MARINE MAMMAL SCIENCE, 19(1):20–37 (January 2003)
© 2003 by the Society for Marine Mammalogy

ACOUSTIC IDENTIFICATION OF NINE DELPHINID SPECIES IN THE EASTERN TROPICAL PACIFIC OCEAN

Julie N. Oswald
Scripps Institution of Oceanography,
University of California San Diego,
La Jolla, California 92093, U.S.A.
E-mail: joswald@ucsd.edu

Jay Barlow
National Marine Fisheries Service,
Southwest Fisheries Science Center,
La Jolla, California 92037, U.S.A.

Thomas F. Norris
SAIC Maritime Division,
3990 Old Town Avenue, Suite 105A,
San Diego, California 92110, U.S.A.

Abstract

Acoustic methods may improve the ability to identify cetacean species during shipboard surveys. Whistles were recorded from nine odontocete species in the eastern tropical Pacific to determine how reliably these vocalizations can be classified to species based on simple spectrographic measurements. Twelve variables were measured from each whistle (n = 908). Parametric multivariate discriminant function analysis (DFA) correctly classified 41.1% of whistles to species. Non-parametric classification and regression tree (CART) analysis resulted in 51.4% correct classification. Striped dolphin whistles were most difficult to classify. Whistles of bottlenose dolphins, false killer whales, and pilot whales were most distinctive. Correct classification scores may be improved by adding prior probabilities that reflect species distribution to classification models, by measuring alternative whistle variables, using alternative classification techniques, and by localizing vocalizing dolphins when collecting data for classification models.

Key words: species identification, towed hydrophone array, sonobuoy, discriminant function analysis, decision tree, dolphin, whistle, acoustic, *Stenella longirostris*, *Stenella attenuata*, *Stenella coeruleoalba*, *Delphinus delphis*, *Delphinus capensis*, *Tursiops truncatus*, *Steno bredanensis*, *Globicephala macrorhynchus*, *Pseudorca crassidens*.

Visual detection and identification of cetaceans during shipboard surveys is often constrained by inclement weather, darkness, and animal behavior. Sound propagates long distances in the ocean (Medwin and Clay 1998) and many cetaceans are extremely vocal (Richardson *et al.* 1995). As a result, acoustic techniques can augment visual surveys by providing methods for detection and identification of cetaceans when they are likely to be missed by visual observers. The use of acoustic techniques to complement visual efforts has increased rates and distances of detection for several cetacean species, including: humpback whales (*Megaptera novaeangliae*, Winn *et al.* 1975), sperm whales (*Physeter macrocephalus*, Leaper *et al.* 1992), blue and fin whales (*Balaenoptera musculus* and *B. physalus*, Clark and Fristrup 1997), bowhead whales (*Balaena mysticetus*, Clark and Ellison 2000), striped dolphins (*Stenella coeruleoalba*, Gordon *et al.* 2000), and other delphinids (Thomas *et al.* 1986). While the use of acoustic techniques to detect marine mammals is becoming an increasingly common element of shipboard surveys, acoustic species identification has, until recently, received less attention (Steiner 1981, Potter *et al.* 1994, Schultz and Corkeron 1994, Wang *et al.* 1995, Matthews *et al.* 1999, Rendell *et al.* 1999, Mellinger and Clark 2000).

Using multivariate discriminant function analysis, Steiner (1981) correctly classified the whistles of five western North Atlantic odontocete species 70% of the time. Wang *et al.* (1995) correctly classified 65% of the whistles of seven odontocete species from diverse geographic locations. Rendell *et al.* (1999) correctly classified 55% of the whistles of five odontocete species from several geographic locations. In contrast, Matthews *et al.* (1999) examined the potential for acoustic species recognition using published spectrographic measurements for 10 cetacean species (nine odontocetes and one mysticete) and achieved only 28% correct classification.

To facilitate comparisons among studies, Steiner (1981), Wang *et al.* (1995), Rendell *et al.* (1999), and Matthews *et al.* (1999) reported similar spectrographic measurements. These measurements can be taken quickly and reliably in the field, which is advantageous if the goal is to aid visual observers with real-time species identification. As an alternative approach, Fristrup and Watkins (1993) devised a number of statistical measures to resolve the many acoustic features used to describe sounds. When these measures were taken from the vocalizations of 53 marine mammal species (including mysticetes, odontocetes, and pinnipeds) and linear classification techniques were applied, a correct classification score of 50% was obtained. Fristrup and Watkins (1993) also used tree-based classification models, which classified 66% of vocalizations to the correct species.

Correct classification scores obtained in most whistle classification studies have been significantly greater than would be expected by chance alone, suggesting that differences in whistle structures can be used to identify species. However, in most cases whistles were recorded from only a few different groups of animals. As a result, high correct classification scores could be biased by over-sampling groups or individuals and not controlling for group composition or behavioral variation in call types.

With the exception of Steiner (1981), the aforementioned studies classified the vocalizations of species recorded in widely separated geographic locations. The correct classification scores in these studies may therefore be a function of geographic differences as well as interspecies differences. To determine whether acoustic signals can be useful for species identification during marine mammal surveys, many recordings from a single study area should be classified. In this study two different statistical methods are used to develop classification systems for the tonal whistles of nine odontocete species recorded in the eastern tropical Pacific Ocean (ETP).

