Development of a Tool for Acoustic Identification of Killer Whale Communities in the Pacific Ocean

Prepared By:

Cetos Research Organization
181 Beachview Ave.
Pacifica, CA 94044
(510) 393-5906

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

October 14, 2015

Suggested Citation:

Oswald, J.N., Yack, T.M., Dunleavy, K.D., and Zoidis, A.M. 2015. *Development of a tool for acoustic identification of killer whale communities in the Pacific Ocean*. Prepared for SeaWorld Busch Gardens 2014 Conservation Fund Grant award, Pacifica, CA.

Photo: Killer whale photograph taken under NMFS scientific research permit 16163.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Development of a tool for acoustic identification of killer whale communities in the Pacific Ocean

1. ABSTRACT

The northeastern Pacific Ocean is home to three ecotypes of killer whales: resident, transient, and offshore. Resident killer whale ecotypes consist of northern and southern communities. These communities are further organized into pods, which are defined as groups of matrilines that are regularly observed together. Different killer whale ecotypes, communities and pods exhibit differences in their behaviors, prey preferences, genetics, and acoustic call repertoires, which include pulsed calls, whistles, and echolocation clicks. Currently, biologists who are familiar with the calls produced by each group can use pulsed calls to acoustically classify killer whales to ecotype, community, or pod. Whistles and clicks may also carry information that could be used to identify ecotypes, pods or communities, however this has not yet been investigated. In the work presented here, we used visually validated recordings from towed hydrophone arrays made in the presence of four killer whale communities (northern and southern residents, offshores, and west coast transients) that were collected during the Pacific Orca Distribution Surveys (PODS) conducted by National Oceanic and Atmospheric Administration (NOAA) Fisheries Northwest Fisheries Science Center (NWFSC). These surveys occurred along the coasts of Oregon, Washington, and northern California during the winters of 2006-2009, 2012, 2013, and 2015. Recordings of killer whale encounters were divided into 'sub-encounters' based on the relative abundance of whistles, pulsed calls and echolocation clicks in the recordings. Pulsed calls, whistles and echolocation clicks were measured from each sub-encounter in order to compare characteristics among communities. PAMGuard's automated click detector module was used to automatically detect echolocation clicks. Newly developed tools within the Real-time Odontocete Call Classification Algorithm (ROCCA) module in PAMGuard were used to automatically measure five click parameters: duration, center frequency, peak frequency, sweep rate, and number of zero crossings. ROCCA was also used to measure up to 50 variables from whistles and pulsed calls. These measurements include: frequency (for whistles only), slope and duration variables, as well as other variables describing the shape of time-frequency contours. These measurements were used to train three random forest classifiers to classify sounds to community: one for whistles, one for pulsed calls and one for echolocation clicks. Classifier performance was evaluated using four-fold cross validation. Overall correct classification scores were 70%, 49% and 42% for whistles, pulsed calls and echolocation clicks, respectively. Random forest votes for each signal type were then summed to produce a final classification for each killer whale encounter. Overall, 80% of sub-encounters were correctly classified using this method. Correct classification scores were significantly greater than expected by chance (Fisher's exact test, $\alpha = 0.05$) for every community except transient killer whales (likely due to small sample size). These new classifiers will allow researchers to more effectively and efficiently identify call types to the ecotype or community level and will allow researchers to better characterize the occurrence, range and distribution of killer whales in the northeastern Pacific Ocean.

2. INTRODUCTION

Three distinct ecotypes of killer whales (residents, offshores, and transients) are recognized in the northeastern Pacific Ocean (Ford 1989). Within the resident and transient ecotypes, multiple communities have been identified: northern resident, southern resident, southern Alaska resident, western Alaska north Pacific resident, west coast transient, and AT1 transient. Southern resident killer whales are listed as endangered under the Endangered Species Act and as a "strategic" stock under the Marine Mammal Protection Act (Carretta et al. 2010). It is crucial to understand the movements and occurrence patterns of these animals in order to create and implement effective conservation strategies. During spring, summer and fall southern resident killer whales spend a significant amount of time in the inland waterways of the Salish Sea (Puget Sound, WA and Georgia Basin, BC). They also occur in coastal waters off Washington, Oregon, and British Columbia and have been encountered as far south as central California (Krahn et al. 2004). In these areas their range overlaps with northern resident, transient and offshore killer whale ecotypes. Resident killer whale communities can be further organized into pods, which consist of groups of matrilines that are regularly observed together (Ford and Fisher 1983, Ford 1989).

Killer whale ecotypes, communities and pods exhibit differences in their behaviors, prey preferences, genetics, and acoustic repertoires (Baird 2000). It is well known that resident killer whales have pod-specific vocal repertoires (also known as dialects) consisting of 'discrete' pulsed calls and whistles (Ford 1989, 1991). These distinctive vocal repertoires make it possible to use passive acoustic methods such as dipping hydrophones, towed hydrophone arrays and autonomous recorders to monitor the occurrence, distribution patterns, and habitat use of killer whales. The use of passive acoustic methods has greatly increased knowledge of the distribution and range of endangered southern resident killer whales. For example, during the 28 years between 1976 and 2004, there were only 11 documented visual sightings of southern resident killer whales in U.S. coastal waters (Krahn et al. 2004). In contrast, Hanson et al. (2013) documented 131 acoustic detections of southern resident killer whales during a five year period (2006-2011), using autonomous acoustic recorders.

In order to obtain information on killer whale distribution, behavior, and habitat use from passive acoustic data, it is necessary to identify vocalizations in recordings and classify them to ecotype, community and, ideally, pod. Communities and pods of resident killer whales can be identified by experienced biologists based on their pulsed calls. However, these methods are time, cost, and personnel intensive and require a high level of training and experience. As passive acoustic monitoring technologies continue to produce greater volumes of data, reducing the time, effort, and expertise necessary for the classification of killer whale calls is becoming essential for efficient analysis of these extensive datasets.

The variability in sounds produced by killer whales and most other delphinids makes them challenging to automatically detect and classify, and much research has been focused on these tasks in recent years. Early work on delphinid whistle classifiers focused on time-frequency characteristics measured from spectrograms and used classification algorithms such as discriminant-function analysis (DFA) and classification tree analysis (e.g., Steiner 1981, Fristrup and Watkins 1993, Wang et al. 1995, Matthews et al. 1999, Rendell et al. 1999, Oswald et al. 2003). More recently, other classification algorithms such as Gaussian mixture models (Roch et al. 2007), hidden Markov models (Brown and Smaragdis 2009) and random forests (Oswald et al. 2013) have been used, each with varying degrees of success.

Random forest classifiers have shown promise when used to identify dolphin whistles to species in the tropical Pacific and northwest Atlantic oceans (Oswald et al. 2013, Oswald 2013). For example, the Real-time Call Classification Algorithm (ROCCA, a user-friendly software module in the freely available acoustic analysis software platform, PAMGuard [Gillespie et al. 2008, www.pamguard.org]) uses random forest analysis to classify delphinid species based on spectrographic measurements extracted from whistle contours. ROCCA currently contains classifiers for eight species in the tropical Pacific Ocean and five species in the northwest Atlantic Ocean, with overall correct classification scores of 60% (Oswald et al. 2013) and 86% (Oswald 2013), respectively. Earlier versions of ROCCA's tropical Pacific classifier used DFA and classification tree analysis, but only 46% of schools were correctly classified using those methods (Oswald et al. 2007). The use of random forest analysis has resulted in significant increases in classification success (Oswald et al. 2013). In this study, we have developed a random forest based classifier for northeast Pacific killer whale communities that incorporates measures from whistles, clicks and pulsed calls.

