Compensating for the effects of site and equipment variation on delphinid species identification from their echolocation clicks

Marie A. Roch[a) and Johanna Stinner-Sloan

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

Simone Baumann-Pickering and Sean M. Wiggins

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

A concern for applications of machine learning techniques to bioacoustics is whether or not classifiers learn the categories for which they were trained. Unfortunately, information such as characteristics of specific recording equipment or noise environments can also be learned. This question is examined in the context of identifying delphinid species by their echolocation clicks. To reduce the ambiguity between species classification performance and other confounding factors, species whose clicks can be readily distinguished were used in this study: Pacific white-sided and Risso's dolphins. A subset of data from autonomous acoustic recorders located at seven sites in the Southern California Bight collected between 2006 and 2012 was selected. Cepstral-based features were extracted for each echolocation click and Gaussian mixture models were used to classify groups of 100 clicks. One hundred Monte-Carlo three-fold experiments were conducted to examine classification performance where fold composition was determined by acoustic encounter, recorder characteristics, or recording site. The error rate increased from 6.1% when grouped by acoustic encounter to 18.1%, 46.2%, and 33.2% for grouping by equipment, equipment category, and site, respectively. A noise compensation technique reduced error for these grouping schemes to 2.7%, 4.4%, 6.7%, and 11.4%, respectively, a reduction in error rate of 56%–86%.
© 2015 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4904507]

I. INTRODUCTION

The application of passive acoustic monitoring for marine mammals is widespread (e.g., Au and Hastings, 2008; Mellinger *et al.*, 2007; Zimmer, 2011) and is a component of many studies for impact and mitigation assessment (e.g., Blackwell *et al.*, 2013; Moretti *et al.*, 2014). For studies that use machine learning techniques on data from multiple sites, over long periods, or that were acquired using different equipment, researchers must ask the question: Did the pattern recognition algorithms learn characteristics of the calls that were studied, or were they confounded by variations in noise, propagation environment, recording equipment, or other non-call related parameters?

Researchers in other fields have had to address this same issue. In telephone-based speaker recognition, it has been shown that carbon button versus electret microphones, which have nonlinear and linear transfer functions, respectively, can impact classification performance (Reynolds, 1996). Similarly, Whitman *et al.* (2001) demonstrated that music artist recognition algorithms frequently learned characteristics of a recording studio session, the so-called "album effect," rather than the artist. Various techniques have been proposed to cope with sources of convolutional and additive noise. Spectral means subtraction (Boll, 1979; see Huang *et al.*, 2001, p. 516 for a more general description) is a common method to deal with additive noise where the noise spectrum is estimated from periods without the signal of interest or over very long periods when the signal of interest is nonstationary. For classifiers that use cepstral features, subtracting the mean in the cepstral domain can remove convolutional noise (Atal, 1974) with similar restrictions on the mean estimation. Adaptive methods have been proposed (e.g., Weiner filtering, see Huang *et al.*, 2001, pp. 520–522 for an overview) as well as statistical techniques that normalize the distribution of likelihood scores after classification by a probabilistic model (e.g., Auckenthaler *et al.*, 2000; Dunn *et al.*, 2001). In this study, we show that straightforward techniques can produce dramatic reductions in the increased error rate that occurs with instrument and site mismatch between training and test data.

Our study of odontocete species identification from their echolocation clicks examines this question by partitioning training and test data such that all test data are novel with respect to the partitioning criterion, e.g., if the criterion is recording location, models trained using data from a specific set of sites are tested with data from alternative sites. To reduce confounding factors, we focus this study on the classification of two species, Pacific white-sided dolphins (*Lagenorhynchus obliquidens*) and Risso's dolphins (*Grampus griseus*). Soldevilla *et al.* (2008) showed that

---

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

encounters from these species could be readily distinguished by spectral patterns of peaks and troughs of what are likely to be off-axis echolocation clicks. This knowledge is used to ground truth data from deployments of autonomous acoustic recorders at seven locations in the Southern California Bight.

We show that these species can be reliably separated using methods similar to those in Roch *et al.* (2011) when splitting training and test data so that each acoustic encounter is entirely in the training or test partition, and that this performance severely degrades when partitioned by equipment type or site. We introduce modifications to our signal processing chain that provide large reductions in error rate for cross-equipment and cross-site test configurations.