METHODS

Data Collection

Acoustic recordings were made from 31 July through 9 December 1998 and from 28 July through 9 December 1999 during a marine mammal survey conducted in the ETP. The study area extended from the United States/Mexico border to the territorial waters of Peru, and from the continental shores of the Americas to the longitude of Hawaii (Fig. 1). Visual line-transect methods were used to survey all cetaceans encountered in the study area.[1]

During the 1998 survey, a hydrophone array was towed during daylight hours at a depth of 4–6 m, approximately 200 m behind the 56-m NSF/UNOLS research vessel *Endeavor* while traveling at a speed of 10 kn. The depth of the array was periodically monitored using a Suunto "Solution Nitrox" dive computer. A three element array (SonaTech Inc., flat frequency response ± 3 dB from 500 Hz to 150 kHz at -163 dB re 1v/μPa after internal amplification) was used for the majority of the survey. A five element array (Innovative Transducers Inc., flat frequency response ± 3 dB from 32 Hz to 25 kHz at -173 dB re 1v/μPa after internal amplification) was used for approximately one month of the survey. An acoustic technician monitored signals from two hydrophones in the array using a stereo headset and custom-written software that displayed real-time spectrograms from a single channel. Signals were high-pass filtered at 500 Hz to 2 kHz to reduce system, ship, and flow noise and were low-pass filtered at 20 kHz to prevent aliasing. Signals of interest were recorded onto digital audio tape (DAT) using Sony TCD-D7 and TCD-D8 DAT recorders (20 Hz to 22 kHz $\pm$ 1 dB).

During the 1999 survey sonobuoys (type 57A) were deployed when dolphins were sighted. These sonobuoys had a flat frequency response from approximately 2 kHz to 20 kHz, and were deployed at a hydrophone depth setting of either 18 or 27 m. Sonobuoy signals were transmitted to a multichannel receiver aboard the research vessel (NOAA ships *McArthur* or *David Starr Jordan*) and were recorded onto DAT using Sony TCD-D7 DAT recorders.

Spectrographic Analysis

Recordings of dolphins that had been visually identified to species by experienced marine mammal observers were digitized (44.1 kHz sample rate, 16 bit precision) using a Pentium III dual-processor personal computer and the commercially available software packages *Spectrogram 4.2.8* (R. S. Horne) and *Cool Edit 96* (Syntrillium Corp.). Only recordings of groups that had been observed to contain a single species were digitized. Because it is possible that some recordings identified as "single species" may contain distant faint vocalizations produced by other species in the area, only "loud and clear" whistles were analyzed. Whistles were considered to be "loud and clear" if they were easily detected aurally and by visual inspection of the spectrogram. Richardson *et al.* (1995) suggest that the maximum detection range for many delphinid species is

[1] Kinzey, D., T. Gerrodette, J. Barlow, A. Dizon, W. Perryman, P. Olson and A. von Saunder. 1999. Marine mammal data collected during a survey in the eastern tropical Pacific Ocean aboard the NOAA ships *McArthur* and *David Starr Jordan* and the UNOLS ship *Endeavor* 31 July–9 December 1998. NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-283. 113 pp.

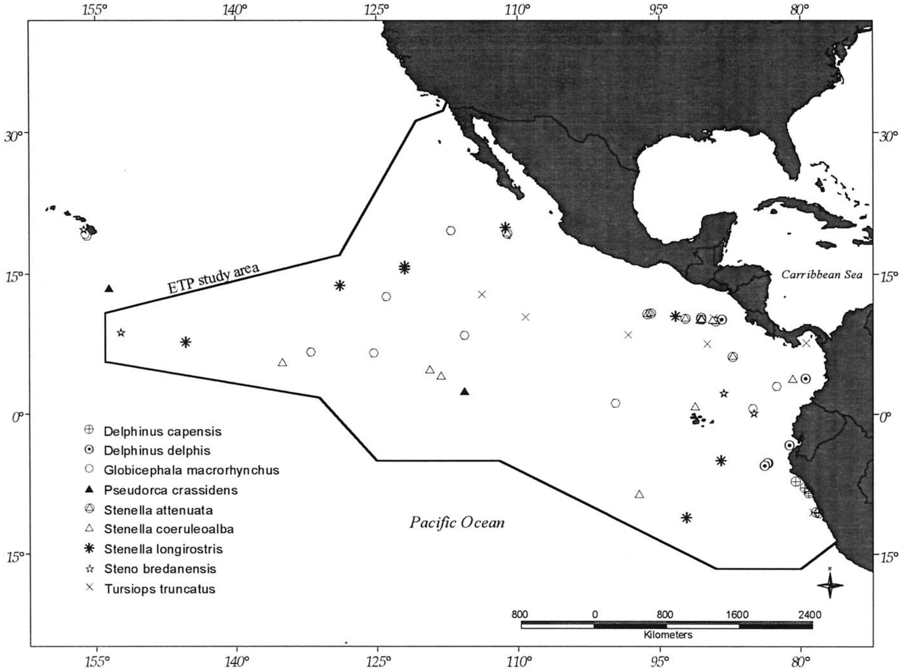

Figure 1. Eastern tropical Pacific study area. Locations of all recordings included in analysis are indicated, with each species represented by a different symbol.

on the order of 1 km (0.54 nmi). To be conservative, recordings made within 3 km (1.62 nmi) of any other sightings were excluded from the analysis. Distance to the next sighting was calculated as the distance between the location of the ship at the beginning of the acoustic recording session and the location of the next group of dolphins when initially sighted (based on angle and reticle measurements read from binoculars). Distance to the previous sighting was calculated as the distance between the location of the ship at the beginning of the acoustic recording session and the location of the previous group of dolphins when last seen.

Spectrograms (512 point FFT, 20 kHz bandwidth) were produced using *Spectrogram 4.2.8* software. Loud and clear tonal whistles that did not overlap extensively with other whistles were randomly chosen for analysis. To avoid oversampling groups or individuals (which can lead to non-independence of data) a maximum of 35 randomly selected whistles were analyzed from each recording session.

Twelve variables were measured from each whistle: (1) beginning frequency (Hz), (2) end frequency (Hz), (3) minimum frequency (Hz), (4) maximum frequency (Hz), (5) duration (msec), (6) slope of the beginning sweep (positive, negative, or zero), (7) slope of the end sweep (positive, negative, or zero), (8) number of inflection points (defined as a change from positive to negative or negative to positive slope), (9) number of steps (defined as a portion of the whistle with zero slope lasting at least 20 msec that separates two portions of similar slope. Similar slope refers to direction, not necessarily magnitude. The angles between the sloped portions and the zero slope portion must lie between 90° and 135°), (10) presence/absence of harmonics (a binary variable), (11) off-scale (a binary variable, indicating whether any portion of the whistle extended beyond the 20 kHz upper limit of the spectrogram), and (12) frequency range (Hz, determined by subtracting minimum frequency from maximum frequency). These variables were chosen because they can be easily measured from a spectrogram and to allow comparisons with results of previous studies.