3. METHODS

3.1 Data Collection

Data were collected during the Pacific Orca Distribution Surveys (PODS) conducted by National Oceanic and Atmospheric Administration (NOAA) Fisheries Northwest Fisheries Science Center (NWFSC). These surveys occurred along coastal Oregon, Washington, and northern California in the winters of 2006-2009, 2012, 2013, and 2015 (**Table 1, Figure 1**). During these surveys, a two-element hydrophone array was towed 200–300 m behind a NOAA research vessel. **Table 1** provides information about the hydrophone array used on each survey. Recordings were made to a computer hard drive using a 96 kHz sample rate during every survey. This gave an effective analysis bandwidth of 48 kHz. Although various hydrophone configurations were used in the arrays during different survey years, there were consistently two mid-frequency (100 Hz – 40 kHz) hydrophones recorded from each survey and these data were selected for use in this analysis. Visual observers scanned the area with high-powered and hand-held binoculars and provided information regarding the species, ecotype, community and pod being recorded acoustically.

Table 1. PODS Survey and recording information.

Survey	Start Date	End Date	Hydrophone Frequency Range
PODS 2006	03/13/2006	03/30/2006	100 Hz – 40 kHz
PODS 2007	05/03/2007	05/15/2007	100 Hz – 40 kHz
PODS 2008	03/17/2008	03/26/2008	100 Hz – 40 kHz
PODS 2009	03/23/2009	04/09/2009	100 Hz – 40 kHz
PODS 2012	02/16/2012	03/05/2012	100 Hz – 40 kHz
PODS 2013	03/02/2013	03/09/2013	1 kHz – 50k Hz
PODS 2015	02/11/2015	03/02/2015	1 kHz – 50k Hz

Figure 1. Map of the PODS study area. The area surveyed is indicated by red hatch marks.

3.2 Data Analysis

3.2.1 Vocal behavioral state identification

Based on previous work (ex. Ford 1989, Thomsen et al. 2002, Miller 2006), it was assumed that different vocal activity levels were associated with different behavioral states, and that behavioral state could affect characteristics of vocalizations. To ensure that vocalizations from a range of behavioral states were included in the classifier training dataset, all killer whale acoustic encounters were divided into 'sub-encounters' based on vocal activity levels. Acoustic activity levels were assessed and logged using Long-Term Spectral Averages (LTSAs) produced using Triton software (Wiggins 2007). Sub-encounters were categorized into periods of high, medium, and low levels of vocal activity. Periods categorized as low vocal activity had less than one vocalization and/or short click bouts during a ten second period (**Figure 2**). Periods categorized as medium vocal activity had between two and three vocalizations and/or short click bouts during a ten second period (**Figure 3**). Periods of high vocal activity had almost continuous vocalizations or echolocation during a ten second period (**Figure 4**).

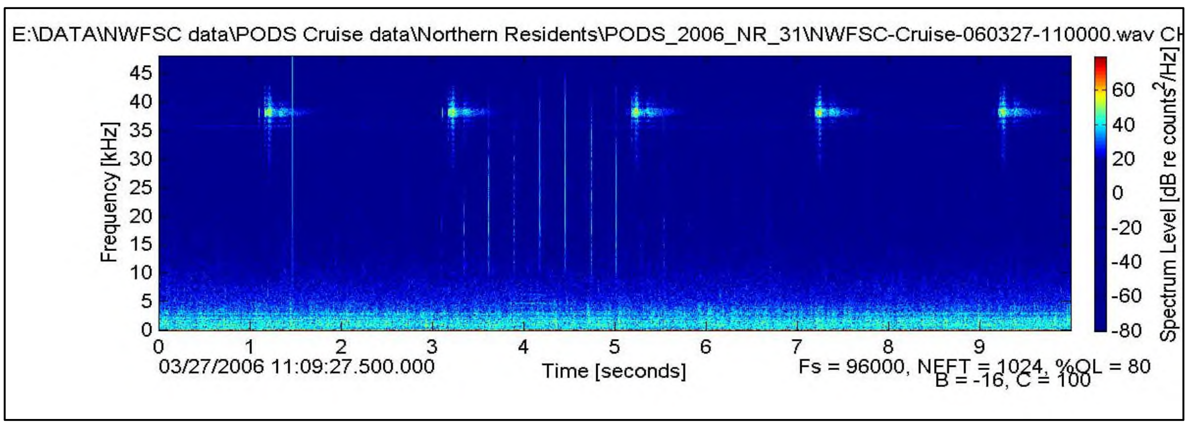

Figure 2. Example of period of low vocal activity

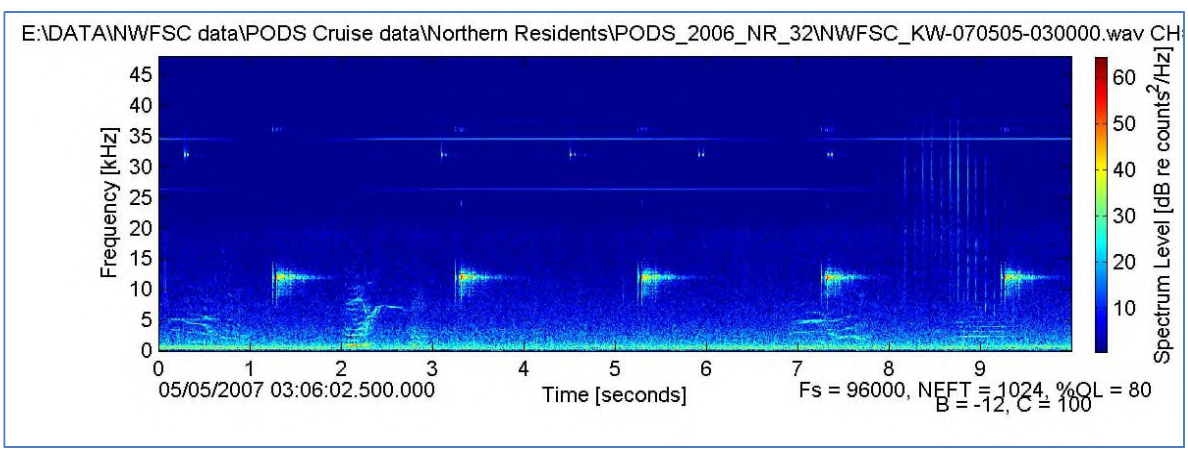

Figure 3. Example of period of medium vocal activity

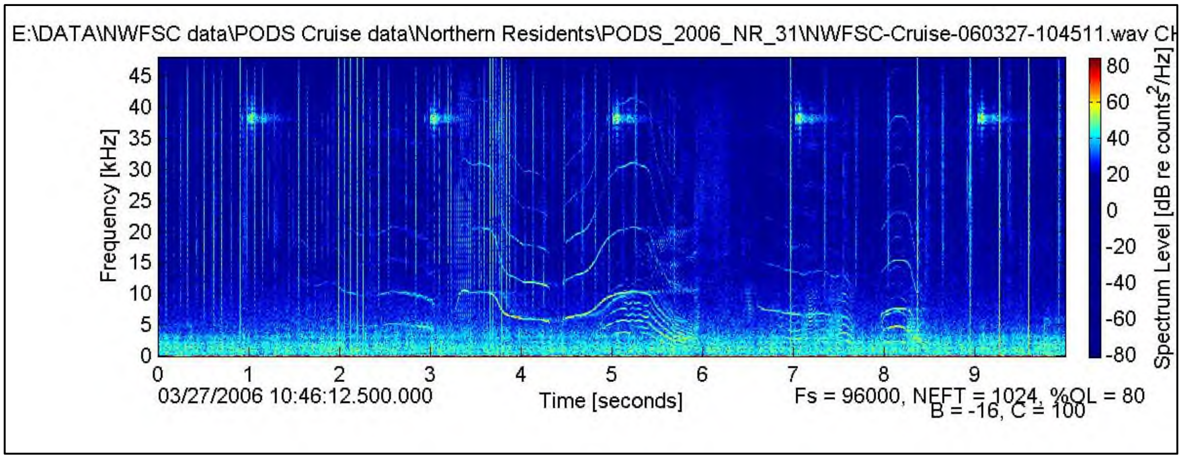

Figure 4. Example of period of high vocal activity.

