Final Report

Does Depth Matter? Examining Factors That Could Influence the Acoustic Identification of Odontocete Species on Bottom-moored Recorders

Submitted to:

Naval Facilities Engineering Command Atlantic under HDR Environmental, Operations and Construction, Inc. Contract No. N62470-10-D-3011, Task Order 21, Issued to HDR, Inc.

Prepared by:

Bio-Waves, Inc.
364 2nd Street, Suite #3
Encinitas, CA 92024
(760) 452-2575

and

Oceanwide Science Institute
P.O. Box 61692
Honolulu, HI 96739

Submitted by:

HDR
Virginia Beach, VA

May 2017

Suggested Citation:

Oswald, J.N., M.O. Lammers, R. Walker, A. Kügler, C. Hom-Weaver, and S. Coates. 2016. Does depth matter? Examining factors that could influence the acoustic identification of odontocete species on bottom-moored recorders. Final Report. Submitted to Naval Facilities Engineering Command Atlantic, Norfolk, Virginia, under Contract No. N62470-10-3011, Task Order 21, issued to HDR, Inc.

Cover Photo Credits:

Pantropical spotted dolphins *(Stenella attenuata)* taken by Julie Oswald, Bio-Waves, Inc.

Work conducted under following contract between Bio-Waves, Inc. and HDR
MSA #: CON-005-4394-009
Subproject #164744, Agreement #105067, Task CTO21

This project is co-funded by US Fleet Forces Command and ONR Marine Mammals and Biology Program, and managed by Naval Facilities Engineering Command Atlantic as part of the US Navy's marine species monitoring program

Table of Contents

Figures

Tables

Abbreviations and Acronyms

dB	decibel(s)
CTD	conductivity, temperature, depth
EAR2	second generation Ecological Acoustic Recorder
LMR	Living Marine Resources
LTSA	Long-term Spectral Average
kHz	kilohertz
NAVFAC LANT	Naval Facilities Engineering Command, Atlantic
ONR	Office of Naval Research
OSI	Oceanwide Science Institute
PAM	passive acoustic monitoring
ROCCA	Real-time Odontocete Call Classification Algorithm

1. Introduction

Over the past decade, passive acoustic monitoring (PAM) methods have been adopted for obtaining information about the occurrence, distribution and behavior of marine mammals. Seafloor recorders are increasingly being used to monitor marine mammals because they allow continuous data to be collected for long periods of time without requiring the presence of a human operator. However, recordings collected using instruments on the sea floor do not typically have associated visual observations and so species must be identified based on the characteristics of the sounds that they produce. Acoustic-based species identification can be difficult because variability in the acoustic repertoire of most marine mammal species makes them challenging to classify. The sounds produced by delphinids exhibit a large amount of overlap in time-frequency characteristics among species, making them particularly difficult for use in identifying species (Oswald et al. 2007, Azzolin et al. 2014). Most delphinids produce two general types of vocalizations: 1) tonal, frequency modulated whistles, and; 2) short, broadband clicks. Traditionally, whistles have been used for delphinid acoustic species identification (e.g., Sturtivant and Datta 1997, Rendell et al. 1999, Oswald et al. 2013, Azzolin et al. 2014), but it has also been proposed that echolocation clicks can be used to identify species such as Risso's dolphins (*Grampus griseus*), Pacific white-sided dolphins (*Lagenorhynchus obliquidens*), short-beaked common dolphins (*Delphinus delphis*), melon-headed whales (*Peponocephala electra*), false killer whales (*Pseudorca crassidens*), and short-finned pilot whales (*Globicephala macrorhynchus*; Soldevilla et al. 2008; Jarvis et al. 2008; Baumann-Pickering et al. 2010, 2011; Roch et al. 2011).

Considerable research has focused on marine mammal acoustic species identification over the past several years and, as a result, several delphinid whistle classifiers with relatively high correct classification scores have been developed. For example, the Real-time Odontocete Call Classification Algorithm (ROCCA), a whistle classifier that has been incorporated into the PAMGuard acoustic data processing software platform, contains classifiers for species in the tropical Pacific and northwest Atlantic oceans, with temperate Pacific and Hawaii classifiers currently in development. Overall correct classification scores are 86 percent and 60 percent for the northwest Atlantic and tropical Pacific classifiers, respectively (Oswald 2013, Oswald et al. 2013). These classifiers were trained and tested using data collected near the sea surface using towed hydrophone arrays. However, marine mammals are commonly monitored using moored acoustic recorders deployed at depths ranging from tens to thousands of meters. As sound travels through the water column, physical processes such as transmission loss and multi-path propagation can cause localized maxima and minima in acoustic intensity (Au and Hastings 2008) and change the characteristics of sounds arriving at the receiver. The characteristics of sounds that arrive at a receiver are also affected by the directionality of the sound producing mechanisms, the relative position and distance of the signaler to the receiver, the seafloor type, bathymetry, sea surface roughness, and variables such as temperature, salinity and pressure gradients in the water column. It has also been suggested that the whistles produced by dolphins may change with depth due to the effects of pressure on lung volume and the sound production structures (Ridgway et al. 2001). Because of these factors, sounds arriving at a seafloor recorder may have different characteristics from those observed at the sea surface.

This in turn could have an effect on the performance (i.e., correct classification scores) of classifiers trained using data obtained at the sea surface.

This report describes the results of a study designed to explore some of the factors that may contribute to ambiguity in species identification using delphinid whistles. We have obtained field recordings of wild and trained (captive) odontocetes in their natural environment under a variety of scenarios, including at the sea surface, at multiple depths in the water column, in different geographic locations, and in different behavioral states. The characteristics of whistles recorded at varying depths have been compared to answer the following questions:

1. Are the same whistles detected in acoustic recordings collected at different depths?
2. Does the depth at which dolphin whistles are recorded affect the received signal characteristics?
3. If received whistle characteristics are different when recorded at different depths, does the performance of species classifiers developed using whistles recorded at the surface change when applied to data from different recording depths?

This project is a collaboration between Bio-Waves, Inc. and Oceanwide Science Institute (OSI), with help from the National Marine Mammal Foundation, and is being jointly funded by US Fleet Forces Command under the US Navy's Marine Species Monitoring Program, the Office of Naval Research (ONR) and the US Navy's Living Marine Resources Program (LMR). This report focused on the analysis of whistles, which is the portion of the project funded by US Fleet Forces Command and led by Bio-Waves, Inc. An analysis of the effect of recording depth on characteristics of echolocation clicks is being led by Oceanwide Science Institute under an ONR-funded portion of the project.

2. Statement of Navy Relevance

PAM is used extensively to collect information regarding marine mammal occurrence, distribution and behavior in areas with high naval activity, and mitigation efforts rely heavily on data obtained by seafloor recorders. However, the suitability of using species classifiers trained with surface data for analyzing recordings obtained at depth is currently unknown. If classifiers perform differently on data recorded at depth, it may be necessary to re-train them or develop new classifiers to ensure accurate results. Similarly, if the behavior of animals or signal propagation affects the identification of species using classifiers developed for echolocation clicks, this must be understood and integrated into analysis methods. It is important to accurately identify odontocete species in acoustic recordings because different species are known to react differently to naval activities, and understanding species-specific responses is important for implementing effective mitigation measures. This study tests previously unexamined assumptions associated with the acoustic identification of odontocete species. The results of this effort will ultimately provide a better understanding of the methods presently being employed for marine mammal monitoring and mitigation, and will lead to greater confidence in their application.

3. Methods

3.1 Data Collection

Two types of vertical arrays were used to obtain recordings of whistles and echolocation clicks at different depths: 1) a surface array of microMARS recorders and four broadband hydrophones used to collect data for localization analysis, and 2) a bottom array of second generation Ecological Acoustic Recorders (EAR2s).

3.1.1 Surface Array

The surface microMARS array was deployed from a small (approximately 24-foot) vessel and was composed of two vertical sub-arrays: a localization array with four broadband hydrophones (Cetacean Research Technology C75) spaced 10 meters apart, and a line array made up of five microMARS recorders spaced 50 meters apart (**Figure 1**). The four hydrophones in the localization array were sampled simultaneously on a high-resolution portable recorder. This allowed fine-scale localization of phonating animals so their depth and distance relative to the vertical array could be established. The array of microMARS recorders extended beyond the localization array with individual recorders separated by 50 meters to a maximum depth of 250 meters. The microMARS (http://www.desertstar.com/acoustic-recorders.html) is a low-cost acoustic recorder with a relatively high maximum sampling rate of 250 kilohertz (kHz) and up to 512 gigabytes of storage space. It measures only 19.5 cm x 6.5 cm, so it can be configured into a hand- deployable/retrievable vertical line array (**Figure 2**). The microMARS array was used to record visually detected schools of odontocetes at multiple depths. When a group of odontocetes was sighted by the observers on the vessel, the observers spent 30 minutes to 1 hour observing the animals to determine their behavior, direction and rate of travel. When the observers determined the behavioral state and movement patterns of the school, the vessel moved approximately 1 kilometer ahead of the school and the surface array was deployed. Recordings were made as the school approached the boat until the animals moved out of acoustic detection range. If time and conditions allowed, the array was recovered and the process was repeated. While in range of the animals, acoustic data were recorded, as well as visual data such as animal location, behavior and species identification. Four of the five microMARS were paired with a Star-Oddi DST tilt sensor, which measured and recorded temperature, depth (pressure) and tilt (in three directions). The fifth (deepest) microMARS was paired with a Star-Oddi conductivity, temperature, depth (CTD) probe, which is a miniature salinity, temperature and depth data logger. This probe was fixed to the bottom of the array to obtain profiles upon deployment and recovery. These data will be used to calculate sound speed profiles, in order to investigate the effects of sound propagation on received signal properties.

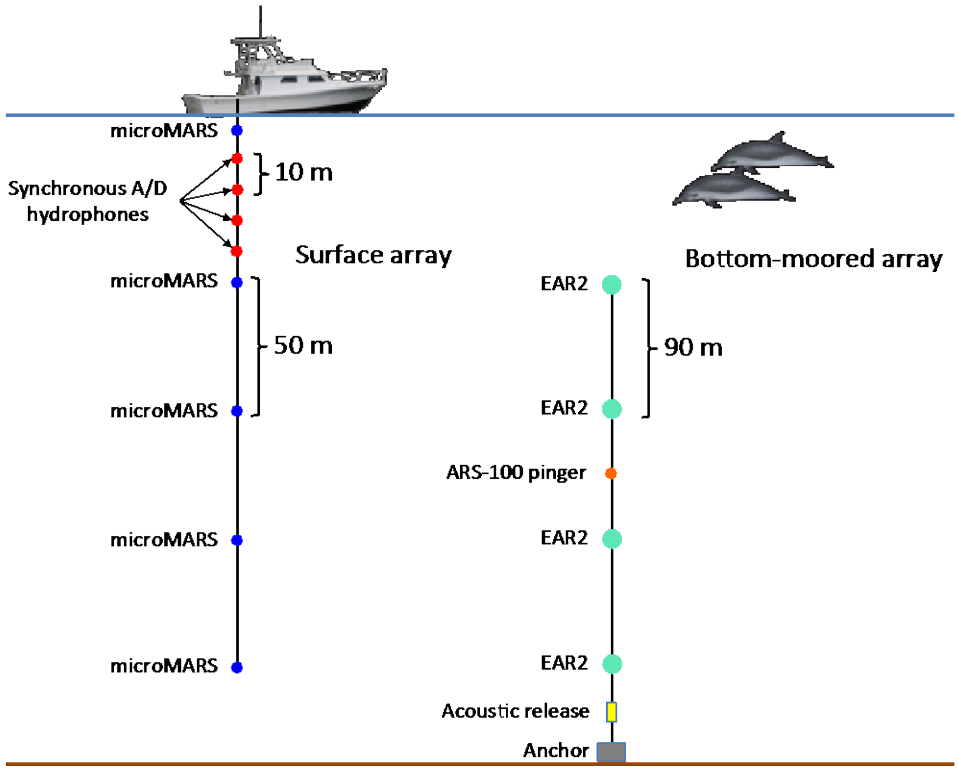


Figure 1. Schematic diagram of the surface microMARS array and bottom-moored EAR2 array.

Figure 2. MicroMARS recorder included in the surface array.

3.1.2 Bottom Array

The bottom-moored vertical array was made up of four EAR2s spaced 90 meters apart (**Figures 1** and **3**). The EAR2 is a redesigned version of the EAR and has a maximum sampling rate of 125 kHz, up to 1 terabyte of storage space and is able to sample continuously. The array also included a RJE International ARS-100 pinger, which provided a 4 to 7 kHz synchronization pulse every 30 minutes. This pulse was recorded on the four EAR2 recorders and was used to precisely time-align recordings during post-processing and analysis in order to localize signaling animals and determine their range and depth. A Star-Oddi DST centi miniature temperature and depth data logger was paired with the shallowest EAR2. The middle two EAR2s were paired with Star-Oddi DST-tilt sensors. A Star-Oddi miniature conductivity, temperature, depth probe was fixed to the bottom of the moored array to allow calculation of the sound speed profile at the deployment site at the time of deployment and the time of recovery. The EAR2 array was deployed on the sea floor at locations of known high odontocete activity at bottom depths between 500 and 950 meters for periods of 1 to 2 weeks during all field work efforts. The EAR2s were set to record on a 33 percent duty cycle (2 minutes every 6 minutes during the pilot work in Lanai, and 10 minutes every 30 minutes during the Kona and San Diego deployments.) This ensured that a sufficiently large sample size of odontocete encounters was obtained for performing statistically meaningful comparisons.

Figure 3. Bottom-moored EAR2 array showing four EAR2s, an acoustic release, 300 meters of line, and anchor (concrete blocks and sand bags).

3.1.3 Pilot Study

A pilot field effort was conducted in Hawaii from 2 to 12 August 2015. At the beginning of this effort, the bottom-moored array was deployed approximately 3 nautical miles south of the island of Lanai in waters 355 meters deep (**Figure 4**). Approximately 400 pounds of weight (sand bags and a concrete block) were used to keep the array moored in place until it was recovered on 12 August. The bottom array was set to record for 2 minutes every 6 minutes. After deployment of the bottom array, visual surveys were conducted off the islands of Maui and Lanai using a 26-foot research vessel (the *Aloha Kai*) and a 21-foot research vessel (the *Coho*; due to engine problems on the *Aloha Kai*). The surface microMARS array was used to record groups of odontocetes that were encountered during these surveys.