Statistical Analysis

Multivariate discriminant function analysis (DFA) was used to classify whistles within and among species. Prior to running DFA, continuous variables (frequency variables, duration, and number of steps and inflection points) were tested for normality and were square root or log transformed as necessary. Binary and categorical variables were coded as dummy variables. Frequency variables with values above 20 kHz were assigned a value of 22 kHz. Assigning the same value to all off-scale cases reduced the variability of the data, however omitting these cases resulted in lower overall means and a loss of information regarding which portions of the whistles extended beyond 20 kHz.

Discriminant function analysis classified whistles to prespecified groups based on orthogonal linear functions derived from the measured variables. Some whistles were missing measurements for one or more variables because a portion of the whistle was higher than the maximum recorded frequency of 20 kHz. Whistles that were missing measurements were excluded from the DFA. A series of DFAs was run using the statistical software package *SPSS 7.0* (SPSS Inc.). Within each species, the presence of group-specific whistle patterns was examined by using DFA to predict group membership from whistle characters (where a group is defined as a "recording session" at one time and location). Only recording sessions containing at least three whistles were included in this analysis. Differences among species were examined by using a DFA to predict species from whistle characters.

The jackknife, or cross-validation, method was used to calculate percent correct classification for within-species DFAs. Each whistle was omitted from the total sample and new discriminant functions were calculated for classification of the omitted whistle. A modified jackknife method, omitting entire recording sessions instead of individual whistles, was used to calculate percent correct classification for among-species DFAs. The discriminant functions calculated using this method were created therefore from data independent of the whistles being classified. This helped ensure that whistles were classified based on species-specific characteristics rather than group- or individual-specific characteristics. To evaluate correct classification scores, it is necessary to compare them to what would be expected by chance alone (50% for two species, 33% for three species, 11% for nine species). Chi-square was used to test whether correct classification was significantly greater than expected by chance alone. Statistical significance was evaluated at $\alpha = 0.05$ without corrections for multiple testing.

Tree structured, non-parametric data analysis was performed using CART (Classification And Regression Trees) software (Salford Systems). CART "grows" the largest possible decision tree by separating data into groups (nodes) through a series of binary splits. Each split is based on a value for a single variable, and the criteria used for making splits are known as primary splitting rules. Surrogate splitters are provided at each node. Surrogate splitters closely mimic the action of primary splitting rules and can be used in cases when the primary splitting variable is missing. As a result, all whistles with missing values were included in this analysis. Nodes are labeled based on the number of whistles of each species in the node. "Pure" nodes are nodes that contain the whistles of only one species. Final classification is reached at terminal nodes. When the maximal tree has been grown, CART removes branches and examines the error rates of smaller trees. The smallest tree with the highest predictive accuracy is considered to be the optimal tree. The misclassification rate is estimated using a cross-validation technique similar to the modified jackknife method used in DFA. CART software, however, is not sufficiently flexible to allow the use of recording sessions as the unit for cross-validation. In CART analysis, data are divided into ten roughly equal subsets, each created by random sampling stratified on the dependent variable. These subsets are the units used in cross validation (Breiman *et al.* 1984, Steinberg and Colla 1995). Because classification trees are built using whistles recorded from the same group and possibly the same individual, percent correct classification of the CART analysis is likely to be exaggerated.

Because CART is a non-parametric technique, it was not necessary to assume normality or transform data. For the reasons cited earlier, off-scale variables were assigned a value of 22 kHz before running the analysis. Initially, a decision tree was constructed using all twelve variables; however, a decision tree requiring fewer variables would increase efficiency in the field. A series of trees were constructed using different subsets of the twelve variables in order to find the smallest subset with acceptable predictive accuracy.

Results

A hydrophone array was towed and monitored for approximately 17,980 km (9,702 nmi) and a total of 38 sonobuoys were deployed. Single species recordings were made of nine species including: spinner dolphins (*Stenella longirostris*), striped dolphins (*S. coeruleoalba*), pantropical spotted dolphins (*S. attenuata*), long-beaked common dolphins (*Delphinus capensis*), short-beaked common dolphins (*D. delphis*), rough-toothed dolphins (*Steno bredanensis*), bottlenose dolphins (*Tursiops truncatus*), short-finned pilot whales (*Globicephala macrorhynchus*), and false killer whales (*Pseudorca crassidens*).

A total of 908 whistles recorded in 62 locations were included in the analysis (Table 1, Fig. 1). Recordings from at least two and up to ten different locations were analyzed for each species. Descriptive statistics for the eight continuous whistle variables are presented in Table 2. Number of inflection points and number of steps had the highest coefficients of variation for every species. Of the nine species, short-finned pilot whales and rough-toothed dolphins generally had the highest coefficients of variation for all variables. Whistles of false killer whales have a markedly narrow frequency range and, similar to short-finned pilot whales, relatively few inflection points and steps. In contrast, whistles of pantropical spotted dolphins and bottlenose dolphins contain a relatively large number of steps. Bottlenose dolphins also produce whistles with distinctively long durations and numerous inflection points.

Table 1. Number of recording sessions and whistles analyzed for each species. Different recording sessions separated by time and geographic location.

Species	# Recording sessions	# Whistles
Bottlenose dolphin	7	157
Short-beaked common dolphin	7	88
False killer whale	2	69
Pantropical spotted dolphin	7	97
Long-beaked common dolphin	6	73
Short-finned pilot whale	10	153
Rough-toothed dolphin	5	68
Striped dolphin	10	91
Spinner dolphin	8	112
Total	**62**	**908**

Discriminant Function Analysis

Within-species—The percentage of whistles classified to the correct recording session was significantly greater than expected by chance alone for every species (χ^2 test, $P < 0.05$; Table 3). Correct classification compared to chance alone was particularly high for short-finned pilot whales.