3.2.2 Measuring pulsed calls and whistles

Pulsed calls and whistles were extracted and measured from encounters using ROCCA. In order to ensure that pulsed calls and whistles were measured from all behavioral states, an equal number of pulsed calls and whistles were randomly selected from each sub-encounter. Whistles and pulsed calls were traced manually on the ROCCA spectrographic display using a mouse. ROCCA then automatically extracted time-frequency contours based on the manual traces and measured 50 variables from each (see **Appendix A** for a complete list and description of variables measured). In the case of pulsed calls with multiple frequency bands, the contour with the highest signal to noise ratio (SNR) was chosen for measurement (**Figure 5**). Because of the inherent bias in frequency measurements resulting from this approach, only the pulsed call shape variables (e.g., number of inflection points, slopes, percent of the contour that was upswept, downswept and flat) were used for classifier development.

a)

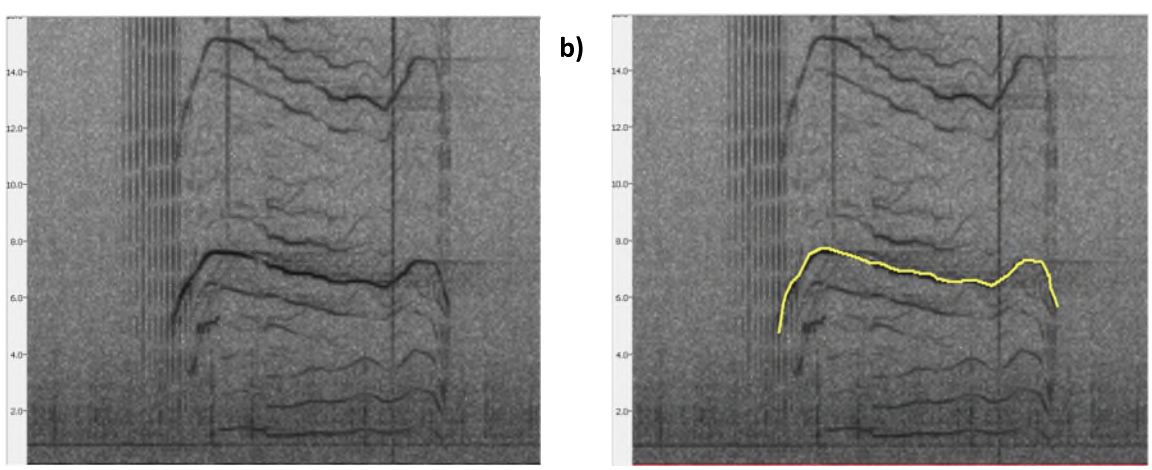

Figure 5. Example of a pulsed call (a), with the highest amplitude time-frequency contour traced and extracted using ROCCA (b).

3.2.3 Echolocation click measurement

PAMGuard's automated click detector module was used to detect echolocation clicks, and newly developed PAMGuard ROCCA tools were used to automatically measure five click parameters: duration, center frequency, peak frequency, sweep rate, and number of zero crossings. All clicks with an SNR less than 5 dB or greater than 25 dB were pruned from the data set in order to remove faint and clipped clicks.

3.2.4 Classification analysis

Random forest classifiers were trained to classify encounters to community (offshore, transient, northern resident, southern resident). Sample sizes for pulsed calls and whistles were not large enough to classify to pod level. A random forest is a predictive model containing 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 (in this case, a whistle, pulsed call or echolocation click feature; Breiman 2001). The goal for each binary partition is to divide the data into two nodes, each as homogeneous as possible (i.e., containing calls from the smallest number of communities 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 nodes is chosen at each partition. When calls are analyzed using a random forest, each of the trees in the forest produces a community classification. Classifications are then tallied over all trees in the forest and the call is classified as the community that received the highest proportion of classifications. Three different random forests were trained; one that classified whistles, one that classified pulsed calls and one that classified echolocation clicks.

To create the random forest classifiers, the datasets for each call type were first randomly sub-sampled so that there were equal sample sizes for each community. This prevented measurements from any one community from dominating the data and skewing the classification results. Four-fold cross-validation was used to test the performance of the classifiers. To accomplish this, each test dataset was randomly divided into four subsets of data. Three quarters of the data were then used to train the classifier, while the fourth quarter was used to test the classifier. The datasets were then swapped so that each subset was used for both training and testing. This procedure was repeated 100 times. Classification success was evaluated by examining the average percent of encounters that were correctly classified for each community and comparing that to the classification score that would be expected by chance alone (e.g., 25% for 4 communities).

In addition to classifying individual calls, acoustic sub-encounters were classified based on the average proportion of tree classifications for each community, calculated over all of the calls that were analyzed for that encounter. Each encounter was classified based on whistles, echolocation clicks and pulsed calls. This information was then combined to produce a final classification for each sub-encounter. To produce a final classification, the average proportion of tree classifications for each community was summed over all three call types. The sub-encounter was then classified as the community that had the largest sum of tree classifications (see **Table 2** for an example).

Table 2. Example of overall classifications for sub-encounters. The average proportion of tree classifications summed over all calls in the sub-encounter are given for each call type (whistles, echolocation clicks and pulsed calls). These proportions are then summed over all three call types and given in the 'summed classifications' columns. The sub-encounter is classified as the community with the highest sum of tree classifications. OF = offshore community, TR = transient community, NR = northern resident community, SR = southern resident community.

Sub-encounter	average proportion of tree votes whistles				average proportion of tree votes echolocation clicks				average proportion of tree votes pulsed calls				Sum OF	Sum TR	Sum NR	Sum SR	Classified as
	OF	TR	NR	SF	OF	TR	NR	SR	OF	TR	NR	SR					
OF54A	**0.90**	0.00	0.03	0.08	**0.35**	0.12	0.19	0.34	**0.54**	0.11	0.10	0.24	1.79	0.23	0.32	0.66	Offshore
NR31G	0.02	0.00	0.49	0.49	0.16	0.37	0.30	0.17	0.17	0.22	0.26	0.35	0.36	0.60	1.05	1.00	Northern resident

4. RESULTS

Whistles, pulsed calls, and echolocation clicks were measured from visually confirmed encounters of southern resident, northern residents, offshore and transient killer whales (**Table 3**).

Table 3: Number of encounters (with the number of sub-encounters in parentheses) and number of individual calls measured for whistles, pulsed calls (PC) and echolocation clicks. Number of echolocation clicks refers to the number of clicks after low and high SNR clicks were pruned from the dataset.