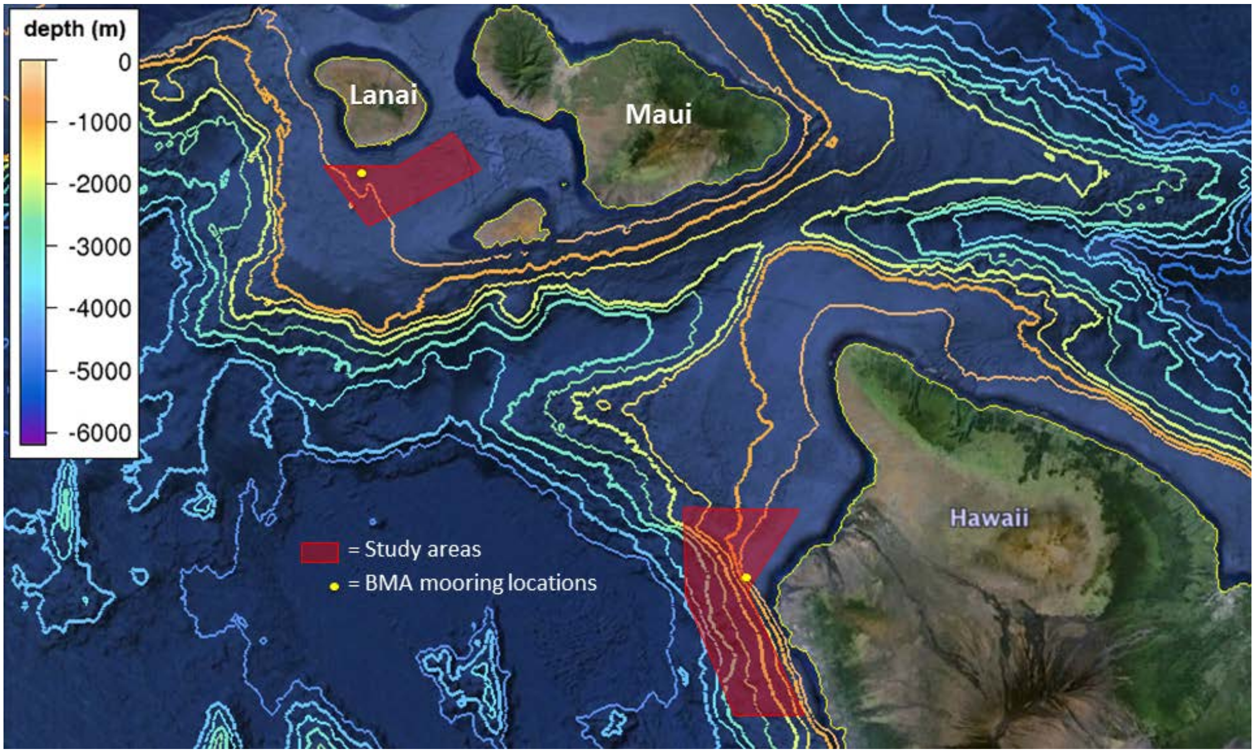


Figure 4. Lanai and Kona coast study areas, which bottom-moored array locations shown as yellow circles.

Three different microMARS hydrophones with different sensitivities and frequency ranges were compared to determine which would be best suited for this project. Two of the hydrophones had flat frequency responses up to 33 kHz and one had a flat frequency response up to 125 kHz (**Figures 5** and **6**). The configuration of microMARS hydrophones on the array was changed several times during the field testing period to compare their performance relative to one another. Various strategies for deployment and recovery of the surface array were also tested. Finally, the sub-array hydrophones and broadband recorder were tested using different gain levels.

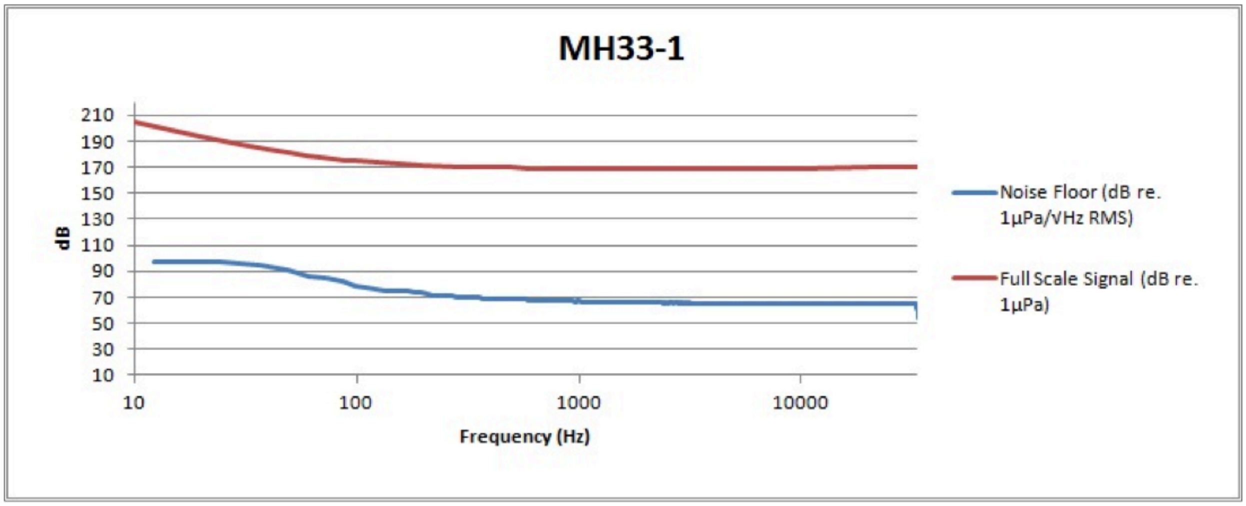


Figure 5. Generic sensitivity curve for MH33-1 microMARS hydrophone used in pilot work.

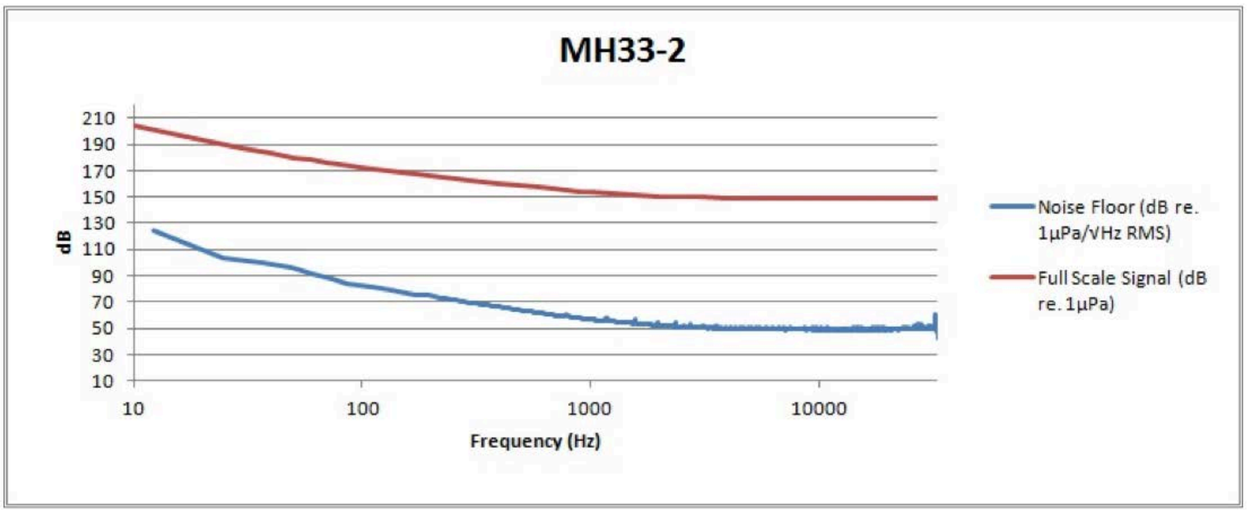


Figure 6. Generic sensitivity curve for MH33-2 microMARS hydrophone used in pilot work.

Because of the differences in microMARS sensitivities and configurations throughout the pilot study work, it was not possible to compare whistles and clicks among depths for this particular dataset.

3.1.4 Kona Data Collection

A field data collection effort was conducted off Kona, Hawaii from 2 to 14 November 2015. The bottom-moored EAR2 array was deployed north of Kona in a water depth of 400 meters for the duration of this field effort. The array was set to record for 10 minutes every 30 minutes. Visual surveys of marine mammals were conducted along the Kona coast (**Figure 4**) using a 26-foot research vessel (the *Hopena*). The surface array of microMARS was deployed near groups of odontocetes that were encountered.

3.1.5 San Diego Data Collection

A 2-week field data collection effort was conducted off San Diego, California, from 15 to 27 May 2016. During this effort, the EAR2 array was moored on the sea floor approximately 7 miles off La Jolla, California (**Figure 7**), in a water depth of 465 meters, and recorded for 10 minutes every 30 minutes. Visual surveys were conducted between Point Loma and La Jolla, and on the leeward (east) side of Catalina Island using one of two 28-foot sport fishing vessels (the *Seasons* and the *Ugly Guy)*. The surface microMARS array was deployed near groups of odontocetes that were encountered during these surveys.

In addition to the vessel surveys, a controlled data collection experiment was conducted in collaboration with the Navy's Marine Mammal Program. During this effort, a trained Navy bottlenose dolphin was transported by boat to an open water location (approximately 1,000 meters in depth) off Point Loma. The animal was instructed to swim to and station on a bite-plate positioned at a depth of 5 meters. The horizontal distance and orientation of the dolphin relative to the surface array were recorded at the beginning and ending of each trial. The dolphin was instructed to produce whistles while on station at the bite-plate. Whistles were recorded using the surface microMARS array at horizontal distances of 50, 100, 250 and 400 meters away from the dolphin. The dolphin was oriented in two positions: 1) facing directly towards the array and, 2) facing directly away from the array. The magnetic bearing of the dolphin's orientation was measured and recorded using a compass app (SensorLog) on an iPad by orienting the iPad's edge along the posterior/anterior axis of the animal. The horizontal distance between the dolphin and the surface array was measured using a range finder.

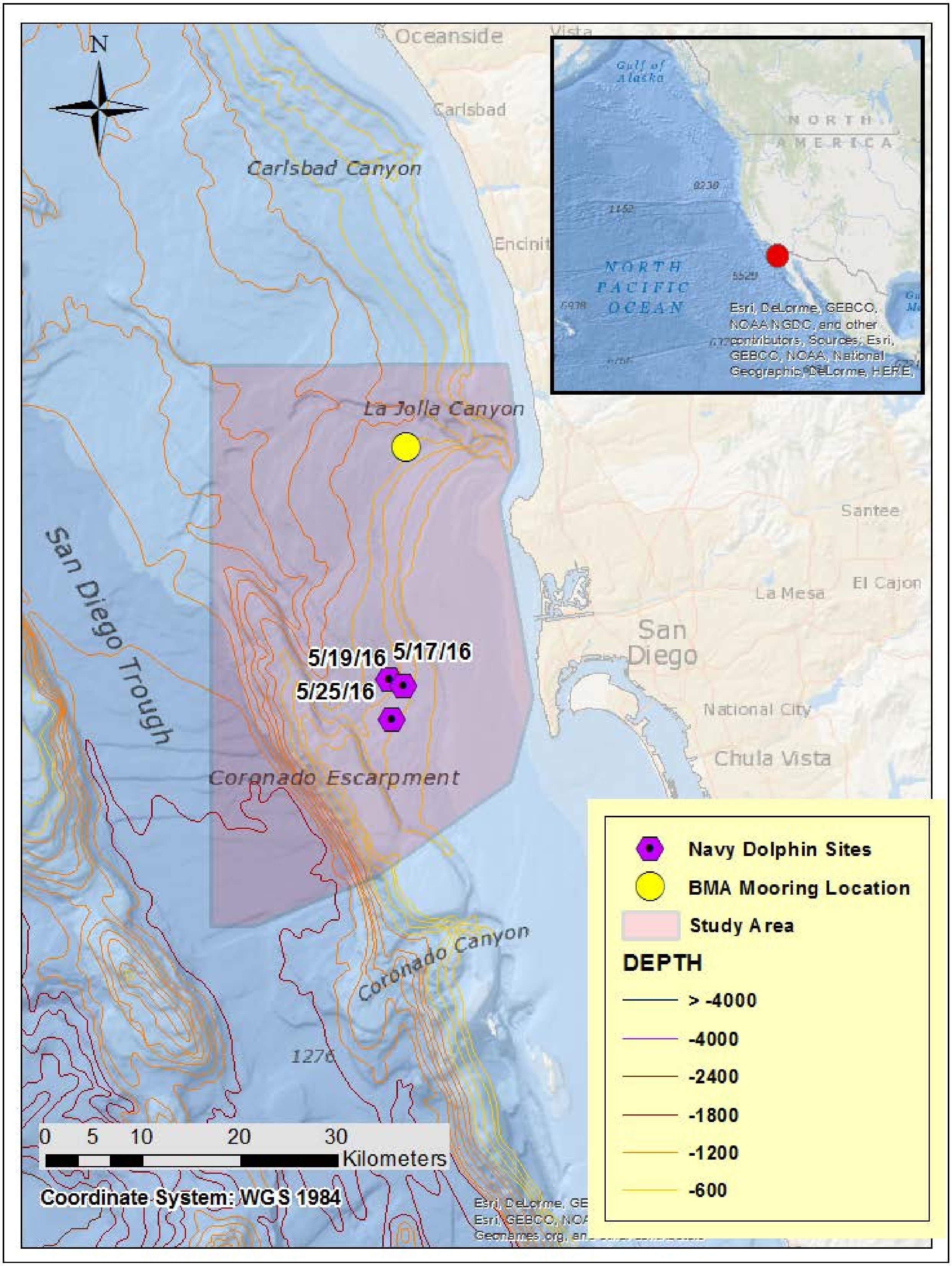


Figure 7. Map of San Diego study area with location of the bottom-mounted EAR2 array (yellow circle), and locations of Navy dolphin trials (purple hexagons).

3.2 Whistle Analysis

For the Lanai data, Triton software (Wiggins 2007) was used to create long-term spectral averages (LTSAs) from recordings from all four EAR2s in the array. An analyst reviewed these LTSAs manually and created a log of all delphinid acoustic encounters in the data. Because there were no visual observations associated with EAR2 recordings, it was not possible to determine when one school left the area (or stopped vocalizing) and another school entered (or started vocalizing). As a proxy, for the Lanai data, an encounter was defined based on elapsed time between vocalizations. A new encounter was delineated when 30 or more minutes had elapsed between whistles or clicks. This was chosen based on our previous experience observing behavior and recording vocalizations of delphinids. A slightly different definition of encounter had to be used for the Kona and San Diego data because of the duty cycle used during these deployment (10 minutes on, 20 minutes off). Since the EAR2s recorded every 30 minutes, which was also the time gap used to denote separate encounters in the Lanai data, each 10-minute sound file was treated as a separate encounter. Analysts used Triton to examine each sound file from the shallowest EAR2, and selected a subset of 19 events from the Kona dataset and 20 events from the San Diego dataset that contained sufficient whistles and clicks for analysis.

For each encounter, analysts used Raven software to examine spectrograms from each EAR2. Up to 25 whistles were randomly selected from the shallowest EAR2 (EAR2-1) and up to 25 different whistles were randomly selected from the deepest EAR2 (EAR2-4). These same whistles (25 from EAR2-1 and 25 from EAR2-4) were then selected from all other EAR2s, if they were present in those recordings. This resulted in a sample size of up to 50 whistles per EAR2 per encounter. Selecting whistles from both EAR2-1 and EAR2-4 reduced the risk of missing whistles that might only be heard on the shallowest or deepest EAR2s due to the position of the animal relative to the array. Only encounters that contained at least 10 whistles with moderate to good signal-to-noise ratios (i.e., at least 3 dB) from EAR2-1 or EAR2-4 were included in the analyses.

Analysts then extracted time-frequency contours from the selected whistles in each encounter using the ROCCA (Oswald et al. 2013) module in the acoustic data processing software platform, PAMGuard (Gillespie et al. 2008). To extract time-frequency contours, the analyst manually traced contours on ROCCA's spectrographic display using a computer touch-pad (**Figure 8**). ROCCA then automatically measured 50 variables from each extracted contour, including duration, frequencies (e.g., minimum, maximum, beginning, ending, and at various points along the whistle), slopes, and variables describing the shape of the whistles (e.g., number of inflection points and steps; see Barkley et al. 2011 for a complete list and description of variables measured). Whistle variables were compared across depths using Kruskall-Wallis tests and post-hoc Dunn's tests with Bonferroni correction. Since not all selected whistles appeared on all four EAR2s (e.g., **Figure 9**), two different analyses were conducted. In the first analysis, whistle variables for all whistles detected on each EAR2 for each encounter. The second analysis included only whistles that were detected on all four EAR2s.

