

Development of Automated Whistle and Click Classifiers for Odontocete Species in the Northwest Atlantic Ocean and the Waters Surrounding the Hawaiian Islands

Julie N. Oswald & Tina M. Yack
Bio-Waves, Inc.
364 2nd Street, Suite #3
Encinitas, CA 92024
phone: (760) 452-2575 fax: (760) 652-4878 email: Julie.oswald@bio-waves.net

Award Number: N00014-14-1-0413
www.bio-waves.net

LONG-TERM GOALS

1. Create combined whistle and click classifiers for odontocetes in the northwest Atlantic Ocean, the temperate Pacific Ocean and the waters surrounding the Hawaiian Islands. These classifiers will also incorporate ancillary information about acoustic encounters.
2. Examine geographic variation in the characteristics of whistles and clicks produced by species that are found in two or more of the three areas included in this effort.

OBJECTIVES

The primary objective of this effort is to develop combined automated whistle and click classifiers for odontocete species in the northwest Atlantic Ocean and the waters surrounding the Hawaiian Islands. We will also incorporate ancillary information (e.g. vocalization rate, relative abundance of whistles and clicks, latitude of acoustic encounter) about acoustic encounters into the classifiers in an effort to improve classification the performance. These classifiers will be implemented in an existing whistle classifier, known as the Real-time Odontocete Call Classification Algorithm (ROCCA). ROCCA currently is available to users as a module in the acoustic processing software platform, PAMGuard (Gillespie et al. 2008).

We will integrate new classifiers with fully automated detectors and feature extractors in PAMGuard to provide a complete tool for efficiently and automatically processing the large datasets generated from Passive Acoustic Monitoring (PAM). When these tasks have been completed, we will use archival acoustic data recorded in the temperate Pacific Ocean to create whistle, click and ancillary information feature vectors. These feature vectors will be used to train a temperate Pacific classifier under an LMR-funded effort.

Additionally, we will examine geographic variation in the characteristics of whistles and clicks produced by species that are found in two or more of the three areas included in this effort. These comparisons will provide important information about the necessity of creating different classifiers

for different geographic areas. Furthermore, these comparisons may allow us to use vocalizations recorded in one location to increase the sample size of the training dataset for another location.

APPROACH

Task 1: Automated collection of ancillary feature vectors

Ancillary variables such as the relative number of clicks and whistles present in an acoustic encounter and the click and/or whistle rate can be calculated based on the output of the automated detectors. We will create ‘ancillary information’ feature vectors based on automated detector output.

Task 2: Train and test classifiers

Task 2.1: Creation of northwest Atlantic feature vectors

Three different feature vectors will be created using variables measured from clicks, whistles and ancillary information related to the acoustic encounter. The click feature vector will include variables such as peak frequency, center frequency, 3 dB and 10 dB bandwidth, duration, inter-click interval, number of zero crossings and sweep rate (similar to those described and recommended by Au 1993, Soldevilla et al. 2008, and Baumann-Pickering et al. 2010). These variables will be measured automatically by ROCCA using information passed from PAMGuard’s automated click detector. Implementation of the ROCCA click measurement modules was completed during a related LMR funded effort. The whistle feature vector will include 56 variables that are measured automatically by ROCCA (see Oswald 2013 for a description of these variables). The variables will be measured using the output of PAMGuards automated tonal detector. The automated detector will pass time-frequency contours to ROCCA, and ROCCA will automatically measure features from the contours.

For this task, feature vectors will be created using visually validated, single species acoustic recordings collected in the northwest Atlantic Ocean. Table 1 lists data currently available for training and testing the random forest classification model. Data have been provided by the Southeast Fisheries Science Center (SEFSC), the Northeast Fisheries Science Center (NEFSC) and Duke University. In addition to the data listed in Table 1, acoustic recordings from other concurrent visual and acoustic marine mammal survey completed by NEFSC and SEFSC in the northwest Atlantic will be provided.

