Classification of echolocation clicks from odontocetes in the Southern California Bight

Marie A. Roch[a)]
San Diego State University, Department of Computer Science, 5500 Campanile Drive, San Diego, California 92182-7720

Holger Klinck
Cooperative Institute for Marine Resources Studies, Oregon State University, 2030 South East Marine Science Drive Newport, Oregon 97365

Simone Baumann-Pickering
Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0205

David K. Mellinger
Cooperative Institute for Marine Resources Studies, Oregon State University, 2030 South East Marine Science Drive, Newport, Oregon 97365

Simon Qui
San Diego State University, Department of Computer Science, 5500 Campanile Drive, San Diego, California 92182-7720

Melissa S. Soldevilla
Duke University Marine Laboratory, 135 Duke Marine Laboratory Road, Beaufort, North Carolina 28516

John A. Hildebrand
Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0205

(Received 11 May 2010; revised 7 October 2010; accepted 7 October 2010)

This study presents a system for classifying echolocation clicks of six species of odontocetes in the Southern California Bight: Visually confirmed bottlenose dolphins, short- and long-beaked common dolphins, Pacific white-sided dolphins, Risso's dolphins, and presumed Cuvier's beaked whales. Echolocation clicks are represented by cepstral feature vectors that are classified by Gaussian mixture models. A randomized cross-validation experiment is designed to provide conditions similar to those found in a field-deployed system. To prevent matched conditions from inappropriately lowering the error rate, echolocation clicks associated with a single sighting are never split across the training and test data. Sightings are randomly permuted before assignment to folds in the experiment. This allows different combinations of the training and test data to be used while keeping data from each sighting entirely in the training or test set. The system achieves a mean error rate of 22% across 100 randomized three-fold cross-validation experiments. Four of the six species had mean error rates lower than the overall mean, with the presumed Cuvier's beaked whale clicks showing the best performance (<2% error rate). Long-beaked common and bottlenose dolphins proved the most difficult to classify, with mean error rates of 53% and 68%, respectively.
 [DOI: 10.1121/1.3514383]

PACS number(s): 43.80.Ev, 43.60.Uv [WWA] Pages: 467–475

I. INTRODUCTION

Passive acoustic monitoring (Mellinger *et al.*, 2007) provides an opportunity to study cetaceans in a non-invasive manner and can provide insight into presence/absence and seasonality (Sirovic *et al.*, 2004), population structure (Deecke *et al.*, 1999), and abundance (Marques *et al.*, 2009). An important precondition of these applications is the ability to identify which species produced a given sound. Acoustic species identification relies on extracting relevant information, or features, from the acoustic record and using techniques from statistics and pattern recognition to decide which species, if any, is present. Early work in this area was limited by low sampling rates and focused primarily on the analysis of whistles (e.g., Steiner, 1981). A brief history of different techniques used for species identification using passive acoustic monitoring can be found in Roch *et al.* (2007), a study which extracted features without identifying them as whistles, echolocation burst pulses, or click trains from 24 kHz band limited data. This study focuses on identifying echolocation clicks of six species of odontocetes in the Southern California

[a)]Author to whom correspondence should be addressed. Electronic mail: marie.roch@sdsu.edu

Bight: Bottlenose dolphins (*Tursiops truncatus*), short- and long-beaked common dolphins (*Delphinus delphis* and *D. capensis*, respectively), Pacific white-sided dolphins (*Lagenorhynchus obliquidens*), Risso's dolphins (*Grampus griseus*), and presumed Cuvier's beaked whales (*Ziphius cavirostris*).

Echolocation clicks do not typically propagate as well as tonal vocalizations (Oswald *et al.*, 2007). Echolocation clicks of Cuvier's beaked whales, for example, have been predicted to be detectable with high probability at ranges up to 0.7 km and very low probability beyond 4 km (Zimmer *et al.*, 2008). The energy in echolocation clicks is narrowly focused along the longitudinal axis of the echolocating animal. As the angle between this axis and the hydrophone increases, the signal is increasingly attenuated and distorted (Au, 1993, pp. 104–108; Zimmer *et al.*, 2005b; Zimmer *et al.*, 2008; Lammers and Castellote, 2009). Signal distortion is further complicated by high frequencies attenuating faster than low frequencies as the animal-to-hydrophone distance increases. Finally, odontocetes have been shown to vary their echolocation clicks in different environments (Au, 1993, Chap. 7) as well as under different behavioral conditions (Madsen *et al.*, 2005). Much of the variation in the spectral bandwidth and peak frequency of these echolocation clicks is likely to be due to the animal's distance and orientation toward the hydrophone, both of which can result in the higher frequencies being attenuated. Other factors, such as the animals' ability to vary peak frequency (Au, 1993, p. 120), are also likely to play a role. An example of this can be seen in Fig. 1, in which echolocation clicks from long-beaked common dolphins recorded on a single sighting during a California cooperative oceanic fisheries investigations (CalCOFI) cruise have been sorted by peak frequency.

These variables make for a challenging environment in which to perform species identification using echolocation clicks. Nonetheless, echolocation clicks are shaped by complex anatomical structures (Cranford, 2000) that affect the characteristics of the clicks. Thus in theory, it is plausible that echolocation clicks can be classified to species, provided that features related to the underlying production system can be extracted or enough of the off-axis and acoustic propagation effects can be captured to model the feature distribution. Echolocation clicks, if identifiable to species, could provide useful data for passive acoustic surveys, as odontocetes produce them quite frequently for foraging, navigation, and communication. Indeed, some genera of odontocetes, including *Physeter*, *Phocoena*, and *Cephalorhynchus*, are thought to vocalize using only clicks and not other sounds such as whistles. Finally, in some behavioral situations such as foraging, species known to produce whistles have been known to favor the use of echolocation clicks over whistles (Benoit-Bird and Au, 2009).

