A tool for real-time acoustic species identification of delphinid whistles

Julie N. Oswald[a)]
Scripps Institution of Oceanography, University of California—San Diego, San Diego, California 92038

Shannon Rankin and Jay Barlow
NOAA-NMFS Southwest Fisheries Science Center, La Jolla Shores Drive, La Jolla, California 92037

Marc O. Lammers
Hawaii Institute of Marine Biology, Kailua, Hawaii 96834

(Received 10 October 2006; revised 1 May 2007; accepted 1 May 2007)

The ability to identify delphinid vocalizations to species in real-time would be an asset during shipboard surveys. An automated system, Real-time Odontocete Call Classification Algorithm (ROCCA), is being developed to allow real-time acoustic species identification in the field. This Matlab-based tool automatically extracts ten variables (beginning, end, minimum and maximum frequencies, duration, slope of the beginning and end sweep, number of inflection points, number of steps, and presence/absence of harmonics) from whistles selected from a real-time scrolling spectrograph (ISHMAEL). It uses classification and regression tree analysis (CART) and discriminant function analysis (DFA) to identify whistles to species. Schools are classified based on running tallies of individual whistle classifications. Overall, 46% of schools were correctly classified for seven species and one genus (*Tursiops truncatus, Stenella attenuata, S. longirostris, S. coeruleoalba, Steno bredanensis, Delphinus* species, *Pseudorca crassidens*, and *Globicephala macrorhynchus*), with correct classification as high as 80% for some species. If classification success can be increased, this tool will provide a method for identifying schools that are difficult to approach and observe, will allow species distribution data to be collected when visual efforts are compromised, and will reduce the time necessary for post-cruise data analysis. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2743157]

PACS number(s): 43.80.Ka [WWA] Pages: 587–595

I. INTRODUCTION

Acoustic techniques have been used to monitor a variety of species, ranging from birds (Mills, 1995; Chesmore, 2001), bats (Vaughan *et al.*, 1997; Parsons and Jones, 2000), and fallow deer (Reby *et al.*, 1997) to insects such as crickets and grasshoppers (Chesmore, 2001). Increasingly, acoustic techniques are being used to monitor marine mammal populations (Leaper *et al.*, 1992; Clark *et al.*, 1996; Stafford *et al.*, 2001; Gordon *et al.*, 2000; van Parijs *et al.*, 2002; Wang *et al.*, 2005). Many marine mammals produce distinctive sounds and therefore acoustic techniques can be used to detect not only their presence, but also species identity. Several species of large whales produce calls that are easily recognized (Thompson *et al.*, 1992; Goold and Jones, 1995; Thompson *et al.*, 1996), but the calls produced by many delphinid species are highly variable and overlap in frequency characteristics, making them more challenging to identify to species.

The calls produced by delphinids have generally been placed into three distinct categories: echolocation clicks, burst pulse sounds, and whistles. Echolocation clicks are short, broadband pulses with peak frequencies that vary from tens of kilohertz to well over 100 kHz (Norris and Evans, 1966; Au, 1980). Echolocation clicks generally occur in trains containing few to hundreds of clicks and are used for navigation and object detection and discrimination (Au, 1993). Burst pulse signals are broadband click trains with very short interclick intervals. These sounds take on a tonal quality to human ears because the clicks are repeated at such high rates that the rate itself, rather than the individual clicks, is audible (Watkins, 1967; Herzing, 2000). It is thought that these signals play a role in social interactions, although they may also function in echolocation tasks. Whistles are continuous, narrow band, frequency modulated signals that often have harmonic components. They range in duration from several tenths of a second to several seconds (Tyack and Clark, 2000). The fundamental frequency of most whistles ranges from 2 to 30 kHz (Lammers *et al.*, 2003; Oswald *et al.*, 2004). Whistles are believed to function as social signals (Janik and Slater, 1998; Herzing, 2000; Lammers *et al.*, 2003).

Delphinid species identification studies have commonly focused on whistle characteristics to develop classification algorithms (Steiner, 1981; Wang *et al.*, 1995; Matthews *et al.*, 1999; Rendell *et al.*, 1999; Oswald *et al.*, 2003). Correct classification scores obtained in these studies are generally significantly greater than expected by chance alone, suggesting that whistles contain information that could be used to identify delphinid species. Roch *et al.*, (2007) included both whistles and clicks in their species identification algorithms.

[a)]Electronic mail: oswald.jn@gmail.com

Their correct classification scores of 67% to 75% for four species suggest that clicks may also be useful for species identification, however, their method does not distinguish the relative contribution of whistles and clicks to the classifier. In most areas, such as the eastern tropical Pacific Ocean (ETP), whistle sounds propagate much farther than echolocation clicks, so whistles are likely to be more useful for long-distance species identification. For example, whistles were the only sounds detected from 60% of 1867 schools detected acoustically during five visual and acoustic surveys in the ETP (unpublished data). Therefore, species identification algorithms based on whistles are likely to be more useful than those that depend on clicks.

Species identification algorithms are developed by post-processing of field recordings. While postprocessing provides valuable information, the ability to identify vocalizations to species in real time would be a great asset during shipboard marine mammal abundance surveys. Traditionally during these surveys, a team of visual observers searches for marine mammals and then directs the ship towards them for school size estimation and species identification (Wade and Gerrodette, 1993; Barlow, 1995). Recently, methods have been developed to tow a hydrophone array behind research vessels during standard visual cetacean surveys (Fristrup and Clark, 1997; Barlow and Taylor, 2005). While the addition of acoustic techniques has been shown to increase rates and distances of detection (Leaper *et al.*, 1992, Clark and Fristrup, 1997; Gordon *et al.*, 2000; Barlow and Taylor, 2005), real-time acoustic species identification would provide several further advantages. This capability would allow the acoustic team to aid visual observers with the identification of groups that are difficult to approach and observe due to factors such as animal behavior, inclement weather, and reduced visibility. In addition, because marine mammals spend much of their lives under water, schools are often detected acoustically but not visually (Barlow and Taylor, 2005). The ability to identify these detections would allow the use of acoustic methods when conditions do not allow for visual observations. Finally, real-time acoustic species identification can reduce the bottleneck of post-cruise data analysis.