II. METHODS

A. Data acquisition and selection

High-frequency acoustic recording packages (HARPs, Wiggins and Hildebrand, 2007) were deployed at seven sites in the Southern California Bight from 2006 to 2012 (Fig. 1). While the instruments all used the same type of calibrated hydrophone sensor (ITC 1042, International Transducer Corp., Santa Barbara, CA) with a flat response across the analysis bandwidth, there were frequency dependent differences in the 15 preamplifier gains used in this study (Fig. 2) and possible differences in electronics noise between preamplifiers. Recording was continuous at a sample rate of 200 kHz. The instrument's high frequency data stream from the ITC 1024 was added to the stream from a set of low frequency sensors and quantized with 16 bits. The frequency response of the low frequency sensors lied outside the frequency range of echolocation click spectra.

Trained analysts examined acoustic recordings from these deployments for echolocation clicks with spectra characteristic of Pacific white-sided dolphins and Risso's dolphins and identified 71 days with data that were used in this study. Specifically, the analysts looked for the distinguishing spectral banding patterns (Fig. 3) of these species (Soldevilla *et al.*, 2008, 2010): Peaks near 22, 25, 31, and
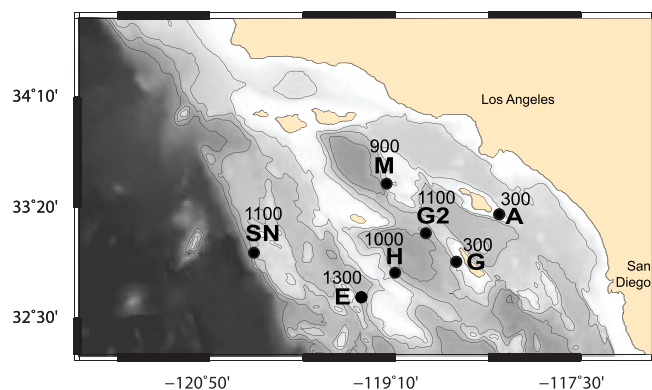

FIG. 1. (Color online) Seven bottom-moored acoustic recorder deployment locations in the Southern California Bight. Each deployment location is labeled with a site name and approximate depth in meters. Bathymetric contour lines are at 500 m increments with darker gray shading representing deeper depths. Land masses, including islands, are shaded areas located inside white regions or indicated with city names.
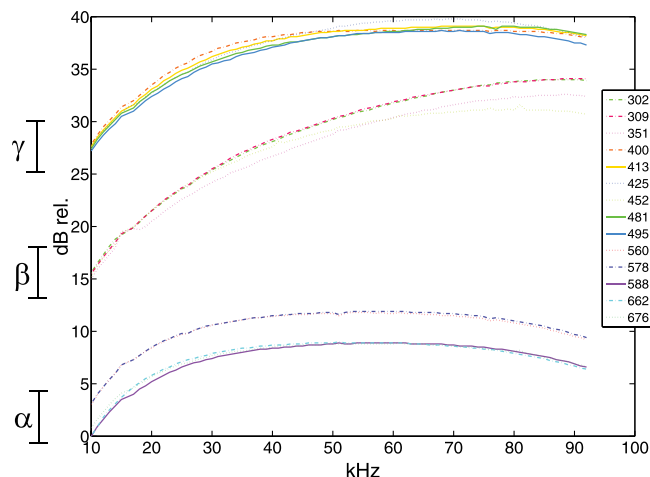

FIG. 2. (Color online) Fourteen preamplifier transfer function calibrations for the instruments used in this study. The calibration for the 15th preamplifier (306) was corrupted and the transfer function of preamplifier 309 which had similar characteristics was used as a proxy. Symbols α, β, and γ show groups of preamplifiers with similar transfer functions.

39 kHz for Risso's dolphins, and either at 22, 27.5 and 39 or 22, 26, and 37 kHz for Pacific white-sided dolphins. Search for the peaks was done by examining long-term spectral average spectrograms (Wiggins and Hildebrand, 2007) consisting of 5 s averages of spectra with 100 Hz resolution. Each 5 s average was generated by the Welch (1967) method with no overlap and a Hann window. Analysts labeled the start and end of acoustic encounters which were defined as the occurrence of echolocation click trains being more than 1 h apart. Analysts were conservative in their labeling, and groups containing what appeared to be multiple types of echolocation click spectra or strong noise (e.g., ship traffic) were not selected.

