doi:10.7289/V5/TM-SWFSC-552

NOAA Technical Memorandum NMFS

OCTOBER 2015

WHISTLE CLASSIFICATION IN THE CALIFORNIA CURRENT: A COMPLETE WHISTLE CLASSIFIER FOR A LARGE GEOGRAPHIC REGION WITH HIGH SPECIES DIVERSITY

Jennifer L. Keating[1], Julie N. Oswald[2]
Shannon Rankin[1], Jay Barlow[1]

NOAA-TM-NMFS-SWFSC-552

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Marine Fisheries Service
Southwest Fisheries Science Center

doi:10.7289/V5/TM-SWFSC-552

NOAA Technical Memorandum NMFS
This TM series is used for documentation and timely communication of preliminary results, interim reports, or special purpose information. The TMs have not received complete formal review, editorial control, or detailed editing

OCTOBER 2015

WHISTLE CLASSIFICATION IN THE CALIFORNIA CURRENT: A COMPLETE WHISTLE CLASSIFIER FOR A LARGE GEOGRAPHIC REGION WITH HIGH SPECIES DIVERSITY

Jennifer L. Keating[1], Julie N. Oswald[2]
Shannon Rankin[1], Jay Barlow[1]

[1]Marine Mammal and Turtle Division
Southwest Fisheries Science Center
NOAA National Marine Fisheries Service
8901 La Jolla Shores Dr.
La Jolla, CA 92037, USA

[2]Bio-Waves Inc
364 2nd Street, Suite #3
Encinitas, CA 92024, USA

NOAA-TM-NMFS-SWFSC-552

U.S. DEPARTMENT OF COMMERCE
Penny S. Pritzker, Secretary of Commerce
National Oceanic and Atmospheric Administration
Dr. Kathryn D. Sullivan, Administrator
National Marine Fisheries Service
Eileen Sobeck, Assistant Administrator for Fisheries

INTRODUCTION

The value of passive acoustics as a tool for studying marine mammals relies on the ability to detect and classify sounds associated with these species. Classification efforts typically focus on a single species, or occasionally on a few species found within a geographic region. In an effort to improve our ability to classify species during shipboard population surveys and other passive acoustic monitoring needs, we have developed an automated whistle classification algorithm that includes all whistling species found within a large geographic region, the California Current.

In order to train this whistle classifier, single-species encounters within the California Current and near (Eastern Tropical Pacific) were compiled from archived recordings. Whistles were automatically detected using the 'Whistle and Moan Detector' (WMD) within the acoustic data processing software platform, PAMGUARD (Gillespie 2008). Measurements from each whistle contour were automatically extracted using the ROCCA (Real-time Odontocete Call Classification Algorithm) module in PAMGUARD. These data were used to train and test a region specific ROCCA whistle classifier to be used on future surveys. This report presents the data used to develop this California Current whistle classifier and the results from testing the classifier on this training data.

METHODS

Whistles were extracted from archived recordings collected by Southwest Fisheries Science Center (SWFSC), Northwest Fisheries Science Center (NWFSC), Cascadia Research Collective (CRC), and Scripps Institution of Oceanography (SIO) (Tables 1 & 2). Only recordings that were collected in or near the California Current Study area (Fig. 1) were considered to minimize the potential effects of geographic variation in whistle structure on classifier performance. SWFSC and NWFSC data were collected with towed hydrophone arrays during multiple cruises (Archer *et al*. 2008, Barlow *et al.* 2014, Hanson *et al*. 2008a & 2008b, Rankin *et al.* 2008). Hand-held dipping hydrophones were used collect SIO (Campbell *et al.* 2011) and opportunistic CRC recordings. Hydrophone sensitivity and band pass filter specifications varied by recording; details are provided in cited reports. Sampling rates varied greatly across the archived recordings (Table 1) and this was taken into account during analysis by adjusting the Fast Fourier Transformation (FFT) to maximize the resolution possible for spectrograms displayed within PAMGUARD.