Among-species—Overall, 41.1% of whistles were classified to the correct species. Correct classification scores for individual species ranged from 6.7% for striped dolphins to 66.0% for short-finned pilot whales (Table 4). Only false killer whales, striped dolphins and short-beaked common dolphins had correct classification scores that were not significantly greater than expected by chance alone (false killer whales: $\chi^2{}_8 = 0.0$, $P = 1.0$; striped dolphins: $\chi^2{}_8 = 1.52$, $P = 0.99$; short-beaked common dolphins: $\chi^2{}_8 = 2.75$, $P = 0.95$). An examination of misclassification scores in Table 4 and the plot of group centroids for the first two canonical discriminant functions (Fig. 2) suggests similarities in whistles among several species. For example, striped dolphin whistles were not accurately classified by the DFA, and misclassifications as bottlenose dolphin, short-beaked common dolphin, long-beaked common dolphin, pantropical spotted dolphin, or spinner dolphin were more likely than correct classification. These facts indicate that striped dolphin whistles lie between those five species (as seen on the group centroid plot) and may be more variable than those of the other species.

Classification Trees

Using all 12 variables, the optimal classification tree consisted of 70 terminal nodes and produced an overall correct classification score of 51.4%. In subsequent CART runs, the tree that provided the best trade-off between number of variables and predictive accuracy included seven of the original 12 variables: beginning frequency, end frequency, minimum frequency, maximum frequency, duration, number of inflection points, and number of steps. Using these seven variables resulted in an optimal tree with 66 terminal nodes and a correct classification score of 53.1%. Correct classification scores for individual species ranged from 24.7% for long-beaked common dolphins to 88.4% for false killer whales (Table 5). All correct classification scores were significantly greater than the 11%

expected by chance alone except for long-beaked common dolphins ($\chi^2_8 = 12.4$, $P = 0.13$). Classification errors followed similar patterns to those in DFA.

The four frequency variables (beginning, end, minimum, maximum) were the most important discriminating variables in the seven variable tree, as judged by their performance as both primary and secondary splitters. Number of inflection points was the least important discriminating variable. Note that the importance of a variable pertains only to that variable's performance in the tree in question and cannot necessarily be generalized to the performance of that variable in any other model.

Discussion

Within-species

The percentage of whistles classified to the correct recording session in within-species comparisons was high for every species (Table 3). Our ability, within a species, to correctly associate a whistle with other whistles from the same recording session may indicate geographic variation in whistle patterns; however, it may also be attributable to other sources of variation, such as behavior, group composition, or distinctive individual vocal characteristics. An attempt was made to analyze as many different recording sessions as possible to obtain a representative sample of the vocal repertoire of each species, but behavioral data and group composition were not recorded. It would be valuable to collect such data during future recording sessions in order to determine the relative contributions of social context, geographic separation, and differences among individuals.

Among-species

The results of both DFA and the classification tree suggest that whistles may be useful for the identification of delphinid species during marine mammal surveys. Overall, correct classification of whistles was between 40% and 50% for both types of analyses, much greater than the 11% correct classification expected by chance alone. Whistles of individual species were correctly classified significantly more often than expected by chance alone, with only a few exceptions. At least one of these exceptions is likely due to sample size; the low correct classification score for false killer whales may be due to the fact that there were only two false killer whale recording sessions in the analysis. Thus, when DFA classification functions were created using the modified jackknife method, they were based on one recording session at a time. Using whistles from only one recording session is not likely to allow a complete representation of the whistle repertoire of a species, especially if that species produces whistles containing pod specific characteristics. Future collection of false killer whale whistles in the eastern tropical Pacific will allow an examination of pod- and species-specific characteristics for this species.

Similarity in overall correct classification scores from a parametric statistical method (DFA) and a non-parametric method (CART) supports the use of either technique for species identification. One beneficial feature of CART is that surrogate splitters are available at each node so whistles can be classified even if primary splitting variables are missing. Surrogate splitters closely mimic the actions of primary splitters so there is little, if any, loss in accuracy when surrogate splitters are used (Breiman *et al.* 1984). A classification tree also provides an

Table 2. Means, standard deviations, and coefficients of variation for measured whistle variables. All frequency measurements expressed in kHz and time measurements expressed in seconds.

Species	Beginning frequency	End frequency	Minimum frequency	Maximum frequency	Frequency range	Duration	No. of inflection points	No. of Steps
Bottlenose dolphin								
Mean	11.2	9.0	7.4	17.2	10.0	1.4	3.7	3.1
SD	4.6	3.7	2.2	3.1	3.5	0.7	3.0	4.3
CV	3.3	3.2	2.3	1.4	2.8	4.4	6.5	11.1
Short-beaked common dolphin								
Mean	9.8	11.4	7.4	13.6	6.3	0.8	1.2	1.0
SD	3.9	3.9	2.3	3.4	3.3	0.4	1.3	1.6
CV	4.2	3.6	3.3	2.7	5.6	5.3	11.7	16.9
False killer whale								
Mean	5.2	5.8	4.7	6.1	1.4	0.4	0.5	0.1
SD	2.2	1.5	1.2	1.5	1.3	0.2	0.7	0.3
CV	5.4	2.9	3.2	2.9	10.9	7.3	17.1	43.4
Pantropical spotted dolphin								
Mean	9.5	15.3	8.2	18.7	10.6	0.9	1.9	4.3
SD	2.9	5.2	1.7	3.0	3.3	0.4	1.8	4.5
CV	3.1	3.4	2.1	1.7	3.2	4.5	9.7	10.7
Long-beaked common dolphin								
Mean	10.1	14.1	7.7	15.5	7.9	0.7	1.3	1.5
SD	3.9	4.5	2.2	4.2	4.3	0.4	1.1	2.4
CV	4.5	3.7	3.4	3.2	6.3	7.6	10.0	19.4
Short-finned pilot whale								
Mean	4.4	5.5	3.6	6.1	2.5	0.4	0.7	0.1
SD	3.1	4.3	2.3	4.2	3.2	0.3	0.9	0.4
CV	5.7	6.3	5.0	5.6	10.4	7.5	11.0	25.1

Table 2. Continued.