B. Signal processing chain and classification

1. Echolocation click detection and feature extraction

Echolocation clicks were found in the recordings using a two-pass process similar to that described in Roch *et al.* (2011). The first pass identified regions containing potential clicks based on the energy spread in spectra created with 10 ms Kaiser windows advanced every 5 ms. Spectral means estimated for each frequency bin over long data blocks (approximately 4 min) and the mean resulting spectra was subtracted from each individual spectra to estimate the signal-to-noise ratio (SNR). Regions for further exploration were identified when 10% or more of the frequency bins above 10 kHz exceeded an SNR of 13 dB, a high threshold designed to identify regions with a strong potential for echolocation activity. Regions of potential echolocation activity less than 0.5 s from one another were combined into longer regions, and isolated regions discarded.

Operating on approximately 30 s blocks of data, the second pass detected individual echolocation clicks within the regions of interest identified in the first pass. The time-series waveform data were high-pass filtered with an equiripple filter and a transition band between 3 and 8 kHz. Detections were identified by examining the Teager energy
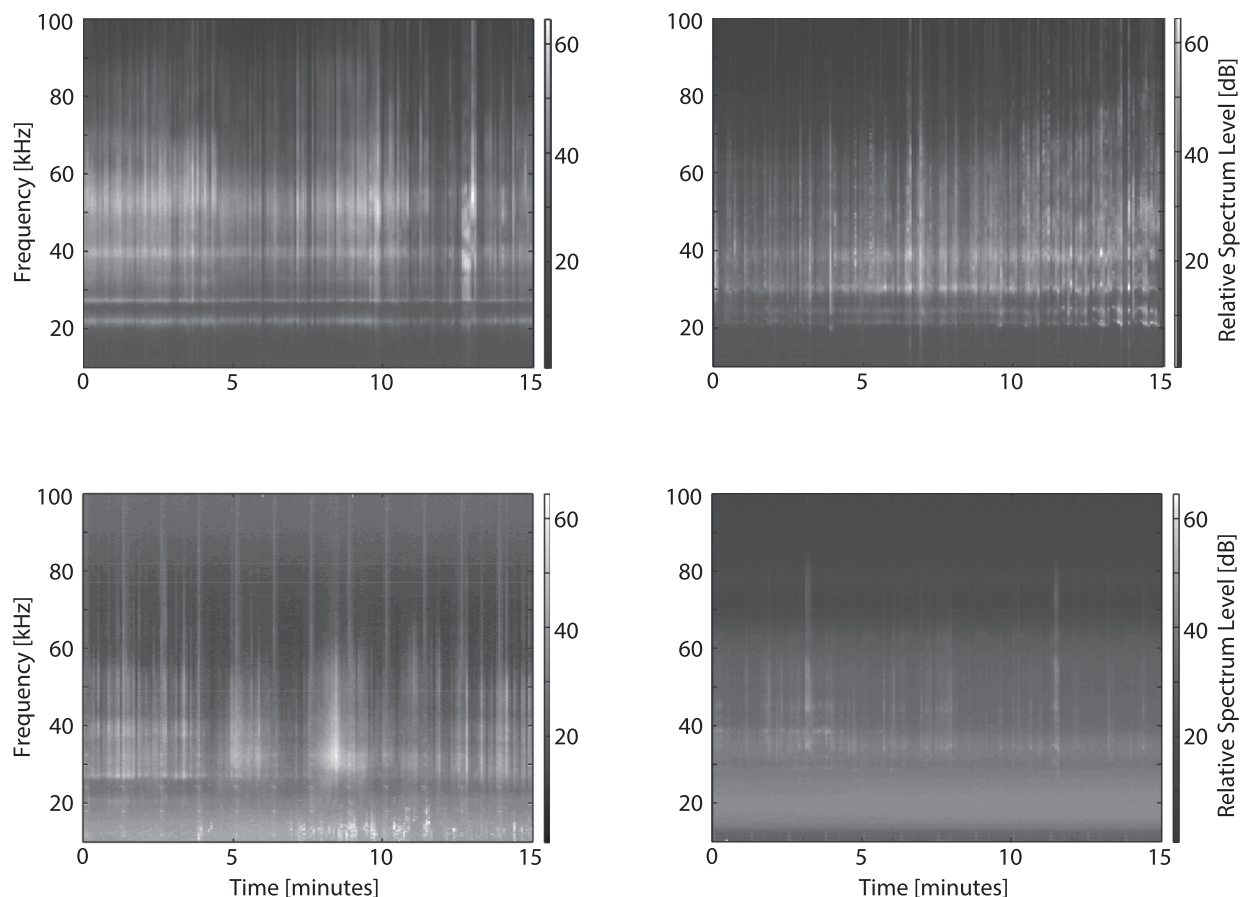
FIG. 3. Examples of 15 min long-term spectral averages uncorrected for hydrophone transfer function. The upper-left and -right panels show examples of presumed Pacific white-sided and Risso dolphin echolocation clicks, respectively, based on spectral peak structure. The lower two panels are examples of suspected encounters of the same species that were rejected by analysts. The Pacific white-sided dolphin encounter in the lower-left was rejected as possibly containing mixed species (change in click structure near the middle of displayed spectra accompanied by the onset of whistles between 10 and 20 kHz) and the Risso's dolphin encounter on the right was rejected as being too faint to show the peak structure. Note that brightness and contrast have been adjusted for each panel to best display the spectra.