Until now, no methods have been available for real time acoustic identification of delphinid whistles. In this paper we present a new software tool that is being developed for this task: Real-time Odontocete Call Classification Algorithm (ROCCA). ROCCA is a Matlab-based tool that extracts, measures, and classifies whistles to species in real-time.

II. METHODOLOGY

A. Data collection

Acoustic recordings were made during six shipboard marine mammal abundance surveys conducted by the Southwest Fisheries Science Center (NOAA, NMFS). Each 4-month survey occurred between the months of July and December. Four of the surveys took place in the ETP: *Stenella* Population and Abundance Monitoring (SPAM) 1998, and *Stenella* Abundance Research (STAR) 1999, 2000, and 2003. This study area extended from the United States/Mexico border southward to the territorial waters of Peru,

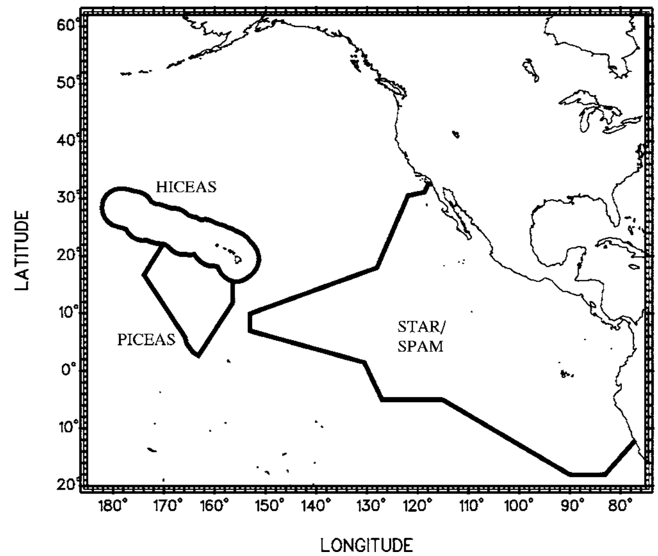

FIG. 1. Pacific Ocean study area boundaries for the HICEAS 2002, PICEAS 2005, SPAM 1998, and STAR 1999, 2000, and 2003 Southwest Fisheries Science Center (NOAA, NMFS) marine mammal abundance surveys.

and from the continental shores of the Americas west to the longitude of Hawaii (Fig. 1). The Hawaiian Islands Cetacean and Ecosystem Assessment Survey (HICEAS 2002) study area covered the waters within the 200 nmi Exclusive Economic Zone (EEZ) around the Hawaiian Island chain from the island of Hawaii in the southeast to Kure Atoll in the northwest (Fig. 1). The Pacific Islands Cetacean Ecosystem Assessment Survey (PICEAS 2005) took place in the US EEZ waters of Palmyra and Johnston Atoll and adjacent waters south of Hawaii (Fig. 1).

During all surveys, a team of three experienced visual observers actively searched for marine mammals using two 25×150 binoculars and by naked eye (Kinzey *et al.*, 2001). When cetaceans were sighted, they were approached for species identification and school size estimation. Cetacean vocalizations were monitored and recorded using a towed hydrophone array and Type SSQ-57 sonobuoys. The array was towed 200 m behind the research vessel during daylight hours at a depth of approximately 4–6 m. Table I gives the frequency response characteristics of the arrays used during the surveys. During the SPAM 1998 survey, signals from the array were recorded onto digital audio tape (DAT) using Sony TCD-D7 and TCD-D8 DAT recorders (48 kHz sampling rate). During the STAR 2000 survey and all subsequent surveys, signals from the array were sent through a Mackie CR1604-VLZ mixer for equalization and were recorded directly to computer hard drive via an analog-to-digital conversion card (National Instruments BNC-2110 and DAQCard-6062E) using a 200-kHz sampling rate.

An acoustic technician monitored signals from two hydrophones in the array using a stereo headset and real-time scrolling spectrographic software (ISHMAEL, Mellinger, 2001). Whaltrak, a mapping program with a GPS interface, automatically logged time and position every 5 min while the array was being monitored. During the STAR 2000 survey and all subsequent surveys, acoustic detections were localized using ISHMAEL and Whaltrak. Bearing angles were

TABLE I. Frequency response and gain characteristics of hydrophone arrays used during Southwest Fisheries Science Center (NOAA, NMFS) marine mammal abundance surveys. Two arrays were used during the SPAM 1998 and STAR 2000 surveys. The array used during the PICEAS 2005 survey had four elements, three relatively narrow band and one relatively broadband.

Survey	Array frequency response	No. of hydrophone elements
SPAM 1998	500 Hz to 150 kHz±3 dB at −163 dB *re* 1 V/mPa	3
	32 Hz to 25 kHz±3 dB at −173 dB *re* 1 V/mPa	5
STAR 2000	2–45 kHz±4 dB at −132 dB *re* 1 V/mPa	5
	2–120 kHz±3 dB at −164 dB *re* 1 V/mPa	3
HICEAS 2002	500 Hz to 30 kHz±5 dB at −155 dB *re* 1 V/mPa	2
STAR 2003	500 Hz to 30 kHz±5 dB at −155 dB *re* 1 V/mPa	3
PICEAS 2005	1–40 kHz±5 dB at −150 dB *re* 1 V/mPa	3
	2–150 kHz±2 dB at −166 dB *re* 1 V/mPa	1

determined using phone-pair cross-correlation algorithms in ISHMAEL, and distance was determined by examining the convergence of bearing angles plotted on Whaltrak. Comparisons of the angle and distance to the acoustic detection with the location of the sighting allowed confirmation that the vocalizations detected were produced by the sighted dolphins.

A hydrophone array was not towed during the STAR 1999 survey. Instead, U.S. Navy sonobuoys (type SSQ-57) were deployed in close proximity to dolphin sightings. These sonobuoys had a flat frequency response from approximately 2 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 and were recorded onto DAT using Sony TCD-D7 DAT recorders.