Species	Beginning frequency	End frequency	Minimum frequency	Maximum frequency	Frequency range	Duration	No. of inflection points	No. of Steps
Rough-toothed dolphin								
Mean	6.8	8.5	6.3	9.1	2.8	0.6	1.3	1.3
SD	2.9	3.1	2.5	3.0	2.1	0.4	2.8	1.6
CV	5.2	4.3	4.9	4.1	8.8	9.1	26.3	15.0
Striped dolphin								
Mean	10.2	12.0	8.1	14.8	6.8	0.8	1.9	2.0
SD	3.7	2.8	1.6	3.5	3.7	0.3	2.1	2.5
CV	3.8	2.4	2.1	2.6	5.8	3.8	11.2	13.3
Spinner dolphin								
Mean	10.4	12.4	9.1	13.7	4.6	0.6	1.9	0.7
SD	3.4	3.6	2.5	3.5	3.4	0.4	4.1	1.4
CV	3.1	2.7	2.5	2.4	7.1	6.7	20.8	20.2

Table 3. Results of within-species discriminant function analysis (DFA). Only recording sessions containing at least three whistles included in the analysis. Fourth column lists percent of whistles classified to correct recording session in within-species DFAs. Column labelled "chance" lists correct classification scores that would be expected by chance alone. Correct classification was significantly greater than expected by chance alone for every species (χ^2 test, $P < 0.05$).

Species	# Recording sessions	# Whistles	% Correct classification	Chance (%)
Bottlenose dolphin	7	151	36.4	14.3
Short-beaked common dolphin	7	88	47.7	14.3
False killer whale	2	68	91.2	50.0
Pantropical spotted dolphin	5	81	37.5	20.0
Long-beaked common dolphin	5	64	40.9	20.0
Short-finned pilot whale	10	149	41.6	10.0
Rough-toothed dolphin	4	64	64.2	25.0
Striped dolphin	8	87	29.9	12.5
Spinner dolphin	6	107	45.8	16.7
Total	54	859		

intuitive diagrammatic representation of the classification process. It displays patterns in the data that may not be apparent using techniques such as DFA. A disadvantage to using CART is that the software is not flexible enough to allow the use of recording sessions as the unit for cross-validation. As a result, percent correct classification of the CART analysis is likely to be exaggerated.

Based on the seven variable classification tree and the 12-variable DFA, false killer whales, pilot whales, and bottlenose dolphins have the most distinctive whistles. These three species lie apart from the others on the plot of group centroids (Fig. 2), and have a small number of relatively pure terminal nodes in the decision tree (Fig. 3), resulting in high correct classification scores (Table 5). The species with the lowest correct classification scores (short-beaked common, long-beaked common, and spinner dolphins) cluster together on the plot of group centroids (Fig. 2), and have many terminal nodes that are generally not very pure.

Although our results show that dolphin whistles contain species-specific information, our correct classification scores are much lower than the usual standards applied to visual identification (*i.e.*, near certainty). Additional research is needed before whistle classification can be used routinely as a field identification tool. We note, however, that the task of classifying species from a single whistle is a difficult challenge. It might be analogous to asking a visual observer to determine species from a single random surfacing of a single individual. It may prove to be an easier task to determine species from the collection of all whistles recorded during a recording session.

A potential method for increasing the probability of correctly identifying whistles in the field is the use of classification models that take species distribution into account. In the current DFA and CART models, each whistle was assigned to species without considering whether that species is common, rare, or even absent in the specific area where the whistle was recorded. Some species are more common in the study area than others and their distributions are not uniform across these waters. Long-beaked common dolphins were seen only in coastal waters during the 1998 survey, while short-beaked common dolphins ranged

Table 4. Results of among-species discriminant function analyses (DFA) (overall correct classification = 41.4%, n = 869). Numbers in parentheses are Chi-square P-values testing whether correct classification is greater than expected by chance. Bold-face numbers are percent correct classification scores; others are percentages of whistles classified incorrectly.

	Classified as								
Actual Species	Bottlenose dolphin	Short-beaked common dolphin	False killer whale	Pantropical spotted dolphin	Long-beaked common dolphin	Short-finned pilot whale	Rough-toothed dolphin	Striped dolphin	Spinner dolphin
Bottlenose dolphin	**64.2** (<0.05)	7.9	2.6	10.6	6.6	0.0	0.7	0.7	6.6
Short-beaked common dolphin	14.8	**17.0** (<0.95)*	6.8	9.1	18.2	1.1	6.8	6.8	19.3
False killer whale	1.5	2.9	**17.6** (1.0)*	0.0	0.0	25.0	50.0	1.5	1.5
Pantropical spotted dolphin	23.2	1.2	0.0	**50.0** (<0.05)	15.9	0.0	3.7	2.4	3.7
Long-beaked common dolphin	4.5	18.2	6.1	12.1	**30.3** (<0.05)	0.0	6.1	6.1	16.7
Short-finned pilot whale	1.3	2.0	19.3	0.0	2.7	**66.0** (<0.05)	4.7	0.0	4.0
Rough-toothed dolphin	1.5	3.0	22.4	1.5	1.5	10.4	**35.8** (<0.05)	3.0	20.9
Striped dolphin	15.7	14.6	1.1	18.0	16.9	0.0	7.9	**6.7** (<0.99)*	19.1
Spinner dolphin	6.5	8.3	6.5	0.9	13.0	0.0	13.0	11.1	**40.7** (<0.05)

