most highly correlated with the replicas were used to build the training sets.

Nine dimensional feature vectors were formed using the times between 6 zero-crossings about the peak and three normalized envelope amplitude peaks. The resulting feature sets from the *Globicephala* and *Grampus* did not cluster as compactly as those from *Mesoplodon*. There was also significant overlap of the features for *Mesoplodon* and *Grampus* in the feature space (figure 6).

One binary SVM was constructed for each signal class versus an ambient noise reference class. The training set T_j for the jth CS-SVM was defined as

$$T_j = \{(\mathbf{x}_j, y)\} = \{(\mathbf{s}_j + \mathbf{n}, 1), (\mathbf{n}, -1)\} \text{ for } j = 1 \text{ to } 4.$$

The training sets for each of the four class-specific SVM consisted of approximately 250 signal-present vectors and a similar number of noise-only vectors. These training sets were used in the optimization (3) to find the optimal hyperplane $f_j(\mathbf{x})$ for each class. A Gaussian radial basis function was used as the kernel in (2) and (3) yielding

$$f_j(\mathbf{x}) = \sum \alpha_k y_k \exp\left(-\|\mathbf{x} - \mathbf{x}_k\|^2 / 2\sigma\right) + b$$

where $k \in S_j$ the set of support vectors class j.

The four class CS-SVM classifier was then tested using additional clicks extracted from the training data files. These test sets also consisted of approximately 250 signal-present vectors and 250 noise-only vectors (Table 3). The classification performance for the test sets was evaluated using the following metrics, P_{cc} = fraction correctly classified (signal present), P_{miss} = fraction misclassified (signal present) and P_{nse} = fraction correctly classified (noise-only). The results were encouraging, especially for the Meso-plodon foraging click class. Note that the poorer performance of the buzz class was attributed to changes in buzz click waveform observed as the inter-click interval decreases during prey capture.

3.2   CS-SVM Results for the Test Data Files

Finally, the CS-SVM classifier was tested using the unlabeled test files. A cursory manual review of the test data prior to processing indicated the presence of sperm whales clicks in multiple test files. Although not part of Workshop training data, a class-specific SVM for sperm whale clicks was trained using additional labeled sperm whale data. The ability to add a class without affecting SVM design for the other classes highlights one of the

strengths of the CS-SVM approach. However, to be consistent with the data conditioning stream for the other classes, the sperm whale data was passed through the same 12 KHz high pass filter. This was known to be a suboptimal processing step as most of the energy in sperm whale clicks is typically below 12 KHz. A better solution for a CS-SVM classifier that includes a sperm whale class would be either to lower the frequency of the high-pass filter or to process in multiple frequency bands.
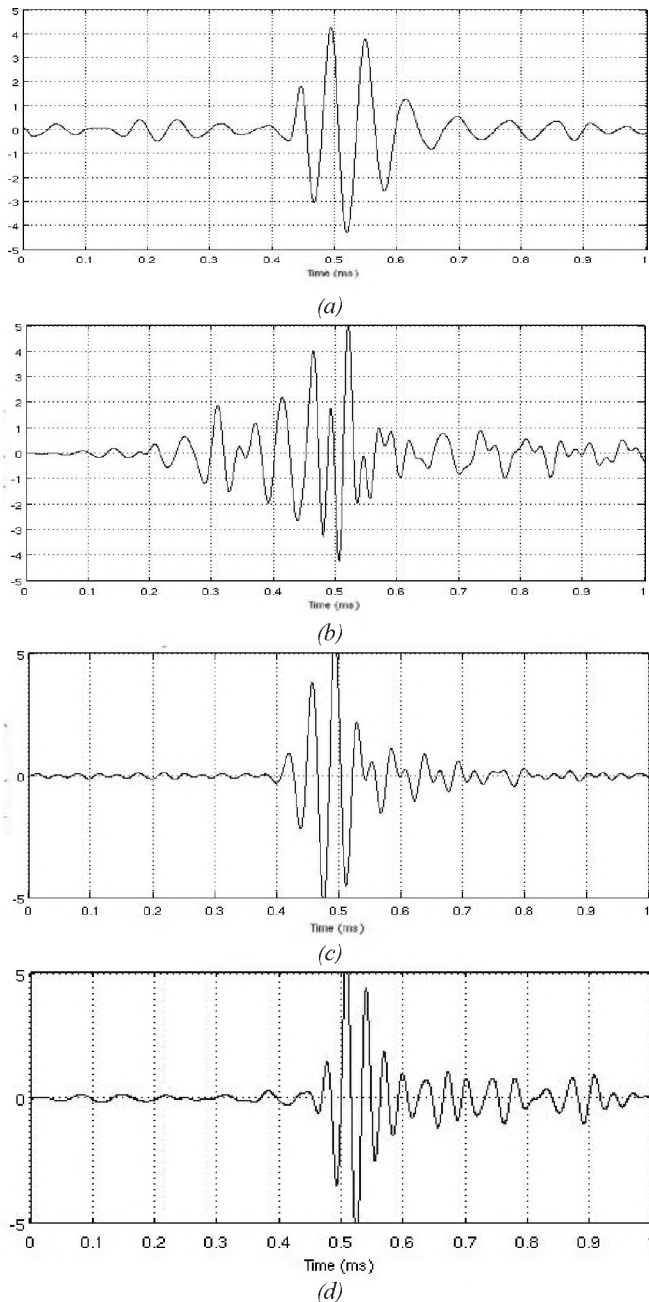
Class	P_{cc}	P_{miss}	P_{nse}
Mesoplodon (forage)	1.0000	0.0000	0.9680
Mesoplodon (buzz)	0.7900	0.2100	0.9600
Globicephela	0.9380	0.0620	0.9682
Grampus	0.9451	0.0549	0.9721