Whistle variables were also used to classify whistles to species using a random forest classifier. A random forest is a collection of decision trees grown using binary partitioning of the data.

Each binary partition is based on the value of one whistle variable (Breiman 2001). Randomness is introduced into the tree-growing process by examining a random subsample of all of the variables at each node. The variable that produces the most homogeneous split is chosen at each partition. When whistle variables are run through a random forest, each of the trees in the forest produces a species classification. Each tree can be considered one 'vote' for a given species classification. Votes are then tallied over all trees and the whistle classification is based on the species with the most 'votes.' In addition to classifying individual whistles, entire acoustic encounters are classified based on the number of tree classifications for each species, summed over all of the whistles that were analyzed for that encounter.

The random forest classifier used to analyze the Lanai and Kona data was a two-stage classifier trained using whistles recorded from single-species schools in the tropical Pacific Ocean. Six species were included in the model: short-finned pilot whales (*Globicephala macrorhynchus*), false killer whales (*Pseudorca crassidens*), pantropical spotted dolphins (*Stenella attenuata*), bottlenose dolphins (*Tursiops truncatus*), rough-toothed dolphins (*Steno bredanensis*), and spinner dolphins (*Stenella longirostris*). The first stage consisted of classifying whistles to one of two categories: 'large delphinids-*Steno'* (including false killer whales, pilot whales and rough-toothed dolphins) and '*Stenella-Tursiops*' (including spinner, spotted, and bottlenose dolphins). In stage two of the model, whistles within each category were then classified to species (**Figure 10**). The random forest classifier used to analyze the San Diego data was a one-stage classifier trained using whistles recorded from single-species schools in the temperate Pacific Ocean. Three species (short-finned pilot whales; striped dolphins, *Stenella coeruleoalba*; bottlenose dolphins) and one species-group (common dolphins, which included *Delphinus delphis and D. capensis)* were included in the model.

Four-fold cross-validation was used to test the performance of both classifiers. To accomplish this, each training dataset was randomly divided into four subsets of data, with whistles from the same encounter kept together in the same dataset. One dataset was used to train the classifier and the other three were used to test the classifier. The datasets were then swapped so that each was used as both a training and a testing dataset. This procedure was repeated 50 times and the results were compiled (**Tables 1 and 2**). Classification success was evaluated by examining the average percent of encounters that were correctly classified for each species and comparing that to the classification score that would be expected by chance (17 percent for six species and 25 percent for four species). Overall, 58 percent of encounters were correctly classified in the tropical Pacific training dataset and 70 percent of encounters were correctly classified in the temperate Pacific training dataset. All correct classification scores were significantly greater than expected by chance (Fisher's exact test, $\alpha = 0.05$).

The EAR2 encounters were analyzed with these classifiers and encounters were classified based on classification results summed over all whistles in the encounter. Individual whistle classifications and overall encounter classifications were compared among EAR2s for each encounter. The classification analysis was performed on two different sets of whistles for each encounter. First, all whistles recorded on each EAR2 were included to examine whether the whistles available for analysis affected the results. Second, only whistles recorded on all four EAR2s were included to examine how differences in whistle structure affected classification results.

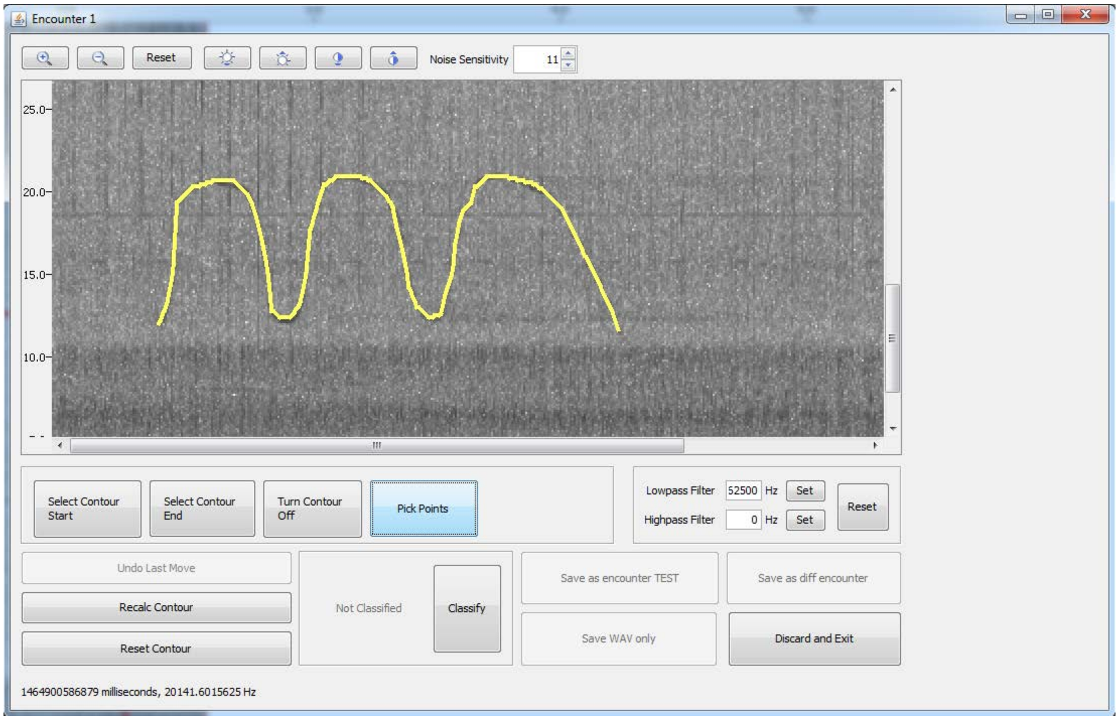


Figure 8. Example of whistle contour traced in ROCCA module in PAMGuard.

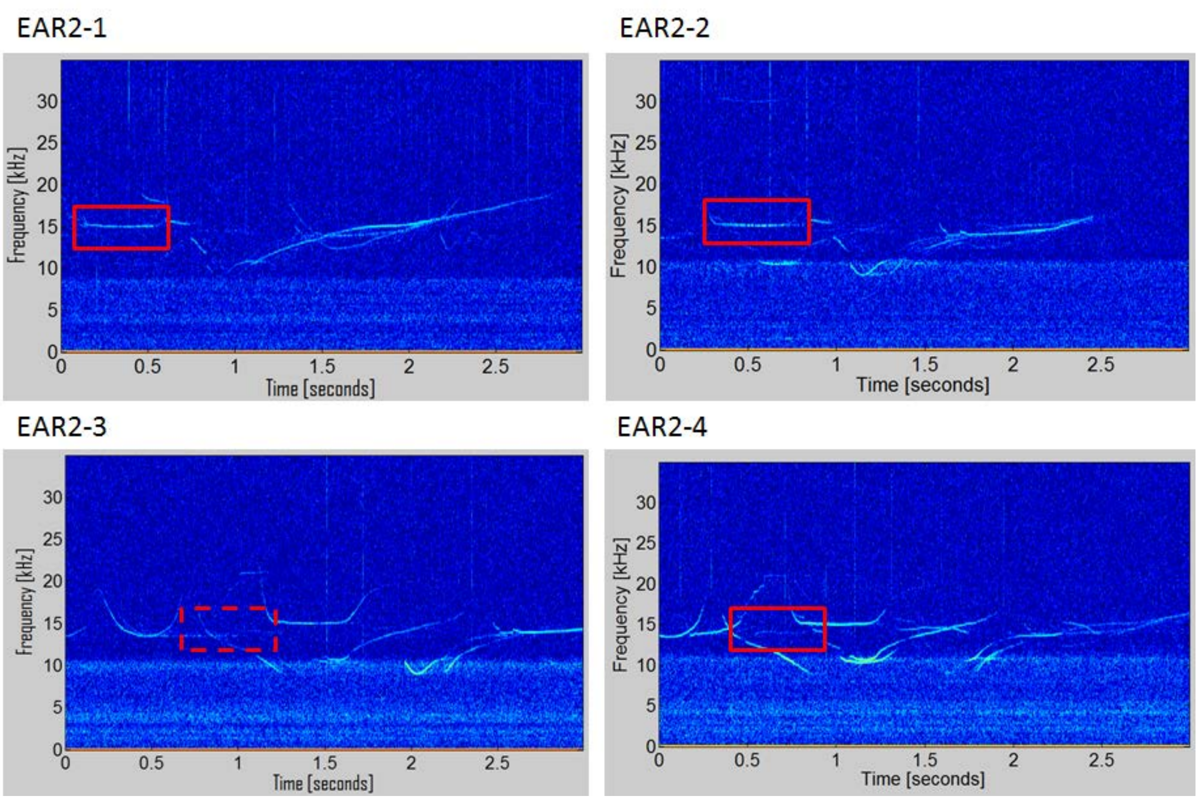


Figure 9. Example of whistle contour not appearing on all EAR2s. **Solid red lines show EAR2s where the whistle contour was visible, dashed red line shows where contour should appear but was not visible.**

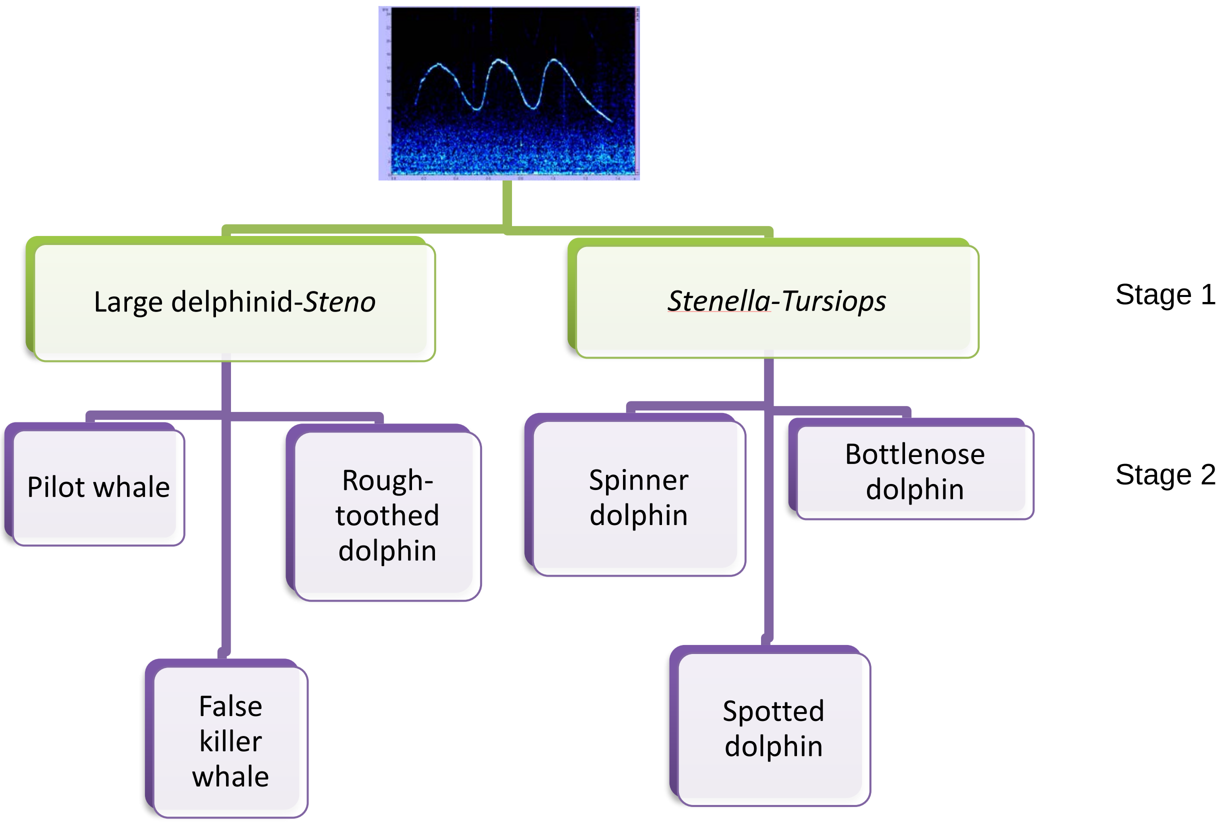


Figure 10. Schematic diagram of the two-stage tropical Pacific random forest classifier. **In stage one, whistles are classified to one of two broad categories (‘large delphinid-*Steno*’ or ‘*Stenella-Tursiops*’). In stage two, whistles within each category are classified to individual species or species-group.**

Table 1. Confusion matrix for the two-stage tropical Pacific classifier used to classify encounters recorded in Kona and Lanai. The percentage of encounters correctly classified for each species is in bold, with standard deviations in parentheses. The number of encounters in the training dataset are given for each species.

Actual species	Percent classified as						Number of encounters
	Pilot whale	False killer whale	Spotted dolphin	Rough-toothed dolphin	Spinner dolphin	Bottlenose dolphin	
Pilot whale	**59 (10)**	24 (10)	0 (4)	0.1 (2)	0 (4)	0 (4)	12
False killer whale	16 (5)	**72 (10)**	0 (0)	0.1 (10)	0 (0)	0 (0)	9
Spotted dolphin	0 (0)	0 (4)	**58 (10)**	0 (1)	20 (6)	19 (10)	17
Rough-toothed dolphin	0 (5)	13 (7)	0 (1)	**57 (9)**	13 (4)	0 (3)	12
Spinner dolphin	0 (5)	0 (3)	0 (7)	0 (2)	**56 (10)**	20 (7)	14
Bottlenose dolphin	0 (3)	0 (4)	16 (9)	15 (8)	19 (10)	**47 (10)**	8

Table 2. Confusion matrix for the temperate Pacific classifier used to classify encounters recorded in San Diego. The percentage of encounters correctly classified for each species is in bold, with standard deviations in parentheses. The number of encounters in the training dataset are given for each species.

Actual species	Percent classified as				Number of encounters
	Common dolphin	Pilot whale	Striped dolphin	Bottlenose dolphin	
Common dolphin	**56 (5)**	3 (2)	28 (6)	13 (4)	35
Pilot whale	2 (6)	**83 (1)**	9 (5)	5 (4)	12
Striped dolphin	17 (6)	2 (s)	**66 (6)**	16 (4)	30
Bottlenose dolphin	6 (7)	10 (5)	10 (8)	**74 (9)**	8

4. Results

4.1 Pilot Field Effort

A total of 31 hours of field survey effort were spent searching for and/or recording odontocetes over 6 days. Odontocetes were encountered and recorded with the surface microMARS array on four occasions. These encounters included two groups of spotted dolphins, one group of spinner dolphins, and one group of short-finned pilot whales.

All EAR2s recorded successfully during the deployment period and the ARS-100 pinger transmitted signals every 30 minutes as expected. The average depth of each EAR2, determined from depth recorder measurements, is provided in **Table 3**. The EAR2 array recordings yielded 2,063 2-minute files, or approximately 68.7 hours of data per recorder. These recordings contained 28 delphinid acoustic encounters.