Task 2.2: Train northwest Atlantic classifiers

Three random forest classifiers will be trained: one for whistles, one for clicks and one for ancillary information pertaining to the acoustic encounters. The Gini variable importance output of the random forest analysis, which provides a measure of how strongly each variable contributes to the model predictions (Breiman 2001), will be used to determine which features have the greatest potential for distinguishing among species. Multiple random forest models will be created, each containing different combinations of variables that contribute strongly to the models. The models will be tested using novel data and the whistle, click and ancillary classifiers that produce the most accurate results will be selected. Once three individual classifiers have been trained and tested, a second step will be implemented where the output of the three classifiers are combined and evaluated to determine a final classification for each odontocete encounter. Not every acoustic encounter will contain both whistles and clicks. As such, classifiers will be built to classify encounters based on any combination of the three feature vectors. The northwest Atlantic dataset described in Task 2.1 will be used for testing and training the classifiers

Table 1. Species and number of encounters per species available for development of northwest Atlantic classifier. Recordings have been provided by the Southeast Fisheries Science Center, the Northeast Fisheries Science Center and Duke University. Whistles from a subset of recordings have been measured using PAMGuard's automated tonal detector as part of a previous effort. All recordings are single species recordings and have visual confirmation of species identity.

Species	Number of encounters
<i>Delphinus delphis</i>	9
<i>Grampus griseus</i>	8
<i>Globicephala macrorhynchus</i>	17
<i>Steno bredanensis</i>	3
<i>Stenella coeruleoalba</i>	7
<i>Stenella frontalis</i>	40
<i>Tursiops truncatus</i>	66
<i>Stenella clymene</i>	2
<i>Pseudorca crassidens</i>	2
Total	154

Task 2.3: Create Hawaii feature vectors

Echolocation clicks, whistles, and ancillary variables will be measured from recordings made in the waters surrounding the Hawaiian Islands using the automated whistle and click detectors and extractors in PAMGuard/ROCCA. The recordings in this dataset were collected using towed hydrophone arrays during concurrent visual and acoustic surveys conducted by the Southwest Fisheries Science Center (SWFSC), the Pacific Islands Fisheries Science Center (PIFSC), and Woods Hole Oceanographic Institution (WHOI). Species identity of all acoustic recordings to be included has been verified by visual observers. The species that will be included and the number of recordings for each are given in Table 2.

Task 2.4: Train and test Hawaii classifiers

The feature vectors created in Task 2.3 will be used to train and test Hawaii random forest classifiers. The feature vectors and classification approach determined in tasks 2.1 and 2.2 will be implemented first, but other feature vectors and approaches will also be tested, as the most effective feature vectors and approach may vary with geographic location and species included in the classifier.

Task 2.5: Create temperate Pacific feature vectors

Clicks and whistles recorded in the temperate Pacific using the automated whistle and click detectors and extractors will be measured. Ancillary encounter variables will also be measured. The recordings in this dataset were collected using towed hydrophone arrays during combined visual and acoustic surveys run by the SWFSC and the species identity of all acoustic recordings has been verified by visual observers. The species that will be included and the number of recordings for each are given in Table 3.

Table 2. Species and number of individual acoustic encounters available for development of Hawaii classifiers. Recordings will be provided by the Southwest Fisheries Science Center, the Pacific Islands Fisheries Science Center and Woods Hole Oceanographic Institution. All recordings are single-species recordings and have visual confirmation of species identity.

Species	Number of encounters
<i>Tursiops truncatus</i>	24
<i>Pseudorca crassidens</i>	34
<i>Stenella attenuata</i>	19
<i>Globicephala macrorhynchus</i>	30
<i>Steno bredanensis</i>	30
<i>Stenella coeruleoalba</i>	31
<i>Stenella longirostris</i>	9
<i>Peponocephala electra</i>	17
<i>Grampus griseus</i>	9
<i>Lagenodelphis hosei</i>	4
<i>Feresa attenuata</i>	3
<i>Physeter macrocephalus</i>	134
<i>Indopacetus pacificus</i>	1
<i>Ziphius cavirostris</i>	5
<i>Mesoplodon</i> sp.	2
TOTAL	352

Table 3. Species and number of encounters per species available for development of temperate Pacific classifier. Recordings will be provided by the Southwest Fisheries Science Center. All recordings are single species recordings and have visual confirmation of species identity.