Whistles used to train the classifier were taken from visually confirmed single-species encounters that were at least 3 nmi from visual or acoustic detections of other whistling species. The WMD was used to automatically detect and extract whistle contours from recordings displayed through spectrograms (Appendix). Extracted contours from the spectrograms were automatically sent to the ROCCA (Oswald *et al*. 2013) module in PAMGUARD for feature measurement and extraction (Appendix and Barkley *et al*. 2011).

Fifty-three whistle contour measurements (Barkley *et al*. 2011) were automatically extracted from whistles contours produced by the following species: striped dolphin (*Stenella coeruleoalba*), long-beaked common dolphin (*Delphinus capensis*), short-beaked common dolphin (*Delphinus delphis*), bottlenose dolphin (*Tursiops truncatus*), Risso's dolphin (*Grampus griseus*), killer whale (*Orcinus orca*), and Baird's beaked whale (*Berardius bairdii*) (Table 3).

The California Current classification algorithm was developed using random forest statistical classification methods. A random forest is a collection of decision trees grown using binary partitioning of the data. Each binary partition of the data is based on the value of one feature (or in this case, a whistle variable; Breiman 2001). The goal for each split is to divide the data into two nodes, each containing whistles from the smallest number of species possible. Randomness is introduced into the tree-growing process by examining a random subsample of all of the features at each node. The feature that produces the most homogeneous split is chosen at each partition. When whistle features are run through a random forest, each of the trees in the forest produces a species classification. Each tree can be considered 1 '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 number of tree classifications for the predicted species was also used as a measure of the certainty of the classification. It was assumed that if a greater percentage of trees classified the whistle as a particular species, then that classification had a higher degree of certainty. Based on this assumption, a 'strong whistle threshold' was defined. If the percentage of trees that classified the whistle as a particular species was greater than this strong whistle threshold, the whistle was considered strongly classified, or simply 'strong' (Oswald et al. 2011). If the percentage of trees that classified the whistle as a particular species did not exceed the strong whistle threshold, then the classification was considered unreliable and the whistle was labeled as 'ambiguous.' Higher strong whistle thresholds generally result in higher correct classification scores, but also result in more whistles being labeled as ambiguous. If all of the whistles within a single encounter were labeled as ambiguous, then that encounter was also classified as ambiguous and could not be classified. The strong whistle threshold was chosen to maximize correct classification scores while minimizing the number of encounters labeled as ambiguous.

To test the random forest model, the training dataset was randomly subsampled so that it contained an equal number of contours per species. This avoided any one species dominating the data and skewing the results. A two-fold cross-validation method was employed, where the subsampled dataset was randomly divided in two, with whistles from the same encounter kept together in the same dataset. One dataset was used to train the model, while the other was used to test the model. The datasets were then swapped so that each was used both as training and testing sets, and a confusion matrix was created by combining the results of the two iterations.

RESULTS

A total of 109 independent acoustic encounters (schools) were included in the training dataset and included the following species: striped dolphin, long-beaked common dolphin, short-beaked common dolphin, bottlenose dolphin, Risso's dolphin, killer whale, and Baird's beaked whale (Table 3). This dataset was subsampled so that there was a total of 364 whistles per species. Sample size was determined by the species with the lowest number of whistles (striped dolphins).

Results of the two-fold cross-validation test showed that correct classification results for all species were greater than the 14% (1/7) expected by chance (Tables 4a & 4b). Correct classification results for individual whistles ranged from a low of 38% for bottlenose dolphin and long-beak common dolphin to a high of 75% for killer whale (Table 4a). Correct classification results for group encounters ranged from a low of 55% for short-beaked common dolphins to a high of 100% for Baird's beaked whale (Table 4b).

DISCUSSION

The California Current ROCCA classifier includes all whistling species with a known distribution in our study area at the time of development. Classification results for the test dataset were well above that expected by chance, even for difficult species such as short-beaked common dolphins. Several species that had not been detected in the California Current in over a decade at the time of development of the classifier were not included in the training dataset. These species included false killer whales (*Pseudorca crassidens*), short-finned pilot whales (*Globicephala macrorhynchus*), and rough-toothed dolphin (*Steno bredanensis*). During the summer of 2014, these species, as well as pygmy killer whale (*Feresa attenuata*) were reported in the California Current suggesting that these species should be added to the classifier. Pacific white-sided dolphin (*Lagenorhynchus obliquidens*) was also excluded from the classifier because we have not found evidence of whistles in our recordings of this species. There is some debate over whether Pacific white-sided dolphins produce whistles (Henderson *et al*. 2011). If they are found to be a whistle-producing species, effort should be made to record their whistles and they should be added to the classifier.