Several groups have recently tested classification methods on a common dataset consisting of clicks of Blainville's beaked whales (*Mesoplodon densirostris*), Risso's dolphins (*Grampus griseus*), and short-finned pilot whales (*Globicephala macrorhynchus*) (Moretti *et al.*, 2008). Gerard *et al.* (2008) assigned scores based on hand-selected characteristics of echolocation spectra and integrated this into a model which took into account inter-click intervals (ICIs). Gillespie and Caillat (2008) pursued two separate techniques. In an extension of earlier work (Gillespie, 2004), a series of features were extracted from a high-energy region of the Wigner-Ville distribution of each click. Parameters included ridge slope, duration, bandwidth, etc., and classification trees were used to decide species identity. Their second method used spectral features from a uniformly spaced filter bank that were classified by a linear discriminant function derived from a one-way multivariate analysis of variance. Harland (2008) analyzed the data using spectrogram correlation (Mellinger and Clark, 2000). Jarvis *et al.* (2008) used zero-crossing intervals and peak frequency with support vector machines (SVMs). Rather than training each binary-decision SVM with one species versus all others, they trained with one species versus a noise class. Roch *et al.* (2008) used cepstral feature vectors and compared the performance of Gaussian mixture model (GMM) and SVM classifiers. The

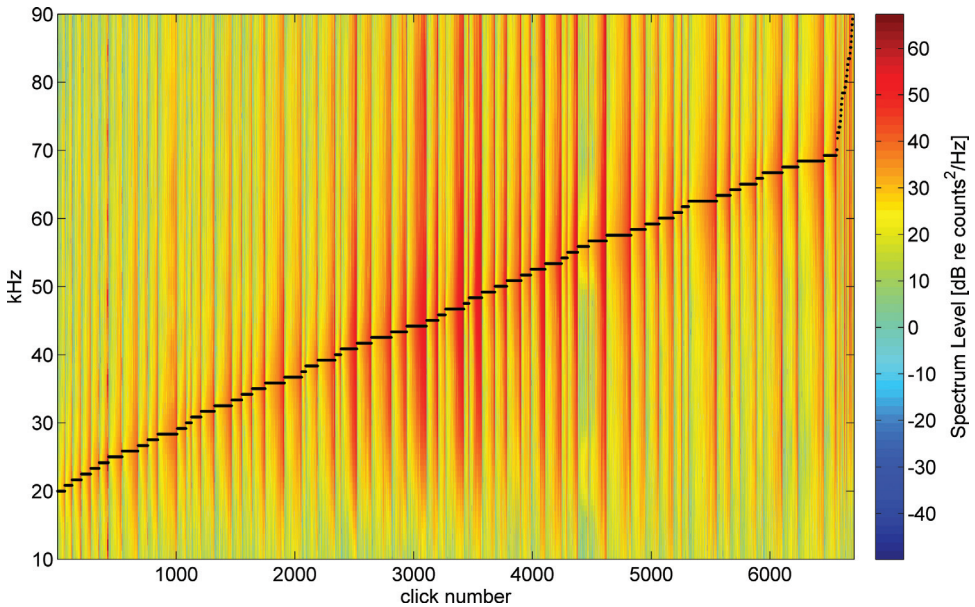

FIG. 1. (Color online) Concatenated spectra of detected events recorded in the presence of long-beaked common dolphins during a CalCOFI cruise. Spectra are sorted by peak frequency which are highlighted by black points.

TABLE I. Summary of data used for the species identification task showing the recording platform and year, number of sightings (sight) and echolocation clicks used, and the custom preamp board. Abbreviations: CalCOFI, California Cooperative Oceanic Fisheries Investigations oceanographic survey; SCI, San Clemente Island rigid hull inflatable boat survey, SC, Southern California Instrumentation cruises, FLIP, R/V Flip moored recordings, and HARP, High-frequency Acoustic Recording Package deployment.

		Bottlenose dolphin		Cuvier's beaked whale		Long-beaked common dolphin		Pacific white-sided dolphin		Risso's dolphin		Short-beaked common dolphin	
Cruise/Platform	Preamp	Sight	Clicks	Sight	Clicks	Sight	Clicks	Sight	Clicks	Sight	Clicks	Sight	Clicks
SC 2003	A100	1	563			1	1994					2	786
CalCOFI 2004	A103					1	2863						
CalCOFI 2006	R100	1	561			1	220						
FLIP 2006	H300					1	4376	6	46 535				
SCI 2006	R100	2	9670									6	23 995
SCI 2007	H300 & HF338							1	617	3	5742		
HARP 2009 Site M	480			4	4498								
HARP 2009 Site N	452			5	18 164								
	Total	4	10 794	8	22 662	4	9453	8	47 152	3	5742	8	24 781

GMM classifiers produced slightly lower error rates than the SVM classifiers. While the system of Roch *et al.* was declared to have the best performance (Moretti *et al.*, 2008), the overall performance of most systems was excellent.

The species selected for the common dataset had very different click spectra and were recorded in different environments using different equipment. Post-filtering was applied by the Navy to remove sensitive signals. When this is done randomly with respect to the classes, it is a problem with which the recognition system must cope. Unfortunately, when the channel conditions correspond to a specific species as in the aforementioned dataset, it has the potential to artificially improve results. A possible result of this is that classification decisions may have been influenced by environment and channel conditions as well as the properties of the echolocation clicks. This problem occurs in many pattern recognition tasks and is known by various names such as channel variation or mismatch in speech and speaker recognition (Bimbot *et al.*, 2004), or the album effect in music identification (Downie, 2008).

The goals of this study are to examine classification performance on a species identification task for acoustic data collected within a single geographic location and to examine how differences in the partitioning of a dataset into training and test data can affect overall error rate. In addition, most of the species in this study are more similar in morphology than in the previously mentioned species identification studies using echolocation clicks, potentially making the clicks more similar and hence the classification task more difficult.

II. MATERIALS AND METHODS

A. Data collection

Acoustic data from multiple surveys in the Southern California Bight were used in this study. The data for all species except for Cuvier's beaked whales were recorded using towed and dipped hydrophone arrays and collected in the presence of single-species schools as determined by teams of experienced visual observers. The Cuvier's beaked whale data were collected by a high-frequency acoustic recording package known as a HARP (Wiggins and Hildebrand, 2007) without visual observation.

A detailed description of the towed and dipped hydrophone dataset was published previously (Soldevilla *et al.*, 2008). Briefly, acoustic data sampled with 16-bit quantization at a rate of 192 kHz were collected at various offshore regions in the Southern California Bight. The quantity of data varied by species and is summarized in Table I. Data have been pooled from multiple surveys between 2004 and 2007: CalCOFI oceanographic surveys, San Clemente Island (SCI) small boat operations, Scripps Institution of Oceanography (SIO) instrumentation servicing cruises (SoCal) on the R/V Robert Gordon Sproul, and moored observations from the R/P FLIP (Fisher and Spiess, 1963). Two types of hydrophones were used, the ITC 1042 (International Transducer Corporation, Santa Barbara, CA), and the HS150 (Sonar Research and Development Ltd., Beverly, UK), both of which have flat frequency responses (± 3 dB) between 1 and 100 kHz. Several series of custom preamplifiers were employed which had different frequency response curves. The preamplifiers were designed to whiten ambient ocean noise, and the transfer function was calibrated by recording the gain of known input signals.