Species	Number of acoustic encounters
<i>Delphinus delphis</i>	106
<i>Delphinus capensis</i>	26
<i>Lagenorhynchus obliquidens</i>	9
<i>Orcinus orca</i>	4
<i>Grampus griseus</i>	20
<i>Tursiops truncatus</i>	17
<i>Berardius bairdii</i>	7
<i>Ziphius cavirostris</i>	7
<i>Lissodelphis borealis</i>	9
<i>Physeter macrocephalus</i>	13
Total	218

Task 2.6: Prepare peer reviewed journal article describing the classification approach and results

A manuscript describing the classification approach and results will be prepared and submitted to a journal such as the Journal of the Acoustical Society of America. All P.I.'s will be co-authors.

Task 3: Geographic comparison of whistles and clicks

Task 3.1: Geographic comparison of whistles

We will include species that occur in multiple study areas in the geographic comparison. Species will be chosen based on sample size and occurrence in the study areas. Features measured from whistles recorded in the northwest Atlantic Ocean, Hawaii, and the temperate Pacific Ocean will be compared using statistics such as multi-variate analysis of variance, Students t-tests and Kruskal-Wallis tests. This task will be completed after measurement of whistles from all three locations, but before training classifiers for these regions. This will allow us to include whistles from multiple locations in a single classifier if significant differences are not found among locations.

Task 3.2: Geographic comparison of clicks

Features measured from clicks recorded in the northwest Atlantic Ocean, Hawaii, and the temperate Pacific Ocean will be compared using statistics such as multi-variate analysis of variance, Students t-tests and Kruskal-Wallis tests. Species will be chosen based on sample size and occurrence in the study areas. This task will be completed after measurement of clicks from all three locations, but before training classifiers for these regions. This will allow us to include whistles from multiple locations in a single classifier if significant differences are not found among locations.

Task 3.3: Prepare peer-reviewed journal article presenting results of geographic comparisons

A manuscript describing the results and implications of the geographic comparisons will be prepared and submitted to a peer-reviewed journal such as the Journal of the Acoustical Society of America or Applied Acoustics. All PIs will be co-authors.

WORK COMPLETED

Task 1: Automate collection of 'ancillary' variables

- Computer programming is still in progress

Task 2.1: Creation of northwest Atlantic feature vectors

- Work on this task is still in progress. We have measured whistles from all of the 154 encounters that we currently have data for (Table 4) and have measured clicks from 80% of currently available encounters (Table 5). Because of restrictions on the use of the SEFSC data, we are currently permitted use those recordings only for the whistle and ancillary feature vectors. We expect to receive additional recordings from NWFS and SEFSC within the next month and will create whistle and click feature vectors for those encounters when we receive them. Ancillary feature vectors will be created during FY15.

All other tasks are scheduled for completion during FY15 and FY16.

Table 4. Whistles measured to date for the northwest Atlantic classifier. Number of independent acoustic encounters analyzed with the total number of whistles measured in parentheses.

Species	SEFSC Encounters (Whistles)	NEFSC Encounters (Whistles)	Duke Encounters (Whistles)	Total Encounters per species	Total Whistles per species
<i>Stenella frontalis</i>	10 (314)	4 (4190)	26 (2016)	40	6520
<i>Tursiops truncatus</i>	26 (651)	2 (109)	38 (3427)	66	4187
<i>Stenella clymene</i>	2 (87)	-	-	2	87
<i>Pseudorca crassidens</i>	2 (255)	-	-	2	255
<i>Globicephala macrorhynchus</i>	-	2 (287)	15 (25410)	17	25697
<i>Grampus griseus</i>	2 (15)	6 (143)	-	8	158
<i>Steno bredanensis</i>	1 (9)	1 (416)	1 (85)	3	510
<i>Delphinus delphis</i>	4 (301)	5 (1651)	-	9	1952
<i>Stenella coeruleoalba</i>	3 (771)	4 (1140)	-	7	1911
Total encounters	50	24	80	154	-
Total whistles	2403	7936	30938	-	41277

Table 5. Number of northwest Atlantic acoustic encounters currently available for click analysis. Red font indicates encounters that have been measured.