Single-species acoustic recordings were obtained from nine delphinid species during the six surveys: bottlenose dolphins (*Tursiops truncatus*), spotted dolphins (*Stenella attenuata*), spinner dolphins (*S. longirostris*), striped dolphins (*S. coeruleoalba*) rough-toothed dolphins (*Steno bredanensis*), short-beaked common (*Delphinus delphis*), long-beaked common dolphins (*D. capensis*), false killer whales (*Pseudorca crassidens*), and short-finned pilot whales (*Globicephala macrorhynchus*). A total of 2606 whistles from 176 schools were included in the analysis. Table II lists the number of whistles analyzed for each species.

B. Spectrographic analysis

Recordings of schools that had been visually identified to species and confirmed to contain only one species were chosen for analysis. Recordings were included only if the school was sighted at least 3 nmi from any other school in the area. This helped to ensure that the whistles analyzed were produced by the school being observed and not another nearby school. This was especially important during the SPAM 1998 and STAR 1999 surveys, when acoustic localization techniques were not available.

We randomly selected 50% of loud and clear whistles

TABLE II. Total number of whistles analyzed for each species, number of schools that those whistles were recorded from, and means and standard deviations (in parentheses) for continuous variables measured from the whistles. Frequency variables are given in kHz, duration is given in seconds, and number of inflection points (IP) and number of steps are count data.

Species	No. of whistles	No. of schools	Beginning frequency	Ending frequency	Minimum frequency	Maximum frequency	Duration	No. of IP	No. of steps
Bottlenose dolphin	306	15	11.61 (5.11)	10.24 (4.78)	7.92 (2.49)	17.07 (4.55)	1.11 (0.69)	2.85 (2.67)	2.17 (3.61)
Spotted dolphin	399	26	9.92 (3.94)	14.92 (5.66)	8.41 (2.39)	17.99 (4.69)	0.75 (0.38)	1.29 (1.45)	3.06 (3.84)
Striped dolphin	401	38	10.80 (3.96)	12.01 (3.40)	8.48 (2.21)	14.98 (3.61)	0.69 (0.35)	1.84 (1.82)	2.36 (3.19)
Spinner dolphin	259	19	11.85 (4.42)	12.94 (4.33)	9.99 (3.18)	15.09 (4.57)	0.61 (0.42)	1.89 (3.53)	0.83 (1.64)
Rough-toothed dolphin	192	14	7.41 (3.15)	8.33 (2.95)	6.46 (2.33)	9.53 (2.97)	0.64 (0.36)	2.56 (3.00)	1.51 (1.84)
Short-beaked common dolphin	314	25	11.63 (4.84)	12.18 (4.38)	8.30 (2.69)	15.04 (4.39)	0.70 (0.39)	1.64 (1.87)	1.76 (2.31)
Long-beaked common dolphin	174	11	10.87 (4.89)	14.46 (5.12)	8.48 (2.70)	16.21 (4.94)	0.62 (0.34)	1.59 (3.29)	1.74 (2.19)
False killer whale	340	10	5.77 (1.62)	6.27 (1.52)	5.28 (1.23)	6.95 (1.83)	0.44 (0.22)	0.85 (0.90)	0.03 (0.18)
Short-finned pilot whale	221	18	4.40 (2.72)	5.59 (3.60)	3.73 (2.04)	6.39 (3.89)	0.48 (0.35)	0.86 (1.58)	0.21 (0.81)

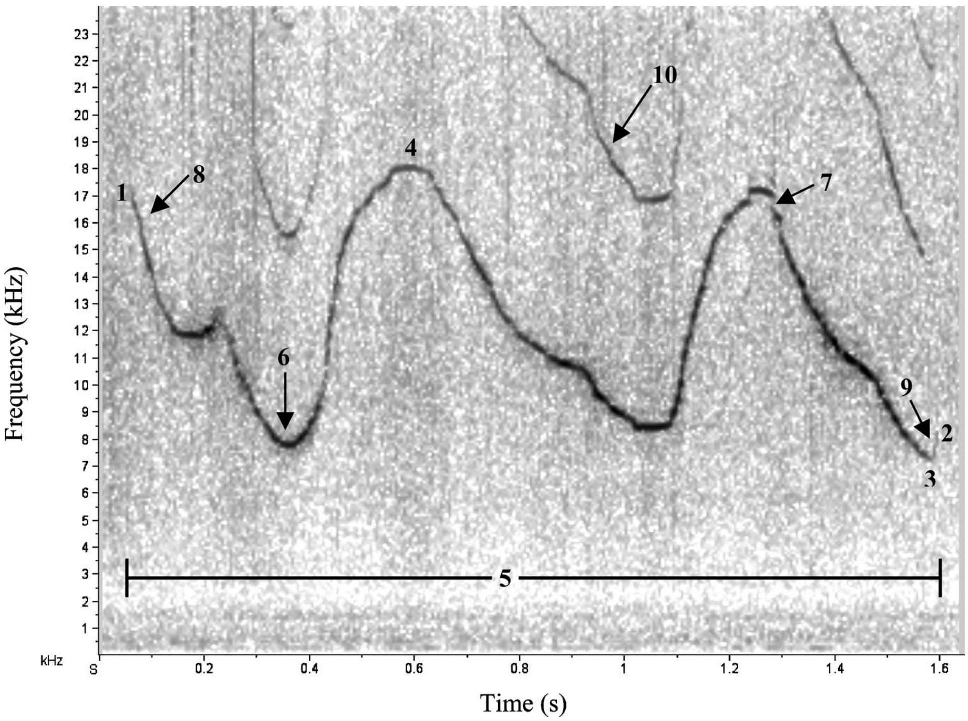

FIG. 2. Spectrogram of a bottlenose dolphin whistle (512-point FFT, Hanning window), showing the ten variables that were measured from each whistle. (1) Beginning frequency (kHz), (2) end frequency (kHz), (3) minimum frequency (kHz), (4) maximum frequency (kHz), (5) duration (s), (6) number of inflection points (defined as a change from positive to negative or negative to positive slope), (7) number of steps (defined as a 10% or greater increase or decrease in frequency over two contour points), (8) slope of the beginning sweep (positive, negative, or zero), (9) slope of the end sweep (positive, negative, or zero), and (10) presence/absence of harmonics (a binary variable).

from each recording session, up to a total of 35 whistles per recording session. This level of subsampling was chosen in order to obtain a sufficient sample size while avoiding oversampling of groups or individuals (which can lead to nonindependence of data). Overlapping whistles were included only if each individual whistle contour could be discerned without question. Whistles were considered to be "loud and clear" if they were at least 9 dB above ambient noise.