Multiple versions of PAMGUARD were used throughout development and testing. It has now become clear that this classifier might not function in some older versions of PAMGUARD and future use of the California Current ROCCA classifier should be done in version 1.13.05 or newer.

ACKNOWLEDGEMENTS

We thank Greg Campbell (Scripps Institute of Oceanography), Marla Holt (Northwest Fisheries Science Center), and Greg Schorr (Cascadia Research Collective) for archive recordings. We also thank Michael Oswald for coding modifications to PAMGUARD. Funding was provided by Advanced Science & Technology Working Group (ASTWG), NOAA.

REFERENCES

Archer, F. I, Henry, A.E., Ballance, L.T. (**2008**). "*Stenella* abundance research line transect and ecosystem (STAR-LITE) 2007 cruise report," NOAA Tech. Memo. NOAA-TM-NMFS-SWFSC-433, 19p

Barkley, Y., Oswald, J. N., Carretta, J. V., Rankin, S., Rudd, A., and Lammers, M. O. (**2011**). "Comparison of real-time and post-cruise acoustic species identification of dolphin whistles using ROCCA (Real-time Odontocete Call Classification Algorithm)," U. S. Dep. Commer., NOAA Tech. Memo. NOAA-TM-NMFS-SWFSC-473, 29p.

Barlow, J., Keating, J. L., and Rankin, S. (**2014**). "Cruise report for passive acoustic monitoring component of the 2012 SOCAL-Behavioral Response Study," U. S. Dep. Commer., NOAA Admin. Report. LJ-14-02, 13p.

Breiman, L. (**2001**). “Random forests,” Machine Learn. **45**, 5-32.

Campbell, G. S., Weller, D. W., Cummins, A. J., and Hildebrand, J. A. (**2011**). “SIO small boat based marine mammal surveys in Southern California: Report of results for August 2010 – July 2011,” MPL Tech. Report, MPL-TM-532, 22p.

Gillespie, D., Gordon, J., McHugh, R., McLaren, D., Mellinger, D. K., Redmond, P., Thode, A., Trinder, P., and Deng, X.-Y. (**2008**). “PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localisation of cetaceans,” Proc. Inst. Acoustics.

Hanson, M. B., Noren, D. P., Norris, T. F., Emmons, C. K., Guy, T. G., and Zamon, J. E. (**2008a**). “Pacific ocean killer whale and other cetaceans distribution survey, March 2006 (PODs 2006) conducted aboard the NOAA ship McArthur II,” U. S. Dep. Commer., NOAA Tech. Memo, 26p.

Hanson, M. B., Noren, D. P., Norris, T. F., Emmons, C. K., Holt, M. M., Guy, T. G., and Zamon, J. E. (**2008b**). “Pacific ocean killer whale and other cetaceans distribution survey, May 2007 (PODs 2007) conducted aboard the NOAA ship McArthur II,” U. S. Dep. Commer., NOAA Tech. Memo, 23p.

Henderson, E. E., Hildebrand, J. A., and Smith, M. H. (**2011**). “Classification of behavior using vocalizations of Pacific white-sided dolphins (*Lagenorhynchus obliquidens*),” J. Acoust. Soc. Am. **130**, 557-567.

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.

Oswald, J.N., J.V. Carretta, M. Oswald, S. Rankin and W.W.L. Au. (**2011**). “Seeing the species through the trees: Using random forest classification trees to identify species-specific whistle types,” J. Acoust. Soc. Am. **129**, 2639.

Rankin, S., Barlow, J., Oswald, J., and Ballance, L. (**2008**). “Acoustic studies of marine mammals during seven years of combined visual and acoustic line-transect surveys for cetaceans in the eastern and central Pacific Ocean,” U. S. Dep. Commer., NOAA Tech. Memo. NOAA-TM-NMFS-SWFSC-429, 58p.