Presumed Cuvier's beaked whale data were collected from two HARP deployments in known beaked whale habitat (Falcone *et al.*, 2009) located to the north and south of SCI. A pair of low- and high-frequency channels were quantized to 15-bit signals and subsequently summed into a single 16-bit channel. The high-frequency hydrophone was ITC 1042, the same model as previously described for the array recordings. While visual confirmation is not feasible for seafloor instruments, trained analysts located sections of the recording where echolocation clicks matched published descriptions of Cuvier's beaked whale clicks (Johnson *et al.*, 2004; Zimmer *et al.*, 2005a). For brevity, the word presumed will be omitted except in the conclusions.

B. Signal treatment

Echolocation clicks were detected in a two stage-process similar to that described by Soldevilla *et al.* (2008). The first stage searches for groups of echolocation clicks. Fourier transforms are computed from 10 ms Hann-windowed signal frames with a 5 ms advance (50% overlap) between frames. Blocks of 3 s were used to estimate the noise floor in each frequency bin, and echolocation clicks were considered to be detected when 12.5% of the frequency bins had signal-to-noise ratios (SNRs) of 13 dB above the noise floor across the 15–95 kHz bandwidth. The moored bottom recorders provided lower-noise recordings and the SNR threshold was dropped to 8 dB to detect fainter echolocation clicks.

Individual clicks were identified using a variation on the method proposed by Kandia and Stylianou (2006). The signal was high-pass filtered with an equiripple finite impulse response filter with a transition band between 3 and 8 kHz to remove confounding sounds such as ship noise. The filter provided 64 dB of attenuation in the stop band and had 3 dB of attenuation at 7.8 kHz. The filter was appropriate for the species of interest, but would need to be modified for odontocetes such as sperm whales (*Physeter macrocephalus*) that have significant low-frequency energy in their clicks (Møhl *et al.*, 2003). Peaks in the Teager energy (Kaiser, 1990) of the high-passed signal were used to identify regions of rapid change where potential echolocation clicks occurred. The noise floor was estimated at the 40th percentile of the energy distribution. Regions were grown about the peak using a smoothed version of the Teager energy envelope. Smoothing was accomplished using a zero-phase smoothing filter with coefficients $[0.\overline{1}, 0.\overline{2}, 0.\overline{3}, 0.\overline{2}, 0.\overline{1}]$, and a greedy region-growing procedure similar to the technique presented by Fristrup and Watkins (2004) was used to identify the start and end of the echolocation click. This consisted of locating the largest interior Teager energy peak within regions that exceeded the noise floor by a factor of 50. The region was grown in the forward and backward directions until the outermost smoothed Teager energy sample was less than three times the noise floor. The departure from traditional techniques developed for measurements of on-axis clicks was used to permit trailing energy from presumed off-axis clicks be included. To prevent close reflections from being merged into the detected click, click growth was also terminated when it reached the midpoint between the largest successive energy peaks. When clicks were less than 500 μs apart, it was assumed that they were likely to be reflections and only the one with the strongest Teager energy was used. Fourier transforms of each click were computed after applying a Hann window of the click duration and zero-padding to a standard length of 1200 μs. This resulted in a standard interpolated bandwidth regardless of the click length or sample rate.

Echolocation click spectra with peaks under 20 kHz or above 70 kHz were discarded. Manual inspection showed that many clicks whose peak frequency fell in this excluded range were either clipped or of poor quality.

Echosounders were present in many recordings and, with the exception of the CalCOFI cruise data, were removed by an analyst examining the first-stage click detector output. An echosounder, with an approximate peak frequency of 56 kHz operated on the CalCOFI cruises, was removed by looking for narrowband signals near that frequency. Duration was not used, as the click detector distinguished individual pulses of the echosounder, thus underreporting the duration.

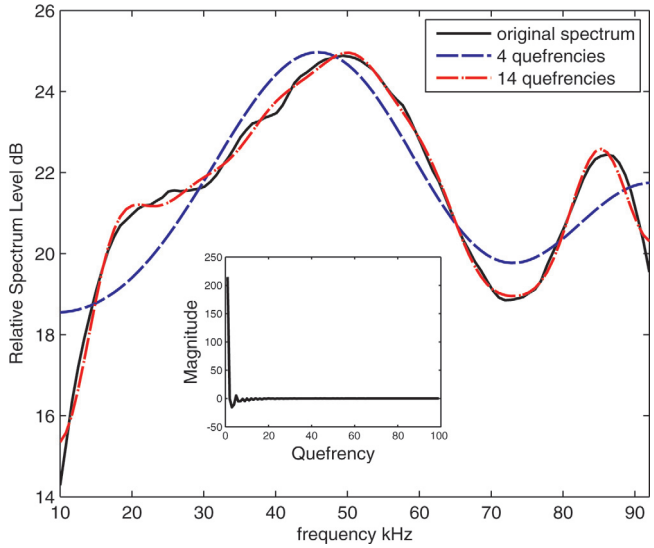

FIG. 2. (Color online) Original and reconstructed spectrum of a long-beaked common dolphin click. The inner figure shows a portion of the cepstrum associated with the echolocation click. The outer figure shows the 99 bin original spectrum (solid black line) derived from the time series and two zero-padded reconstructed spectra using the first 4 (dash) and 14 points (dashed-dotted line) of the cepstrum. The 0th cepstral coefficient was retained to preserve the frequency offset for illustrative purposes.