Ten variables were measured from the fundamental contour of each whistle: (1) beginning frequency (kHz), (2) end frequency (kHz), (3) minimum frequency (kHz), (4) maximum frequency (kHz), (5) duration (s), (6) number of inflection points (defined as a change from positive to negative or negative to positive slope), (7) number of steps (defined as a 10% or greater increase or decrease in frequency over two contour points), (8) slope of the beginning sweep (positive, negative, or zero), (9) slope of the end sweep (positive, negative, or zero), and (10) presence/absence of harmonics (a binary variable). These variables are shown in Fig. 2. Some whistles from the SPAM 1998 and STAR 1999 surveys were missing measurements for one or more variables because a portion of the whistle extended beyond the upper bandwidth limit of the recording equipment. These whistles were excluded from the analysis. Otherwise, all whistles from all years were combined by species for analysis. Table II gives descriptive statistics for the seven continuous whistle variables.

C. Classification algorithms

Following Oswald *et al.* (2003), classification algorithms were created using multivariate discriminant function analysis (DFA) and classification and regression tree analysis (CART). Discriminant function analysis classifies whistles to prespecified groups based on orthogonal linear functions derived from the ten variables listed above. Mahalanobis distances were calculated for each whistle being classified. The Mahalanobis distance is a measure of the distance in multivariate space of the whistle in question to the group centroid of each species in the analysis. A whistle was classified as the species that it was closest to in multivariate space. Prior to running DFA, continuous variables (frequency variables, duration, and number of steps and inflection points) were tested for normality and log or square-root transformed as necessary. Classification and regression tree analysis creates decision trees by separating data into groups known as nodes through a series of binary splits. Each split is based on the value of a single variable. Final classification is reached at terminal nodes. Terminal node probabilities reflect the certainty of the classification based on the purity of the node. Because CART is nonparametric, it was not necessary to transform variables for normality.

Two different methods of classification using DFA and CART were evaluated. In the first method, whistles were classified directly to species level. This will be referred to as the "direct" method for the remainder of this paper. The second method was hierarchical. Whistles were first classified to the broad categories of "blackfish" or "delphinid." The blackfish category consisted of two species: false killer whales and short-finned pilot whales. The delphinid category consisted of five species and one genus: bottlenose, spotted, spinner, striped, rough-toothed, and common dolphins. Common dolphin species (*Delphinus delphis* and *D. capensis*) were pooled in this analysis (see Sec. III). Once classified to category using the hierarchical approach, whistles within each category were then classified to species.

A jackknife method was used to calculate correct classification scores for DFA and CART. Six versions of the DFA and CART classification algorithms were created, each omitting all of the whistles from one cruise (one year of sampling). Whistles were classified using the algorithms that did not include them. In this way, classification algorithms were created from data that were 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. Fisher's exact test was used to test whether correct classification scores were significantly greater than expected by chance alone. Statistical significance was evaluated at $\alpha = 0.05$ without corrections for multiple testing.

D. ROCCA

ROCCA was created using Matlab and interfaces with real-time scrolling spectrograph software, ISHMAEL (Mellinger, 2001). ISHMAEL is used to monitor signals detected by the hydrophone array. When a whistle of interest is detected, the user stops the scrolling spectrograph and selects the whistle. A Matlab routine called through ISHMAEL opens Matlab and saves the selection as a wav file. ROCCA is then run through Matlab. First, ROCCA automatically extracts the whistle contour from the wav file by stepping through the file one FFT window at a time. The fundamental frequency of the whistle contour is selected based on the peak frequency in each window. A routine within ROCCA ensures that random transient peaks in the spectrum are not mistaken for the fundamental peak frequency. For this study, a Hanning window was used. Window size was set at 1024 points and window overlap was set at 0.25.

When the whistle contour has been extracted, ROCCA automatically measures the ten variables described previously from the fundamental frequency contour of the whistle. The ten variables are then processed using DFA and CART classification algorithms and ROCCA outputs two predicted species, one based on each analysis. It takes approximately 20–45 s for ROCCA to extract a whistle contour, measure whistle variables, and provide an estimate of species identification. As multiple whistles from a single school of dolphins are processed, ROCCA keeps a running tally of species predictions. When all of the whistles from a school have been analyzed, ROCCA classifies the school as the species that the majority of whistles were predicted to be. When DFA and CART results do not agree, the algorithm that resulted in the greatest number of whistles classified as one species is chosen. For example, if DFA classifies 65% of the whistles in a school as bottlenose dolphins and CART classifies 58% of the same whistles as spotted dolphins, the school is classified as bottlenose dolphins. When the same numbers of whistles are classified as two or more species within DFA or CART, the species with the smallest mean Mahalanobis distance is chosen for DFA and the species with the largest mean terminal node probability is chosen for CART. In order to ensure that all whistles analyzed are from one discrete school, whistles are localized using ISHMAEL and Whaltrak before they are analyzed using ROCCA.

As whistles are analyzed, ROCCA creates three text files for each school. One contains the extracted whistle contours (time, frequency, and intensity of the peak frequency in each window). The second contains the ten whistle variables measured from each whistle in the school. The third contains DFA and CART predicted species, as well as Mahalanobis distances (DFA) and terminal node probabilities (CART) for each whistle in the school.