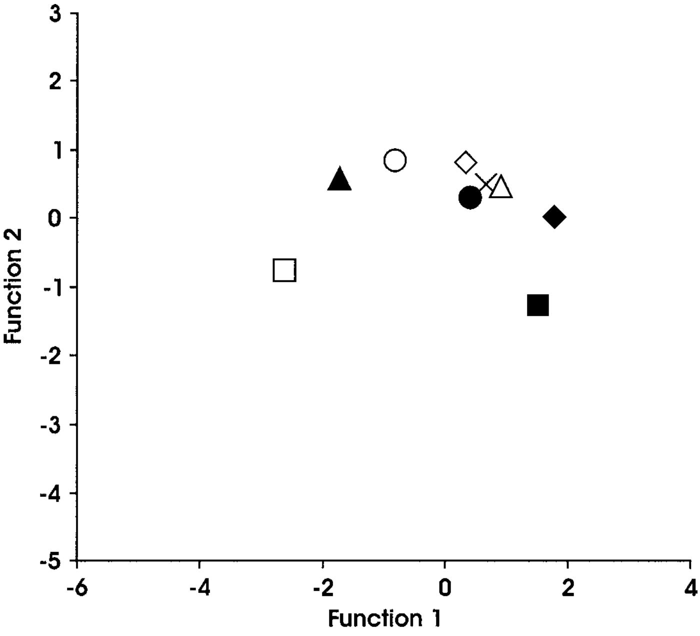

Figure 2. Plot of group centroids for first two canonical discriminant functions in nine-species comparison. × Long-beaked common dolphin (*Delphinus capensis*), ■ bottlenose dolphin (*Tursiops truncatus*), ● short-beaked common dolphin (*Delphinus delphis*), ▲ false killer whale (*Pseudorca crassidens*), ◆ pantropical spotted dolphin (*Stenella attenuata*), △ striped dolphin (*S. coeruleoalba*), ◇ spinner dolphin (*S. longirostris*), ○ rough-toothed dolphin (*Steno bredanensis*), □ short-finned pilot whale (*Globicephala macrorhynchus*).

much farther offshore.[1] Wade and Gerrodette (1993) observed that pantropical spotted and spinner dolphins were most abundant in the warm tropical waters of the eastern tropical Pacific, short-beaked common dolphins were most abundant in cold upwelling-modified waters, and striped dolphins were most abundant where the other three species were not. To take species distribution into account, the study area should be divided into strata and classification models built using prior probabilities based on sighting frequencies in each stratum.

Lower than desired correct classification scores may also be a result of the variables measured. The twelve variables used in this study were chosen due to their compatibility with previous work, allowing for comparisons among studies. They are variables that can be measured relatively easily and reliably in the field and do not require extensive training of operators. These variables, however, do not provide a complete representation of dolphin whistles. Additionally, it is difficult to make biological interpretations based on these variables, as they are simply a representation of the way humans perceive whistles and may not reflect whistle characters actually utilized by dolphins. Measuring additional or alternative variables (such as frequency at intervals along a whistle) may provide a more accurate representation of whistles and lead to higher correct classification scores.

The fact that the variables in this study are measured by human operators reduces the need for special programs or hardware; however, it introduces an element of subjectivity to the measurements. It can also create a bottleneck when

Table 5. Results of 66 terminal node classification tree grown using seven variables (beginning frequency, end frequency, minimum and maximum frequency, duration, number of inflection points, number of steps). Overall correct classification = 53.1%, n = 908. Bold-face numbers are percent correct classification scores; others are percentages of whistles classified incorrectly. Numbers in parentheses are Chi-square P-values testing whether correct classification is greater than expected by chance.

	Classified as								
Actual species	Bottlenose dolphin	Short-beaked common dolphin	False killer whale	Pantropical spotted dolphin	Long-beaked common dolphin	Short-finned pilot whale	Rough-toothed dolphin	Striped dolphin	Spinner dolphin
Bottlenose dolphin	**60.3 (<0.05)**	7.7	0.6	7.1	7.1	0.6	1.3	11.5	3.8
Short-beaked common dolphin	12.5	**28.4 (<0.05)**	5.7	5.7	10.2	2.3	8.0	15.9	11.4
False killer whale	0.0	1.4	**88.4 (<0.05)**	0.0	0.0	4.3	2.9	1.4	1.4
Pantropical spotted dolphin	10.3	9.3	0.0	**48.5 (<0.05)**	12.4	0.0	2.1	12.4	5.2
Long-beaked common dolphin	5.5	5.5	4.1	19.2	**24.7 (<0.2)***	0.0	9.6	20.5	11.0
Short-finned pilot whale	2.0	2.6	11.8	1.3	0.7	**68.0 (<0.05)**	7.2	3.3	3.3
Rough-toothed dolphin	2.9	5.9	16.2	0.0	7.4	11.8	**45.6 (<0.05)**	4.4	5.9
Striped dolphin	2.2	14.3	1.1	15.4	4.4	1.1	7.7	**40.7 (<0.05)**	13.2
Spinner dolphin	7.1	11.6	6.3	8.0	10.7	3.6	7.1	14.3	**31.3 (<0.05)**