The real cepstrum was computed as the logarithm of the magnitude spectra followed by a discrete cosine transform of the frequency bins between 10 and 92 kHz. For the very rare occasions where a frequency bin had zero energy, the energy was raised to a value just above the machine's floating point precision to prevent the log operation from producing negative infinity and unduly biasing subsequent processing. The cepstrum is a homomorphic transform that replaces the convolution operator with addition and can be viewed as a method of blind source-filter separation (Picone, 1993). The production of echolocation clicks is not well enough understood to determine whether or not its use is justified on this basis. An alternative view is that the spectrum is being treated as a signal and the cepstrum provides a low-dimensional characterization of its general shape through spectral analysis. The 0th coefficient (representative of the overall energy in the signal) was discarded and only coefficients 1–14 were retained. Higher order coefficients are representative of finer detail in the spectrum and increase the model order without necessarily increasing classification performance. An example for a long-beaked common dolphin click is shown in Fig. 2, demonstrating the cepstrum's ability to model a complex spectrum with a small number of points.

C. Classification

For each species, cepstral click features were grouped by sighting and placed entirely in one of the three groups

used in a three-fold cross-validation experiment (Duda *et al.*, 2001, pp. 483–485). For the Cuvier's beaked whale data from the seafloor instrument, a sighting was defined as a set of call bouts separated from other bouts by at least 2 h. Data from a single sighting are likely to be more similar than data collected when instrumentation, individual animals, or conditions such as bathymetry, behavioral state, and sea state differ. Splitting data from a single sighting across folds is not representative of field conditions and therefore prohibited in our experimental design.

Overfitting is a common problem in the application of machine learning algorithms. This occurs when a model is constrained to represent the training data so tightly that it fails to generalize well to an independent test set. Overfit models are said to exhibit high variance, which means that minor variations in the training data can produce radically different models. An alternate view of overfitting is that a *specific* model is too closely tied to the training data and lacks sufficient variability to accurately model the distribution. To measure the variability of the system, sightings for each species were randomly shuffled before being distributed to one of the three-folds. When the number of sightings was not a multiple of 3, remaining sightings were put in the last fold. As an example, Pacific white-sided dolphins, for which there were eight sightings would have the data from two sightings in each of the first two folds and four in the last one. This shuffling was repeated 100 times to provide a better understanding of system performance by exploring both fortuitous and disastrous partitioning of the training and test data.

Each species was modeled with a 16-mixture GMM. GMMs model arbitrary distributions by scaling a set of normal distributions such that the integration of all distributions over the vector space equals 1. Complete descriptions and derivations of GMMs are readily available elsewhere (Dempster *et al.*, 1977; Duda *et al.*, 2001, pp. 524–526; Huang *et al.*, 2001, pp. 172–175). The expectation maximization (EM) algorithm was used to iteratively improve the likelihood of an initial model derived from the sample statistics of a partitioning induced by 16 clusters. Clustering for the initial model was accomplished with the Linde-Buzo-Gray (LBG) algorithm (Linde *et al.*, 1980), which starts with a single cluster and iteratively performs binary splits followed by k-means clustering. Initial parameters were established from sample statistics of the training data partitioned by the LBG clusters. Once the initial model was formed, the expected contribution of each mixture to the likelihood of the training data was calculated using the model. The expected value along with the training data was then used to estimate an improved model using a maximum likelihood estimator, and the process was repeated. The EM algorithm guarantees convergence, and iterations were computed until either 15 iterations were reached or the likelihood of training data given successive models improved by no more than 10%. Although some effort was placed into the investigation of the number of mixtures, earlier work (Roch *et al.*, 2008) showed 16 mixtures to be reasonable for estimating odontocete echolocation click distributions with cepstral features, and selection of the number of mixtures was not a major component of this study.

During testing, species were assumed to have a uniform prior distribution and successive clicks were assumed to be independent for the purposes of computing their joint posterior probability. Classification decisions were made for groups of 100 clicks. Consequently, the log class conditional probability

$$P_{\log}(\text{group}_k|M_{\text{species}}) = \sum_{i=(k-1)100+1}^{100k} \log \text{P}(\text{click}_i|M_{\text{species}})$$

is proportional to the posterior probability and is computed for each model ($M_{\text{bottlenose}}$, M_{Cuvier}, $M_{\text{long-beaked}}$, …). The probability of each click is then

$$P(\text{click}_i|M_{\text{species}}) = \sum_{m=1}^{16} c_m \frac{1}{(2\pi)^{d/2}|\Sigma_m|^{1/2}} \times \exp\left(\frac{-1}{2}(\text{click}_i - \mu_m)^t \Sigma_m^{-1}(\text{click}_i - \mu_m)\right)$$

where μ_m and Σ_m are the species-specific mean and diagonal covariance matrices for the mth normal distribution, c_m is the species-specific prior probability of the mixture, and d is the dimensionality of the feature space. The highest class conditional score represented the optimal decision rule for this set of data and models when one does not assume *a-priori* knowledge of the probability of acoustic encounters for each species.

III. RESULTS

A mean error rate of $0.22 \pm 0.11\sigma$ was obtained across the six species and a histogram of overall error rates is shown in Fig. 3. As multiple trials were conducted, the standard confusion matrix metric became less meaningful than usual, so instead we report the mean correct/incorrect performance for each species across the 100 three-fold cross-validation trials in Table II. As the mean error rate for

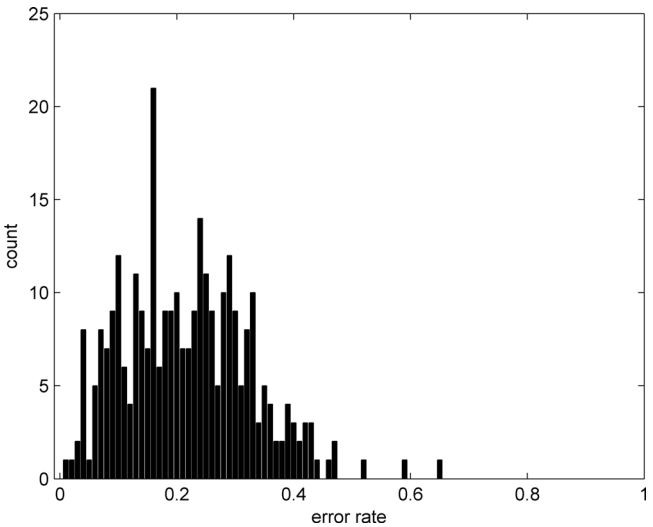

FIG. 3. Histogram showing distribution of overall error rate in 100 trials of a randomized three-fold cross-validation experiment (total $N = 300$). The mean error is $0.22 \pm 0.11\sigma$ with a median of 0.21.