III. RESULTS

When ROCCA was run on all nine species, only 17.8% of short-beaked common dolphin whistles were correctly classified by DFA and 5.7% by CART. Similarly, 6.1% of long-beaked common dolphin whistles were correctly classified by DFA and 2.9% by CART. To explore the possibility that this result was caused by an inability to differentiate between the two common dolphin species, a version of ROCCA was created that included only short-beaked and long-beaked common dolphins. Overall correct classification scores in this analysis were not significantly greater than the 50% expected by chance (DFA: 49.7%, $p = 1$, CART: 46.8%, $p = 0.45$). Because the two common dolphin species could not be distinguished reliably from one another, they were pooled in subsequent analyses.

When the direct version of ROCCA was run on seven species and the pooled common dolphin species, DFA correctly classified 33.5% of whistles and CART correctly classified 33.6% of whistles. These correct classification scores are significantly greater than the 12.5% expected by chance ($p < 0.0001$ for both DFA and CART). For individual species, DFA correct classification scores ranged from 14.7% (striped dolphins) to 63.8% (short-finned pilot whales). Correct classification scores were significantly greater than expected by chance for every species except striped dolphins ($p = 0.41$). Correct classification scores for CART ranged from 18.5% (spinner dolphins) to 57.1% (false killer whales). All correct classification scores were significantly greater than chance with the exception of spinner dolphins ($p = 0.07$). Based on the pooled tallies of individual whistle classifications, 43.8% of schools were correctly classified by DFA and CART combined. Correct classification scores for schools ranged from 31.6% (spinner dolphins) to 73.3% (bottlenose dolphins). Half were significantly greater than chance, with the exceptions being common dolphins ($p = 0.24$), spinner dolphins ($p = 0.23$), spotted dolphins ($p = 0.10$), and striped dolphins ($p = 0.06$).

The hierarchical version of ROCCA resulted in no significant difference in the overall correct classification of either individual whistles or schools compared to the direct version of ROCCA (individual whistles: DFA $p = 0.23$; CART $p = 0.31$; schools $p = 0.75$). However, several significant differences were found for individual species when the hierarchical version of ROCCA was run. The correct classification of individual whistles increased significantly from 22.6% to 35.9% for false killer whales (DFA, $p < 0.001$) and from 33.8% to 41.4% for spotted dolphins (CART, $p = 0.03$) and decreased significantly from 26.2% to 14.7% for striped dolphins (CART, $p < 0.001$). Correct classification scores for individual whistles were significantly greater than chance for every species with the exception of striped dolphins (CART, $p = 0.41$).

The hierarchical version of ROCCA resulted in a significant difference in the correct classification of schools for

TABLE III. Classification results for schools classified based on multiple whistles using the hierarchical DFA and CART method. Percent of schools correctly classified for each species are in bold. Correct classification scores that are significantly greater (Fisher's exact test, $\alpha=0.05$) than the 12.5% expected by chance are marked by an asterisk, and p values are given in the last column. The number of schools included in the analysis for each species (n) is given in the second to last column. Overall, 46.0% of schools were classified to the correct species. This is significantly greater ($p<0.0001$) than expected by chance.

	% Classified as									
Actual species	Bottlenose dolphin	Spotted dolphin	Striped dolphin	Spinner dolphin	Rough-toothed dolphin	Common dolphin	False killer whale	Short-finned pilot whale	n	p
Bottlenose dolphin	**80.0***	0.0	0.0	6.7	6.7	6.7	0.0	0.0	15	<0.001
Spotted dolphin	23.1	**50.0***	7.7	7.7	0.0	7.7	3.8	0.0	26	0.006
Striped dolphin	26.3	23.7	**15.8**	7.9	0.0	26.3	0.0	0.0	38	0.76
Spinner dolphin	21.1	5.3	21.1	**26.3**	10.5	15.8	0.0	0.0	19	0.96
Rough-toothed dolphin	14.3	0.0	7.1	0.0	**64.3***	7.1	0.0	7.1	14	0.02
Common dolphin	22.2	13.9	5.6	0.0	2.8	**55.6***	0.0	0.0	36	<0.001
False killer whale	10.0	0.0	0.0	10.0	0.0	0.0	**80.0***	0.0	10	0.005
Short-finned pilot whale	5.6	0.0	0.0	5.6	0.0	5.6	38.9	**44.4**	18	0.06

only one species. Correct classification of common dolphin schools increased significantly from 27.8% to 55.6% ($p=0.03$). Schools of all species were correctly classified significantly more often than expected by chance, with the exception of short-finned pilot whales ($p=0.06$), striped dolphins ($p=0.76$), and spinner dolphins ($p=0.96$). The confusion matrix for this analysis is given in Table III as an example of classification errors. Overall, 46.0% of whistles were classified to the correct species. With the exception of false killer whales and short-finned pilot whales, whistles from most species were commonly misclassified as bottlenose dolphins. Short-finned pilot whales were most commonly misclassified as false killer whales. Striped dolphins were misclassified as bottlenose dolphins, common dolphins, and spotted dolphins more often than they were correctly classified.

To evaluate the effect of combining DFA and CART predictions and basing classification decisions on all whistles analyzed during an encounter rather than on individual whistles, correct classification scores were compared for three approaches: (1) classifying one whistle at a time, (2) classifying schools based on tallies of species predictions for DFA and CART individually, and (3) classifying schools based on a combination of DFA and CART predictions. These comparisons were made for both the direct version of ROCCA and the hierarchical version. Correct classification scores and p values for these comparisons are given in Tables IV and V. Basing classification decisions on schools rather than individual whistles for DFA and CART individually resulted in no significant differences in correct classification scores in the direct version of ROCCA. Correct classification of common dolphins increased significantly for both DFA and CART ($p=0.004$ and $p=0.006$, respectively) in the hierarchical version of ROCCA. When classification decisions were based on a combination of DFA and CART predictions, rather than on individual whistles, overall cor-

TABLE IV. Correct classification scores for individual whistles classified to species by DFA and CART individually and for schools classified based on multiple whistles using DFA and CART individually and DFA and CART combined. p values are given for comparisons of correct classification scores for individual whistles versus schools (DFA and CART individually) and for individual whistles versus schools (DFA and CART combined). Significant differences are marked by asterisks.