Community	Number of Whistle Encounters Measured	Total Number of Whistles Measured	Number of PC Encounters Measured	Total Number of PC Measured	Number of Click Encounters Measured	Total Number of Clicks Measured
Southern Resident	10 (45)	150	20 (81)	150	26 (91)	22483
Northern Resident	4 (17)	150	4 (20)	166	4 (18)	5080
Offshore	1 (4)	148	1 (4)	132	1 (5)	9729
Transient	2 (5)	44	2 (5)	101	2 (7)	254
Total	17 (71)	492	27 (110)	549	33 (121)	37546

4.1 Pulsed Calls

Eight variables were examined for pulsed calls: mean slope, mean positive slope, mean negative slope, number of inflection points, duration, and percent of the contour that was upswept, downswept and flat (**Figure 6**). Non-parametric Kruskal-Wallis and post-hoc Dunn's tests showed significant differences in variables among all communities (**Table 4**). Pulsed calls produced by offshore killer whales showed the greatest number of significant differences when compared to all other communities.

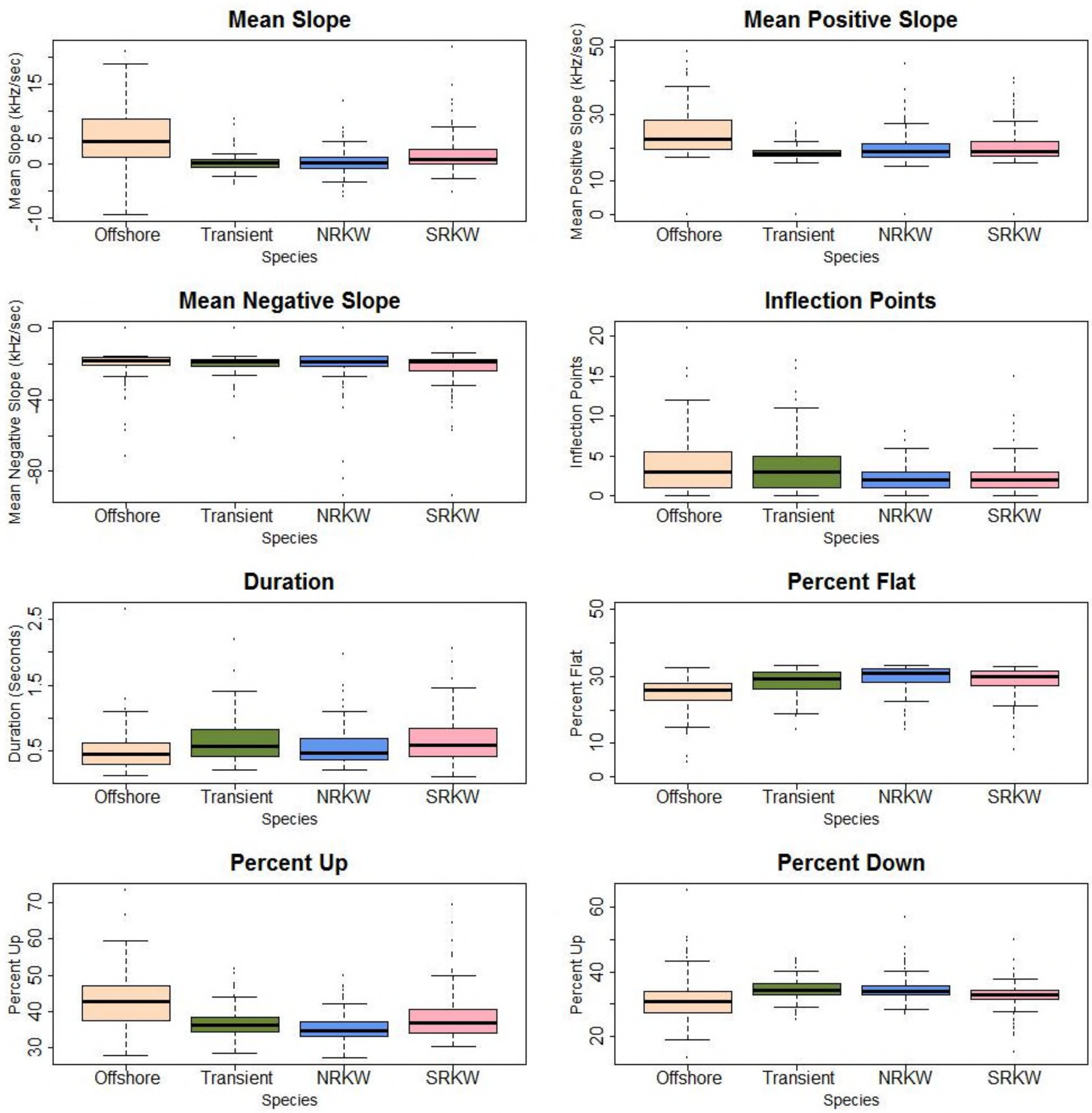

Figure 6. Box plots displaying the median, first and third quartiles for each pulsed call parameter with community along the x-axis and each parameter value along the y-axis (NRKW = northern resident killer whales, SRKW = southern resident killer whale). Sample sizes are 132, 101, 166 and 150 pulsed calls for offshore, transient, NRKW and SRKW, respectively.

Table 4. Matrix displaying pulsed call parameters that were significantly different when compared between communities (Kruskal-Wallis and Dunn's tests with Bonferroni adjustment, α<0.05) (NRKW = northern resident killer whales, SRKW = southern resident killer whale)

	Offshore	**Transient**	**NRKW**
Transient	Duration Mean Slope Mean Positive Slope Percent Flat Percent Up Percent Down		
NRKW	Mean Slope Mean Positive Slope Percent Flat Percent Up Percent Down Inflection Points	Duration Inflection Points Percent Flat Percent Up	
SRKW	Duration Mean Slope Mean Positive Slope Inflection Points Percent Flat Percent Up Percent Down	Mean Slope Mean Positive Slope Percent Down Inflection Points	Duration Mean Slope Percent Flat Percent Up Percent Down

4.2 Whistles

Nine whistle variables were statistically compared among communities: beginning frequency, end frequency, maximum frequency, minimum frequency, mean frequency, duration, mean slope, mean positive slope and mean negative slope (**Figures 7 and 8**). Non-parametric Kruskal-Wallis and post-hoc Dunn's tests showed significant differences in whistle variables among all communities (**Table 5**). The whistles produced by offshore and transient killer whales showed the greatest number of significant differences when compared to all other communities.