TABLE II. Error rate statistics summarizing 100 three-fold cross-validation trials. The mean error (μ), standard deviation (σ), and median for each species represents single-species identification rate over all trials. In contrast, the mean overall error rate is computed by determining the percentage of misclassified click groups across all species. Cuvier's beaked whale, Risso's dolphin, and Pacific white-sided dolphins all had mean error rates below 7%.

Species	μ	σ	Median
Bottlenose dolphin	0.682	0.322	0.845
Cuvier's beaked whale	0.013	0.022	0.000
Long-beaked common dolphin	0.531	0.354	0.606
Pacific white-sided dolphin	0.063	0.069	0.034
Risso's dolphin	0.044	0.047	0.056
Short-beaked common dolphin	0.216	0.185	0.182
Overall	0.222	0.109	0.213

each species is selected, it is representative of a single-species identification error rate over all trials. In contrast, the mean overall error rate is computed by determining the percentage of misclassified click groups across all species, and the summary statistics are thus different. Classification performance for Cuvier's beaked whales, Pacific white-sided dolphins, and Risso's dolphins was exemplary with mean error rates of 0.01 $\pm$ 0.06σ, 0.06 $\pm$ 0.07σ, and 0.04 $\pm$ 0.05σ, respectively. Short-beaked common dolphins were distinguishable from the other species (0.22 $\pm$ 0.19σ), but the system had difficulties distinguishing long-beaked common (0.53 $\pm$ 0.35σ) and bottlenose dolphins (0.68 $\pm$ 0.32σ). Histograms showing the distribution of errors by species are shown in Fig. 4. A normalized correlation analysis was performed to determine how species-specific error varied across the 100 three-fold cross-validation trials. Negative correlations between species are indicative of a tendency to confuse the two species. Table III shows that the species with the best performances are relatively uncorrelated from each other and from the poorer performing species, with the exception of Pacific white-sided dolphins which showed a -0.18 correlation with long-beaked common dolphins. Bottlenose, short-beaked, and long-beaked dolphins exhibited variable performance.

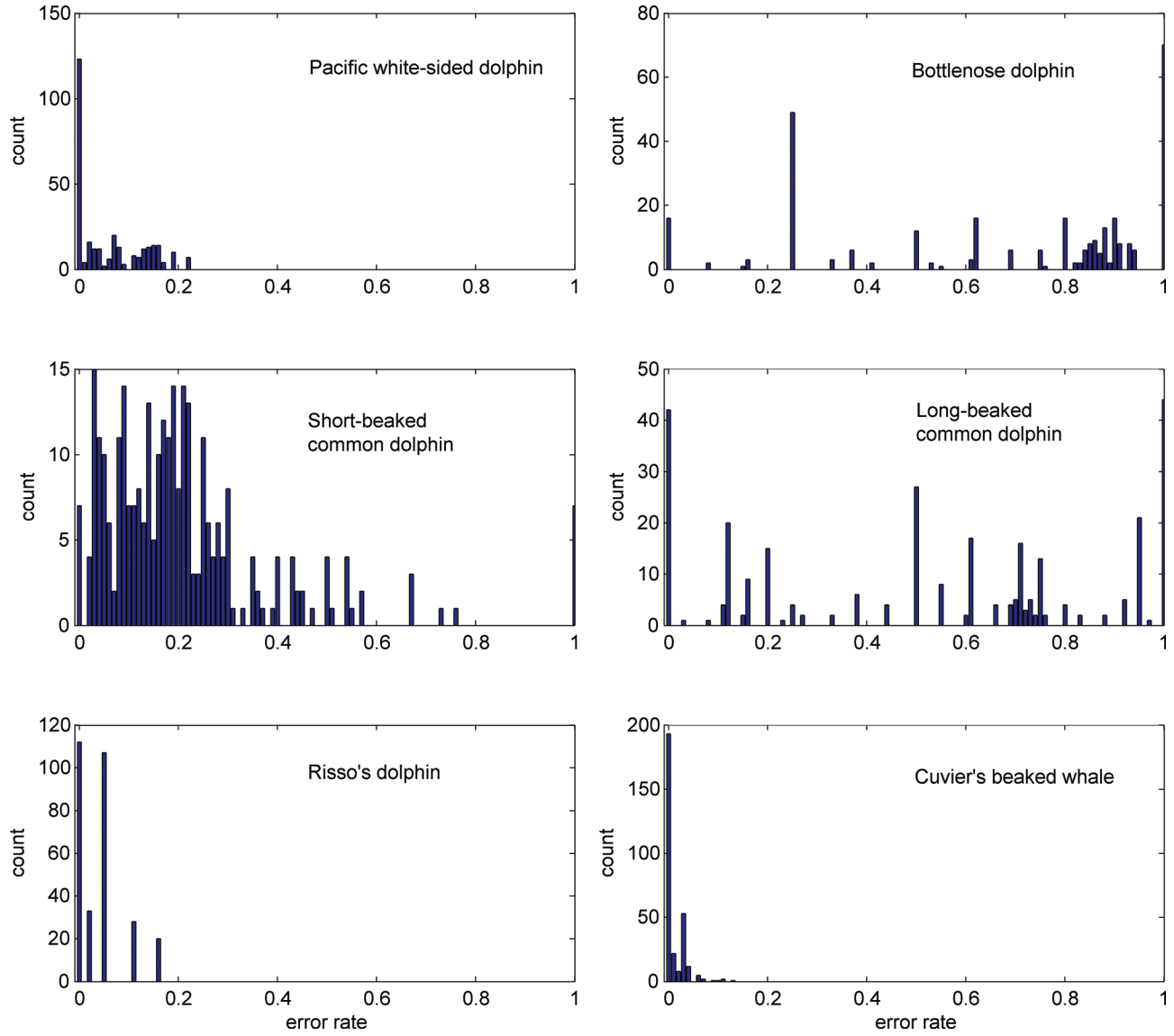

FIG. 4. (Color online) Distribution of error in 100 randomized three-fold cross-validation experiments reported by species. Mean, standard deviation, and median error rates are given in Table 2.

TABLE III. Correlation between per species error rates across 100 three-fold cross-validation trials. More negative values indicate higher levels of misclassification.