Species	NEFSC	Duke	Total
<i>Stenella frontalis</i>	4	22	26
<i>Tursiops truncatus</i>	2	37	39
<i>Stenella clymene</i>	TBD	NA	0
<i>Delphinus delphis</i>	3	NA	3
<i>Pseudorca crassidens</i>	TBD	NA	0
<i>Globicephala macrorhynchus</i>	2	6	8
<i>Grampus griseus</i>	2	4	6
<i>Steno bredanensis</i>	1	1	2
<i>Stenella coreuleoalba</i>	3	NA	3
<i>Physeter macrocephalus</i>	TBD	TBD	TBD
Total	17	9	87

RESULTS

Click Measurements

Click measurements have been obtained for all of the data provided by Duke. Figures 1 and 2 show species comparisons for the variables that have been measured. The plots show that there are detectable differences in some click characteristics (e.g. peak frequency, center frequency, -3dB and -10 dB bandwidth, etc) that can be used improve species classifications, especially when used in conjunction with whistle features.

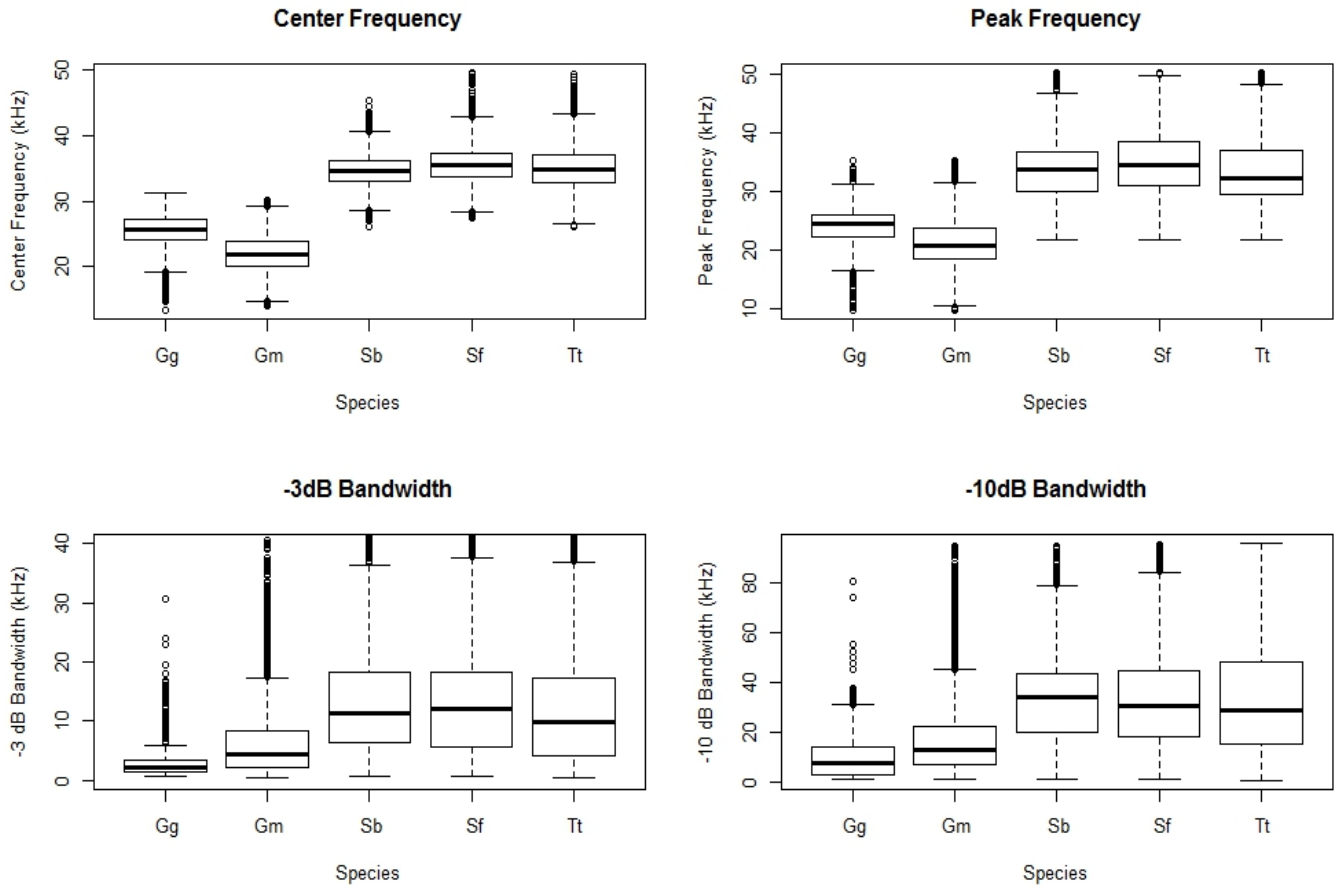


Figure 1. Frequency and bandwidth variables measured from clicks recorded by Duke University in the northwest Atlantic ocean. All variables are shown by species as box plots. The boxes represent the upper 75% quartile and the lower 25% quartile, with the solid black line within each box indicating the median. The hinges show the maximum and minimum values excluding outliers and the stars represent outliers that are more than or less than 3/2 times the quartile ranges, respectively. Species listed on the x-axis are: *Grampus griseus* (Gg), *Globecephala macrorhynchus* (Gm), *Stenobredanensis* (Sb), *Stenella frontalis* (Sf), and *Tursiops truncatus* (Tt).