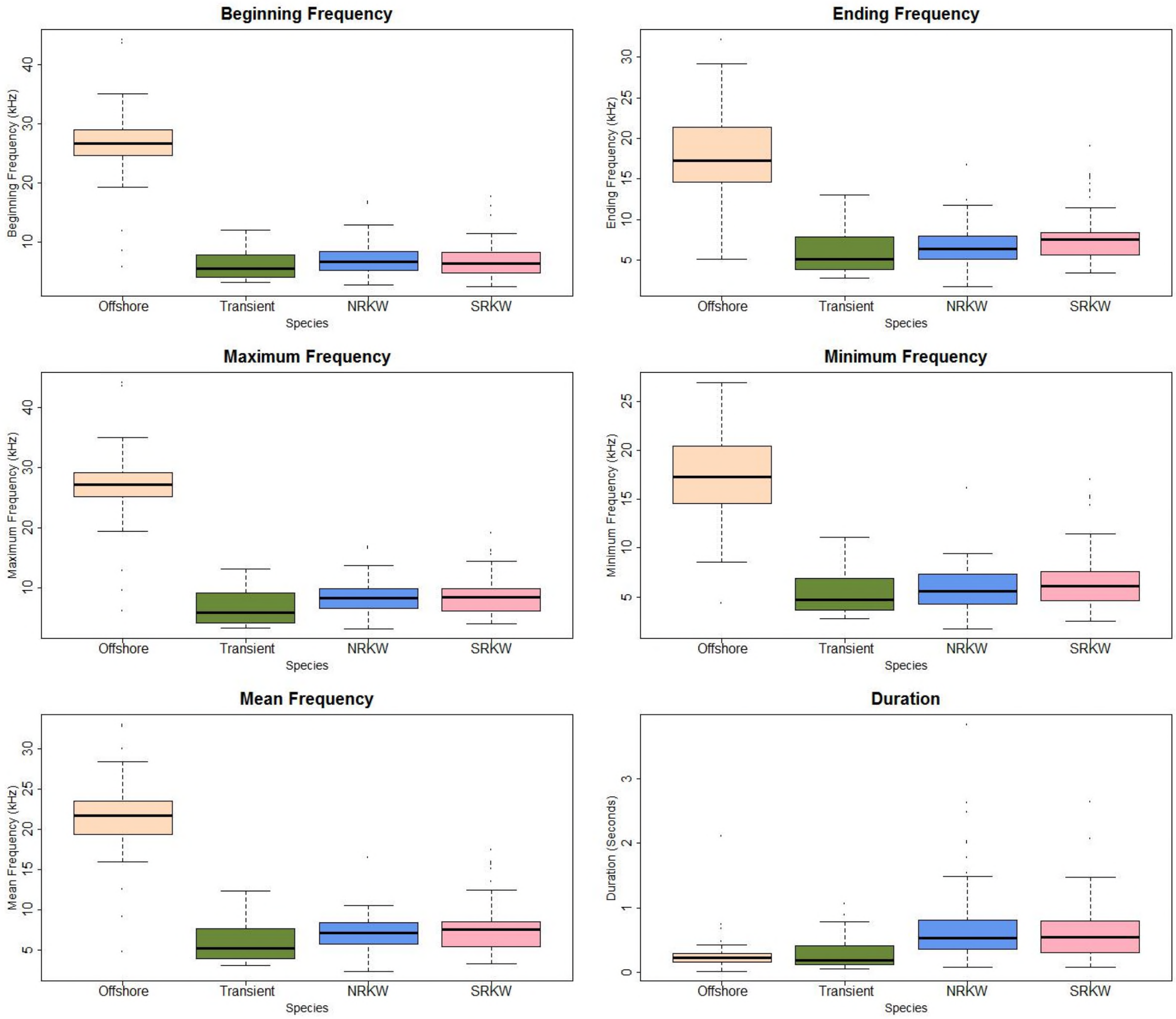

Figure 7. Box plots displaying the median, first and third quartiles for the whistle frequency parameters with community along the x-axis and parameters along the y-axis (NRKW = northern resident killer whales, SRKW = southern resident killer whale). Sample sizes are 148, 44, 150 and 150 whistles for offshore, transient, NRKW and SRKW, respectively.

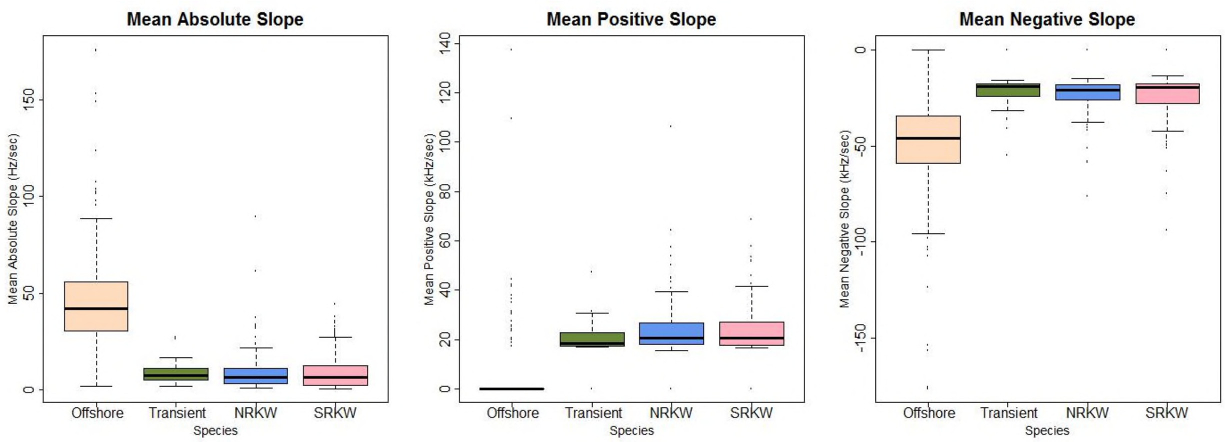

Figure 8. Box plots displaying the median, first and third quartiles for the whistle slope parameters with community along the x-axis and parameters along the y-axis (NRKW = northern resident killer whales, SRKW = southern resident killer whale). Sample sizes are 148, 44, 150 and 150 whistles for offshore, transient, NRKW and SRKW, respectively.

Table 5. Matrix displaying whistle parameters that were significantly different between communities (Kruskal-Wallis and Dunn's tests with Bonferroni adjustment, α<0.05) (NRKW = northern resident killer whales, SRKW = southern resident killer whale).

	Offshore	**Transient**	**NRKW**
Transient	Mean Frequency Mean Slope Mean Positive Slope Mean Negative Slope Beginning Frequency Ending Frequency Maximum Frequency Minimum Frequency		
NRKW	All Parameters Significant	Mean Frequency Mean Positive Slope Maximum Frequency Duration	
SRKW	All Parameters Significant	Mean Frequency Ending Frequency Maximum Frequency Duration	Mean Slope

4.3 Echolocation Clicks

Five echolocation click variables were examined and statistically compared: duration, center frequency, peak frequency, sweep rate, and number of zero crossings (**Figure 9**). Kruskal-Wallis and post-hoc Dunn's tests showed significant differences for variables among all communities (**Table 6**). Significant differences were identified for almost every variable in all pairwise comparisons except for transients versus northern residents. No variables were found to be significantly different when those two communities were compared.