Table 3: Performance of the 4 class CS-SVM on test sets of approximately 250 signal-present vectors and 250 noise-only vectors drawn from the training data files for each clas

The nine test data files were processed in a fully automated fashion. The classifier program automatic-ally read the data from the .wav files, filtered it, and performed time domain energy detection to identify click events. Time series data about the energy detector peaks were used to construct feature vectors. The features vectors x extracted for each click event were then used to evaluate the class-specific decision functions, $f_i(x)$. The click event was assigned membership in the class whose decision function was maximum or to the noise-only reference class when $max\{f_i(x)\} \leq 0$. Figure 7 shows the output of the 5 class-specific decision functions for data in the neighborhood of a *Mesoplodon* foraging click (from Test File 1). Although $f_4(x)$ associated with Risso's dolphin also peaked, $f_1(x)$ for the foraging click was maximum. Figure 8 shows the class decision output of the CS-SVM for Test File 1. This test file contained clicks from both *Mesoplodon* and *Globicephela*. Table 4 summarizes the performance of the 5-class CS-SVM for all of the Workshop test data.

(a)

(b)

(c)

(d)

Figure 5: Clicks showing some of the variability in the *Globicephala macrorhynchus* (a-b) and *Grampus griseus* (c-d) training data.

Amb. Noise
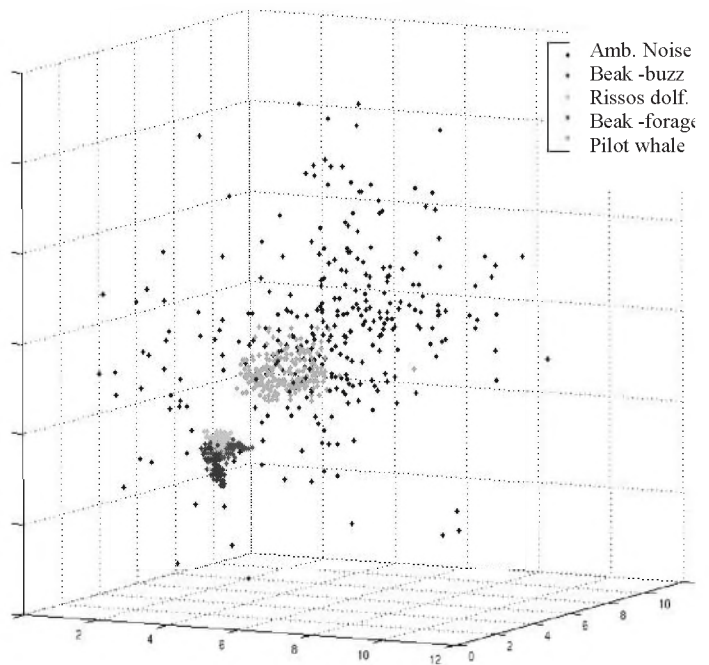Beak -buzz
Rissos dolf.
Beak -forage
Pilot whale

Figure 6: Distribution of the periods between three consecutive zero-crossings for ambient noise, *M. desirostris* (buzz), *Grampus griseus*, *M. desirostris* (forage), and *Globicephala macrorhynchus*.