Table 3. Average depth (standard deviation in parentheses) of each EAR2. No depth sensors were used for the pilot work, so depths are estimated based on the depth of deployment and the spacing between EAR2s. Depths for the Kona field effort are based on depth recorder readings throughout the deployment. Depths for the San Diego field effort are based on the deployment depth of the array and the distance between EAR2s.

EAR2 number	Average Depth (m)		
	Pilot work	Kona	San Diego
EAR2-1	70	118 (1.2)	177
EAR2-2	160	209 (0.9)	267
EAR2-3	250	289 (0.6)	357
EAR2-4	340	389 (1.8)	447

4.2 Kona Field Effort

A total of 80 hours of survey effort were spent searching for and/or recording odontocetes over 8 days. Odontocetes were encountered and recorded with the surface microMARS array on 14 occasions. These encounters are summarized in **Table 4**. The EAR2 recordings yielded 538 10-minute files, or approximately 90 hours of data per recorder. Each 10-minute recording was treated as a separate encounter, resulting in 538 encounters. The average depth of each EAR2, based on depth recorder readings throughout the deployment is given in **Table 3**.

Table 4. Number of encounters per species recorded with the microMARS surface array during the Kona field effort.

Species	Number of encounters
Spotted dolphin	5
Short-finned pilot whale	2
Rough-toothed dolphin	1
False killer whale/rough-toothed dolphin	1
Pilot whale/bottlenose dolphin	1

4.3 San Diego Field Effort

A total of 75 hours of survey effort were spent searching for and/or recording odontocetes over 11 days. Odontocetes were encountered and recorded with the surface microMARS array on 16 occasions. These encounters are summarized in **Table 5**. The EAR2s yielded 521 10-minute files, or approximately 87 hours of data per recorder. The average depth of each EAR2, based on deployment depth and distance between EAR2s, is given in **Table 3**.

Table 5. Number of encounters per species recorded with the microMARS surface array during the San Diego field effort.

Species	Number of encounters
Common dolphin (sp.)	11
Short-beaked common dolphin	3
Bottlenose dolphin	2

Approximately 8 hours of the 75 total survey hours (10.6 percent) were spent working with the trained Navy dolphin to collect controlled data.

4.4 Lanai Pilot Study MicroMARS Analysis

The surface microMARS array that was used during the pilot field effort contained different hydrophones deployed at different depths. These hydrophones had different sensitivities and frequency responses and so these data cannot be used to compare signal characteristics among depths. These data did, however, provide valuable information regarding which microMARS hydrophone is most appropriate for this work. Based on examinations of the microMARS recordings, it was decided that the broadband hydrophones (125 kHz) with the highest sensitivity are necessary to capture the most whistles and echolocation clicks. These hydrophones were subsequently used in all microMARS during the Kona and San Diego field work.

4.5 EAR Analyses

4.5.1 Datasets

Whistles were measured and classified from 20 of the 28 encounters recorded with the Lanai EAR2 array (**Table 6**). Eight of the 28 encounters did not contain enough whistles to be included in the analysis.

Table 6. Date, start time, end time for each acoustic encounter recorded on the EAR2 array during the pilot work off Lanai. **Detections with an asterisk were used to compare whistle contours that appeared on all four EAR2s.**

Encounter	Date	Start time	End time
1	8/4/2015	4:24:13	6:25:52
2*	8/4/2015	12:52:42	13:30:04
3*	8/4/2015-8/5/2015	18:42:51	3:30:03
5	8/5/2015	7:48:03	8:19:05
7	8/6/2015	3:54:55	4:13:03
9	8/6/2015	10:42:05	11:06:04
10	8/6/2015	12:43:04	13:07:04
11	8/6/2015	17:00:03	17:30:07
12*	8/6/2015-8/7/2015	18:42:34	1:49:39
13*	8/7/2015	3:48:04	6:07:44
14*	8/7/2015-8/8/2015	20:06:35	1:36:03
16*	8/8/2015	10:54:04	11:48:03
18*	8/8/2015-8/9/2015	20:12:34	7:54:04
19*	8/9/2015	8:36:04	12:12:04
21*	8/9/2015-8/10/2015	16:43:05	0:54:04
22*	8/10/2015	2:55:34	7:30:04
23	8/10/2015	11:48:03	13:31:12
25*	8/10/2015-8/11/2015	22:00:03	9:24:34
27*	8/11/2015	17:24:59	20:54:35
28*	8/11/2015-8/12/2015	21:55:43	6:36:04

Whistles were measured and classified from 19 encounters recorded with the Kona EAR2 array (**Table 7**) and from 20 encounters recorded with the San Diego array (**Table 8**). The encounters included in the analysis were randomly selected from all of the encounters in each dataset. Encounters were only included if they contained at least 10 whistles with a SNR of at least 3dB in recordings from the shallowest EAR2.

Table 7. Date, start time and end time for each acoustic encounter recorded on the EAR2 array off Kona and included in the analysis. **All detections were used to compare whistle contours that appeared on all four EAR2s.**

Encounter	Date	Start time	End time
1	11/3/2015	18:30:01	18:40:01
2	11/3/2015	22:30:02	22:40:02
3	11/4/2015	1:30:02	1:40:02
4	11/4/2015	4:00:02	4:10:02
5	11/4/2015	4:30:00	4:40:00
6	11/5/2015	4:30:03	4:40:03
7	11/5/2015	6:00:02	6:10:02
8	11/6/2015	0:30:02	0:40:02
9	11/6/2015	3:00:02	3:10:02
10	11/6/2015	3:30:02	3:40:02
11	11/6/2015	4:30:02	4:40:02

12	11/6/2015	5:00:02	5:10:02
13	11/6/2015	5:30:02	5:40:02
14	11/7/2015	4:30:02	4:40:02
15	11/7/2015	5:00:02	5:10:02
16	11/7/2015	5:30:03	5:40:03
17	11/7/2015	6:00:02	6:10:02
18	11/7/2015	19:00:02	19:10:02
19	11/8/2015	19:30:02	19:40:02

Table 8. Date, start time and end time for each acoustic encounter recorded on the EAR2 array off San Diego and included in the analysis. **Detections with an asterisk were used to compare whistle contours that appeared on all four EAR2s.**

Encounter	Date	Start time	End time
1	5/17/2016	3:30:02	3:40:02
2	5/17/2016	4:30:02	4:40:02
3	5/17/2016	5:00:02	5:10:02
4	5/17/2016	6:00:02	6:10:02
5	5/17/2016	19:00:01	19:10:01
6	5/17/2016	20:00:01	20:10:01
7	5/18/2016	2:30:01	2:40:01
8	5/18/2016	3:30:02	3:40:02
9	5/18/2016	4:30:02	4:40:02
10	5/18/2016	15:00:01	15:10:01
11	5/18/2016	20:30:02	20:40:02
12	5/18/2016	21:00:01	21:10:01
13	5/19/2016	1:00:02	1:10:02
14	5/19/2016	19:30:02	19:40:02
15	5/20/2016	19:30:03	19:40:03
16	5/21/2016	21:30:02	21:40:02
17	5/23/2016	19:00:02	19:10:02
18	5/23/2016	20:00:02	20:10:02
19	5/24/2016	5:30:02	5:40:02
20	5/24/2016	22:30:02	22:40:02

4.5.2 Percent of whistles detected at each depth

In the 20 Lanai encounters that were included in the analysis, not all whistles were detected on all EAR2s (**Figure 11a**). For some encounters (e.g., encounters 2, 3, and 18) a high percentage of whistles were detected on all four EAR2s, but for others (e.g., encounters 9, 10, and 11), the majority of whistles were detected on only one EAR2. In most cases (16 out of 20 encounters), the greatest number of whistles was detected on the deepest EAR2 (at a depth of 340m; **Table 9**). In contrast, for most Kona EAR2 encounters (84%), at least 90% of whistles were detected on all four EAR2s. Only three encounters had any whistles that were detected on only one EAR2 (**Figure 11b**). Greater than 50% of whistles were detected on all four EAR2s in San Diego, however most encounters contained a significant number of whistles that were not detected on all four EAR2s (**Figure 11c**).

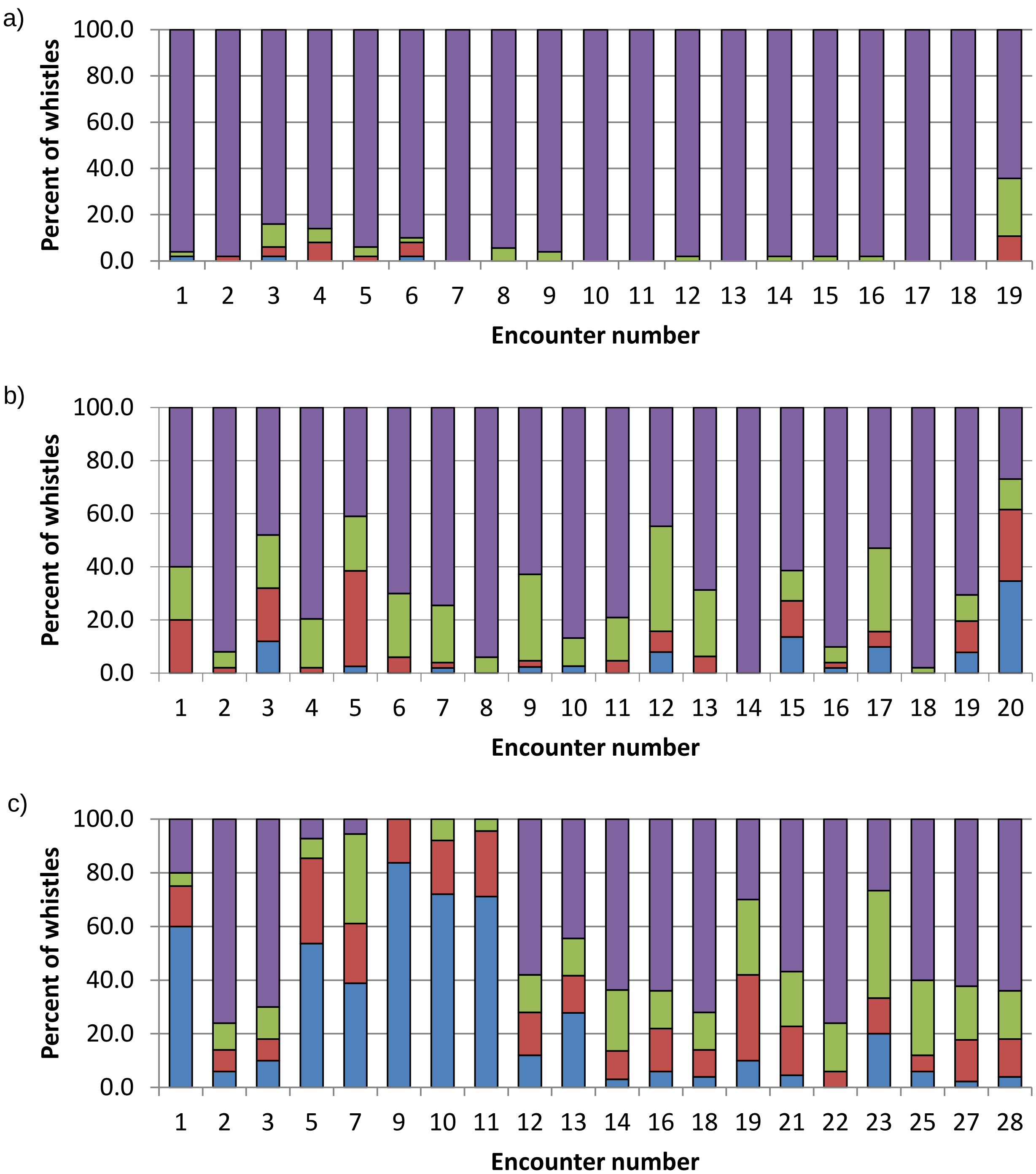


Figure 11. Percentage of whistles detected on only one EAR2 (blue), on two EAR2s (red), on three EAR2s (green) and on all four EAR2s (purple) for the Lanai (a), Kona (b) and San Diego (c) EAR2 arrays

4.5.3 Variable comparisons

Results of statistical comparisons among Lanai EAR2 depths are provided in **Table 9** for all whistles measured for each encounter. One or more variables were significantly different (Kruskall-Wallis test and post-hoc Dunn's tests with Bonferonni correction, α=0.05) when compared among EAR2s for 8 of the 20 encounters (40 percent). Most of the significant differences were for variables measuring frequency characteristics; mean, median and center frequency were significantly different among EAR2s for six of the encounters that had significant differences (**Table 9**). Slope variables, duration, and other frequency variables were also significantly different for some of the encounters. **Figure 12** shows an example of a whistle that was detected on more than one EAR2, but with only a portion of the contour evident in some of the recordings. For only those whistles that appeared on all four EAR2s, only 1 of the 13 encounters (8 percent) showed a significant difference (Kruskall-Wallis test and post-hoc Dunn's tests with Bonferonni correction, α=0.05). Minimum frequency was significantly different between the EAR2 at depth 70m and the EAR2 at depth 250m (p=0.04).

No variables were significantly different (Kruskall-Wallis test and post-hoc Dunn's tests with Bonferonni correction, α=0.05) between depths for the Kona EAR2 recordings, either when all whistles were included or when only whistles detected on all four EAR2s were included.

Four out of twenty San Diego encounters had variables that were significantly different (Kruskall-Wallis test and post-hoc Dunn's tests with Bonferonni correction, α=0.05) between depths when all whistles were included. For encounter 12, duration was significantly different when compared between the EAR2 at depth 177m and the EAR2 at depth 357m (p<0.001). For encounter 17, three variables were significantly different (positive slope, 177m EAR2 vs 357m EAR2, p=0.008; duration, 267m Ear2 vs 357m EAR2, p=0.03; number of inflection points, 267m EAR2 vs 357m EAR2, p=0.005, 357m EAR2 vs 447m EAR2, p=0.01). For encounter 19, duration was significantly different for 177m EAR2 vs 447m EAR2 (p=0.03). For encounter 20, number of steps was significantly different for 177m EAR2 vs 357m EAR2 (p=0.02), 267m EAR2 vs 357m EAR2 (p=0.03) and 357m EAR2 vs 447m EAR2 (p=0.04). No variables were significantly different for any encounter when only whistles that were detected on all four EAR2s were included in the analysis.

Table 9. Variables that were significantly different for whistles that were measured from Lanai EAR2 recordings made at different depths, including all whistles (i.e. including those that did not appear on all EAR2s). Species that encounters were classified as based on those same whistles are given for each EAR2, with number of whistles included in the analyses in parentheses.