	Bottlenose dolphin	Cuvier's beaked whale	Long-beaked common dolphin	Pacific white-sided dolphin	Risso's dolphin	Short-beaked common dolphin
Bottlenose dolphin	1.00					
Cuvier's beaked whale	0.02	1.00				
Long-beaked common dolphin	0.08	0.09	1.00			
Pacific white-sided dolphin	−0.06	0.02	−0.18	1.00		
Risso's dolphin	−0.09	0.00	0.06	0.05	1.00	
Short-beaked common dolphin	−0.17	−0.02	−0.18	0.20	0.06	1.00

IV. DISCUSSION

Good feature extraction techniques are critically important for proper classification techniques. Without removal of confounding sources (e.g., echosounders), systems are likely to learn characteristics other than what the experimenter intended. While the cepstral features have been shown empirically to be an effective representation of the spectral shape, features which do not represent energy outside of the echolocation click bandwidth or exploit any potential timing differences in the signal may be fruitful areas for exploration.

Although not reported in detail here, we also conducted experiments which subtracted the spectral mean from the spectral magnitude before computing the discrete cosine transform which is the final step in cepstral feature extraction. This resulted in a slight degradation in performance which suggests that our noise estimator may be contaminated. While we provide no formal measure of our click detector performance, it has been designed to detect clicks with high confidence and to minimize the number of false positives. It may be that introduction of clicks rejected by the detector is responsible for the failure of spectral means subtraction to yield better results.

It should be noted that we have not used ICI, or time between clicks, as a component of our feature vectors. ICI can be indicative of species, but reliable extraction of ICI can be difficult when examining species that aggregate into large groups. Such species include common dolphins where mean group sizes are measured in the hundreds (Barlow and Forney, 2007), and groups of thousands sometimes occur. As a consequence, in this study we chose not to use ICI, although we do believe that for smaller group sizes this may be an appropriate component of the feature vector. Any such study would need to take into account ICI variability (Au *et al.*, 1974), which is likely to depend on both behavioral (Johnson *et al.*, 2004) and environmental (Akamatsu *et al.*, 2000; Simard *et al.*, 2010) factors.

While overall results showed the ability to distinguish four of the six species well, the high confusion rate between long-beaked common dolphins and bottlenose dolphins was unexpected. We had envisioned that distinguishing between long- and short-beaked common dolphins would be the most difficult task due to their similar morphology (Heyning and Perrin, 1994); visual field identification of the two is difficult enough that group misidentification or failure to detect mixed species groups is possible.

After examining low- versus high-error-rate experiments, it was discovered that much of the error could be accounted for by the FLIP sighting of long-beaked common dolphins. Manual inspection of the selected echolocation clicks did not show inordinate signs of bad detections such as clipped calls. While the sighting reports described behavior that was consistent with common dolphins and did not note the presence of any other groups, it is impossible to completely rule out the possibility that bottlenose dolphins producing echolocation clicks may have been within acoustic range. It is unlikely that this can be explained entirely by channel mismatch. While the FLIP long-beaked common dolphin recordings used the 300 series preamp which was significantly different than the 100 series preamps used in the systems that provided training data for the long-beaked common dolphin models, all of the bottlenose dolphin data were recorded with systems that used 100 series preamps. Acoustic data associated with two pairs of bottlenose and long-beaked common dolphin sightings were even recorded on the same cruise. Other explanations include the possibility that data associated with bottlenose dolphin models was not parameterized effectively. Bottlenose dolphin clicks are known to have peak frequencies of up to 130 kHz (Au *et al.*, 1974), and the 92 kHz upper frequency limit along with the rejection of clicks with peak levels greater than 70 kHz may have limited the ability to distinguish these two species.

The 100 randomized three-fold cross-validation trials were selected from a set of over 282 million different possible combinations of training and test data. Due to the method in which sightings were first shuffled then assigned to folds with the last fold having left over sightings [e.g., for four sightings, the folds might be (3), (1), (2,4)], each fold contains either $\lfloor \frac{N}{3} \rfloor$ or $\lfloor \frac{N}{3} \rfloor + \mathrm{mod}(N,3)$ sightings where N is the number of sightings, $\lfloor \cdot \rfloor$ denotes the floor operator, and mod (a,b) is the remainder of a/b. It is possible for different permutations to result in some of the cross-fold tests having identical training and test partitions. To continue with the four sighting example, if the sightings assigned to folds were (1), (2), (3,4), both this example and the previous one would have one of their trials use training data from sightings (2,3,4) and test on sighting (1). Once either the training or test set is decided, the other is implied. Thus, to count the unique number of combinations, it suffices to count the number of possible ways that sightings can be assigned to the test set. Consequently, the number of unique partitions of the training and test data per species is:

TABLE IV. Mean group size rounded to the nearest integer and the number of visual observations contributing to the mean as reported by Barlow and Forney (2007). Species with high mean group sizes aggregate in large groups at times.

Species	Mean group size	Sightings
Bottlenose dolphin	13	31
Cuvier's beaked whale	3	3
Long-beaked common dolphin	287	16
Pacific white-sided dolphin	34	15
Risso's dolphin	15	50
Short-beaked common dolphin	168	239

$$TestPartitions(N) = \begin{cases} \binom{N}{\lfloor \frac{N}{3} \rfloor} & \mathrm{mod}(N,3) = 0 \\ \binom{N}{\lfloor \frac{N}{3} \rfloor} + \binom{N}{\lfloor \frac{N}{3} \rfloor + \mathrm{mod}(N,3)} & \mathrm{mod}(N,3) > 0 \end{cases}$$

where (N/k) is the binomial coefficient, $N!/k!(N-k)!$. As the partitioning for each species is independent of all the others, the total number of permutations of training and test data can be obtained by taking the product of the number of ways to choose one train/test split from each species, or:

$$\left(\binom{4}{1} + \binom{4}{2}\right)^2 \left(\binom{8}{2} + \binom{8}{4}\right)^3 \binom{3}{1} = 282\,357\,600.$$

The number of individuals represented in each training fold is dependent on the group size and what portion of the group is vocalizing. Table IV provides the mean group size for the species in this study based on the data of Barlow and Forney (2007). As can be seen, group sizes vary greatly and a high number of sightings for species with small group sizes such as Cuvier's beaked whales can still result in a relatively low number of overall animals. In contrast, there were only four sightings of long-beaked common dolphins used in this study, yet the average group size of visual observations reported for this species is 287 animals which is likely to result in a high number of individuals being sampled. There is also a possibility that the same animal may be encountered multiple times. In spite of the limitations of the dataset, the authors believe this study to represent a reasonable estimation of what one might expect for field performance within this geographic region.