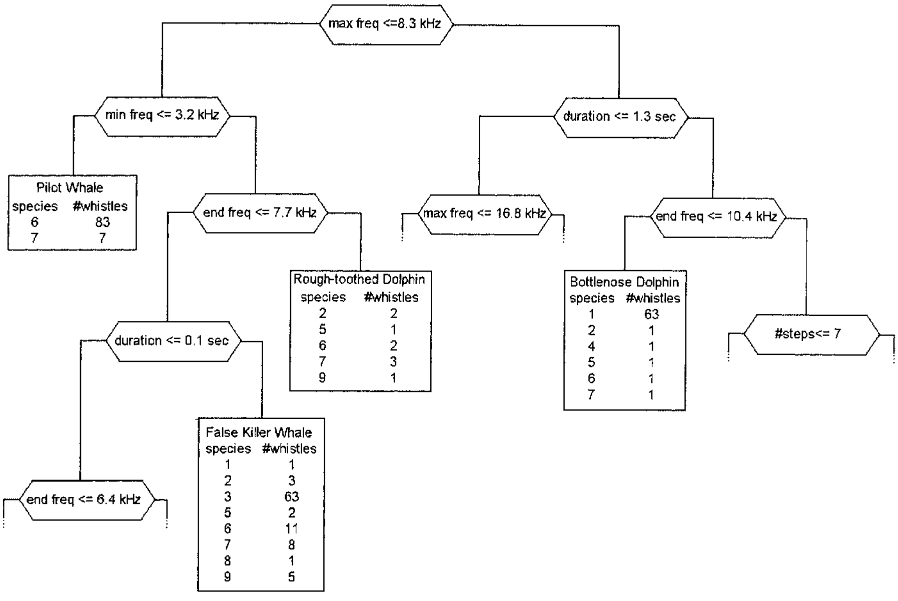

Figure 3. Seven variable classification tree constructed using CART software. For brevity, only initial portion of 66 terminal node tree presented. Squares represent terminal nodes and labeled based on species with greatest number of whistles in that node. Species designation: 1 = bottlenose dolphin (*Tursiops truncatus*), 2 = short-beaked common dolphin (*Delphinus delphis*), 3 = false killer whale (*Pseudorca crassidens*), 4 = pantropical spotted dolphin (*Stenella attenuata*), 5 = long-beaked common dolphin (*D. capensis*), 6 = short-finned pilot whale (*Globicephala macrorhynchus*), 7 = rough-toothed dolphin (*Steno bredanensis*), 8 = striped dolphin (*Stenella coeruleoalba*), 9 = spinner dolphin (*S. longirostris*).

there are large volumes of data to analyze, and may make the measurement of additional or alternative variables difficult. An automated feature extraction system could be implemented in order to reduce subjectivity and make the measurement of additional variables more feasible.

The use of alternative classification methods, such as artificial neural networks, may be another way to increase the accuracy of whistle classification. Artificial neural networks operate in a non-linear, self-organizing way and therefore may be able to detect differences among species that would be missed by other statistical methods (Deecke *et al.* 1999). Artificial neural networks have been successfully utilized to recognize the calls of bowhead whales (Potter *et al.* 1994) and to measure the similarity of discrete calls of killer whales (Deecke *et al.* 1999).

Another consideration that must be taken into account before the classification system can be used in the field is that it currently includes only 9 of the 16 delphinid species encountered in the ETP.[2] Adding the missing species (Risso's

[2] Kinzey, D., T. Gerrodette, J. Barlow, A. Dizon, W. Perryman and P. Olson. 2000. Marine mammal data collected during a survey in the eastern tropical Pacific ocean aboard the NOAA ships *McArthur* and *David Starr Jordan* 28 July–9 December 1999. NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-293. 89 pp.

dolphins, *Grampus griseus*; killer whales, *Orcinus orca;* pygmy killer whales, *Feresa attenuata*; dusky dolphins, *Lagenorhynchus obscurus*; Pacific white-sided dolphins, *L. obliquidens*; Fraser's dolphins, *L. hosei*; and melon-headed whales, *Peponocephala electra*) will make the system complete and ensure that every whistle has a chance of being correctly classified. It is important to note, however, that adding species is likely to decrease correct classification because the structure of the DFA and classification tree will change as variable space becomes more crowded.

Not every school encountered is a single species school. During the 1998 and 1999 surveys, 11% and 12% of all sightings were mixed species schools.[1,2] Mixed species schools present a challenge because it is difficult to determine whether whistles have been classified as multiple species due to classification errors or due to the actual presence of multiple species in the group being recorded. Knowledge of which species commonly associate with each other will help with these decisions. For example, mixed schools composed of spinner and spotted dolphins were the most commonly sighted mixed species schools during both the 1998 and 1999 surveys (30% and 43% of the mixed species schools, respectively).[1, 2] If whistles are being classified as spinner dolphins and spotted dolphins consistently during a sighting, it is likely to be a mixed school. Whistles from known mixed species schools should be run through the classification system and confusion matrices for these schools compared to confusion matrices for single species schools. Perhaps patterns exist that would aid in discerning actual mixed species schools from classification errors.

There are two additional issues that must be addressed when developing a classification system based on whistles recorded at sea. The first is the statistical assumption of independent data. Using a towed array, it is currently not possible to precisely locate individual animals that are being recorded. Therefore, it is not possible to ensure that each whistle included in the analysis is produced by a different individual. We attempted to avoid over-sampling groups or individuals by randomly selecting a small subsample of whistles from each recording session, and by analyzing as many different recording sessions as possible for each species.

The second obstacle inherent to recording animals at sea is ensuring that each recording session included in the analysis contains only whistles produced by a single species. If a group is detected both acoustically and visually, it can usually be identified as a single species school by experienced marine mammal observers, but whistles of other species present in the area may also be detected by the array. Recent observations suggest that whistles can be heard at distances much greater than 3 km (1.6 nmi) (Janik 2000), and hence, it is possible that the recordings used in our analysis may include vocalizations produced by species other than those seen by the visual observers.

The ability to localize dolphins detected using a towed hydrophone array could aid in the resolution of both issues. Differences in the arrival times of sperm whale clicks at two hydrophones in a towed array have been used to estimate bearing angles to vocalizing animals in order to track them during dives (Leaper *et al.* 1992). Miller and Tyack (1998) used frequency domain beamforming techniques to localize individual killer whales detected using a small towed array. Thode *et al.* (2000) obtained bearing angles to whistling dolphins using a three-element towed array and frequency domain beamforming techniques. These bearings were not precise enough to allow the identification of individual animals. Beamforming techniques may, however, be used to reduce over-sampling individuals. Whistles originating from widely spaced bearing angles at similar times are likely to have been produced by different individuals. Including such whistles in the analysis would ensure that a wider cross-section of the school is sampled. Similarly, determining the location of vocalizing dolphins makes it possible to discern whether whistles are being produced by the school seen and identified by visual observers or by some other school in the area. This will reduce the chance of mislabeling recordings and should result in a more accurate classification system. Localization techniques are currently being developed and tested for use during future acoustic surveys.