4. CONCLUSION

This paper has presented a novel multi-class support vector machine classifier, the class-specific SVM. The new classifier consists of v binary SVMs where each SVM discriminates between one of v classes of interest and a common reference class. Test inputs are assigned membership in either the class whose decision function is maximized or the reference class if all decision functions are negative. A five class CS-SVM was created to classify broadband click vocalizations from several species of odontocetes using data provided by the 3rd International Workshop on Detection and Classification of Marine Mammals Using Passive Acoustics. While the CS-SVM's classification performance was quite good for species specific test cases drawn from the labeled training data, its classification performance was not as good for the unlabeled test data files. Some classes, like the *Mesoplodon densirostris* foraging class and the Sperm whale class, performed well on the test files but the performance for the other classes not as reliable.
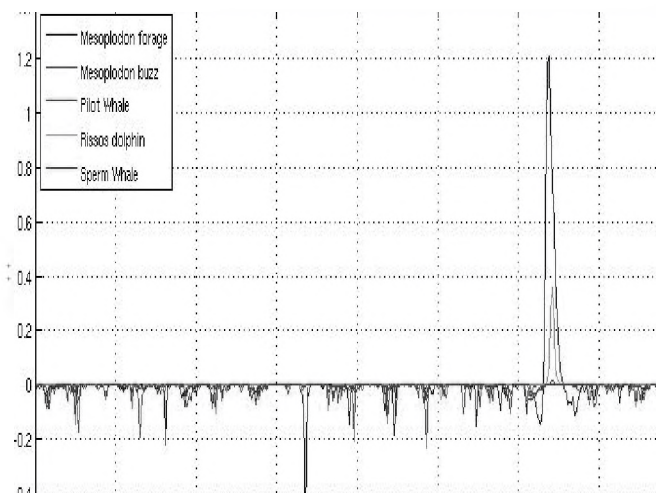
Figure 7: **Output of the decision functions $f_i(x)$ vs time for the 5-class CS-SVM processing a data stream containing a *Mesoplodon* foraging click.**

This difference in classification performance for the test data files is most likely a reflection of the feature sets chosen. The zero crossing and amplitude features used were very distinctive for the sterotypical *Mesoplodon* foraging clicks, but less distinctive for the other classes. In particular, there was significant overlap in the feature space between the *Mesoplodon* foraging clicks and the *Grampus* clicks. As a result, many *Grampus* clicks were misidentified as foraging clicks. Further analysis of *Grampus* vocalizations and modification of the feature set is recommended. Another issue in classification of the test data was the lack of a "none of the above" designation. The CS-SVM always assigned detected events to one of the 5 classes or to the noise-only class. The addition of a second level of thresholding of the decision functions, such as requiring $max\{f_i(x)\}>L$, would reduce misclassification of unknown signal-present events.

5. ACKNOWLEDGEMENTS

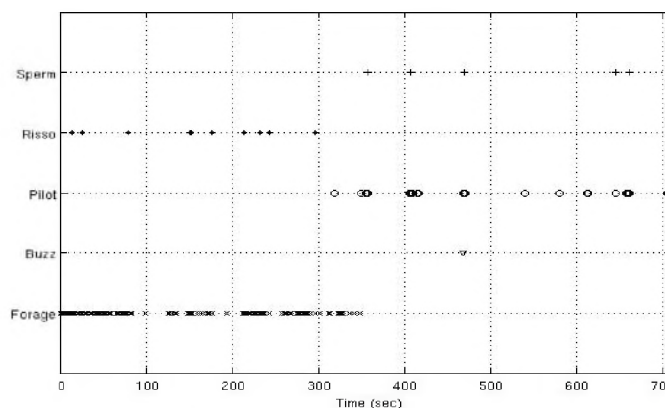***Figure 8***: **Class decisions from the 5-class CS-SVM for 579 click events from Test File 1..**