V. CONCLUSIONS

We have demonstrated that echolocation clicks can be used to distinguish six species of odontocetes within the Southern California Bight with a mean error rate of $0.22 \pm 0.11\sigma$ (median 21%) using cepstral features of echolocation clicks and GMMs. While this is competitive with other methods, direct comparison is not possible due to differences in datasets, recording equipment, etc. In spite of the high variability of free-ranging odontocete echolocation clicks due to off-axis effects, high-frequency attenuation at distance, and the ability of animals to control aspects of click production, the relationship between echolocation click properties and animal morphology makes these vocalizations an excellent target for bioacoustics work that determines an animal's species from the sounds that they produce.

Performance varied by species, with the presumed Cuvier's beaked whale clicks being best classified followed by Risso's and Pacific white-sided dolphins. All three species had error rates under 0.07. Short-beaked common dolphins exhibited significantly higher error rates ($\mu = 0.22$, median 0.18), while the long-beaked and bottlenose dolphins were difficult to distinguish. These results hold across randomizations of training and test data and across recording systems and environmental conditions, ensuring that data from the same sighting were not used for both training and testing. In many cases, it is likely that different individual animals were recorded, but as individuals were not identified this cannot be stated with certainty.

It is the authors' opinion that the most fruitful area for continued work in this area is to work toward appropriate low-dimensional characterizations of echolocation clicks. The sensitivity of the echolocation click signal to animal orientation with respect to the hydrophone and high-frequency attenuation contributes to making this a challenging problem. Techniques such as cepstral analysis, which captures spectral shape, or Gillespie and Caillat's (2008) use of the spectrum and measures derived from it, are steps toward this, but much work remains to be done.

ACKNOWLEDGMENTS

We would like to thank our colleagues at Cascadia Research Collective and the Scripps Whale Acoustics Lab who provided visual confirmations on our sightings, especially John Calambokidis, Dominique Camacho, Stephen Claussen, Annie Douglas, Erin Falcone, Greg Falxa, Andrea Havron, Allan Ligon, Megan McKenna, Greg Campbell, Jen Quan, Greg Schorr, and Michael Smith, also the crews of Cal-COFI, the R/V Sproul, and the R/P Flip. We thank Brent Hurley for calibrating our preamps, Greg Campbell and Liz Henderson for their help with sighting data and array configurations. Yoav Freund was kind enough to verify our combinatorial analysis. Thanks are also due to Ted Cranford, Mark McDonald, and Sean Wiggins, who have always been generous with their time whenever we had questions about click production or our recording systems as well as the two anonymous reviewers for their helpful comments on a previous version of this manuscript. This work was sponsored by the U.S. Navy Chief of Naval Operations N45, Curt Collins, Frank Stone, and Ernie Young, under projects N00244-07-1-0005, N00244-07-1-0011, N00244-08-1-0029, N00244-09-1-0079, and N00244-10-1-0047, and the Office of Naval Research, Jim Eckman and Michael Weise, under projects N00014-08-1-119, N00244-08-1-0029, N00014-08-1-1082, and N00244-09-1-0079, and N00014-10-1-0387. This is National Oceanic and Atmospheric Administration Pacific Marine Environmental Laboratory contribution #3584.