Encounter	Significant Variable	Depths	p	Classified as			
				70m	160m	250m	340m
1	Mean frequency	70m vs. 340m	0.008	Spinner (5)	Bottlenose (6)	Bottlenose (7)	Bottlenose (19)
		160m vs. 340m	0.04				
		250m vs. 340m	0.05				
	Median frequency	250m vs. 340m	0.08				
	Maximum frequency	70m vs. 340m	0.02				
		160m vs. 340m	0.04				

Encounter	Significant Variable	Depths	p	Classified as			
				70m	160m	250m	340m
	Minimum frequency	70m vs. 340m	0.04				
	Center frequency	70m vs. 340m	0.004				
		250m vs. 340m	0.02				
2	None			Bottlenose (44)	Bottlenose (42)	Bottlenose (43)	Bottlenose (49)
3	None			Rough-toothed (40)	Rough-toothed (40)	Bottlenose (43)	Rough-toothed (48)
5	Mean frequency	70m vs. 340m	0.01	Spinner (16)	Spinner (12)	Spinner (20)	Spinner (21)
	Median frequency	70m vs. 340m	0.02				
	Beginning frequency	70m vs. 340m	0.02				
	Minimum frequency	70m vs. 250m	0.02				
		70m vs. 340m	0.04				
	Center frequency	70m vs. 250m	0.02				
		70m vs. 340m	0.006				
	Mean slope	70m vs. 340m	0.02				
7	Median frequency	70m vs. 340m	0.02	Spinner (12)	Spinner (10)	Spotted (9)	Spotted (6)
	Center frequency	70m vs. 340m	0.04				
	Mean negative slope	70m vs. 340m	0.04				
9	none			Spinner (1)	Bottlenose (7)	Spinner (13)	Spinner (29)
10	Mean frequency	160m vs. 340m	0.04	n/a (0)	Spotted (2)	Spinner (12)	Spinner (20)
		250m vs. 340m	0.02				
	Median frequency	160m vs. 340m	0.04				
		250m vs. 340m	0.02				
	Maximum frequency	160m vs. 340m	0.04				
	Center frequency	160m vs. 340m	0.04				
		250m vs. 340m	0.04				
11	Mean frequency	160m vs. 250m	0.008	Spinner (6)	Rough-toothed (14)	Spinner (27)	Spinner (13)
	Median frequency	70m vs. 160m	0.04				
		160m vs. 250m	0.04				
	Ending frequency	70m vs. 160m	0.004				
	Maximum frequency	160m vs. 250m	0.01				

Encounter	Significant Variable	Depths	p	Classified as			
				70m	160m	250m	340m
	Center frequency	160m vs. 250m	0.01				
	Duration	160m vs. 250m	0.04				
		160m vs. 340m	0.006				
	Percent flat	70m vs. 160m	0.001				
		160m vs. 340m	0.01				
	Mean absolute slope	70m vs. 160m	0.02				
12	None			Spinner (38)	Bottlenose (37)	Bottlenose (41)	Spinner (43)
13	Maximum frequency	70m vs. 340m	0.01	Spinner (25)	Spinner (20)	Spinner (25)	Spinner (29)
14	None			Pilot whale (55)	Pilot whale (54)	Rough-toothed (57)	Rough-toothed (63)
16	None			Pilot whale (39)	Pilot whale (38)	Pilot whale (44)	Pilot whale (47)
18	None			Bottlenose (46)	Bottlenose (42)	Bottlenose (44)	Bottlenose (45)
19	Mean frequency	70m vs. 160m	0.02	Spinner (35)	Spinner (23)	Spinner (38)	Spinner (43)
		70m vs. 340m	0.01				
	Median frequency	70m vs. 160m	0.02				
		70m vs. 340m	0.01				
	Maximum frequency	70m vs. 160m	0.04				
21	None			Bottlenose (34)	Bottlenose (30)	Bottlenose (40)	Bottlenose (41)
22	Minimum frequency	70m vs. 250m	0.04	Bottlenose (43)	Bottlenose (45)	Bottlenose (48)	Bottlenose (49)
23	None			Spinner (5)	Spotted (10)	Bottlenose (12)	Bottlenose (14)
25	None			Bottlenose (46)	Bottlenose (38)	Bottlenose (44)	Bottlenose (43)
27	None			Bottlenose (34)	Spinner (40)	Spinner (40)	Bottlenose (40)
28	None			Bottlenose (42)	Bottlenose (40)	Bottlenose (42)	Bottlenose (47)

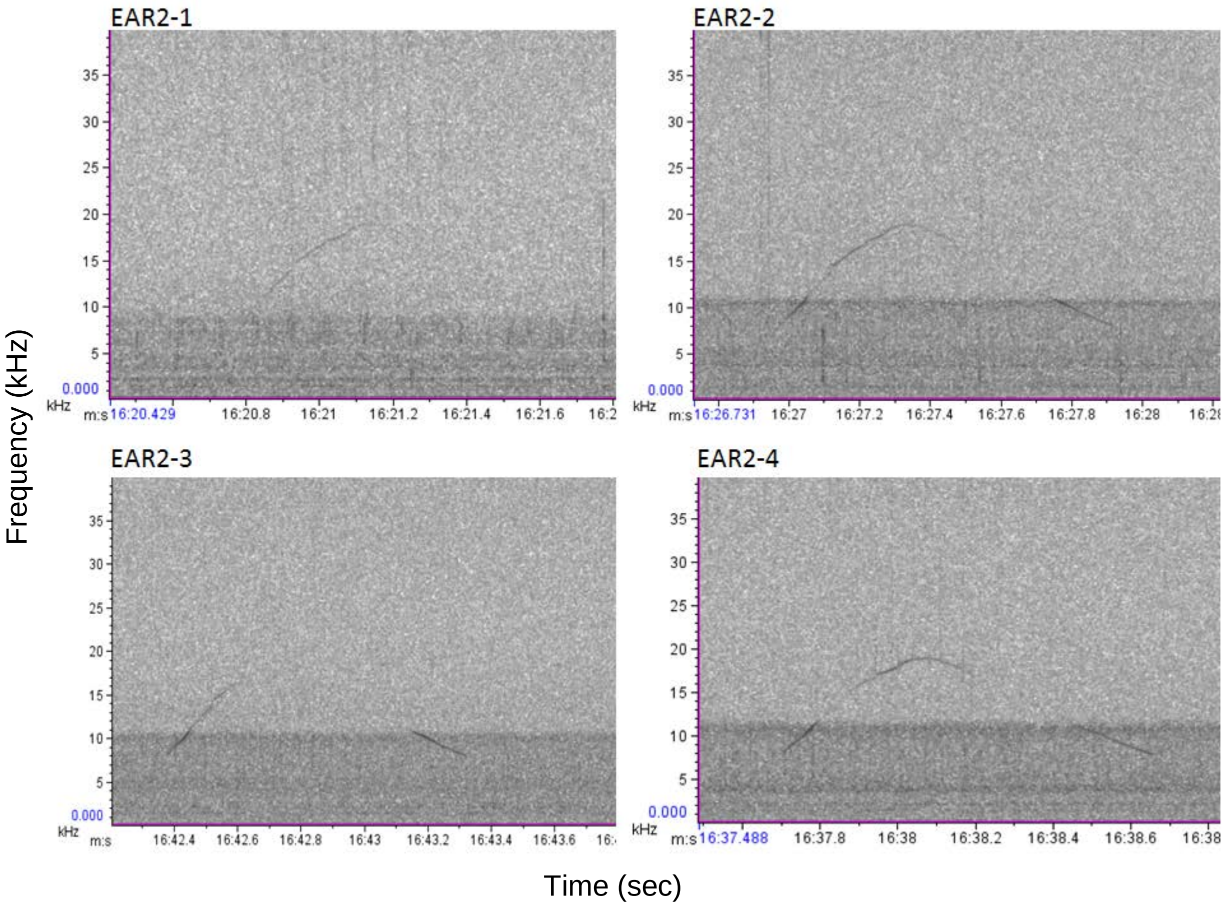


Figure 12. Example of a whistle detected on more than one EAR2, but with only a portion of the contour evident in some of the recordings.

4.5.4 Classification

4.5.4.1 LANAI

For 10 of the 20 encounters (50 percent), ROCCA classified the encounter as a different species based on whistles recorded at different depths when all whistles were included in the analysis (**Table 9**). Usually, one out of the four EAR2s had a different classification result. Four of the 10 encounters (40 percent) that were classified differently on different EAR2s had significant differences in whistle variables and six (60 percent) did not. Of the 10 encounters that were classified as the same species on all four EAR2s, four (40 percent) had significant differences in whistle variables among EAR2s. When classification results were different on one or more of the EAR2s, the percent of tree votes was often similar for the two species that the encounter was classified as on the different EAR2s.

Thirteen of the 20 (65 percent) encounters included a sufficient number of whistles that were detected on all four EAR2s (at least 10 whistles) to be included in the classification analysis. Four of the encounters (31 percent) were classified as one species for three of the EAR2s and as a different species on the fourth EAR2 (**Table 10**). The EAR2 that differed was not consistent among encounters. For all four of these encounters, the percent of trees votes was similar for the two species in question. The remaining nine encounters (69 percent) were classified as the same species on all four EAR2s.

Table 10. Species classification results by encounter for whistles appearing on all four EAR2s for Lanai dataset.

Encounter	Classified as				Number of whistles per EAR2
	70m depth	160m depth	250m depth	340m depth	
2	Bottlenose dolphin	Bottlenose dolphin	Bottlenose dolphin	Bottlenose dolphin	38
3	Rough-toothed dolphin	Rough-toothed dolphin	Rough-toothed dolphin	Rough-toothed dolphin	35
12	Spinner dolphin	Bottlenose dolphin	Bottlenose dolphin	Bottlenose dolphin	29
13	Spinner dolphin	Spinner dolphin	Spinner dolphin	Spinner dolphin	16
14	Pilot whale	Pilot whale	Rough-toothed dolphin	Pilot whale	42
16	Pilot whale	Pilot whale	Pilot whale	Pilot whale	32
18	Bottlenose dolphin	Bottlenose dolphin	Bottlenose dolphin	Bottlenose dolphin	36
19	Spinner dolphin	Spinner dolphin	Spinner dolphin	Bottlenose dolphin	15
21	Bottlenose dolphin	Bottlenose dolphin	Bottlenose dolphin	Bottlenose dolphin	25
22	Bottlenose dolphin	Bottlenose dolphin	Bottlenose dolphin	Bottlenose dolphin	38
25	Bottlenose dolphin	Bottlenose dolphin	Bottlenose dolphin	Bottlenose dolphin	30
27	Spinner dolphin	Spinner dolphin	Spinner dolphin	Bottlenose dolphin	28
28	Bottlenose dolphin	Bottlenose dolphin	Bottlenose dolphin	Bottlenose dolphin	32

4.5.4.2 KONA

A total of 892 whistles that were detected on all four EAR2s were analyzed from all Kona EAR2 recordings. Most (65 percent) of these whistles were classified as the same species on all four EAR2s and some (34 percent) were classified as two different species on different EAR2s. A small percentage (1 percent) of whistles were classified as three different species and no whistles were classified as four different species on four different EAR2s.

When all whistles were included in the classification analysis, 5 of 19 encounters (26 percent) were classified as different species at different depths (**Table 11**). When only whistles that were detected on all four EAR2s were included in the classification analysis, 3 out of 19 encounters (16 percent) were classified as different species at different depths (**Table 12**). For both analyses, the EAR2s that differed varied by encounter. Only encounter 1 was classified as different species at different depths both when all whistles were included and when only whistles detected on all four EAR2s were included.

Table 11. Species classification results by encounter including all whistles (i.e. including those that did not appear on all four EAR2s) for Kona dataset. **Number of whistles included in the analyses is in parentheses.**

Encounter	Classified as			
	118m depth	209m depth	289m depth	389m depth
1	Spinner (49)	Spinner (48)	Bottlenose (49)	Bottlenose (50)
2	Bottlenose (49)	Bottlenose (49)	Bottlenose (50)	Bottlenose (50)
3	Spinner (45)	Spinner (44)	Spinner (49)	Spinner (50)
4	Bottlenose (45)	Bottlenose (45)	Bottlenose (50)	Bottlenose (49)
5	Bottlenose (49)	Bottlenose (49)	Bottlenose (49)	Bottlenose (49)
6	Bottlenose (46)	Bottlenose (45)	Bottlenose (49)	Bottlenose (50)
7	Bottlenose (50)	Bottlenose (50)	Bottlenose (50)	Bottlenose (50)
8	Bottlenose (36)	Bottlenose (34)	Bottlenose (36)	Bottlenose (36)
9	Bottlenose (50)	Rough-toothed (49)	Rough-toothed (50)	Rough-toothed (49)
10	Bottlenose (50)	Bottlenose (50)	Bottlenose (50)	Bottlenose (50)
11	Bottlenose (50)	Bottlenose (50)	Bottlenose (50)	Bottlenose (50)
12	Bottlenose (50)	Bottlenose (49)	Bottlenose (50)	Bottlenose (50)
13	Spinner (50)	Bottlenose (50)	Spinner (50)	Bottlenose (50)
14	Bottlenose (50)	Bottlenose (49)	Bottlenose (50)	Bottlenose (50)
15	Spinner (50)	Spinner (50)	Spinner (49)	Spinner (50)
16	Spinner (50)	Bottlenose (49)	Bottlenose (50)	Bottlenose (50)
17	Bottlenose (50)	Bottlenose (50)	Bottlenose (50)	Bottlenose (50)
18	Bottlenose (28)	Bottlenose (28)	Bottlenose (28)	Spinner (28)
19	Bottlenose (24)	Bottlenose (20)	Bottlenose (27)	Bottlenose (28)

Table 12. Species classification results by encounter for whistles appearing on all four EAR2s for Kona dataset.

Encounter	Classified as				Number of whistles per EAR2
	118m depth	209m depth	289m depth	389m depth	
1	Spinner	Spinner	Bottlenose	Bottlenose	47
2	Bottlenose	Bottlenose	Bottlenose	Bottlenose	49
3	Spotted	Spotted	Spotted	Spotted	42
4	Bottlenose	Bottlenose	Bottlenose	Bottlenose	43
5	Bottlenose	Bottlenose	Bottlenose	Bottlenose	47
6	Bottlenose	Bottlenose	Bottlenose	Bottlenose	45
7	Bottlenose	Bottlenose	Bottlenose	Bottlenose	50
8	Spinner	Bottlenose	Bottlenose	Bottlenose	34
9	Rough-toothed	Rough-toothed	Rough-toothed	Rough-toothed	48
10	Bottlenose	Bottlenose	Bottlenose	Bottlenose	50
11	Bottlenose	Bottlenose	Bottlenose	Bottlenose	50
12	Bottlenose	Bottlenose	Bottlenose	Bottlenose	49
13	Bottlenose	Bottlenose	Bottlenose	Bottlenose	50
14	Bottlenose	Bottlenose	Bottlenose	Bottlenose	49
15	Spinner	Spinner	Bottlenose	Spinner	49
16	Bottlenose	Bottlenose	Bottlenose	Bottlenose	49
17	Bottlenose	Bottlenose	Bottlenose	Bottlenose	50
18	Bottlenose	Bottlenose	Bottlenose	Bottlenose	28
19	Bottlenose	Bottlenose	Bottlenose	Bottlenose	18

4.5.4.3 SAN DIEGO

A total of 665 whistles that were detected on all four EAR2s were analyzed from all San Diego EAR2 encounters. Most of these whistles (74 percent) were classified as the same species on all four EAR2s. A quarter of whistles (25 percent) were classified as 2 different species on different EAR2s and only 1 percent were classified as 3 different species on different EAR2s. No whistles were classified as four different species on all four EAR2s.

When all whistles were included in the analysis, 5 out of 20 encounters (25 percent) were classified as different species at different depths (**Table 13**). For three of these encounters, the classification was different for the EAR2 at depth 357m. These three encounters also had significant differences in whistle variables between depths. For two of the encounters the classification was different for the EAR2 at depth 177m. These encounters did not have significant differences in whistle variables between depths. When only whistles that were detected on all four EAR2s were included in the analysis, every encounter was classified as the same species on all four EAR2s.