(Kaiser, 1990) which was first applied to the detection of echolocation clicks by Kandia and Stylianou (2006). The Teager energy was smoothed with a non-causal 55 μs moving average filter centered about the current energy sample (11 samples at 200 kHz sampling rate).

The distribution of the Teager energy is skewed toward higher energy when echolocation clicks are present (Kandia and Stylianou, 2006), and a detection threshold was set based on an outlier test (Emerson and Strenio, 1983, pp. 59–60) on the natural logarithmic transformed smoothed Teager energy

$$\text{outlier}(x) = \begin{cases} \text{true} & x \geq Q_3 + k(Q_3 - Q_1) \\ \text{false} & \text{otherwise,} \end{cases} \quad (1)$$

where Q_i is the ith quartile of the data and k is a scale factor. A value of $k = 3$ was used to find extreme outliers ($k = 1.5$ is the canonical outlier test) and regions that exceeded this threshold for 10 μs or more were subjected to further analysis. These regions were grown both forward and backward until the Teager energy no longer exceeded a threshold that was slightly outside the third quartile [Eq. (1), $k = 0.5$] or they reached the midpoint between the current potential click and its predecessor or successor. When the preceding or

subsequent region abutted the current one, they were merged. Detections from this two-stage process were spot checked at each site to ensure a low incidence of false positives, but false positive rates were not estimated as the rates fluctuate both spatially and temporally due to environmental factors such as propagation environment and noise. To provide an effective estimate of the false positive rate would require an experimental design that takes into account variables such as site and seasonality over the 7 years during which data were collected and is beyond the purview of this article.

The unfiltered time series of each detected click was windowed with a Blackman-harris window and zero-padded to a standard length. The discrete Fourier transform spectra had a standard interpolated frequency resolution of 240 Hz and transformed to relative dB. Spectra were adjusted for the preamplifier's transfer function (Fig. 2). The calibration of one preamplifier board, 306, was corrupted and the calibration from another board in the same series (309) was used as a proxy. Except for very close on-axis echolocation clicks which occur infrequently and low SNR clicks that would not be detected, the peak frequency of the echolocation clicks for Risso's and Pacific white-sided dolphins was typically between 15 and 50 kHz. All echolocation clicks with peaks

outside of this range were removed from analysis and the first 14 points of the cepstrum (Roch *et al.*, 2011) were used as estimates of the spectral shape for the remaining echolocation clicks.

2. Noise estimation

A major concern in the noise estimation was to prevent the energy of weak undetected echolocation clicks from contaminating the noise estimate. As such, a lower Teager energy threshold was established [Eq. (1), $k = 0.5$] and acoustic data corresponding to regions of the Teager energy that fell beneath this threshold were considered to be noise.

Noise within an acoustic encounter was estimated around regions of low echolocation activity as measured by the weak threshold echolocation detector. The regions from which noise was estimated were defined by searching for periods when click rates were low and hence the corresponding interclick interval (ICI) was high as measured by an order-statistic based threshold. ICIs within 30 s blocks were sorted and the threshold was empirically set to three times the ICI value occurring at the 90th percentile position, thus resulting in the detection of very high outliers.

Samples that fell in these regions, free of even weakly detected clicks, were used for noise estimation. Spectral frames were computed from these noise regions using the same signal processing chain as used for spectra of echolocation clicks. The order-truncate average described in Helble *et al.* (2012) was used to estimate the noise from these frames. Each frequency bin in the magnitude-squared transfer function corrected spectra was sorted. Within a sorted frequency bin, a mean was taken such that the span between the lowest and highest value covering 90% of the frames was minimized, thus corresponding to the mean of the most tightly packed values of the distribution.

In experiments that used noise compensation, the magnitude squared noise estimate was converted to dB and subtracted from the echolocation spectra, yielding the SNR spectra of echolocation clicks which were subsequently converted to cepstra.