Table 1. Cruise information, institutions providing recordings, and recording sampling rates. Institutions that contributed recordings: Southwest Fisheries Science Center (SWFSC), Northwest Fisheries Science Center (NWFSC), Cascadia Research Collective (CRC), and Scripps Institution of Oceanography (SIO).

Cruise	**Full Name**	**Institution**	**Sampling Rate (kHz)**	**Literature**
BRS2012_1	Southern California Behavioral Response Study 2012: Leg 1	SWFSC	192	Barlow *et al.* 2014
BRS2012_2	Southern California Behavioral Response Study 2012: Leg 2	SWFSC	192	Barlow *et al.* 2014
CRC_13	Opportunistic Recordings	CRC	192	
CSCAPE_05	Collaborative Survey of Cetacean Abundance and the Pelagic Ecosystem 2005	SWFSC	44.1 & 48	Rankin *et al.* 2008
ORCAWALE_01	Oregon, California, Washington Marine Mammal Survey 2001	SWFSC	200 & 250	Rankin *et al.* 2008
ORCAWALE_08	Oregon, California, Washington Marine Mammal Survey 2008	SWFSC	480	Rankin *et al.* 2008
PODS_06	Pacific Orca Distribution Survey 2006	NWFSC	96	Hanson *et al*. 2008a
PODS_07	Pacific Orca Distribution Survey 2007	NWFSC	96	Hanson *et al*. 2008b
SIO-SB	SIO-Small Boat marine mammal surveys 2009-2011	SIO	192	Campbell *et al.* 2011
STAR_00	Stenella Abundance Research 2000	SWFSC	100 & 250	Rankin *et al.* 2008
STAR_03	Stenella Abundance Research 2003	SWFSC	80 & 200	Rankin *et al.* 2008
STAR_06	Stenella Abundance Research 2006	SWFSC	96	Rankin *et al.* 2008
STAR-LITE	Stenella Abundance Research – Line Transect and Ecosystem Survey 2007	SWFSC	96	Archer *et al*. 2008

Table 2. Summary information related to archived recordings to train and test the whistle classifier. Start and end times are of the recording filenames. (See Table 3 for species codes)