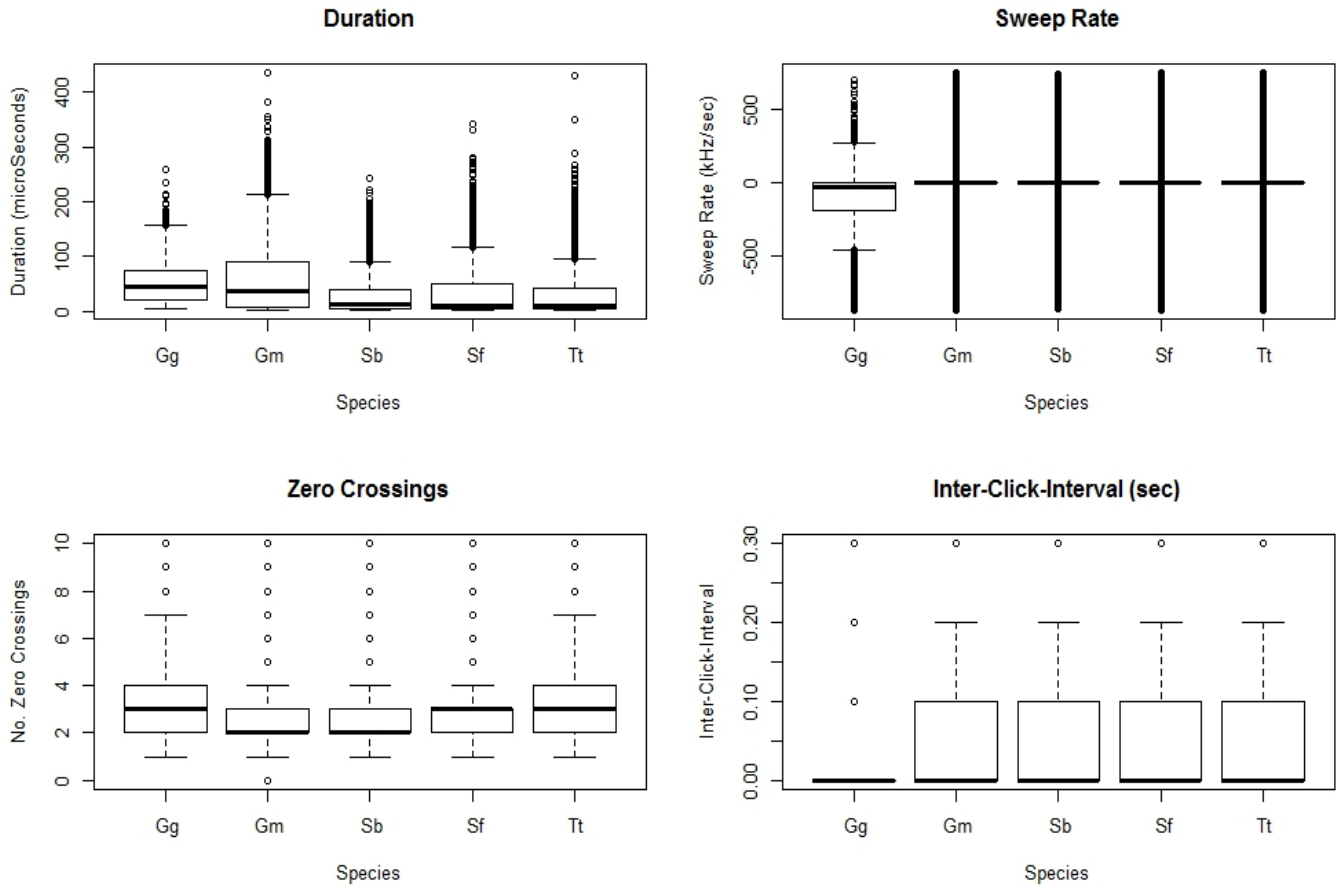


Figure 2. Duration, sweep rate, zero crossings and inter-click intervals measured from clicks recorded by Duke University in the northwest Atlantic ocean. All variables are shown by species as box plots. The boxes represent the upper 75% quartile and the lower 25% quartile, with the solid black line within each box indicating the median. The hinges show the maximum and minimum values excluding outliers and the stars represent outliers that are more than or less than 3/2 times the quartile ranges, respectively. Species listed on the x-axis are: *Grampus griseus* (Gg), *Globecephala macrorhynchus* (Gm), *Steno bredanensis* (Sb), *Stenella frontalis* (Sf), and *Tursiops truncatus* (Tt).

Whistle measurements

A total of 41,277 automatically detected whistle contours have been measured from 154 independent acoustic encounters recorded in the northwest Atlantic oceans. This dataset includes false detections as well as actual whistles. During FY15, we will work to develop methods for automatically ‘pruning’ as many false detections from the dataset as possible. Figures 3 and 4 show a subset of the 52 variables that were measured from the whistle contours. These figures show that differences of varying magnitudes exist among some variables and some species. During FY15, we will evaluate different combinations of variables to determine the optimal feature vector for species identification.