Akamatsu, T., Wang, D., Wang, K. X., and Naito, Y. (**2000**). "A method for individual identification of echolocation signals in free-ranging finless porpoises carrying data loggers," J. Acoust. Soc. Am. **108**, 1353–1356.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York), Chap. 7, pp. 104–108.
Au, W. W. L., Floyd, R. W., Penner, R. H., and Murchison, A. E. (**1974**). "Measurement of echolocation signals of the Atlantic bottlenose dolphin, *Tursiops truncatus* Montagu, in open waters," J. Acoust. Soc. Am. **56**, 1280–1290.
Barlow, J., and Forney, K. A. (**2007**). "Abundance and population density of cetaceans in the California current ecosystem," Fish. Bull. **105**, 509–526.
Benoit-Bird, K. J., and Au, W. W. L. (**2009**). "Phonation behavior of cooperatively foraging spinner dolphins," J. Acoust. Soc. Am. **125**, 539–546.
Bimbot, F., Bonastre, J. F., Fredouille C., Gravier, G., Magrin-Chagnolleau, I., Meignier, S., Merlin, T., Ortega-Garcia, J., Petrovska-Delacretaz, D., and Reynolds, D. A. (**2004**). "A tutorial on text-independent speaker verification," EURASIP J. Appl. Signal Process. **2004**, 430–451.
Cranford, T. W. (**2000**). "In search of impulse sound sources in odontocetes," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 109–155.
Deecke, V. B., Ford, J. K. B., and Spong, P. (**1999**). "Quantifying complex patterns of bioacoustic variation: Use of a neural network to compare killer whale (*Orcinus orca*) dialects," J. Acoust. Soc. Am. **105**, 2499–2507.
Dempster, A. P., Laird, N. M., and Rubin, D. B. (**1977**). "Maximum likelihood from incomplete data via the EM algorithm," J. R. Stat. Soc. Ser. B (Methodol.) **39**, 1–38.
Downie, J. S. (**2008**). "The music information retrieval evaluation exchange (2005–2007): A window into music information retrieval research," Acoust. Sci. Tech. **29**, 247–255.
Duda, R. O., Hart, P. E., and Stork, D. G. (**2001**). *Pattern Classification* (Wiley-Interscience, New York), pp. 524–526.
Falcone, E. A., Schorr, G. S., Douglas, A. B., Calambokidis, J., Henderson, E., McKenna, M. F., Hildebrand, J., and Moretti, D. (**2009**). "Sighting characteristics and photo-identification of Cuvier's beaked whales (*Ziphius cavirostris*) near San Clemente Island, California: A key area for beaked whales and the military?," Mar. Biol. (Berlin) **156**, 2631–2640.
Fisher, F. H., and Spiess, F. N. (**1963**). "FLIP-floating instrument platform," J. Acoust. Soc. Am. **35**, 1633–1644.
Fristrup, K. M., and Watkins, W. A. (**2004**). "Characterizing acoustic features of marine mammal sounds," Technical Report No. WHOI-92-04, Woods Hole Oceanographic Institute, Woods Hole, MA.
Gerard, O., Carthel, C., Coraluppi, S., and Wilett, P. (**2008**). "Feature-aided tracking for marine mammal detection and classification," Can. Acoust. **36**, 27–33.
Gillespie, D. (**2004**). "Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram," Can. Acoust. **32**, 39–47.
Gillespie, D., and Caillat, M. (**2008**). "Statistical classification of odontocete clicks," Can. Acoust. **36**, 20–26.
Harland, E. (**2008**). "Processing the workshop datasets using the trud algorithm," Can. Acoust. **36**, 20–26.
Heyning, J. E., and Perrin, W. F. (**1994**). "Evidence for two species of common dolphins (genus Delphinus) from the Eastern North Pacific," Contrib. Sci. (Los Angeles) **442**, 1–35.
Huang, X., Acero, A., and Hon, H. W. (**2001**). *Spoken Language Processing* (Prentice Hall, Upper Saddle River), pp. 172–175.
Jarvis, S. M., DiMarzio, N. A., Morrissey, R. P., and Moretti, D. J. (**2008**). "A novel multi-class support vector machine classifier for automated classification of beaked whales and other small odontocetes," Can. Acoust. **36**, 34–40.
Johnson, M., Madsen, P. T., Zimmer, W. M. X., de Soto, N. A., and Tyack, P. L. (**2004**). "Beaked whales echolocate on prey," Proc. R. Soc. London, Ser. B **271**, S383–S386.
Kaiser, J. F. (**1990**). "On a simple algorithm to calculate the "energy" of a signal," in *Proceedings of the IEEE International Conference of the Acoustical Speech, and Signal Processsing*, Albuquerque, NM, pp. 381–384.
Kandia, V., and Stylianou, Y. (**2006**). "Detection of sperm whale clicks based on the Teager–Kaiser energy operator," Appl. Acoust. **67**, 1144–1163.
Lammers, M. O., and Castellote, M. (**2009**). "The beluga whale produces two pulses to form its sonar signal," Biol. Lett. **5**, 297–301.
Linde, Y., Buzo, A., and Gray, R. M. (**1980**). "An algorithm for vector quantizer design," IEEE Trans. Commun. **28**, 84–95.
Madsen, P. T., Johnson, M., Aguilar de Soto, N., Zimmer, W. M. X., and Tyack, P. (**2005**). "Biosonar performance of foraging beaked whales (*Mesoplodon densirostris*)," J. Exp. Biol. **208**, 108–191.
Marques, T. A., Thomas, L., Ward, J., DiMarzio, N., and Tyack, P. L. (**2009**). "Estimating cetacean population density using fixed passive acoustic sensors: An example with Blainville's beaked whales," J. Acoust. Soc. Am. **125**, 1982–1994.
Mellinger, D. K., and Clark, C. W. (**2000**). "Recognizing transient low-frequency whale sounds by spectrogram correlation," J. Acoust. Soc. Am. **107**, 3518–3529.
Mellinger, D. K., Stafford, K. M., Moore, S. E., Dziak, R. P., and Matsumoto, H. (**2007**). "An overview of fixed passive acoustic observation methods for cetaceans," Oceanography **20**, 36–45.
Møhl, B., Wahlberg, M., Madsen, P. T., Heerfordt, A., and Lund, A. (**2003**). "The monopulsed nature of sperm whale clicks," J. Acoust. Soc. Am. **114**, 1143–1154.
Moretti, D., DiMarzio, N., Morrissey, R., Mellinger, D. K., Heimlich, S., and Pettis, H. (**2008**). "Overview of the 3rd International Workshop on the Detection and Classification of Marine Mammals Using Passive Acoustics," Can. Acoust. **36**, 7–11.
Oswald, J. N., Rankin, S., Barlow, J., and Lammers, M. O. (**2007**). "A tool for real-time acoustic species identification of delphinid whistles," J. Acoust. Soc. Am. **122**, 587–595.
Picone, J. W. (**1993**). "Signal modeling techniques in speech recognition," Proc. IEEE **81**, 1215–1247.
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.
Roch, M. A., Soldevilla, M. S., Hoenigman, R., Wiggins, S. M., and Hildebrand, J. A. (**2008**). "Comparison of machine learning techniques for the classification of echolocation clicks from three species of odontocetes," Can. Acoust. **36**, 41–47.
Simard, P., Hibbard, A. L., McCallister, K. A., Frankel, A. S., Zeddies, D. G., Sisson, G. M., Gowans, S., Forys, E. A., and Mann, D. A. (**2010**). "Depth dependent variation of the echolocation pulse rate of bottlenose dolphins (*Tursiops truncatus*)," J. Acoust. Soc. Am. **127**, 568–578.
Sirovic, A., Hildebrand, J. A., Wiggins, S. M., McDonald, M. A., Moore, S. E., and Thiele, D. (**2004**). "Seasonality of blue and fin whale calls and the influence of sea lee in the Western Antarctic Peninsula," Deep-Sea Res. Part II-Top. Stud. Oceanogr. **51**, 2327–2344.
Soldevilla, M. S., Henderson, E. E., Campbell, G. S., Wiggins, S. M., Hildebrand, J. A., and Roch, M. A. (**2008**). "Classification of Risso's and Pacific white-sided dolphins using spectral properties of echolocation clicks," J. Acoust. Soc. Am. **124**, 609–624.
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.
Wiggins, S. M., and Hildebrand, J. A. (**2007**). "High-frequency acoustic recording package (HARP) for broad-band, long-term marine mammal monitoring," in *Proceedings of the International Symposium on Underwater Technology*, Tokyo, Japan, pp. 551–557.
Zimmer, W. M. X., Harwood, J., Tyack, P. L., Johnson, M. P., and Madsen, P. T. (**2008**). "Passive acoustic detection of deep-diving beaked whales," J. Acoust. Soc. Am. **124**, 2823–2832.
Zimmer, W. M. X., Johnson, M. P., Madsen, P. T., and Tyack, P. L. (**2005a**). "Echolocation clicks of free-ranging Cuvier's beaked whales (*Ziphius cavirostris*)," J. Acoust. Soc. Am. **117**, 3919–3927.
Zimmer, W. M. X., Madsen, P. T., Teloni, V., Johnson, M. P., and Tyack, P. L. (**2005b**). "Off-axis effects on the multipulse structure of sperm whale usual clicks with implications for sound production," J. Acoust. Soc. Am. **118**, 3337–3345.