Test Files	5-class CS-SVM Results
File 1 = Unverified Blainville's Beaked Whale and Short-finned Pilot Whale	*Correctly identified* as Blainville's beaked whale and Short-finned pilot whale
File 2 = Blainville's Beaked Whale	*Correctly identified* as Blainville's beaked whale
File 3 = Pantropical Spotted Dolphin	*Incorrectly classified* as Short-finned pilot whale
File 4 = Pantropical Spotted Dolphin	*Incorrectly classified* as Short-finned pilot whale
File 5 = Risso's Dolphin	*Incorrectly classified* as Blainville's beaked whale
File 6 = Unverified Blainville's Beaked Whale & Sperm Whale	*Correctly identified* as Blainville's beaked whale and sperm whale
File 7 = Risso's Dolphin	*Incorrectly identified* as a mix of Mesoplodon forage & Mesoplodon buzz (only 11.1% of clicks correctly classified as Risso's dolphin).
File 8 =Short-finned pilot whale	*Incorrectly identified* as a mix of Mesoplodon forage, Mesoplodon buzz & Risso's dolphin (only 16% correctly classified as Pilot whale).
File 9 = Sperm Whale	*Correctly classified* as Sperm whale (some confusion with Pilot whale probable caused by 12 KHz HPF)

Table 4: **Results from the 5 class CS-SVM for the Workshop Test Data. The contents of the Test Files were not known prior to the Workshop.**

6. REFERENCES

1. Johnson, M. and Tyack, P.L. (2003), "A digital acoustic recording tag for measuring the response of wild marine mammals to sound", IEEE Journal of Ocean Engineering, vol. 28, pp 3-12.

2. Johnson, M., Madsen, P.T., et al., (2004) , "Beaked whales echolocate on prey", Proceeding of the Royal Society London, B (Suppl.), Vol. 271, S383-386.

3. Moretti, et al, (2005) "Algorithms for passive acoustic detection and localization of marine mammals, including Blainville beaked whales (*Mesoplodon densirostris*), 2nd Int'l Conf. on Detection and Localization of Marine Mammals using Passive Acoustics, 16-18 November, Monoco.

4. Baggenstoss, P. (2003), "The PDF projection theorem and the class-specific method", IEEE Trans. on Signal Processing, vol. 51(3), pp 685.

5. Vapnik, V. N. (1998), *Statistical Learning Theory*, New York, John Wiley & Sons, Inc.

6. Scholkopf, B., Smola, A. (2002), *Learning with Kernels*, Cambridge, MA, MIT Press.

7. Li, Z., Tang, S., and Yan, S., (2002), "Multi-Class SVM Classifier Based on Pairwise Coupling", Proceedings of the 1st Int'l Workshop, SVM 2002, Niagara Falls, Canada, pp. 321.

8. Weston, J. and Watson, C., (1998), "Multi-class Support Vector Machines", Tech. Rep. CSD-TR-98-04, Royal Holloway, Univ. of London, Dept. of Computer Science.

9. Bailey, A. (2001) *Class-dependent features and multicategory classification*, PhD Thesis, University of Southampton.

10. Moretti, et al, (2005), "Passive acoustic detection and tracking of beaked whales in the Tongue of the Ocean", Int'l Conf. on Effects of Sound in the Ocean on Marine Mammals, 2-5 May, Lerici, Italy.

11. Jarvis, et al, (2006), "Automated Classification of Beaked Whalesand Other Small Odontocetes in the Tongue of the Ocean, Bahamas", Proc. of MTS/IEEE Oceans 2006, 18-21 Sep., Boston, MA.

12. 3rd Int'l Workshop on Detection and Classifications of Marine Mammals Using Passive Acoustics, July 24-26, 2007, Boston, MA, http://www.rightwhaleweb.org/workshop/.

13. Johnson, M. et al. (2006) "Foraging Blainville's beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation", Jour. of Exp. Biology, vol. 209.