Cruise	Acoustic Detection #	Date	Start Time	End Time	Species	Latitude	Longitude
STAR_00	1	7/30/2000	19:40:00	N/A	13	26.3387	-120.403
STAR_00	9	8/6/2000	18:37:00	N/A	13	17.0628	-120.1563
STAR_00	23	8/10/2000	07:20:00	N/A	13	9.7638	-126.0828
STAR_03	313	12/6/2003	14:29:00	N/A	13	23.8358	-119.4755
STAR_03	316	12/7/2003	11:30:00	N/A	13	25.9445	-119.1615
STAR_03	s788	12/7/2003	11:38:00	N/A	13	25.9445	-119.1615
STAR_06	738	12/2/2006	08:00:00	8:30:00	13	18.1258	-118.8060
STAR-LITE	87	9/11/2007	16:33:08	16:38:08	13	14.7204	-104.6883
ORCAWALE_01	s398	11/9/2001	10:33:04	10:34:04	16	33.9310	-119.3499
BRS2012_1	67	8/3/2012	22:12:08	22:21:09	16	33.4092	-118.6225
BRS2012_1	97	8/8/2012	15:37:40	15:51:09	16	33.9071	-119.4751
BRS2012_2	7	10/13/2012	22:50:43	23:07:43	16	33.9051	-119.5550
BRS2012_2	12	10/15/2012	18:37:05	19:09:07	16	34.2070	-119.4729
BRS2012_2	13	10/16/2012	20:31:19	20:37:19	16	34.1520	-119.3834
BRS2012_2	15	10/17/2012	15:35:07	15:36:07	16	34.0803	-119.2455
BRS2012_2	18	10/17/2012	16:25:10	16:36:11	16	34.0265	-119.3101
BRS2012_2	32	10/18/2012	14:52:58	14:53:58	16	33.3543	-118.5729
ORCAWALE_01	s267	10/2/2001	13:18:00	13:20:00	17	33.9625	-118.8610
ORCAWALE_01	s268	10/2/2001	13:36:00	13:39:00	17	33.9915	-118.8934
ORCAWALE_01	s280	10/3/2001	07:50:00	07:53:00	17	33.3338	-117.8954
ORCAWALE_01	s289	10/9/2001	08:24:00	08:44:00	17	32.3753	-118.4829
ORCAWALE_01	s301	10/15/2001	12:05:29	12:21:08	17	37.7066	-129.0359
ORCAWALE_01	s319	10/22/2001	09:31:39	09:34:39	17	40.8336	-129.9198
ORCAWALE_01	s320	10/22/2001	12:01:21	12:03:21	17	40.7300	-129.3247
ORCAWALE_01	s322	10/22/2001	15:34:02	15:36:15	17	40.5634	-128.6353
ORCAWALE_01	s366	11/3/2001	11:35:59	11:37:59	17	40.8865	-128.5622
STAR_03	s792	12/17/2003	15:18:51	15.:41:04	17	26.4508	-119.0715
ORCAWALE_08	200	9/13/2008	16:28:37	16:55:37	17	37.3206	-126.3181
ORCAWALE_08	326	9/25/2008	01:32:22	N/A	17	37.8259	-129.4739
ORCAWALE_08	372	9/29/2008	14:17:20	14:26:18	17	36.9904	-125.5795
ORCAWALE_08	376	9/29/2008	16:14:24	16:32:23	17	37.0794	-125.8763
ORCAWALE_08	428	10/3/2008	23:44:34	00:02:33	17	34.9555	-123.1344
ORCAWALE_08	431	10/4/2008	14:40:01	14:48:59	17	34.7233	-122.2840
ORCAWALE_08	481	10/16/2008	18:35:50	18:44:51	17	32.7478	-120.4462
ORCAWALE_08	474	10/17/2008	00:17:52	00:26:50	17	33.0085	-121.2794
ORCAWALE_08	553	10/21/2008	20:28:39	20:37:43	17	39.0785	-127.4709
ORCAWALE_08	567	10/22/2008	14:53:08	15:02:06	17	39.0692	-128.4891
ORCAWALE_08	648	10/27/2008	16:09:36	16:36:36	17	34.2692	-124.1240
ORCAWALE_08	649	10/27/2008	16:54:38	17:12:37	17	34.2022	-124.1589

Table 2. (Recording Summary Continued)