	Individual whistles		Schools						
Species	DFA % correct	CART % correct	DFA % correct	p Whistles versus schools (DFA)	CART % correct	p Whistles versus schools (CART)	DFA and CART % correct	p Whistles (DFA) versus schools (DFA and CART)	p Whistles (CART) versus schools (DFA and CART)
Bottlenose dolphin	60.8	35.9	86.7	0.06	40.0	0.8	73.3	0.4	0.005*
Spotted dolphin	29.6	33.8	38.5	0.4	34.6	1.0	34.6	0.7	1.0
Striped dolphin	14.7	26.2	13.2	1.0	34.2	0.3	34.2	0.005*	0.3
Spinner dolphin	24.7	18.5	26.3	1.0	21.1	0.8	31.6	0.6	0.2
Rough-toothed dolphin	46.9	34.4	64.3	0.3	57.1	0.2	71.4	0.1	0.008*
Common dolphin	28.3	21.7	41.7	0.1	22.2	1.0	27.8	1.0	0.4
False killer whale	22.6	57.1	0.0	0.1	80.0	0.2	70.0	0.002*	0.5
Short-finned pilot whale	63.8	50.7	72.2	0.6	44.4	0.6	61.1	0.8	0.5
Overall	33.5	33.6	39.8	0.1	36.4	0.5	43.8	0.007*	0.007*

TABLE V. Correct classification scores for individual whistles classified to species by hierarchical DFA and CART individually and for schools classified based on multiple whistles using hierarchical DFA and CART individually and DFA and CART combined. p values are given for comparisons of correct classification scores for individual whistles versus schools (DFA and CART individually) and for individual whistles versus schools (DFA and CART combined). Significant differences are marked by asterisks.

	Whistles		Schools						
Species	DFA % correct	CART % correct	DFA % correct	p Whistles versus schools (DFA)	CART % correct	p Whistles versus schools (CART)	DFA and CART % correct	p Whistles (DFA) versus schools (DFA and CART)	p Whistles (CART) versus schools (DFA and CART)
Bottlenose dolphin	60.8	37.9	86.7	0.06	53.3	0.3	80.0	0.2	0.002*
Spotted dolphin	29.6	25.4	34.6	0.7	50.0	0.4	50.0	0.04*	0.4
Striped dolphin	20.2	14.7	10.5	0.2	15.8	0.8	15.8	0.7	0.8
Spinner dolphin	24.3	19.7	15.8	0.6	21.1	0.8	26.3	0.8	0.6
Rough-toothed dolphin	42.7	41.7	57.1	0.4	57.1	0.3	64.3	0.2	0.2
Common dolphin	28.5	25.4	52.8	0.004*	47.2	0.006*	55.6	0.001*	<0.001*
False killer whale	35.9	60.6	40.0	0.8	90.0	0.1	80.0	0.007*	0.3
Short-finned pilot whale	56.1	50.2	55.6	1.0	44.4	0.8	44.4	0.5	0.8
Overall	35.1	35.0	39.8	0.2	40.8	0.09	46.0	0.004*	0.004*

rect classification increased significantly in both versions of ROCCA. Individual species correct classification scores increased significantly for bottlenose dolphins ($p=0.005$ when individual whistles classified by CART), rough-toothed dolphins ($p=0.008$ when individual whistles classified by CART), striped dolphins ($p=0.005$ when individual whistles classified by DFA), and false killer whales ($p=0.002$ when individual whistles classified by DFA) in the direct version of ROCCA, and for bottlenose dolphins ($p=0.002$ when individual whistles classified by CART), spotted dolphins ($p=0.04$ when individual whistles classified by DFA), false killer whales ($p=0.007$ when individual whistles classified by DFA), and common dolphins ($p=0.001$ when individual whistles classified by DFA, $p<0.001$ when individual whistles classified by CART) in the hierarchical version of ROCCA.

IV. DISCUSSION

Traditional visual monitoring techniques during shipboard marine mammal surveys are limited by animal behavior, environmental conditions, and logistical constraints. The addition of a passive acoustic component to these surveys provides a method for overcoming some of these limitations. Real-time acoustic species identification offers an additional tool for identifying schools that are difficult to approach and observe and allows species distribution data to be collected even when visual effort is compromised by factors such as poor visibility, inclement weather, and high sea states.

Real-time acoustic species identification is especially valuable during surveys dedicated to specific species. For example, the focus of the PICEAS 2005 survey was to determine the population status of false killer whales in an area of high fishery bycatch in the central tropical Pacific Ocean. Visual detection of these animals was extremely difficult due to animal behavior and high sea states. During the first month of this survey, five schools of false killer whales were encountered. Three of the five schools were detected and located by the acoustic team and identified as false killer whales using ROCCA. The acoustic identifications were confirmed visually. Time constraints demanded that the ship deviate from the survey trackline for acoustic detections of this focal species only, and therefore real-time acoustic identification was crucial. The combination of high correct classification scores for the species identity of false killer whale whistles, combined with high vocal rates and poor visual detection of this species, created a situation in which passive acoustics played an indispensable role. Without the capability for real-time species identification provided by ROCCA, the acoustic detections would not have been investigated and valuable data would have been lost. ROCCA was also used to estimate the fraction of schools of false killer whales missed by visual methods within 4.5 km of the transect line during the PICEAS 2005 survey (Barlow and Rankin, 2007). This provided a means of ground-truthing estimated line-transect parameters and showed that the fraction of detections missed visually (0.56) was consistent with the expected fraction missed (0.58).

In addition to providing assistance to the visual observers, ROCCA has the advantage of reducing the bottleneck of post-cruise analysis. ROCCA's automated whistle extraction, measurement, and data storage features reduce the time necessary for post-cruise analysis and make ROCCA valuable for other applications such as processing the voluminous amounts of data collected using seafloor mounted acoustic recorders.