3. Classification

The classification methodology is similar to that reported in Roch *et al.* (2011), but was modified to increase the variability of training data and to evaluate the effects of site and equipment variability. Data were grouped according to one of three criteria. The first was a grouping of acoustic encounter, similar to our use of sighting in previous work. The second grouping was based on the custom preamplifier boards used with our acoustic recorders. The preamplifiers were either partitioned by an individual preamplifier or grouped into three sets of preamplifiers with similar characteristics (α, β, and γ in Fig. 2). The final criterion used recording site to group echolocation click features.

Within each grouping criterion, groups were randomly permuted and assigned in a balanced manner to partitions of a three-fold experiment. Two folds were used to train a 16-mixture Gaussian mixture model (GMM) for each species. To increase the variability of the training data, a bootstrap

procedure selected 85% of the echolocation clicks randomly with replacement. All echolocation features in the remaining fold were grouped into trials of 100 sequential click feature vectors without violating the grouping criterion. Extraneous click features that did not form a complete set of 100 feature vectors were discarded. Partitioning was done on a per species basis, so in some of the permutations it might be possible to train a model for one species based on a specific environment and test using the other species in the same environment. As an example for site partitioning one could train a Risso's dolphin model with Risso's dolphin echolocation features from site G and test with Pacific white-sided dolphin echolocation features from site G. This represents a potentially more difficult classification problem than excluding all site G data as the unmodeled species feature data have a greater potential for matching any environmental or equipment conditions that the classifier may have learned.

Each set of feature vectors were scored against the Pacific-white sided and Risso's dolphin GMM models. Logarithm likelihood scores of the clicks were summed and the species associated with the model that produced the highest summed logarithm likelihood score was selected as the species that produced the echolocation click group. This process was repeated with one of the training folds moving to the role of test fold and vice versa until all three folds had been tested. The experiment was repeated 100 times, resulting in 300 different train/test configurations.

III. RESULTS

Over 450 000 echolocation clicks were detected for each species across a variety of sites and instruments (Table I) with site A providing the largest contribution. Performance across training criteria and presence/absence of noise compensation for 100 three-fold experiments is summarized in Table II and Fig. 4. In all experimental cases, the noise compensation provided reductions in error rates. Grouping by preamplifier or site resulted in large increases in error rates that were reduced with noise compensation, but still resulted in higher error rates than when different acoustic encounters from the same recording site or preamplifier were allowed to cross the train/test boundary.

To determine whether or not certain regions of the acoustic record were more difficult to classify, each test token (group of 100 clicks) was analyzed over the 100 classifications obtained from the bootstrapped training models. Errors for Pacific white-sided dolphins (Fig. 5) showed that there were indeed regions that were more difficult to classify than others and that noise compensation greatly reduced these regions of error. Although not shown, errors for Risso's dolphin echolocation clicks contain similar trends.

IV. DISCUSSION

Without noise compensation, error rates in these easily distinguished species were increased by factors of up to 7 when there were differences in site or instrument. In general, Risso's dolphins produced less classification errors than Pacific-white sided dolphins (Table II). A likely cause for part of the performance difference is that some of the

TABLE I. Detected echolocation click distribution of 943 621 echolocation clicks by site and preamplifier board.

| Site | Recorder depth m | Preamplifier board | Pacific white-sided dolphins | | Risso's dolphins | |
			Clicks	Encounters	Clicks	Encounters
A	300	302	0	0	61 460	4
		306	23 145	4	41 135	4
		309	151 056	5	86 100	4
		400	0	0	60 650	3
		425	18 196	3	94 745	4
		Subtotal	192 397	12	344 090	19
E	1300	452	10 291	4	2634	2
G	300	351	148 403	24	9751	1
		413	84 335	7	28 463	2
		Subtotal	232 738	31	38 214	3
G2	1100	481	103	1	1295	2
H	1000	676	11 923	1	0	0
M	900	560	10 686	2	0	0
		578	4388	2	949	1
		588	0	0	58 140	1
		662	0	0	22 719	1
		Subtotal	15 074	4	81 808	3
SN	1100	495	13 054	5	0	0
Total			475 580	68	468 041	29

TABLE II. Classification error rate percentages for 100 three-fold experiments. The overall error rate shows the percentage classification error rate for both species, their standard deviations, and median error rates. The next two groups of numbers show the metrics based on an accept/reject validation task for a specific species and provide an indication of how good the system is at classifying a specific species (numbers are unrelated to the overall error in that they only consider trials from the species in question). Each row represents a different grouping criterion with noise compensation applied or omitted. Application of noise compensation reduces the error rate by: 56% for acoustic encounter, 76% by individual preamplifier, 86% by preamplifier group, and 66% by site.