Cruise	Acoustic Detection #	Date	Start Time	End Time	Species	Latitude	Longitude
ORCAWALE_08	719	10/31/2008	15:31:01	15:49:00	17	33.8355	-120.9907
ORCAWALE_08	752	11/10/2008	23:29:00	23:38:31	17	31.7348	-124.9690
ORCAWALE_08	753	11/11/2008	15:09:55	15:18:56	17	32.2903	-124.8847
ORCAWALE_08	758	11/12/2008	15:13:40	15:40:43	17	33.9278	-125.8630
ORCAWALE_08	760	11/12/2008	19:52:44	20:01:45	17	33.2447	-126.1289
ORCAWALE_08	798	11/16/2008	22:45:26	N/A	17	31.8533	-122.9931
ORCAWALE_08	801	11/17/2008	16:36:57	16:45:55	17	31.6680	-122.1984
ORCAWALE_08	803	11/17/2008	17:30:57	17:39:55	17	31.6370	-122.0490
ORCAWALE_08	805	11/17/2008	18:33:55	18:42:56	17	31.5930	-121.8800
ORCAWALE_08	806	11/17/2008	18:51:57	19:00:55	17	31.5800	-121.7960
ORCAWALE_08	809	11/17/2008	20:39:57	20:48:55	17	31.4701	-121.6951
ORCAWALE_08	810	11/17/2008	21:06:57	N/A	17	31.4119	-121.7211
ORCAWALE_08	816	11/17/2008	22:54:57	N/A	17	31.1610	-121.6840
ORCAWALE_08	857	11/27/2008	00:42:38	01:00:37	17	31.7757	-121.5747
ORCAWALE_08	858	11/27/2008	15:00:00	15:18:02	17	33.1262	-122.8591
ORCAWALE_08	864	11/28/2008	15:23:00	N/A	17	31.5678	-121.6993
ORCAWALE_08	865	11/28/2008	15:58:59	N/A	17	31.5400	-121.5890
ORCAWALE_08	866	11/28/2008	16:08:00	16:34:55	17	31.5070	-121.5230
ORCAWALE_08	867	11/28/2008	16:43:56	17:01:58	17	31.4881	-121.4058
BRS2012_1	3	7/26/2012	14:46:27	15:09:29	17	33.8682	-119.8262
BRS2012_1	42	7/31/2012	16:56:58	17:10:59	17	33.1360	-118.8389
BRS2012_1	50	8/1/2012	16:29:53	16:37:54	17	33.3738	-119.2490
BRS2012_1	69	8/4/2012	15:49:55	16:06:56	17	33.7136	-118.4076
BRS2012_2	23	10/17/2012	23:06:00	23:13:00	17	33.4029	-119.2748
BRS2012_2	24	10/17/2012	23:36:02	23:42:03	17	33.3776	-119.2218
BRS2012_2	44	10/22/2012	15:18:14	15:20:14	17	33.4148	-118.6966
SIO-SB	s4	7/22/2009	10:28:00	10:51:38	18	32.9914	-118.5177
SIO-SB	s6	7/22/2009	11:48:23	11:57:58	18	33.0560	-118.5601
SIO-SB	s3	7/25/2009	09:32:53	N/A	18	33.0606	-118.5512
SIO-SB	s6	7/26/2009	16:57:10	17:11:15	18	32.9843	-118.5170
SIO-SB	s1	6/19/2010	10:18:48	N/A	18	32.9956	-118.5371
SIO-SB	s7	6/20/2010	13:49:30	N/A	18	32.8924	-118.4256
SIO-SB	s2	6/22/2010	12:11:55	N/A	18	33.4413	-118.4403
SIO-SB	s1	1/8/2011	14:01:17	16:05:07	18	33.3139	-118.2897
SIO-SB	s5	1/9/2011	13:09:18	N/A	18	33.4344	-118.4171
SIO-SB	s6	5/4/2011	14:33:33	14:56:11	18	33.3499	-118.4986
BRS2012_1	65	8/3/2012	20:51:54	N/A	18	33.4490	-118.7592
BRS2012_1	86	8/7/2012	13:14:15	13:22:16	18	33.3680	-118.5353
BRS2012_1	93	8/7/2012	22:32:09	22:33:09	18	33.4744	-118.9990
BRS2012_2	9	10/14/2012	21:30:34	21:37:35	18	33.9396	-119.4058

Table 2. (Recording Summary Continued)

Cruise	Acoustic Detection #	Date	Start Time	End Time	Species	Latitude	Longitude
BRS2012_2	10	10/15/2012	15:23:50	15:29:51	18	34.0565	-119.2445
ORCAWALE_01	s34	8/4/2001	19:51:00	N/A	21	37.0710	-126.5711
CSCAPE_05	s1526	11/18/2005	00:10:07	N/A	21	37.6700	-123.0092
BRS2012_1	4	7/26/2012	15:24:30	15:37:31	21	33.8929	-119.8514
BRS2012_1	17	7/27/2012	20:43:00	N/A	21	33.4134	-118.6717
BRS2012_1	24	7/28/2012	18:44:52	18:48:52	21	33.5096	-118.8792
BRS2012_1	36	7/30/2012	15:57:48	16:01:48	21	33.4613	-118.7070
BRS2012_1	49	8/1/2012	15:49:50	15:58:51	21	33.4142	-119.2301
BRS2012_1	56	8/2/2012	13:16:06	13:25:07	21	32.9794	-118.4842
BRS2012_1	57	8/2/2012	18:29:52	18:38:52	21	33.0332	-118.4484
CSCAPE_05	s1449	10/23/2005	20:29:21	20:49:15	37	46.3044	-124.2322
CSCAPE_05	s1451	10/29/2005	17:47:22	18:07:17	37	46.2717	-124.2708
CSCAPE_05	s1451	10/29/2005	21:00:21	21:10:15	37	46.2717	-124.2708
PODS_06	4	3/30/2006	15:30:00	16:30:00	37	46.1650	-124.2862
PODS_07	1	5/5/2007	08:00:00	09:20:00	37	48.6990	-126.0900
PODS_07	3	5/13/2007	16:30:00	17:30:00	37	48.4020	-124.8660
ORCAWALE_08	93	8/25/2008	23:32:17	N/A	37	45.3073	-125.3017
CRC_13	1	1/8/2013	16:44:00	N/A	37	33.1802	-118.6756
CSCAPE_05	s1058	8/15/2005	00:39:57	N/A	63	45.8475	-129.0275
CSCAPE_05	s1112	8/23/2005	17:26:34	17:46:34	63	43.6925	-128.3772
CSCAPE_05	s1112	8/23/2005	19:44:11	20:11:25	63	43.6925	-128.3772