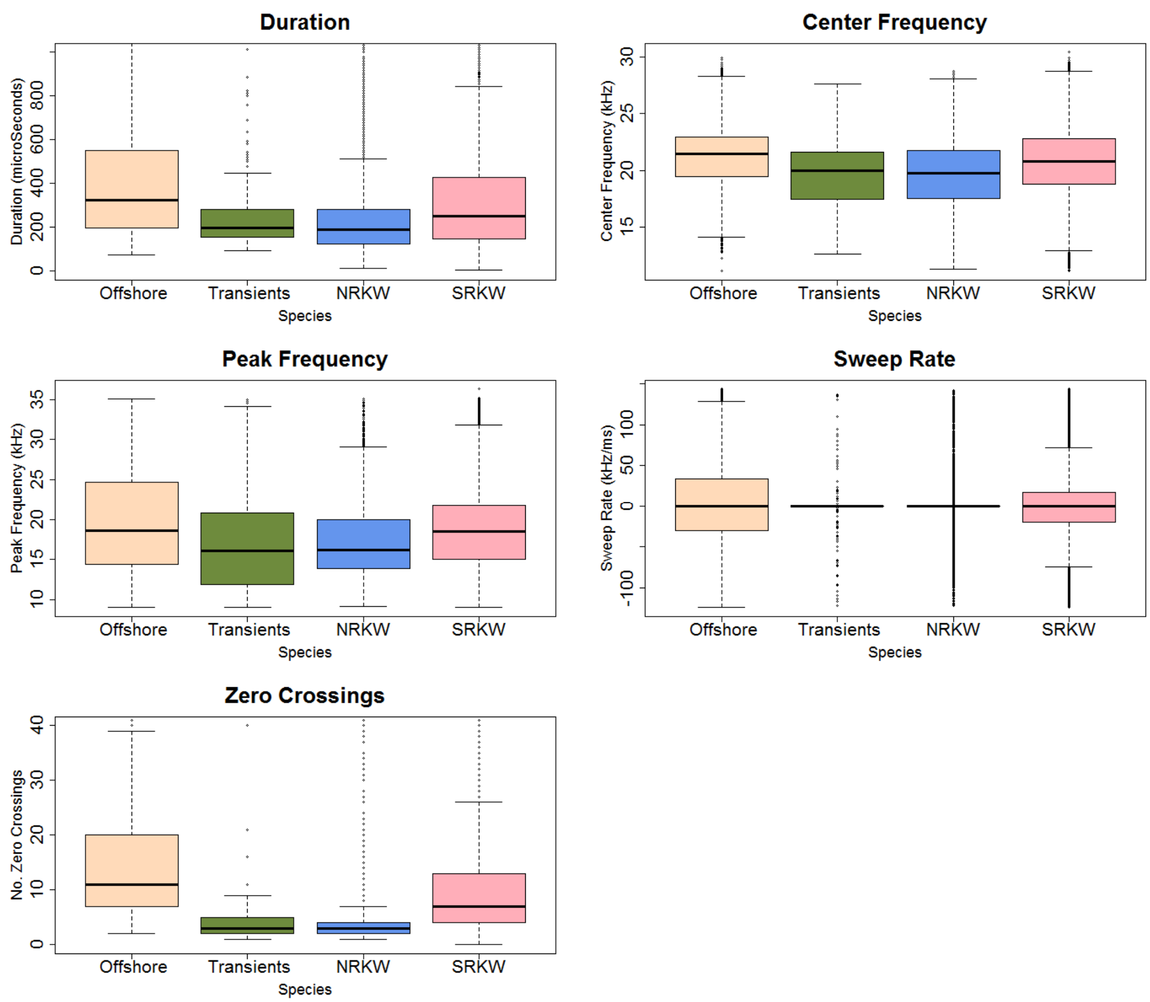

Figure 9. Box plots displaying the median, first and third quartiles for each echolocation click parameter with community along the x-axis and parameters along the y-axis (NRKW = northern resident killer whales, SRKW = southern resident killer whale). Sample sizes are 9729, 254, 5808 and 22483 echolocation clicks for offshore, transient, NRKW and SRKW, respectively.

Table 6. Matrix displaying echolocation click parameters that were significantly different between communities (Kruskal-Wallis and Dunn's tests with Bonferroni adjustments, α<0.05) (NRKW = northern resident killer whales, SRKW = southern resident killer whale).

	Offshore	Transient	NRKW
Transient	All Parameters Significant		
NRKW	Duration Center Frequency Peak Frequency Zero Crossing	No Parameters Significant	
SRKW	Duration Peak Frequency Center Frequency Zero Crossings	All Parameters Significant	All Parameters Significant

4.4 Classification Results

The whistle classifier had an overall correct classification score of 70.4% for individual whistles and 70.6% for sub-encounters. Correct classification scores were significantly greater than the 33.3% expected by chance alone (Fisher's exact test, α = 0.05) for both individual whistles and sub-encounters for all communities, with the exception of northern resident sub-encounters (**Table 7**). Transient killer whales were not included in the whistle classifier because the sample size of whistles for this community was not large enough to adequately train and test a classifier. Overall, 49.0% of individual pulsed calls were classified to the correct community and 58.1% of sub-encounters were correctly classified by the pulsed call classifier. For individual pulsed calls, correct classification scores were significantly greater than the 25% expected by chance (Fisher's exact test, α = 0.05) for every community except transient killer whales. For sub-encounters, correct classification scores were significantly greater than chance (Fisher's exact test, α = 0.05) for northern resident and southern resident communities. Offshore and transient sub-encounter correct classification scores were not significantly greater than chance (Fisher's exact test, α = 0.05), likely due to small sample sizes (**Table 8**). For the echolocation click classifier, 41.8% of individual echolocation clicks were correctly classified and 55.0% of sub-encounters were correctly classified overall. For individual echolocation clicks, correct classification scores were significantly greater than the 25% expected by chance alone (Fisher's exact test, α = 0.05) for every community except for southern resident killer whales. For sub-encounters, only northern resident killer whales had a correct classification score that was significantly greater than expected by chance (Fisher's exact test, α = 0.05). Offshore and transient correct classification scores were likely not significantly greater than chance (Fisher's exact test, α = 0.05) due to small sample sizes (**Table 9**).

When the random forest votes for each signal type were summed to produce a final classification for each sub-encounter, the resultant correct classification score was 80.0%. Correct classification scores were significantly greater than expected by chance (Fisher's exact test, α = 0.05) for every community except transient killer whales (likely due to small sample size, **Table 10**).

Table 7. Confusion matrices for (A) individual whistles and (B) whistle sub-encounters. Percent of whistles or sub-encounters correctly classified for each community is in bold, with standard deviations in parentheses. Correct classification scores that are significantly greater (Fisher's exact test, α = 0.05) than the 33.3% expected by chance are denoted with an asterisk.

A.

	Classified as			
Actual community	**Offshore**	**Northern resident**	**Southern resident**	**n**
Offshore	**98.0* (0.0)**	1.0 (0.1)	1.0 (0.1)	148
Northern resident	0.5 (0.5)	**49.8* (3.5)**	49.9 (3.6)	148
Southern resident	1.4 (0.9)	35.3 (2.9)	**63.4* (3.0)**	148

B.

	Classified as			
Actual community	**Offshore**	**Northern resident**	**Southern resident**	**n**
Offshore	**98.0 * (0.0)**	1.0 (0.1)	1.0 (0.1)	88
Northern resident	1.3 (1.4)	**49.9 (8.4)**	48.9 (8.5)	40
Southern resident	4.3 (2.2)	32.0 (5.6)	**64.0 * (5.6)**	36

Table 8. Confusion matrices for (A) individual pulsed calls and (B) pulsed call sub-encounters. Percent of pulse calls or sub-encounters correctly classified for each community is in bold, with standard deviations in parentheses. Correct classification scores that are significantly greater (Fisher's exact test, α = 0.05) than the 25% expected by chance are denoted with an asterisk.

A.

	Classified as				
Actual community	**Offshore**	**Transient**	**Northern resident**	**Southern resident**	**n**
Offshore	**64.9*** **(3.6)**	9.6 (2.4)	7.9 (2.4)	17.6 (3.6)	101
Transient	10.2 (1.8)	**34.8** **(3.3)**	26.0 (4.0)	30.0 (4.3)	101
northern resident	8.1 (2.4)	19.6 (3.9)	**49.1*** **(4.2)**	23.8 (4.4)	101
Southern resident	15.8 (2.5)	21.9 (3.4)	15.7 (3.3)	**47.3*** **(4.3)**	101

B.

	Classified as				
Actual community	**Offshore**	**Transient**	**Northern resident**	**Southern resident**	**n**
Offshore	**74.0** **(4.9)**	7.0 (11.3)	11.5 (12.5)	7.5 (12.1)	4
Transient	0.0 (0.0)	**45.2** **(14.7)**	25.0 (13.1)	29.8 (14.4)	5
northern resident	3.0 (3.5)	15.3 (7.2)	**64.7*** **(8.3)**	17.0 (6.7)	20
Southern resident	16.7 (3.1)	18.3 (3.9)	16.7 (4.0)	**48.4*** **(5.5)**	60

Table 9. Confusion matrices for (A) individual echolocation clicks and (B) echolocation click sub-encounters. Percent of echolocation clicks or sub-encounters correctly classified for each community is in bold, with standard deviations in parentheses. Correct classification scores that are significantly greater (Fisher's exact test, α = 0.05) than the 25% expected by chance are denoted with an asterisk.