| | Overall | | | Pacific white-sided dolphins | | | Risso's dolphins | | |
	μ	σ	Med.	μ	σ	Med.	μ	σ	Med.
Encounter	6.1	4.7	4.8	6.9	5.7	5.3	5.3	7.5	2.9
Encounter noise compensation	2.7	1.3	2.4	4.5	2.9	4.0	0.9	0.6	0.8
Preamp	18.1	12.1	15.0	26.8	24.6	17.9	14.0	14.2	9.3
Preamp noise compensation	4.4	2.8	3.8	9.9	9.8	5.8	1.4	1.6	0.8
Preamp group	46.2	22.3	59.1	71.5	26.4	75.2	38.5	26.5	45.8
Preamp group noise compensation	6.7	1.8	7.1	18.2	11.0	12.8	1.5	1.5	0.9
Site	33.2	22.8	26.3	40.6	34.0	29.9	23.5	26.0	13.2
Site noise compensation	11.4	14.3	4.4	21.0	19.1	17.4	1.8	2.4	1.0
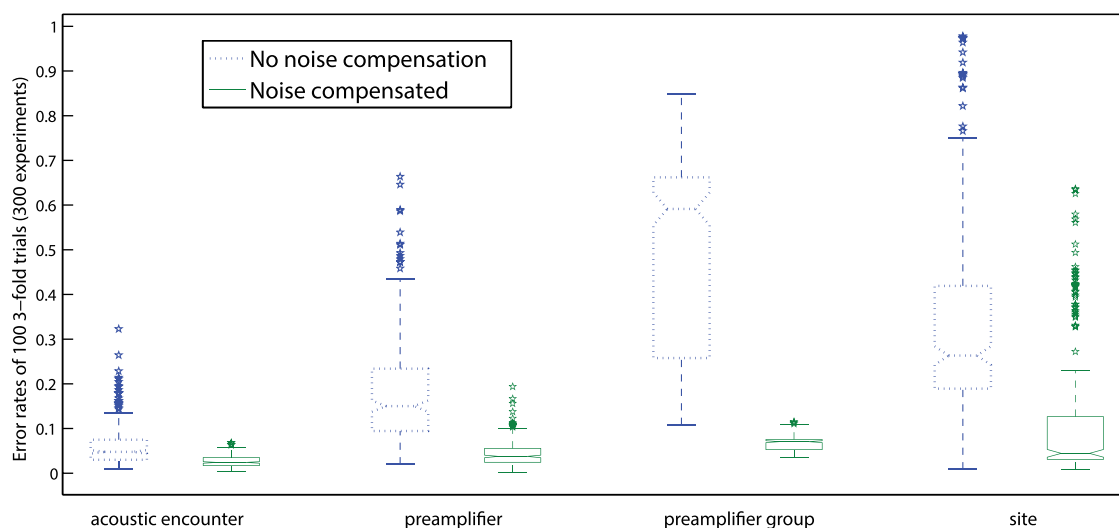
FIG. 4. (Color online) Box plots of error rate distributions for different train/test partitioning criteria over 100 three-fold trials: Acoustic encounter, individual preamplifier, groups of preamplifiers with similar transfer functions, and deployment site. For each criterion, results are shown with noise compensation and without.

FIG. 5. (Color online) Performance of Pacific white-sided dolphin test token groups of 100 echolocation clicks when grouped by acoustic encounter, individual preamplifier board, preamplifier group, and site. Shading represents the per token error rate over the 100 different models that were trained with different training data. The right column reports performance with noise compensation, the left column without. Error rates for Risso's dolphin test tokens exhibit similar patterns.

Pacific-white sided echolocation data from site G (preamplifier 413, preamplifier group γ) were frequently misclassified (Fig. 5). Frequent amplitude saturation (i.e., clipping and waveform distortion) in the acoustic record will be worse for preamplifier group γ because it has the highest gain (Fig. 2) which also will contribute to higher electronic noise levels. Also, while echolocation clicks with peak frequencies above 50 kHz were discarded, there are still differences in frequency distributions due to reduced high-frequency attenuation that are likely to occur when animals are closer to the acoustic sensor or have their highly directional echolocation beam oriented in the sensor direction. For some strong, presumably off-axis clicks in this poorly performing section of the data, the Teager energy associated with the tail structure of echolocation clicks rose above the threshold a second time, creating a second region that was examined for a potential click. When the area between the two detected regions has a high enough

Teager energy, the region growing procedure will grow these two regions together and merge them. In this specific encounter, the Teager energy was low enough that this did not occur and the tail structure was erroneously discarded as a potential multipath detection. In addition, closer examination of regions between clicks noted higher noise in the lower frequencies of the echolocation click band, suggesting possible evidence of distant echolocation activity that was not noted by the analysts selecting data.