Table 3. Species codes for species included in the California Current classifier and number of whistles detected using the PAMGUARD ‘Whistle and Moan Detector.’

Species	Species Code	Species ID#	# of Encounters	# of Whistles
Striped dolphin (*Stenella coeruleoalba*)	Sc	13	8	594
Long-beaked common dolphin (*Delphinus capensis*)	Dc	16	9	40369
Short-beaked common dolphin (*Delphinus delphis*)	Dd	17	57	155164
Bottlenose dolphin (*Tursiops truncatus*)	Tt	18	15	7946
Risso's dolphin (*Grampus griseus*)	Gg	21	9	2111
Killer whale (*Orcinus orca*)	Oo	37	8	4122
Baird’s beaked whale (*Berardius bairdii*)	Bb	63	3	3410

Table 4. Confusion matrices for the California Current classifier using a strong whistle threshold of 35%. The diagonal (values in bold) is the percent of individuals (4a) or encounters (4b) that were correctly classified. Overall, 55% of whistles and 70% of encounters were correctly classified.

a.

	Predicted Species						
Correct Species	**Sc**	**Dc**	**Dd**	**Tt**	**Gg**	**Oo**	**Bb**
Sc	**43**	8	22	8	9	9	1
Dc	5	**38**	9	24	25	0	0
Dd	15	14	**53**	12	6	0	0
Tt	2	34	6	**38**	19	1	0
Gg	3	19	2	14	**61**	1	0
Oo	11	0	3	0	0	**75**	11
Bb	1	0	0	0	0	24	**74**

b.

	% Correct Classification	
Species	**Individual Whistles**	**Encounters**
Sc	43	63
Dc	38	78
Dd	53	55
Tt	38	58
Gg	61	56
Oo	75	83
Bb	74	100
Overall	**55**	**70**