Acknowledgments

We would like to extend our thanks to Karen Fear, Mark McDonald, Laura Morse, Shannon Rankin, Aaron Thode, and Christine Vitulli for their many hours spent at sea recording whistles. We gratefully acknowledge the patience and cooperation of the scientists and crew aboard the NSF/UNOLS R/V *Endeavor* and the NOAA ships *McArthur* and *David Starr Jordan.* This project could not have been completed without the aid of skilled visual observers: David Au, Lisa Ballance, Lisa Baraff, Isabel Beasley, Susan Chivers, James Cotton, Jorge Del Angel, Meghan Donahue, Mike Force, Karin Forney, Gary Friedrichsen, Chris Hoefer, Kathy Hough, Brett Jarrett, Doug Kinzey, Greg Krutzikowsky, Rick LeDuc, Mark Lowry, John Mason, Stephanie Norman, Cornelia Oedekoven, Paula Olson, Robert Pitman, Richard Rowlett, Cheryl Ryder, Juan Carlos Salinas, Larry Spear, Barbara Taylor, Koen Van Waerebeek, Ernesto Vázquez, Suzanne Yin, and Elizabeth Zúñiga. Many thanks are due to Theresa Weber for her assistance in data analysis. The assistance of Christine Loftus is also greatly appreciated. We are grateful to Steve Dawson, Jennifer Marsh, Justin Matthews, Michael Oswald, Kate Stafford, and one anonymous reviewer for their insightful and helpful suggestions on drafts of this manuscript.

Literature Cited

Breiman, L., J. Friedman, R. Olshen and C. Stone. 1984. Classification and regression trees. Wadsworth, Pacific Grove, CA.

Clark, C. W., and W. T. Ellison. 2000. Calibration and comparison of the acoustic location methods used during the spring migration of the bowhead whale, *Balaena mysticetus*, off Pt. Barrow, Alaska, 1984–1993. Journal of the Acoustical Society of America 107:3509–3517.

Clark, C. W., and K. M. Fristrup. 1997. Whales '95: A combined visual and acoustic survey of blue and fin whales off Southern California. Report of the International Whaling Commission 47:583–600.

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. Journal of the Acoustical Society of America 105:2499–2507.

Fristrup, K. M., and W. A. Watkins. 1993. Marine animal sound classification. Woods Hole Oceanographic Institution Technical Report WHOI-94-13. 29 pp.

Gordon, J. C. D., J. N. Matthews, S. Panigada, A. Gannier, J. F. Borsani and G. Notarbartolo Di Sciara. 2000. Distribution and relative abundance of striped dolphins in the Ligurian Sea Cetacean Sanctuary: Results from an acoustic collaboration. Journal of Cetacean Research and Management 2:27–36.

Janik, V. M. 2000. Source levels and the estimated active space of bottlenose dolphin (*Tursiops truncatus*) whistles in the Moray Firth, Scotland. Journal of Comparative Physiology 186:673–680.

LEAPER, R., O. CHAPPELL AND J. GORDON. 1992. The development of practical techniques for surveying sperm whale populations acoustically. Report of the International Whaling Commission 42:549–560.

MATTHEWS, J. N., L. E. RENDELL, J. C. D. GORDON AND D. W. MACDONALD. 1999. A review of frequency and time parameters of cetacean tonal calls. Bioacoustics 10:47–71.

MEDWIN, H., AND C. S. CLAY. 1998. Fundamentals of acoustical oceanography. Academic Press, San Diego, CA.

MELLINGER, D., AND C. CLARK. 2000. Recognizing transient low-frequency whale sounds by spectrogram correlation. Journal of the Acoustical Society of America 107:3518–3529.

MILLER, P. J., AND P. L. TYACK. 1998. A small towed beamforming array to identify vocalizing resident killer whales (*Orcinus orca*) concurrent with focal behavioral observations. Deep-sea Research 45:1389–1405.

POTTER, J. R., D. K. MELLINGER AND C. W. CLARK. 1994. Marine mammal call discrimination using artificial neural networks. Journal of the Acoustical Society of America 96:1255–1262.

RENDELL, L. E., J. N. MATTHEWS, A. GILL, J. C. D. GORDON AND D. W. MACDONALD. 1999. Quantitative analysis of tonal calls from five odontocete species, examining interspecific and intraspecific variation. Journal of Zoology 249:403–410.

RICHARDSON, W. J., C. R. GREEN, C. I. MALME AND D. H. THOMSON. 1995. Marine mammals and noise. Academic Press, San Diego, CA.

SCHULTZ, K. W., AND P. J. CORKERON. 1994. Interspecific differences in whistles produced by inshore dolphins in Moreton Bay, Queensland, Australia. Canadian Journal of Zoology 72:1061–1068.

STEINBERG, D., AND P. COLLA. 1995. CART: Tree-structured non-parametric data analysis. Salford Systems, San Diego, CA.

STEINER, W. W. 1981. Species-specific differences in pure tonal whistle vocalizations of five western north Atlantic dolphin species. Behavioral Ecology and Sociobiology 9:241–246.

THODE, A., T. NORRIS AND J. BARLOW. 2000. Frequency beamforming of dolphin whistles using a sparse three-element towed array. Journal of the Acoustical Society of America 107:3581–3584.

THOMAS, J. A., S. A. FISHER AND L. M. FERM. 1986. Acoustic detection of cetaceans using a towed array of hydrophones. Report of the International Whaling Commission (Special Issue 8):139–148.

WADE, P. R., AND T. GERRODETTE. 1993. Estimates of cetacean abundance and distribution in the eastern tropical Pacific. Report of the International Whaling Commission 43:477–493.

WANG, D., B. WURSIG AND W. EVANS. 1995. Comparisons of whistles among seven odontocete species. Pages 299–323 *in* R. A. Kastelein, J. A. Thomas and P. E. Nachtigall, eds. Sensory Systems of aquatic mammals. De Spil Publishers, Woerden, The Netherlands.

WINN, H. E., R. K. EDEL AND A. G. TARUSKI. 1975. Population estimation of the humpback whale (*Megaptera novaeangliae*) in the West Indies by visual and acoustic techniques. Journal of the Fisheries Research Board of Canada 32:499–506.

Received: 16 July 2001
Accepted: 16 May 2002