Table 13. Species classification results by encounter including all whistles (i.e. including those that did not appear on all four EAR2s) for San Diego dataset. **Number of whistles included in the analyses is in parentheses.**

Encounter	Classified as			
	177m depth	267m depth	357m depth	447m depth
1	Bottlenose (33)	Striped (36)	Striped (39)	Striped (45)
2	Common spp. (47)	Common spp. (50)	Common spp. (48)	Common spp. (50)
3	Common spp. (37)	Common spp. (39)	Common spp. (31)	Common spp. (45)
4	Striped (43)	Striped (49)	Striped (50)	Striped (48)
5	Common spp. (29)	Common spp. (31)	Common spp. (26)	Common spp. (31)
6	Common spp. (44)	Common spp. (48)	Common spp. (40)	Common spp. (50)
7	Common spp. (49)	Common spp. (50)	Common spp. (39)	Common spp. (50)
8	Common spp. (50)	Common spp. (50)	Common spp. (47)	Common spp. (50)
9	Common spp. (40)	Striped (42)	Striped (28)	Striped (43)
10	Common spp. (36)	Common spp. (37)	Common spp. (35)	Common spp. (37)
11	Common spp. (36)	Common spp. (42)	Common spp. (40)	Common spp. (43)
12	Common spp. (20)	Common spp. (31)	Striped (33)	Common spp. (38)
13	Common spp. (37)	Common spp. (46)	Common spp. (44)	Common spp. (47)
14	Common spp. (50)	Common spp. (50)	Common spp. (50)	Common spp. (50)
15	Common spp. (33)	Common spp. (34)	Common spp. (36)	Common spp. (38)
16	Striped (48)	Striped (48)	Striped (50)	Striped (50)
17	Common spp. (43)	Common spp. (45)	Striped (31)	Common spp. (48)
18	Common spp. (50)	Common spp. (50)	Common spp. (49)	Common spp. (50)
19	Common spp. (41)	Common spp. (44)	Common spp. (42)	Common spp. (48)
20	Striped (19)	Striped (14)	Common spp. (13)	Striped (14)

4.6 MicroMARS Analyses

4.6.1 Datasets

Whistles were measured and classified from 6 out of 10 encounters recorded from 5 species with the Kona microMARS array and from 15 out of 16 encounters recorded from 2 species with the San Diego microMARS array (**Table 14**).

Table 14. Species, location, date, start time, end time and duration for each acoustic encounter recorded on the microMARS array off Kona and San Diego. **Date, start time and end time are in Greenwich Mean Time (GMT). Encounters marked with an * did not contain enough whistles to include in the analysis. Due to technical issues, microMARS 2 did not record during encounters marked with a +.**

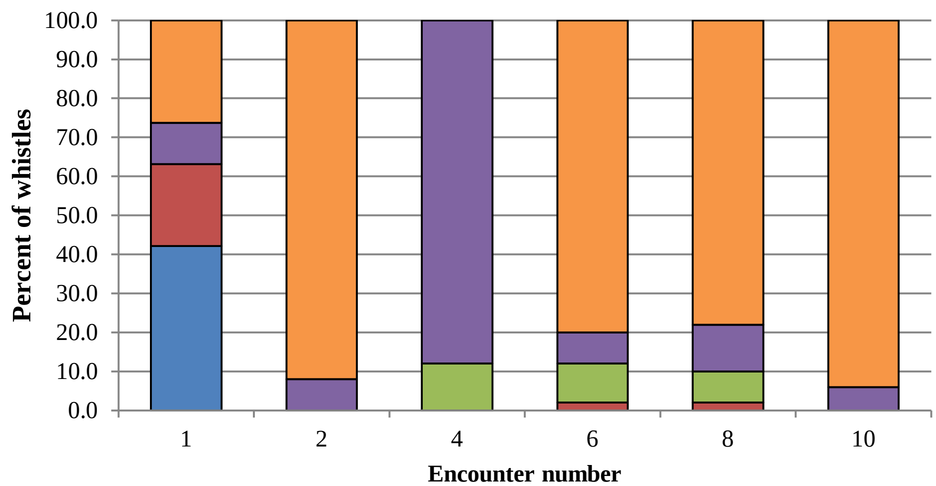
Location	Encounter	Known Species	Latitude	Longitude	Date	Start time	End time	Duration
Kona	1	Spotted dolphin	19.8299	156.1496	11/4/2015	21:12:00	21:28:00	0:16:00
	2	Pilot whale, Bottlenose dolphin	19.8589	156.2063	11/4/2016	22:37:00	23:37:00	1:00:00
	3*+	Spotted dolphin	19.0386	156.404	11/5/2015	18:48:00	19:00:00	0:12:00
	4+	Pilot whale	N/A	N/A	11/5/2015-11/6/2015	22:12:25	0:01:53	1:49:28
	5*	Rough-toothed dolphin	N/A	N/A	11/9/2015	0:00:00	0:18:00	0:18:00
	6	Spotted dolphin	19.5001	156.02173	11/10/2015	19:50:00	22:42:00	2:52:00
	7*	False killer whale, Rough-toothed dolphin	N/A	N/A	11/12/2015	1:10:23	1:40:50	0:30:27
	8	Spotted dolphin	N/A	N/A	11/13/2015	20:29:00	20:58:00	0:29:00
	9*	Pilot whale	19.81215	156.12553	11/14/2015	0:15:00	1:32:00	1:17:00
	10	Spotted dolphins	19.81215	156.12553	11/14/2015	20:06:00	20:41	0:35:00
San Diego	1	Common spp.	32.72451	-117.3785	5/16/2016	15:12:00	15:50:00	0:38:00
	3	Common spp.	32.71416	-117.3574	5/16/2016	19:08:00	19:59:00	0:51:00
	8	Short beaked common dolphin	32.72973	-117.4008	5/17/2016	22:36:00	23:48:00	1:12:00
	9	Common spp.	32.6591	-117.4785	5/18/2016	15:18:00	15:52:00	0:34:00
	10	Common spp.	32.64255	-117.4714	5/18/2016	16:24:00	16:57:00	0:33:00
	11	Common spp.	32.7016	-117.3821	5/19/2016	15:35:00	16:21:00	0:46:00
	15	Common spp.	32.70104	-117.3849	5/19/2016	19:22:00	19:39:00	0:17:00

	16	Short beaked common dolphin	32.51125	-117.3757	5/20/2016	16:51:00	18:11:00	1:20:00
	17	Bottlenose dolphin	33.3804	-118.2733	5/22/2016	15:26:00/16:03:00	15:35:00/ 16:18:00	
	18	Common spp.	33.45015	-118.3755	5/23/2016	14:28:00	15:03:00	0:35:00
	19*	Bottlenose dolphin	33.38665	-118.2994	5/23/2016	16:07:00	16:21:00	0:14:00
	20	Common spp.	33.3884	-118.2221	5/24/2016	18:37:00	19:09:00	0:32:00
	21	Short beaked common dolphin	32.68785	-117.3985	5/25/2016	15:43:00	16:09:00	0:26:00
	25	Common spp.	32.8898	-117.3772	5/26/2016	18:34:00	19:00:00	0:26:00
	26	Common spp.	32.9309	-117.4415	5/26/2016	20:27:00	20:46:00	0:19:00
	27	Common spp.	32.84453	-117.3796	5/26/2016	21:24:00	21:58:00	0:34:00

4.6.2 Percent of whistles detected at each depth

Most whistles were detected on 4 or 5 microMARS in the Kona recordings (**Figure 13a**). At least 88 percent of whistles were detected on 4 or 5 microMARS for all Kona encounters, with the exception of encounter 1. Only spotted dolphin encounters had whistles that were detected on only one or two microMARS. The percent of whistles detected on four or five microMARS was more variable for the San Diego encounters (**Figure 13b**). For 9 encounters (60 percent), at least 85 percent of whistles were detected on four or five microMARS. For three of those encounters, at least 85 percent of whistles were detected on all five microMARS. In contrast, for encounter 27, 42 percent of whistles were detected on only one or two microMARS and for encounters 20 and 26, 42 percent and 57 percent of whistles (respectively) were detected on only 1, 2, or 3 microMARS.

a)



b)

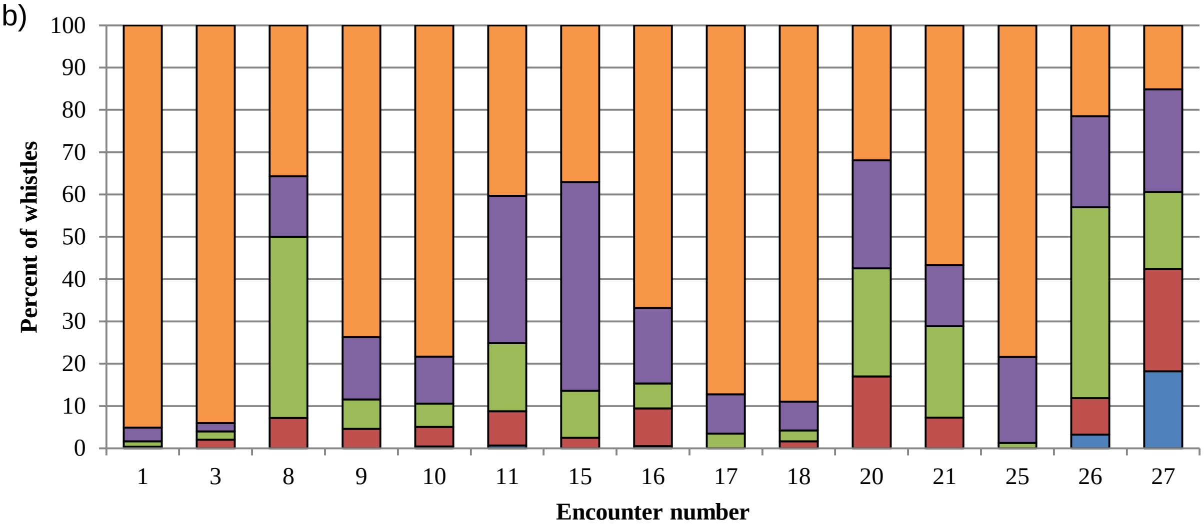


Figure 13. Percentage of whistles detected on only one microMARS (blue), on two microMARS (red), on three microMARS (green), on four microMARS (purple), and on all five microMARS (orange) for the Kona (a) and San Diego (b) microMARS arrays

4.6.3 Variable comparisons

There were no significant differences in whistle variables (Kruskall-Wallis test and post-hoc Dunn's tests with Bonferonni correction, α=0.05) among depths for any of the Kona encounters recorded with the microMARS array.

Three of the fifteen encounters recorded in San Diego with the microMARS array had significant differences in whistle variables among depths when all whistles were included in the analysis. For encounter three, positive slope, negative slope, percent of the whistle with zero slope and percent of the whistle with positive slope were significantly different for the microMARS at 100m depth compared to all other depths (p<0.001 for all comparisons). For encounter nine, minimum frequency was significantly different for 50m depth vs 200m depth (p=0.01) and 50m depth vs 150m depth (p=0.04). For encounter 25, the number of inflection points was significantly different for 50m vs all other depths (vs 100m p=0.04, vs 150m p=0.007, vs 200m p=0.007, 250m p=0.01).

Ten of the fifteen San Diego microMARS encounters contained enough whistles recorded at every depth to be included in the analyses. Of these 10 encounters, 4 had significant differences in whistle variables when only whistles recorded on all 5 microMARS were included. Most of the significant differences involved whistles recorded at 50m or 100m depth. For encounter three, positive slope, negative slope, percent of the whistle with zero slope and percent of the whistle with positive slope were significantly different for 100m depth compared to all other depths (p<0.001 for all comparisons). For encounter nine, minimum frequency was significantly different at 50m depth compared to both 150m depth (p=0.02) and 200m depth (p=0.02). The number of inflection points was significantly different for 50m depth compared to both 150m depth (p=0.004) and 200m depth (p=0.004) for encounter 21. Number of inflection points was also significantly different for encounter 25. This variable was significantly different for 50m depth compared to all other depths (vs 100m depth p=0.02, vs 150m depth p=0.03, vs 200m depth p=0.04, vs 250m depth p=0.02).

4.6.4 Classification

4.6.4.1 KONA

A total of 269 whistles that were detected on all five microMARS were analyzed from the Kona encounters. Approximately half (55 percent) of these whistles were classified as the same species on all five microMARS, 39 percent were classified as 2 different species and 6 percent were classified as 3 different species on different microMARS. No whistles were classified as four or five different species on different microMARS.

When all whistles were included in the classification analysis, all six encounters were classified as the same species at every depth (**Table 15**). Results were the same when only whistles recorded on every microMARS were included, with the exception of encounter two. This encounter was classified as spinner dolphin on all five microMARS when all whistles were included, but it was classified as bottlenose dolphin at 50m depth and 100m depth and as spinner dolphin at all deeper depths when only whistles detected on all microMARS were included.

Table 15. Species classification results by encounter and recording depth including all whistles (i.e. including those that did not appear on all five microMARS) for Kona dataset. Number of whistles included in the analyses is in parentheses.

Encounter	Known Species	Classified as				
		50m depth	100m depth	150m depth	200m depth	250m depth
1	Spotted dolphin	Spinner (16)	Spinner (8)	Spinner (8)	Spinner (9)	Spinner (9)
2	Pilot whale, Bottlenose dolphin	Spinner (50)	Spinner (50)	Spinner (48)	Spinner (49)	Spinner (49)
4	Pilot whale	Pilot Whale (48)	N/A	Pilot Whale (50)	Pilot Whale (50)	Pilot Whale (46)
6	Spotted dolphin	Spinner (49)	Spinner (43)	Spinner (50)	Spinner (46)	Spinner (45)
8	Spotted dolphin	Spinner (48)	Spinner (49)	Spinner (46)	Spinner (48)	Spinner (42)
10	Spotted dolphin	Spinner (48)	Spinner (50)	Spinner (50)	Spinner (50)	Spinner (49)

4.6.4.2 SAN DIEGO

A total of 282 whistles that were detected on all five microMARS were included in the analysis. Of these whistles, approximately half (56 percent) were classified as the same species on all microMARS and slightly less than half (42 percent) were classified as 2 different species on different microMARS. A small percentage (two percent) were classified as three different species and no whistles were classified as four or five different species on different microMARS.

When all whistles were included in the analysis, 4 out of 15 encounters (27 percent) were classified as different species at different depths (**Table 16**). Five of the fifteen encounters did not contain enough whistles detected at every depth to be included in the classification analysis. Of the 10 encounters that were included, the classification results were the same as the 'all whistles' results for most encounters (**Table 17**). Only encounters 3 and 11 differed between the two analyses. Two of the ten encounters were classified as different species at different depths.

Table 16. Species classification results by encounter including all whistles (i.e. including those that did not appear on all five microMARS) for San Diego dataset. Number of whistles included in the analyses is in parentheses.