While correct classification scores obtained using ROCCA are not at the level of near-certainty that would be optimal for shipboard surveys, results are promising as correct classification scores for the individual whistles of most species were significantly greater than expected by chance. Scores did not reach near certainty due to high within-species variability in whistle variables and a large degree of overlap in the time and frequency variables of many species (Table II). Bottlenose dolphins and false killer whales had the highest correct classification scores, with 80% of schools of both species being correctly identified using the hierarchical version of ROCCA (Table III). Examination of descriptive statistics (Table II) shows that some whistle variables for these two species are distinctive. Bottlenose dolphin whistles have a longer mean duration and false killer whale whistles have lower mean frequencies compared to most other species in the analysis. However, while few whistles of species other than short-finned pilot whales and rough-toothed dolphins were misclassified as false killer whales, whistles from most species were commonly misclassified as bottlenose dolphins. This may be due to frequency and duration variables. Bottlenose dolphins had one of the lowest values for mean minimum frequency, one of the highest values for mean maximum frequency, and the longest mean duration. All of these values had high standard deviations, suggesting that, based on the variables measured, many different whistle types would fall into the bottlenose dolphin category. This implies that the variables measured were not sufficient for separating species in this analysis.

Correct classification scores were low for spinner and striped dolphins in all analyses. Striped dolphin classification errors were relatively evenly spread across all species except false killer whales, short-finned pilot whales, and rough-toothed dolphins. Similarly, spinner dolphin classification errors were generally evenly spread across all species other than short-finned pilot whales. These patterns are also likely due to frequency variables. The whistles of the small delphinid species (spotted, striped, spinner, short-beaked common and long-beaked common dolphins) had very similar frequency characteristics for the variables measured (Table II).

Because of the high degree of overlap in frequency characteristics among species, a method for increasing classification success may lie in the exploration of additional whistle variables such as the rate of change in frequency (slope), the locations of steps and inflection points within whistles, and relative intensities of different frequencies. Also, compound variables such as a combined value for slope and frequency range may be more effective for separating species. In addition, alternate classification algorithms such as artificial neural networks and hidden Markov models may be better suited to the task of identifying dolphin whistles. Work is currently under way to explore the effect of alternate whistle variables and classification algorithms on correct classification scores.

The species included in ROCCA are often found in single species schools, but have also been observed mixed with other species. ROCCA was created using recordings of single species schools and therefore does not currently have the capability to identify mixed species schools as such. Future plans for ROCCA include the analysis of recordings of mixed species schools in order to develop decision criteria for identifying schools as mixed versus single species.

ROCCA performed best when classification decisions were based on multiple whistles classified using the hierarchical method and when decisions were based on a combination of DFA and CART results. This approach resulted in an overall correct classification score of 46.0% and some very high correct classification scores of up to 80% for species such as bottlenose dolphins and false killer whales (Table III). Making classification decisions based on multiple whistles rather than individual whistles resulted in slight increases in correct classification scores, but these were not significant for DFA or CART (Tables IV and V). However, when classification decisions for schools were made based on a combination of DFA and CART results, correct classification increased significantly both overall and for several individual species. While the hierarchical version of ROCCA did not result in a significant increase in overall correct classification, it did increase slightly, and correct classification of common dolphin schools in particular increased significantly. Additionally, correct classification scores were significantly greater than chance for five of the eight species in the hierarchical version, compared to four of the eight species in the direct version of ROCCA. This approach shows some promise and may produce more significant results with different species categories or a greater number of levels within the hierarchy.

The results of this study point not only to the benefit of making classification decisions based on multiple whistles, but also to the benefit of using more than one classification algorithm. Different classification algorithms are sensitive to different characteristics of the data set, and the ability to combine the strengths of more than one algorithm can result in higher classification success. When the optimal set of whistle variables and classification algorithms is assembled, ROCCA will be valuable not only for real-time species identification during shipboard surveys, but also for analysis of vocalizations recorded using seafloor-mounted hydrophones. While ROCCA has been created for use in the eastern tropical and temperate Pacific Ocean, it can be modified for use in other regions.

ACKNOWLEDGMENTS

We would like to gratefully acknowledge the tireless efforts of the skilled acousticians, visual observers, and cruise leaders who dedicated months to the surveys. Without them this project could not have been completed. We thank the officers and crew of the NOAA ships *McArthur*, *McArthur II*, and *David Starr Jordan* and the NSF/UNOLS R/V *Endeavor* for their patience and cooperation. Tim Gerrodette and Lisa Balance (and J.B.) were the Chief Scientists for these surveys. Special thanks to Stephanie Grassia and Theresa Weber for their assistance with data analysis. We also thank Michael Oswald and two anonymous reviewers for providing insightful comments on drafts of this manuscript.

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

Au, W. W. L. (**1980**). "Echolocation signals of the Atlantic bottlenose dolphin (*Tursiops truncatus*) in open waters," in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 251–282.

Barlow, J. (**1995**). "The abundance of cetaceans in California waters. Part I: Ship surveys in summer and fall of 1991," Fish. Bull. **93**, 1–14.

Barlow, J., and Rankin, S. (**2007**). "False killer whale abundance and density: preliminary estimates for the PICEAS study area south of Hawaii and new estimates for the US EEZ around Hawaii," NOAA Administrative Report LJ-07-02.

Barlow, J., and Taylor, B. L. (**2005**). "Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey," Marine Mammal Sci. **21**, 429–445.

Chesmore, E. D. (**2001**). "Application of time domain signal coding and artificial neural networks to passive acoustical identification of animals," Appl. Acoust. **62**, 1359–1374.

Clark, C. W., Charif, R., Mitchell, S., and Colby, J. (**1996**). "Distribution and behavior of the bowhead whale, *Balaena mysticetus*, based on analysis of acoustic data collected during the 1993 spring migration off Point Barrow, Alaska," Rep. Int. Whal. Comm. **46**, 541–552.

Clark, C. W., and Fristrup, K. M. (**1997**). "Whales '95: A combined visual and acoustic survey of blue and fin whales off Southern California," Rep. Int. Whal. Comm. **47**, 583–600.