Regardless of the cause, noise compensation reduced much of this error not only for the problematic subset of site G data, but generally for all the data tested. The automatic selection of nearly signal free noise sections from specific sites and noise environments during both training and test was responsible for the improvements seen in this study. For partitioning by acoustic encounter, the best matched scenario error rate was reduced from 6.1% to 2.7%, a 56% reduction

in error rate. Other partitioning scenarios reduced error rates by 66% to 86% by using noise compensation. Noise compensation for the two variants of the preamplifier partitioning criterion was able to reduce the error rate to near or below the error rate of the acoustic encounter scenario when tested without noise compensation. The wider spread in the distribution of error rates from the grouped preamplifier tests can be attributed to a multimodal error distribution with several widely separated clusters. As there are only three preamplifier groups, there are only three configurations for the three-fold tests (different pairings of test groups for each species) and the remaining variability is from the random draw of training features.

It is critical to note that no preamplifier board was deployed at more than one site, so any partitioning by site is also a partitioning by preamplifier and the effects of preamplifier and site cannot be easily disentangled. Furthermore, sites A and G, laying southeast and west of Santa Catalina and San Clemente islands, respectively, are deployed in 300 m of water, a much shallower environment than the remaining sites that are deployed at depths of 900 to 1300 m. This results in a different propagation environment that is also likely to be a cause of mismatch. Finally, the HARP data loggers were not end-to-end calibrated, and while most of the variation should be attributable to the preamplifier gain for which we have compensated, other differences such as system noise may also contribute. Disentangling these issues would require an experimental design with a broader distribution of preamplifiers, instruments, propagation environments, and end-to-end calibration.

In addition to differences in data logger hardware and physical ocean environment, animals vary their echolocation signal types in the face of different acoustic environments (Au *et al.*, 1985), or with respect to differences in population structure (suggested as a possible cause of the two Pacific white-sided dolphin click types in Soldevilla *et al.*, 2010) and behavior (Johnson *et al.*, 2006).

Finally, it is important to remember that these acoustic encounters were selected without visual confirmation due to the autonomous nature of the passive acoustic monitoring instruments used in the study. Efforts were made to avoid selection of questionable data (possible mixed species, ship noise, etc.). While the selected encounters displayed characteristics that matched the peak structure in the literature, the authors' unpublished data from visual marine mammal surveys during California Cooperative Oceanic Fisheries Investigations (CalCOFI) cruises between 2004 and 2013 show that both species occasionally occur in mixed species groups. Pacific white-sided dolphins were observed with common and northern right whale dolphins (*Delphinus spp.* and *Lissodelphis borealis*) in 8.4% of 119 sightings, and Risso's dolphins with bottlenose dolphins (*Tursiops truncatus*) in 6.25% of 96 sightings. In addition to mixed-species groups, separate groups of different species may occur simultaneously within the detection range. A consequence of our efforts to select regions of echolocation signal data that did not appear to contain mixed species data for this study which focuses on the effects of instrumentation and site differences in classification performance is that one should not expect the methods to generalize well in an environment where mixed species groups occur regularly or to operate effectively on long acoustic records without the identification of regions of echolocation activity without additional development.

V. SUMMARY

Variation in instrumentation characteristics and recording location have been shown to have an impact on the ability to correctly classify echolocation clicks to species when features derived from spectra are used. The available data for the study do not permit us to fully separate the contributions of training and testing with different preamplifiers or sites, although the higher error rates with different site propagation characteristics are likely to account for a larger portion of the variation in the error rate. A spectral compensation technique was introduced that provided reductions in error rate from 56% to 86%, although error rates for instrument and site partitioned trials were still higher than when partitioned by acoustic trial.

This work suggests that studies using passive acoustic monitoring may wish to use caution when there is a mismatch between training and test data due to differences in instrumentation or operating environment. As has been shown, it is possible to compensate to some degree for these differences, and continued work in this area has the potential to further mitigate for these differences.

Atal, B. S. (**1974**). "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," J. Acoust. Soc. Am. **55**, 1304–1312.