Encounter	Known Species	Classified as				
		50m depth	100m depth	150m depth	200m depth	250m depth
1	Common spp.	Common spp. (43)	Common spp. (42)	Common spp. (37)	Common spp. (39)	Common spp. (42)
3	Common spp.	Striped (50)	Striped (45)	Striped (50)	Common spp. (38)	Striped (47)
8	Short-beaked common	Striped (14)	Striped (13)	Striped (13)	Common spp. (7)	Common spp. (6)

Encounter	Known Species	Classified as				
		50m depth	100m depth	150m depth	200m depth	250m depth
9	Common spp.	Common spp. (46)	Common spp. (44)	Common spp. (49)	Common spp. (38)	Common spp. (34)
10	Common spp.	Common spp. (47)	Common spp. (44)	Common spp. (49)	Common spp. (39)	Common spp. (37)
11	Common spp.	Common spp. (35)	Common spp. (32)	Common spp. (37)	Common spp. (23)	Striped (12)
15	Common spp.	Common spp. (19)	Common spp. (18)	Common spp. (19)	Common spp. (15)	Common spp. (6)
16	Short-beaked common	Common spp. (44)	Common spp. (40)	Common spp. (42)	Common spp. (36)	Common spp. (38)
17	Bottlenose	Common spp. (34)	Common spp. (33)	Common spp. (34)	Common spp. (33)	Common spp. (31)
18	Common spp.	Common spp. (41)	Common spp. (42)	Common spp. (40)	Common spp. (45)	Common spp. (45)
20	Common spp.	Common spp. (11)	Common spp. (10)	Common spp. (11)	Common spp. (9)	Common spp. (6)
21	Short-beaked common	Common spp. (49)	Common spp. (41)	Common spp. (49)	Common spp. (30)	Common spp. (24)
25	Common spp.	Common spp. (50)	Common spp. (50)	Common spp. (49)	Common spp. (44)	Common spp. (42)
26	Common spp.	Common spp. (25)	Common spp. (24)	Common spp. (28)	Common spp. (7)	Common spp. (9)
27	Common spp.	Common spp. (6)	Common spp. (6)	Common spp. (5)	Common spp. (5)	Striped (11)

Table 17. Species classification results by encounter including only whistles detected on all five microMARS for San Diego dataset.

Encounter	Known Species	Whistles per microMARS	Classified as				
			50m depth	100m depth	150m depth	200m depth	250m depth
1	Common spp.	25	Common spp.	Common spp.	Common spp.	Common spp.	Common spp.
3	Common spp.	37	Common spp.	Common spp.	Striped	Common spp.	Common spp.
9	Common spp.	30	Common spp.	Common spp.	Common spp.	Common spp.	Common spp.
10	Common spp.	34	Common spp.	Common spp.	Common spp.	Common spp.	Common spp.
11	Common spp.	9	Striped	Common spp.	Striped	Common spp.	Striped
16	Short-beaked common	26	Common spp.	Common spp.	Common spp.	Common spp.	Common spp.
17	Bottlenose	25	Common spp.	Common spp.	Common spp.	Common spp.	Common spp.

Encounter	Known Species	Whistles per microMARS	Classified as				
			50m depth	100m depth	150m depth	200m depth	250m depth
18	Common spp.	37	Common spp.	Common spp.	Common spp.	Common spp.	Common spp.
21	Short-beaked common	22	Common spp.	Common spp.	Common spp.	Common spp.	Common spp.
25	Common spp.	37	Common spp.	Common spp.	Common spp.	Common spp.	Common spp.

4.7 Navy dolphin analysis

4.7.1 Dataset

The controlled data collection experiment with the trained Navy bottlenose dolphin was conducted over three days, 17 May, 19 May, 25 May 2016. A total of 11 trials were conducted at different distances and orientations of the dolphin relative to the array (**Table 18**). Unfortunately, wild common dolphins were in the area during all three days (sometimes passing within tens of meters of the Navy dolphin). Because of the proximity of common dolphins during the trials, in many cases it was not possible to confidently determine which whistles were produced by the Navy dolphin. Because of the small sample sizes for individual trails, all trials for each distance were combined and analyzed as one 'encounter'. This resulted in four 'encounters': 50m facing the array, 50m facing away from the array, 100m facing the array, and 250m facing the array. There were not enough whistles detected from the 400m trials or the 100m and 200m facing away from the array trials to be included in the analysis.

Table 18. Trails conducted with trained Navy dolphin stationed at 5m depth. **Distance is distance between the array and the dolphin and orientation is the orientation of the dolphin relative to the array.**

Date	Distance	Orientation	Number of trials
17/05/2016	50m	Towards array	11
17/05/2016	100m	Towards array	10
17/05/2016	400m	Towards array	5
19/05/2016	50m	Towards array	6
19/05/2016	50m	Away from array	10
19/05/2016	100m	Towards array	5
19/05/2016	100m	Away from array	5
19/05/2016	250m	Towards array	10
25/05/2016	100m	Towards array	5
25/05/2016	100m	Away from array	10
25/05/2016	200m	Away from array	5

4.7.2 Percent of whistles detected at each depth

Almost all whistles were detected on all five microMARS at every distance (**Figure 14**). The smallest percentage of whistles detected on all five microMARS was 90 percent. This occurred when the dolphin was stationed 50m away from the array and was facing directly away from the array.

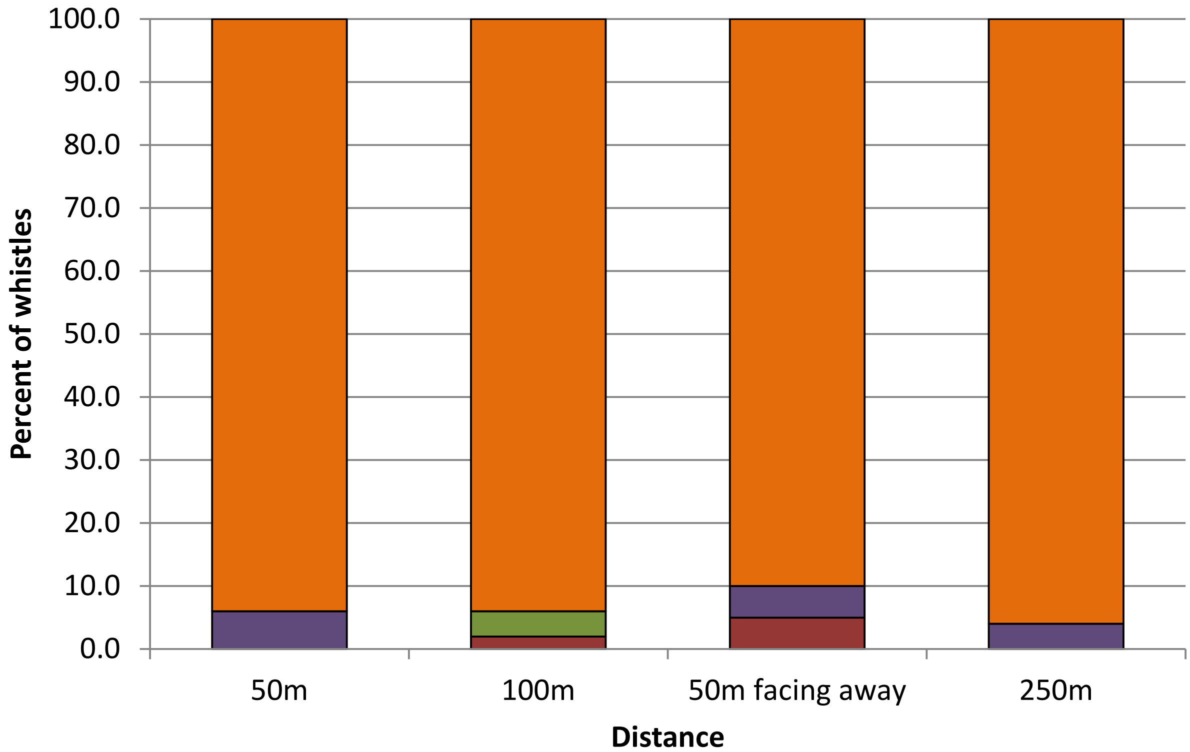


Figure 14. Percentage of whistles detected on two microMARS (red), on three microMARS (green), on four microMARS (purple), and on all five microMARS (orange) for the Navy dolphin trials.

4.7.3 Variable comparisons

4.7.3.1 BY DEPTH

Only beginning frequency and minimum frequency were significantly different (Kruskall-Wallis test and post-hoc Dunn's tests with Bonferonni correction, α=0.05) among depths for the 50m and 100m distance trials and only beginning frequency was significantly different among depths for the 50m facing away trial (**Table 19**). For the 100m trial, a greater number of variables were significantly different among depths. Many of the significant differences at 100m occurred at 50m depth. Results were very similar when only whistles detected at all five depths were included. For the 100m trial, median, end and maximum frequency were not significantly different when only whistles detected at all depths were included (**Table 19**).

Table 19. Variables that were significantly different for whistles were measured from microMARS recordings made at different depths during the Navy dolphin trials, including all whistles (i.e. including those that did not appear on all microMARS). Species that encounters were classified as based on those same whistles are given for each depth, with number of whistles included in the analyses in parentheses.

Encounter Distance	Significant Variable	Depth	p (all whistles)	p (whistles heard on all 5 MM)
50m	Beginning frequency	50m vs 200m	0.0003	0.0005
		50m vs 250m	<0.0001	<0.0001
		100m vs 200m	0.05	N/A
		100m vs 250m	0.0002	0.0004
		150m vs 250m	0.0005	0.0004
	Minimum frequency	50m vs 200m	0.0003	0.0006
		50m vs 250m	<0.0001	<0.0001
		100m vs 200m	0.01	0.03
		100m vs 250m	0.0001	0.0002
		150m vs 250m	0.0009	0.0007
100m	Median Frequency	50m vs 250m	0.02	N/A
	Mean positive slope	50m vs 100m	<0.0001	<0.0001
		50m vs 150m	<0.0001	<0.0001
		50m vs 200m	<0.0001	<0.0001
		50m vs 250m	<0.0001	<0.0001
	Beginning frequency	50m vs 200m	0.01	0.0101
		50m vs 250m	0.0003	0.0002
		100m vs 250m	0.004	0.002
		150m vs 250m	0.02	0.01
	End frequency	100m vs 250m	N/A	0.006
	Maximum frequency	100m vs 250m	N/A	0.006
	Minimum frequency	50m vs 200m	0.002	0.003
		50m vs 250m	<0.0001	<0.0001
		100m vs 200m	<0.0001	0.03
		100m vs 250m	0.0005	0.0003
		150m vs 250m	0.002	0.002
	Duration	50m vs 200m	0.006	0.01
		50m vs 250m	<0.0001	<0.0001
		100m vs 200m	0.0162	0.01
		100m vs 250m	<0.0001	<0.0001
		150m vs 250m	0.0005	0.0004

Encounter Distance	Significant Variable	Depth	p (all whistles)	p (whistles heard on all 5 MM)
	Percent of whistle with zero slope	50m vs 100m	0.0009	0.002
		50m vs 150m	0.0002	0.0002
		50m vs 200m	0.002	0.0009
		50m vs 250m	0.002	0.001
	Percent of whistle with positive slope	50m vs 100m	0.003	0.004
		50m vs 150m	0.002	0.001
		50m vs 200m	0.01	0.004
		50m vs 250m	0.04	0.01
	Percent of whistle with negative slope	50m vs 100m	0.02	0.01
		50m vs 150m	0.01	0.006
		50m vs 200m	0.04	0.01
50m (facing away)	Beginning frequency	100m vs 200m	0.04	0.03
250m	Beginning frequency	50m vs 100m	0.0003	0.0002
		50m vs 150m	0.004	0.004
		100m vs 250m	<0.0001	<0.0001
		150m vs 250m	0.0002	0.0001
	Minimum frequency	50m vs 100m	0.0001	<0.0001
		50m vs 150m	0.004	0.003
		100m vs 200m	0.04	0.02
		100m vs 250m	<0.0001	<0.0001
		150m vs 250m	0.0007	0.0007

4.7.3.2 BY DISTANCE

Variables were also compared between distances for each recording depth. Only encounters with the dolphin facing towards the array were included in this analysis to remove the confounding variable of animal orientation from the comparisons. The number of variables that were significantly different among distances was highest for 50m and 100m depth (12 and 9 variables, respectively). At 150m, 200m and 250m depths, the number of variables that were significantly different decreased to 6 (150m) or 7 (200m and 250m). At 50m depth, significant differences included all distances, but at depths of 100m and greater, all significant differences were between 50m and 100 or 250m distances.

Table 20. Variables that were significantly different between distances for whistles were measured from microMARS recordings made during the Navy dolphin trials, including all whistles (i.e. including those that did not appear on all microMARS). **Species that encounters were classified as based on those same whistles are given for each depth, with number of whistles included in the analyses in parentheses.**

microMARS depth	Significant Variable	Distances	p
50m	Median frequency	50m – 100m	0.04
		100m – 250m	0.002
	Mean slope	50m – 100m	0.001
		50m – 250m	0.005
	Mean positive slope	50m – 100m	0.04
		50m – 250m	<0.0001
		100m – 250m	<0.0001
	Mean negative slope	50m – 100m	0.0001
	Number of inflections	50m – 100m	0.0001
		50m – 250m	<0.0001
	Beginning frequency	50m – 250m	<0.0001
		100m – 250m	0.004
	Minimum frequency	50m – 250m	<0.0001
		100m – 250m	0.0003
	Duration	50m – 100m	0.0007
		50m – 250m	<0.0001
		100m – 250m	0.03
	Percent of whistle with zero slope	50m – 100m	0.0003
		100m – 250m	0.002
	Percent of whistle with positive slope	50m – 100m	<0.0001
		100m – 250m	0.007
	Percent of whistle with negative slope	50m – 100m	<0.0001
		50m – 250m	0.03
	Number of steps	50m – 100m	0.02
		50m – 250m	0.005
100m	Mean slope	50m – 100m	0.0007
		50m – 250m	0.0008
	Mean positive slope	50m – 100m	<0.0001
		50m – 250m	<0.0001
	Number of inflections	50m – 100m	0.0005
		50m – 250m	<0.0001
	End frequency	50m – 100m	0.003
	Duration	50m – 100m	0.0008
		50m – 250m	<0.0001
	Percent of whistle with zero slope	50m – 100m	0.02
		50m – 250m	0.007
	Number of steps	50m – 250m	0.01
	Mean absolute slope	50m – 250m	0.02

microMARS depth	Significant Variable	Distances	p
	Center frequency	50m – 100m	0.02
150m	Mean slope	50m – 100m	0.0006
		50m – 250m	0.0008
	Mean positive slope	50m – 100m	<0.0001
		50m – 250m	<0.0001
	Number of inflections	50m – 100m	<0.0001
		50m – 250m	<0.0001
	Duration	50m – 100m	0.0008
		50m – 250m	0.0002
	Percent of whistle with negative slope	50m – 100m	0.04
		50m – 250m	0.008
	Number of steps	50m – 100m	0.009
		50m – 250m	0.03
200m	Mean slope	50m – 100m	0.0001
		50m – 250m	0.0006
	Mean positive slope	50m – 100m	<0.0001
		50m – 250m	<0.0001
	Number of inflections	50m – 100m	0.002
		50m – 250m	<0.0001
	Duration	50m – 100m	0.0004
		50m – 250m	0.02
	Percent of whistle with zero slope	50m – 100m	0.008
		50m – 250m	0.004
	Mean absolute slope	50m – 100m	0.04
		50m – 250m	0.02
250m	Mean slope	50m – 100m	0.0001
		50m – 250m	0.02
	Mean positive slope	50m – 100m	<0.0001
		50m – 250m	<0.0001
	Mean negative slope	50m – 100m	0.0001
		50m – 250m	0.02
	Number of inflections	50m – 100m	<0.0001
		50m – 250m	0.0008
	Duration	50m – 100m	0.0002
	Percent of whistle with negative slope	50m – 100m	0.002
		50m – 250m	0.02
	Number of steps	50m – 250m	0.03

4.7.4 Classification results

Classification results changed based on the receiver depth as well as the distance and orientation of the dolphin relative to the array (**Table 21**). When the dolphin was stationed 50m away from and facing the array, the encounter was classified as striped dolphin at all depths when all whistles were included in the analysis. In contrast, when the dolphin was stationed at 50m but facing away from the array, the encounter was classified as striped dolphin at 50m and 100m depth and as common dolphin at all deeper depths. At 100m distance, the encounter was classified as common dolphin at 100m depth and as striped dolphin at all other depths. The 250m encounter was classified as common dolphin at all depths. The species classification at a given receiver depth changed with the distance and orientation of the dolphin relative to the array. For example, at 100m receiver depth, encounters were classified as striped dolphin when the dolphin was stationed at a distance of 50m from the array (both facing towards and away from the array) and as common dolphin when the dolphin was stationed at distances of 100m and 200m from the array.