Fristrup, K. M., and Clark, C. W. (**1997**). "Combining visual and acoustic survey data to enhance density information," Rep. Int. Whal. Comm. **47**, 933–936.

Goold, J. C., and Jones, S. E. (**1995**). "Time and frequency domain characteristics of sperm whale clicks," J. Acoust. Soc. Am. **98**, 1279–1291.

Gordon, J. C. D., Matthews, J. N., Panigada, S., Gannier, A., Borsani, J. F., and Notarbartolo di Sciara, G. (**2000**). "Distribution and relative abundance of striped dolphins in the Ligurian Sea Cetacean Sanctuary: Results from an acoustic collaboration," J. Cetacean Res. Manage. **2**, 27–36.

Herzing, D. L. (**2000**). "Acoustics and social behavior of wild dolphins: implications for a sound society," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 225–272.

Janik, V. M., and Slater, P. J. B. (**1998**). "Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls," Anim. Behav. **56**, 829–838.

Kinzey, D., Gerrodette, T., Perryman, W., Olson, P., and Rankin, S. (**2001**). "Marine mammal data collected during a survey in the eastern tropical Pacific ocean aboard the NOAA ships McArthur and David Starr Jordan, July 28–December 9, 2000," NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-303.

Lammers, M. O., Au, W. W. L., and Herzing, D. L. (**2003**). "The broadband social acoustic signaling behavior of spinner and spotted dolphins," J. Acoust. Soc. Am. **114**, 1629–1639.

Leaper, R., Chappell, O., and Gordon, J. (**1992**). "The development of practical techniques for surveying sperm whale populations acoustically," Rep. Int. Whal. Comm. **42**, 549–560.

Matthews, J. N., Rendell, L. E., Gordon, J. C. D., and MacDonald, D. W. (**1999**). "A review of frequency and time parameters of cetacean tonal calls," Bioacoustics **10**, 47–71.

Mellinger, D. K. (**2001**). "ISHMAEL 1.0 User's Guide," NOAA Technical Memorandum OAR PMEL-120. Available from NOAA/PMEL, 7600 Sand Point Way, NE, Seattle, WA 98115-6349.

Mills, H. (**1995**). "Automatic detection and classification of nocturnal migrant bird Calls," J. Acoust. Soc. Am. **97**, 3370.

Norris, K. S., and Evans, W. E. (**1966**). "Directionality of echolocation clicks in the rough-toothed porpoise, *Steno bredanensis* (Lesson)," in *Marine Bio-acoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 305–324.

Oswald, J. N., Barlow, J., and Norris, T. F. (**2003**). "Acoustic identification of nine delphinid species in the eastern tropical Pacific ocean," Marine Mammal Sci. **19**, 20–37.

Oswald, J. N., Rankin, S., and Barlow, J. (**2004**). "The effect of recording and analysis bandwidth on acoustic identification of delphinid species," J. Acoust. Soc. Am. **116**, 3178–3185.

Parsons, S., and Jones, G. (**2000**). "Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks," J. Exp. Biol. **203**, 2641–2656.

Reby, D., Lek, S., Dimopoulos, I., Joachim, J., Lauga, J., and Aulagnier, S. (**1997**). "Artificial neural networks as a classification method in the behavioral sciences," Behav. Processes **40**, 35–43.

Rendell, L. E., Matthews, J. N., Gill, A., Gordon, J. C. D., and MacDonald, D. W. (**1999**). "Quantitative analysis of tonal calls from five odontocete species, examining interspecific and intraspecific variation," J. Zool. **249**, 403–410.

Roch, M. A., Soldevilla, M. S., Burtenshaw, J. C., Henderson, E. E., and Hildebrand, J. A. (**2007**). "Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California," J. Acoust. Soc. Am. **121**, 1737–1748.

Stafford, K. M., Nieukirk, S. L., and Fox, G. C. (**2001**). "Geographic and seasonal variation of blue whale calls in the North Pacific," J. Cetacean Res. Manage. **3**, 65–76.

Steiner, W. W. (**1981**). "Species-specific differences in pure tonal whistle vocalizations of five western north Atlantic dolphin species," Behav. Ecol. Sociobiol. **9**, 241–246.

Thompson, P., Findley, L. T., and Vidal, O. (**1992**). "20 Hz pulses and other vocalizations of fin whales, *Balaenoptera physalus*, in the Gulf of California, Mexico," J. Acoust. Soc. Am. **92**, 3051–3057.

Thompson, P., Findley, L. T., Vidal, O., and Cummings, W. (**1996**). "Underwater sounds of blue whales, *Balaenoptera musculus*, in the Gulf of California, Mexico," Marine Mammal Sci. **12**, 288–292.

Tyack, P. L., and Clark, C. W. (**2000**). "Communication and acoustic behavior of dolphins and whales," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 156–224.

van Parijs, S. M., Smith, J., and Corkeron, P. J. (**2002**). "Using calls to estimate the abundance of inshore dolphins: A case study with Pacific humpback dolphins *Sousa chinensis*," J. Appl. Ecol. **39**, 853–864.

Vaughan, N., Jones, G., and Harris, S. (**1997**). "Habitat use by bats (*Chiroptera*) assessed by means of a broad-band acoustic method," J. Appl. Ecol. **34**, 716–730.

Wade, P. R., and Gerrodette, T. (**1993**). "Estimates of cetacean abundance and distribution in the eastern tropical Pacific," Rep. Int. Whal. Comm. **43**, 477–493.

Wang, D., Wursig, B., and Evans, W. (**1995**). "Comparisons of whistles among seven odontocete species," in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden), pp. 299–323.

Wang, K., Wang, D., Akamatsu, T., Songhai, L., and Xiao, J. (**2005**). "A passive acoustic monitoring method applied to observation and group size estimation of finless porpoise," J. Acoust. Soc. Am. **118**, 1180–1185.

Watkins, W. A. (**1967**). "The harmonic interval: fact or artifact in spectral analysis of pulse trains," in *Marine Bio-acoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 15–42.