Au, W. L., and Hastings, M. C. (**2008**). *Principles of Marine Bioacoustics* (Springer, New York), 679 pp.

Au, W. W. L., Carder, D. A., Penner, R. H., and Scronce, B. L. (**1985**). "Demonstration of adaptation in beluga whale echolocation signals," J. Acoust. Soc. Am. **77**, 726–730.

Auckenthaler, R., Carey, M., and Lloyd-Thomas, H. (**2000**). "Score normalization for text-independent speaker verification systems," Digit. Signal Process. **10**, 42–54.

Blackwell, S. B., Nations, C. S., McDonald, T. L., Greene, C. R., Thode, A. M., Guerra, M., and Macrander, A. M. (**2013**). "Effects of airgun sounds on bowhead whale calling rates in the Alaskan Beaufort Sea," Mar. Mammal Sci. **24**, 342–365.

Boll, S. (**1979**). "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process. **27**, 113–120.

Dunn, R. B., Quatieri, T. F., Reynolds, D. A., and Campbell, J. P. (**2001**). "Speaker recognition from coded speech and the effects of score normalization," in *Proceedings of the Asilomar Signals, Systems, Comp.*, Pacific Grove, CA, Vol. 2, pp. 1562–1567.

Emerson, J. D., and Strenio, J. (**1983**). "Boxplots and batch comparison," in *Understanding Robust and Exploratory Data Analysis*, edited by D. C. Hoaglin, F. Mosteller, and J. W. Tukey (John Wiley & Sons, Inc., New York), pp. 58–96.

Helble, T., Ierley, G. R., D'Spain, G. L., Roch, M. A., and Hildebrand, J. A. (**2012**). "A generalized power-law detection algorithm for humpback whale vocalizations," J. Acoust. Soc. Am. **131**, 2682–2699.

Huang, X., Acero, A., and Hon, H. W. (**2001**). *Spoken Language Processing* (Prentice Hall PTR, Upper Saddle River, NJ), pp. 516, 520–522.

Johnson, M., Madsen, P. T., Zimmer, W. M. X., de Soto, N. A., and Tyack, P. L. (**2006**). "Foraging Blainville's beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation," J. Exp. Biol. **209**, 5038–5050.

Kaiser, J. F. (**1990**). "On a simple algorithm to calculate the 'energy' of a signal," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Albuquerque, NM, Vol. 1, pp. 381.

Kandia, V., and Stylianou, Y. (**2006**). "Detection of sperm whale clicks based on the Teager-Kaiser energy operator," Appl. Acoust. **67**, 1144–1163.

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," J. Oceanogr. **20**, 36–45.

Moretti, D., Thomas, L., Marques, T., Harwood, J., Dilley, A., Neales, B., Shaffer, J., McCarthy, E., New, L., Jarvis, S., and Morrissey, R. (**2014**). "A risk function for behavioral disruption of Blainville's beaked whales (*Mesoplodon densirostris*) from mid-frequency active sonar," PLoS One **9**(1), e85064.

Reynolds, D. A. (**1996**). "The effects of handset variability on speaker recognition performance experiments on the switchboard corpus," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Atlanta, GA, Vol. 1, pp. 113–117.

Roch, M. A., Klinck, H., Baumann-Pickering, S., Mellinger, D. K., Qui, S., Soldevilla, M. S., and Hildebrand, J. A. (**2011**). "Classification of echolocation clicks from odontocetes in the Southern California Bight," J. Acoust. Soc. Am. **129**, 467–475.

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.

Soldevilla, M. S., Wiggins, S. M., and Hildebrand, J. A. (**2010**). "Spatio-temporal comparison of Pacific white-sided dolphin echolocation click types," Aquat. Biol. **9**, 49–62.

Welch, P. D. (**1967**). "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," IEEE Trans. Audio Electroacoust. **15**(2), 70–73.

Whitman, B., Flake, G., and Lawrence, S. (**2001**). "Artist detection in music with Minnowmatch," in *Proceedings of the IEEE Neural Networks Signal Processing XI*, pp. 559–568.

Wiggins, S. M., and Hildebrand, J. A. (**2007**). "High-frequency Acoustic Recording Package (HARP) for broad-band, long-term marine mammal monitoring," in *International Symposium of Underwater Technology*, Tokyo, Japan, pp. 551–557.

Zimmer, W. M. X. (**2011**). *Passive Acoustic Monitoring of Cetaceans* (Cambridge University Press, Cambridge), 356 pp.