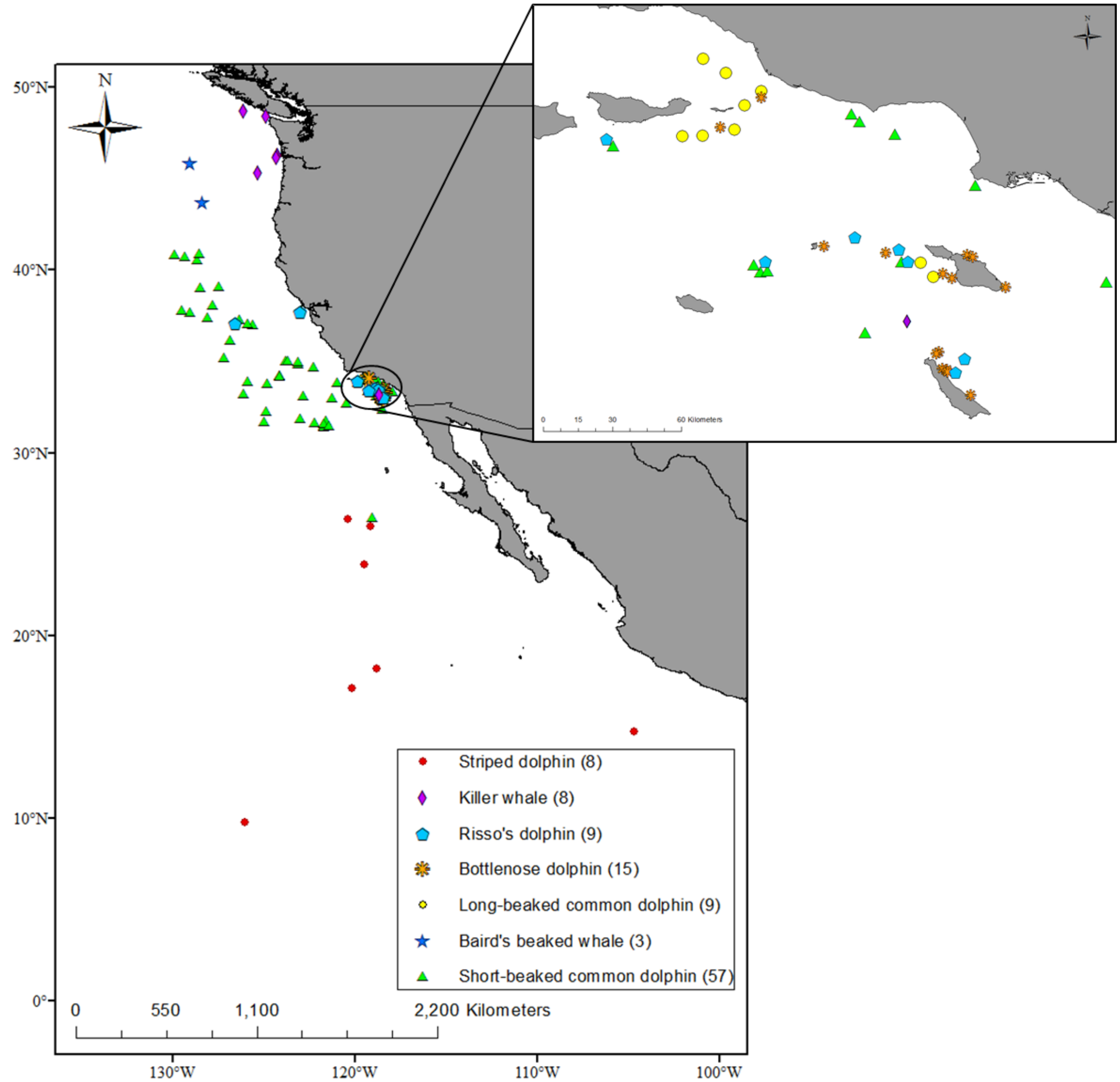

Figure 1. Map of acoustic detections within the California Current used to test and train the whistle classifier. Number of schools for each species is given in parenthesis. (Inset map expands acoustic detections from the Channel Islands, CA)

APPENDIX

PAMGUARD Setup (for version 1.12.05)

- **General Setup:** Create a top level folder to hold all recordings for whistle training and PAMGUARD output files. (ex: Whistle_Detectors)
- Within the top folder (Whistle_Detectors) create a folder for each species (ex: Bottlenose) and copy training recordings to those folders.
- Create a Binaries folder at the same folder level at Whistle_Detectors
- Modules required in PAMGUARD: Database, Binary files, FFT, Spectrogram, Whistle and Moan Detector, Whistle Classifier, and Rocca
- Spectrogram: R-click on the spectrogram and select 'Whistle and moan detector connected regions.' (This allows you to visually see the detector overlaid on the screen in real time).

- **Training Data:** Detection – Whistle Classifier – Settings – Collect Training Data. Leave all settings at default except Classifier Training Parameters, set to the Whistle_Detectors folder for storage. Also check use folder names for species.
- Selecting recordings to classify: Detection – Sound Acquisition – DAQ Settings – Source Type (Audio file folder or multiple files) – Select sound files from species folders.
- FFT: Detection – FFT (Spectrogram) Engine Settings – Raw Data Source for FFT (Raw input data from Sound Acquisition) – Select only the channel training data is being collected from.
- Detection – Whistle and Moan Detector – No grouping and all default settings
- Rocca: Select 'Use Whistle & Moan Source.' Also set Classifier (RF8sp53att.model), Output, and Filename Template (%f-%n_D%X).

Press play