A.

	Classified as				
Actual community	**Offshore**	**Transient**	**Northern resident**	**Southern resident**	**n**
Offshore	**41.5* (6.9)**	9.3 (2.0)	11.9 (3.3)	37.2 (7.0)	254
Transient	6.7 (1.6)	**44.7* (3.0)**	41.1 (3.3)	7.5 (1.8)	254
northern resident	6.5 (1.7)	28.6 (3.0)	**52.7* (3.8)**	12.5 (2.7)	254
Southern resident	33.7 (5.0)	10.9 (2.2)	26.8 (3.4)	**28.5 (5.1)**	254

B.

	Classified as				
Actual community	**Offshore**	**Transient**	**Northern resident**	**Southern resident**	**n**
Offshore	**68.4 (20.8)**	4.0 (9.2)	2.5 (7.9)	25.2 (19.3)	5
Transient	13.1 (5.9)	**52.4 (14.2)**	32.5 (12.5)	2.0 (5.3)	7
northern resident	9.6 (5.9)	21.3 (9.5)	**65.4* (10.5)**	3.7 (5.0)	14
Southern resident	21.9 (5.4)	8.4 (5.4)	35.7 (8.1)	**33.9 (10.8)**	31

Table 10. Confusion matrix for final sub-encounter classification based on results of the whistle, pulsed call and echolocation click classifiers. Percent of sub-encounters correctly classified for each community is in bold. Correct classification scores that are significantly greater (Fisher's exact test, α = 0.05) than the 25% expected by chance are denoted with an asterisk.

	Classified as				
Actual community	**Offshore**	**Transient**	**Northern resident**	**Southern resident**	**n**
Offshore	**100***	0	0	0	5
Transient	0	**71.4**	28.6	0	7
northern resident	0	0	**87.5***	12.5	24
Southern resident	12.2	2	22.4	**63.3***	147

5. CONCLUSIONS AND RECOMMENDATIONS

All three sound-types examined in this study varied among communities. Offshore killer whales produced the most distinctive sounds; almost every measured variable for every sound-type was significantly different when offshore killer whales were compared to every other community. This is perhaps not surprising, given that offshore killer whales differ genetically (reproductively isolated) and morphologically (e.g., smaller, fin structure more rounded, less sexual dimorphism) from resident and transient ecotypes , and also have a larger but overlapping geographic range and more diverse diet than residents and transients (Barret-Leonard and Heise 2006).

Very few pairwise comparisons were not significant for any sound-type, with the exception of northern resident versus southern resident whistles and northern resident versus transient echolocation clicks. These results suggest that northern resident killer whales produce whistles that are similar to those produced by southern resident killer whales, and echolocation clicks that are similar to those produced by transient killer whales.

Due to the large number of significant differences in measured variables, classifiers trained using a single sound-type (whistles, pulsed calls or echolocation clicks) performed relatively well at classifying killer whales for most communities. Not surprisingly, offshore killer whales had the highest correct classification scores for all three classifiers. The whistle classifier produced the highest correct classification scores for offshore and southern resident communities, but did not include transient killer whales because the sample size for whistles was too low for this community. Transient killer whales have been reported to produce very few whistles during most behavior states (Riesch and Deecke 2011) and so a lack of whistling may be a distinguishing characteristic for this community. However, given a sufficient sample size, it would be valuable to include transient killer whale whistles in the classifier. When they do whistle, the whistles produced by this community have been found to be significantly lower in frequency and shorter in duration than whistles produced by northern and southern resident killer whales (Riesch and Deecke 2011). Even with a limited sample size, similar trends were observed in the data analyzed for this study. Differences found by Riesch and Deecke (2011) suggest that the addition of transient killer whale whistles to our classifier would produce a more complete and accurate classification algorithm.

Almost all correct classification scores were significantly greater than chance for individual sounds. The only exceptions were southern resident echolocation clicks and transient pulsed calls. It was difficult to extract time-frequency information from pulsed calls due the presence of side-bands and harmonics and the fact that the fundamental frequency contour was often difficult to identify. Given this, we chose to extract measurements from only the loudest contour. Which contour is the loudest in a given pulsed call may be influenced by sound propagation through the water column, distance and orientation of the killer whale relative to the hydrophone, and other factors such as behavior state. In addition, the choice of the loudest contour was subjective. Because of these factors, we did not include frequency variables in the classification analysis. Adding frequency information may improve classifier performance and future work should consider this and explore approaches that could incorporate frequency information.

While the three independent classifiers resulted in relatively high correct classification scores, the goal of this study was to produce correct classification scores that are as close to 100% as possible. Correct classification scores were improved by basing classification decisions on all three sound-types rather than on only one sound-type. Small sub-encounter sample sizes for offshore and transient communities make it difficult to evaluate the results, however they are promising, with 100% of offshore sub-encounters correctly classified. Future work should include larger sample sizes for all communities to allow for more robust testing and training of the classifiers.

Many of the encounters included in this analysis spanned hours of time and likely included several different behavior states. The relative use of different sound-types by resident killer whales has been found to vary with behavior state (Ford 1989, Thomsen et al. 2002, Miller 2006). In order to ensure that as much variability as possible was captured in whistle, pulsed call and echolocation click characteristics, we divided encounters into sub-encounters based on vocal activity levels. Visual data related to behavior state was not available for this analysis and so these divisions may not accurately reflect changes in behavior state. In addition, because encounters were generally divided into several sub-encounters that included the same individual killer whales, these sub-encounters were not truly independent units. This may have artificially inflated the correct classification results reported here. Future work should include the analysis of recordings that also include information about behavior state from visual observations.

Future work should also include the analysis of additional recordings for all communities. Large volumes of archival data exist for all of the resident killer whale pods, as well as offshore and transient killer whales. Analysis of larger sample sizes would allow for more robust testing and training of the classifiers, as well as exploration of the potential for applying these methods to classifying acoustic encounters to pod level. Pod level classification was not possible during this study due to small sample sizes. With larger training and testing datasets, this tool could be made more broadly applicable and implemented into freely available acoustic data processing software such as PAMGuard for real-time and post-processing analysis.

The classifier developed here represents the first killer whale classifier that uses a combination of information from echolocation clicks, whistles and pulsed calls to identify groups of northeastern Pacific killer whales to the community level. This work provides a significant contribution to the study of killer whales and has important conservation applications. It will allow acoustic data to be analyzed in a more efficient and timely manner, providing important data that can be used to explore questions related to the occurrence, distribution and behavior of killer whales in the northeast Pacific Ocean.

6. ACKNOWLEDGEMENTS

Thank you to our sponsor, the Sea World Busch Gardens Conservation Fund for supporting this work. We would also like to thank the NOAA Northwest Fisheries Science Center for providing us with the data for this project, especially Brad Hanson, Marla Holt and Candice Emmons. We would also like to thank the entire crew of NOAA Ships Bell M. Shimada and McArthur II for their dedicated effort on this project, and the entire PODS science team from all years (Thomas Norris, Alyson Azzara, Shannon Coates, Cory Hom-Weaver, and Dave Haas). A special thanks to Courtney Drake for her help measuring whistles and pulsed calls and to Robyn Walker for report formatting and editing assistance.

7. LITERATURE CITED

Baird, R.W. 2000. The killer whale - foraging specializations and hunting. In: Mann, J., Connor, R.C., Tyack, P.L., Whitehead, H. (Eds.), *Cetacean societies: field studies of dolphins and whales*. Univ. Chicago Press, Chicago, IL, pp. 127-153.