Table 21. Species classification results by encounter and depth including all whistles (i.e. including those that did not appear on all five microMARS) for Navy dolphin trials. **Number of whistles included in the analyses is in parentheses.**

Encounter Distance	Classified as				
	50m depth	**100m depth**	**150m depth**	**200m depth**	**250m depth**
50m facing array	Striped (50)	Striped (50)	Striped (50)	Striped (50)	Striped (50)
50m facing away	Striped (20)	Striped (20)	Common spp. (19)	Common spp. (19)	Common spp. (18)
100m	Striped (50)	Common spp. (50)	Striped (49)	Striped (47)	Striped (47)
250m	Common spp. (50)	Common spp. (50)	Common spp. (50)	Common spp. (50)	Common spp. (48)

5. Discussion

The results of this work paint a complicated picture of the effect of receiver depth on whistle characteristics and species classification. Portions of some whistles that were detected at multiple depths were not visible on spectrograms at every depth (**Figure 12**), however this did not generally have a significant effect on whistle variables in the EAR2 datasets. Only one variable for one Lanai EAR2 encounter was significantly different when compared among depths and no variables were significantly different among depth in the Kona and San Diego datasets. For the microMARS datasets, whistle variables did not change significantly with depth in the Kona recordings, but 4 out of 10 San Diego microMARS encounters did have significant differences in whistle variables. Most of these differences involved whistles recorded at depths of 50m or 100m. This suggests that, at least at the San Diego study site, differences in whistle structure may be more pronounced in the surface waters. This may be due to more complicated sound propagation pathways in the surface waters vs deeper waters. Surface reflection and surface ducting lead to constructive and destructive interference in acoustic signals (Au and Hastings 2008) and could explain some of the differences seen in the surface array recordings. Acoustic propagation modeling to investigate this is currently underway under an ONR-funded portion of this project.

Although there were very few significant differences in whistle variables among depths, for some encounters classification results did vary by depth. A small percentage of both EAR2 and microMARS encounters, ranging from 16 percent for Kona EAR2s to 31 percent for Lanai EAR2s were classified as different species at different depths when only whistles detected at every depth were included in the analyses. The whistle variable comparison included only 12 out of the 50 variables that are included in the random forest classifier, and it is possible that significant differences in the remaining 38 variables contributed to the classification differences. The 12 variables that were compared included the variables that are generally most important in the random forest classification models and we omitted correlated variables such as frequency range (which is correlated with minimum and maximum frequency). None-the-less, the variables that were not compared may have varied with depth and affected classification results.

The compliment of whistles available for analysis is an additional factor that may be influenced by recording depth and affect classification results. The compliment of whistles available for analysis at different depths did appear to have a greater and more variable effect on whistle variables and classification results than did changes in whistle structure with depth. The percent of whistles detected at each depth varied between geographic locations and also between encounters within locations. In Kona, most whistles recorded with the EAR2 and microMARS arrays were detected on all recorders (**Figures 11 and 13**). The percent of whistles detected on all recorders was more variable for the San Diego EAR2 array, ranging from 41 percent to 100 percent and it was still more variable for the Lanai EAR2 array, ranging from 0 percent to 76 percent. Similar trends were evident in the microMARS datasets. These differences between and within locations were reflected in both the whistle variable comparisons and the classification results. For Kona, where most whistles were detected on all recorders, there were no significant differences in whistle variables for either the EAR2 or microMARS recordings when all whistles were included in the analysis. In contrast, in San Diego, fewer whistles were detected on all recorders and several microMARS (20 percent) and EAR2 (20 percent) encounters had significant differences in whistle variables among depths. The percent of microMARS encounters classified as the same species on all recorders was lower in San Diego than in Kona (73 percent vs 100 percent, respectively), which reflects the differences in whistle variable comparisons. However, the percent of EAR2 encounters that were classified as the same species on all recorders was similar between the two locations (San Diego: 25 percent, Kona 26 percent). As mentioned previously, the fact that the Kona EAR2 encounters did not have significant differences in whistle variables but did show some differences in classification results may be caused by variables that were not included in the variable comparison.

Both the compliment of whistles and characteristics of whistle contours received at different depths can be impacted by many factors, including sound propagation through different water columns, the distance and orientation of phonating animals relative to the receivers, and animal behavior. As sound travels through the water column, energy is lost through attenuation, with higher frequencies attenuating more quickly than lower frequencies. In addition, reflection of sound off the sea surface and at boundaries such as the thermocline cause constructive and destructive interference in received signals. Finally, shadow zones occur at certain depths and distances from sound sources (Medwin and Clay 1998). All of these effects are currently being investigated using sound propagation modeling in an ONR-funded portion of this project.

The distance and orientation of the signaling dolphin relative to the receivers is likely to have a significant effect on the signals that are received. For the microMARS recordings, we could assume that the signaling dolphins were relatively close to the surface and within hundreds of meters of the receivers, but it was not possible to determine which specific animals were producing sounds based on field observations. The EAR2 recordings did not have associated visual observations and animals could have been anywhere in the water column and at a range of distances and orientations relative to the recorders. Acoustic localization is currently under-way through an ONR-funded portion of this project, but this analysis will provide only approximate locations of phonating dolphins. The controlled experiment with the Navy dolphin stationed at a known depth, distance and orientation relative to the recorders allowed us to examine these factors in a way that was not possible with data collected from free-swimming dolphins. When the Navy dolphin was facing the mircoMARS array and was stationed at distances of 50m, 100m and 250m from the array, at least 94 percent of whistles were detected at every depth and even when the dolphin was facing directly away from the array at a distance of 50m, 90 percent of whistles were detected at every depth (**Figure 14**). Very few whistles were detected when the dolphin was stationed 400m from the array, and it was not possible to analyze the 100m and 250m trials with the dolphin facing away from the array due to the presence of whistling common dolphins. It would be valuable to conduct this experiment again, at a time when common dolphins are not present.

In contrast to the free-ranging dolphin recordings, in the controlled experiment, changes in whistle structure with depth had a greater effect on whistle variable comparisons and classification results than did the compliment of whistles available for analysis at different depths. Because most whistles in this experiment were detected at all depths, there were few differences between analyses including all whistles and those including only whistles detected at all depths. There were few significant differences in whistle variables at different depths for the 50m trials and the 250 trial, and many significant differences for the 100m trial (**Table 19**). The 100m trial was also classified as different species at different depths. These differences were likely caused by sound propagation effects and this is currently being investigated. Sound propagation effects are also the likely cause of differences in whistle variables between distances within depths. A greater number of variables were significantly different when compared across distances at 50m depth and 100m depth than for the deeper receivers. It is possible that the thermocline was located below 100m, causing more complicated sound propagation in the surface waters and the microMARS located below 100m were receiving more direct-path sounds. We hope to answer those questions with the results of the on-going ONR sound propagation effort.

All of the Navy dolphin trials were misclassified as either common dolphin or striped dolphin. It is possible that, despite our best efforts, some of the whistles included in the analysis were produced by the common dolphins that were in the area. It is more likely, however, that the misclassifications were caused by the fact that the Navy dolphin was producing a trained whistle-type. This whistle was part of the dolphins 'natural' repertoire, but as the dolphin has lived in captivity for many years, this whistle may not be representative of the true 'natural' repertoire of bottlenose dolphin whistles.

In the wild, schools of dolphins are often spread over hundreds of meters and individual animals are oriented in many different directions at any one time and because of this all of the factors described above come into play for a given encounter. Next steps towards a deeper understanding of the interplay between animal location and orientation, receiver depth and sound propagation would include a larger sample size of surface array recordings from known species with detailed behavioral observations. The addition of time-depth and acoustic recording tags such as DTAGs (Johnson and Tyack 2003) would allow us to determine the position and orientation of the whistling dolphin relative to the array with a greater degree of accuracy and would provide insights similar to those provided by the controlled experiment, but under a wider range of conditions and for a greater number of species. Finally, efforts should be made to obtain surface recordings and bottom recordings from individual groups of dolphins. This is logistically challenging, but could be possible with a significant amount of dedicated effort in areas of high dolphin density, such as the Kona coast of Hawaii. This was not possible during the current study, as our focus was on obtaining surface array data from as many schools and species as possible.

Overall, the effect of recording depth on whistle variables and classification results varied by geographic location and encounter. Classification results varied with depth for fewer than half of the EAR2 encounters and for fewer than a quarter of microMARS encounters. Classification results varied to a greater extent when whistles produced by only one dolphin stationed at a fixed and known distance, depth and orientation were analyzed. This suggests that analyzing whistles produced by multiple dolphins at different distances and orientations relative to the receiver provides a significant advantage. The interplay between the location and orientation of whistling dolphins, the sound propagation characteristics of the water column and likely other factors such as group size and behavior is complicated and requires further data collection and analysis for a more complete understanding of the extent to which classifiers can be generalized to different recording scenarios.

6. Acknowledgements

We would like to thank Gabriela Alongi, Mark Cotter, Liz Ferguson, Megan McElligott, Tom Norris, Tina Yack, and Eden Zang for their hard work preparing equipment and for their help in the field. We are grateful to our boat captains, Jamie Thinnes, Pete McCormick and Jeff Stock, for their dedication, patience and skill in getting us to where we needed to be. We thank Dorian Houser, Elaine Allen, Jefferson Filamor, Catlin Rixon and Nicole Simon at the National Marine Mammal Foundation for their efforts in training the Navy dolphin and for all of their help in the field trials. We would also like to thank Elizabeth Henderson and Jay Barlow for logistical assistance. We are grateful to Michael Oswald for assistance with data analysis and helpful comments on drafts of this report. We thank US Fleet Forces Command under the US Navy's Marine Species Monitoring Program, ONR and LMR for funding this project. Project management was provided by NAVFAC Atlantic (NAVFAC LANT). We are especially grateful to Dan Engelhaupt and Michael Richlen (HDR), Joel Bell (NAVFAC LANT), Mike Weise (ONR), Anu Kumar (LMR) and Mandy Shoemaker (LMR) for their assistance.

7. References

Au, W.W.L., and M. Hastings. 2008. Principles of marine bioacoustics. Springer, NY. 679pp.

Azzolin, M., A. Gannier, M.O. Lammers, J.N. Oswald, E. Papale, G. Buscaino, G. Buffa, S. Mazzola and C. Giacoma. 2014. Combining whistle acoustic parameters to discriminate Mediterranean odontocetes during passive acoustic monitoring. *Journal of the Acoustical Society of America* 135:502-512.

Barkley, Y., J.N. Oswald, J.V. Carretta, S. Rankin, A. Rudd, and M.O. Lammers. 2011. Comparison of real-time and post-cruise acoustic species identification of dolphin whistles using ROCCA (Real-time Odontocete Call Classification Algorithm). NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-473.National Marine Fisheries Service, La Jolla, CA. 29 pp.

Baumann-Pickering, S, S. Wiggins, J.A. Hildebrand, M.A. Roch and H.U. Schnitzer. 2010. Discriminating features of echolocation clicks of melon-headed whales (*Peponocephala electra*), bottlenose dolphins (*Tursiops truncatus*), and Gray's spinner dolphins (*Stenella longirostris longirostris*). *Journal of the Acoustical Society of America* 128: 2212-2224.

Bauman-Pickering, S., Simonis, A.E., Oleson, E.M., Rankin, S., Baird, R., Roch, M.A., Wiggins, S.M., and J.A. Hildebrand. 2011. False killer whale and short-finned pilot whale acoustic occurrences around the Hawaiian Islands. Proceedings of 19th Biennial Conference on the Biology of Marine Mammals, November, Tampa, FL.

Breiman, L. (2001). Random forests. *Machine Learning* 45: 5-32.

Gillespie, D., J. Gordon, R. McHugh, D. McLaren, D.K. Mellinger, P. Redmond, A. Thode, P. Trinder, and D. Xiao. 2008. PAMGUARD: Semiautomated, open-source software for real-time acoustic detection and localization of cetaceans. *Proceedings of the Institute of Acoustics* 30 (Part 5): 9 pp.

Jarvis, S., N. DiMarzio, R. Morrissey and D. Moretti. 2008. A novel multi-class support vector machine classifier for automated classification of beaked whales and other small odonotocetes. *Canadian Acoustics* 36: 34–40.

Johnson, M.P., and P.L. Tyack. 2003. A digital acoustic recording tag for measuring the response of wile marine mammals to sound. *IEEE Journal of Oceanic Engineering* 28: 3-12.

Medwin, H. and C.S. Clay. 1998. *Fundamentals of Acoustical Oceanography.* Academic Press, San Diego, 712pp.

Oswald, J.N. 2013. *Development of a Classifier for the Acoustic Identification of Delphinid Species in the Northwest Atlantic Ocean.* Final Report. Submitted to HDR Environmental, Operations and Construction, Inc. Norfolk, Virginia under Contract No. CON005-4394-009, Subproject 164744, Task Order 003, Agreement # 105067. Prepared by Bio-Waves, Inc., Encinitas, California.

Oswald, J.N., S. Rankin, J. Barlow, M. Oswald and M.O. Lammers. 2013. Real-time Call Classification Algorithm (ROCCA): software for species identification of delphinid whistles. In: Detection, Classification and Localization of Marine Mammals using Passive Acoustics, 2003-2013: 10 years of International Research, DIRAC NGO, Paris, France, pp. 245-266.

Rendell, L.E., J.N. Matthews, A. Gill, J.C.D. Gordon, and D.W. Macdonald. 1999. Quantitative analysis of tonal calls from five odontocete species, examining inter-specific and intra-specific variation. *Journal of Zoology* 249:403-410.

Ridgway, S.A., D.A. Carder, T. Kamolnick, R.R. Smith, C.E. Schlundt, and W.R. Elsberry. 2001. Hearing and whistling in the deep sea: depth influences whistle spectra but does not attenuate hearing by white whales (*Delphinapterus leucas*) (Odontoceti, Cetacea). *Journal of Experimental Biology* 204:3829-3841.

Roch M.A., H. Klinck, S. Baumann-Pickering, D.K. Mellinger, S. Qui, M.S. Soldevilla and J.A. Hildebrand. 2011. Classification of echolocation clicks from odontocetes in the Southern California Bight. *Journal of the Acoustical Society of America* 129: 467–475.

Soldevilla, M.S., E.E. Henderson, G.S. Campbell, S.M. Wiggins, J.A. Hildebrand and M.A. Roch. 2008. Classification of Risso's and Pacific white-sided dolphins using spectral properties of echolocation clicks. *Journal of the Acoustical Society of America* 124: 609–624.

Sturtivant, C., and S. Datta. 1997. Automatic dolphin whistle detection, extraction, encoding, and classification. *Proceedings of the Institute of Acoustics* 19: 259-266.

Wiggins, S. 2007. Triton (Version 1.80) [Acoustic Processing Software]. Scripps Institution of Oceanography, UC San Diego, La Jolla, California. Retrieved August 1, 2011. Available from www.cetus.ucsd.edu.