Barrett-Lennard, L.G., Heise, K. 2006 The Natural History and Ecology of Killer Whales: Foraging Specialization in a Generalist Predator. In Estes, J.A., Brownell, R.L., DeMaster, D.P., Doak, D.F., Williams, T.M. (Eds.) Whales, whaling and ocean ecosystems. University of California Press, Berkely, C.A. pp. 163-173.

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

Brown, J.C., and Smaragdis, P. 2009. Hidden Markov and Guassian mixture models for automatic call classification. *The Journal of the Acoustical Society of America.* 125: EL221-EL224.

Carretta, J. V., Forney, K.A., Lowry, M.S., Barlow, J., Baker, J., Johnston, D., Hanson, B., Brownell Jr., R.L., Robbins, J., Mattila, D.K., Ralls, K., Muto, M.M., Lynch, D., and Carswell, L. 2010. U.S. Pacific Marine Mammal Stock Assessments: 2009. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-SWFSC-453. 336 pages.

Ford, J.K.B. 1989. Acoustic behaviour of resident killer whales (*Orcinus orca)* off Vancouver Island, British Columbia. *Canadian Journal of Zoology*. 67: 727–745.

Ford, J.K.B. 1991. Vocal traditions among resident killer whales (*Orcinus orca*) in coastal waters of British Columbia. *Canadian Journal of Zoology*. 69: 1454–1483.

Ford, J.K.B., and Fisher, H.D. 1983. Group-specific dialects of killer whales (*Orcinus orca*) in British Columbia. In: Payne, R. (Ed.), *Communication and Behavior of Whales*. AAAS Selected Symp, 76: 129–161.

Fristrup, K.M., and Watkins, W.A. 1993. Marine animal sound classification. Woods Hole Oceanographic Institution Technical Report WHOI-94-13. pp. 29.

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

Hanson, M. B., Emmons, C. K., Ward, E. J., Nystuen, J. A., and Lammers, M. O. 2013. Assessing the coastal occurrence of endangered killer whales using autonomous passive acoustic recorders. *The Journal of the Acoustical Society of America*. 134(5): 3486-3495.

Krahn, M.M., Ford, M.J., Perrin, W.F., Wade, P.R., Angliss, R.P., Hanson, M.B., Taylor, B.L., Ylitalo, G.M., Dahlheim, M.E., Stein, J.E., and Waples, R.S. 2004. 2004 Status review of southern resident killer whales (*Orcinus orca*) under the Endangered Species Act. U.S. Department of Commerce., NOAA Technical Memorandum NMFS NWFSC-62, pp. 73.

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

Miller, P.J.O. 2006. Diversity in sound pressure levels and estimated active space of resident killer whale vocalizations. *Journal of Comparative Physiology A* 192:449-459.

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., Rankin, S., Barlow, J., Oswald M. and M.O. Lammers. 2013. Real-time Call Classification Algorithm (ROCCA): software for species identification of delphinid whistles. In: Adam, O. and Samaran, F. (Eds.) *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., Rankin, S., Barlow, J., and Lammers, M.O. 2007. A tool for real-time acoustic species identification of delphinid whistles. *The Journal of the Acoustical Society of America.* 122: 587-595.

Oswald, J.N., Barlow, J., and Norris, T.F. 2003. Acoustic identification of nine delphinid species in the eastern tropical Pacific Ocean. *Marine Mammal Science.* 19: 20-37.

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

Riesch, R. and Deecke, V.B. 2011. Whistle communication in mammal-eating killer whales (*Orcinus orca*): further evidence for acoustic divergence between ecotypes. *Behavioral Ecology and Sociobiology*. 65: 1377-1387.

Roch, M.A., Soldevilla, M.S., Burtenshaw, J.C., Henderson, E.E., and Hildebrand, J.A. 2007. Gaussian mixture model classification of odontocetes in the southern California Bight and the Gulf of California. *The Journal of the Acoustical Society of America.* 121: 1737-1748.

Steiner, W.W. 1981. Species-specific differences in pure tonal whistle vocalization of five western North Atlantic dolphin species. *Behavioral Ecology and Sociobiology*. 9: 241-246.

Thomsen, F., Franck, D., and Ford, J.K.B. 2002. On the communicative significance of whistles in wild killer whales (*Orcinus orca*). Naturwissenschaften 89: 404-407.

Wang, D., Würsig, B., and Evans, W. 1995. Comparisons of whistles among seven odontocete species. *In*: R.A. Kastelein, J.A. Thomas, and P.E. Nachtigall (Eds). *Sensory Systems of Aquatic Mammals*. De Spil Publishers, Woerden, Netherlands, pp. 299-323.

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

APPENDIX A:

VARIABLES MEASURED BY ROCCA.

Appendix A

Variables measured by ROCCA

Variable	Explanation
Begsweep	slope of the beginning sweep (1 = positive, -1 = negative, 0 = zero)
Begup	binary variable: 1 = beginning slope is positive, 0 = beginning slope is negative
Begdwn	binary variable: 1 = beginning slope is negative, 0 = beginning slope is positive
Endsweep	slope of the end sweep (1 = positive, -1 = negative, = 0 zero)
Endup	binary variable: 1 = ending slope is positive, 0 = ending slope is negative
Enddwn	binary variable: 1 = ending slope is negative, 0 = ending slope is positive
Beg	beginning frequency (Hertz [Hz])
End	ending frequency (Hz)
Min	minimum frequency (Hz)
Dur	duration (seconds)
Range	maximum frequency - minimum frequency (Hz)
Max	maximum frequency (Hz)
mean freq	mean frequency (Hz)
median freq	median frequency (Hz)
std freq	standard deviation of the frequency (Hz)
Spread	difference between the 75th and the 25th percentiles of the frequency
quart freq	frequency at one-quarter of the duration (Hz)
half freq	frequency at one-half of the duration (Hz)
Threequart	frequency at three-quarters of the duration (Hz)
Centerfreq	(minimum frequency + (maximum frequency-minimum frequency))/2
rel bw	relative bandwidth: (maximum frequency - minimum frequency)/center frequency
Maxmin	maximum frequency/minimum frequency
Begend	beginning frequency/end frequency
Cofm	coefficient of frequency modulation (COFM): take 20 frequency measurements equally spaced in time, then subtract each frequency value from the one before it. COFM is the sum of the absolute values of these differences, all divided by 10,000
tot step	number of steps (10 percent or greater increase or decrease in frequency over two contour points)
tot inflect	number of inflection points (changes from positive to negative or negative to positive slope)
max delta	maximum time between inflection points
min delta	minimum time between inflection points
maxmin delta	maximum delta/minimum delta
mean delta	mean time between inflection points

Variable	Explanation
std delta	standard deviation of the time between inflection points
median delta	median of the time between inflection points
mean slope	overall mean slope
mean pos slope	mean positive slope
mean neg slope	mean negative slope
mean absslope	mean absolute value of the slope
Posneg	mean positive slope/mean negative slope
perc up	percent of the whistle that has a positive slope
perc dwn	percent of the whistle that has a negative slope
perc flt	percent of the whistle that has zero slope
up dwn	number of inflection points that go from positive slope to negative slope
dwn up	number of inflection points that go from negative slope to positive slope
up flt	number of times the slope changes from positive to zero
dwn flt	number of times the slope changes from negative to zero
flt dwn	number of times the slope changes from zero to negative
flt up	number of times the slope changes from zero to positive
step up	number of steps that have increasing frequency
step dwn	number of steps that have decreasing frequency
step.dur	number of steps/duration
inflect.dur	number of inflection points/duration