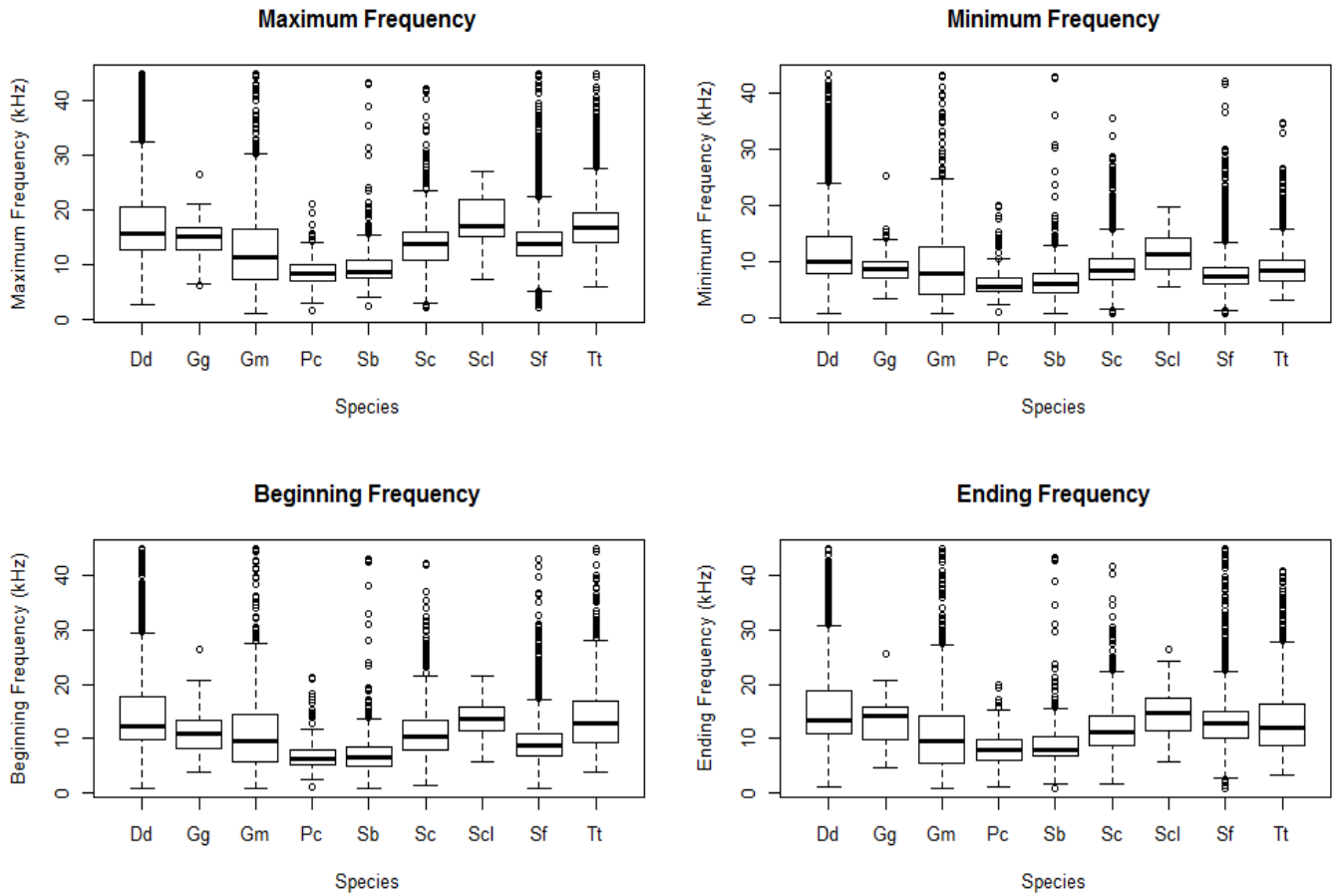


Figure 3. A subset of the frequency variables measured from whistles recorded in the northwest Atlantic ocean. All variables are shown by species as box plots. The boxes represent the upper 75% quartile and the lower 25% quartile, with the solid black line within each box indicating the median. The hinges show the maximum and minimum values excluding outliers and the stars represent outliers that are more than or less than 3/2 times the quartile ranges, respectively. Species listed on the x-axis are: *Delphinus delphis* (Dd), *Grampus griseus* (Gg), *Globecephala macrorhynchus* (Gm), *Pseudorca crassidens* (Pc), *Steno bredanensis* (Sb), *Stenella coeruleoalba* (Sc), *Stenella clymene* (Scl), *Stenella frontalis* (Sf), and *Tursiops truncatus* (Tt).

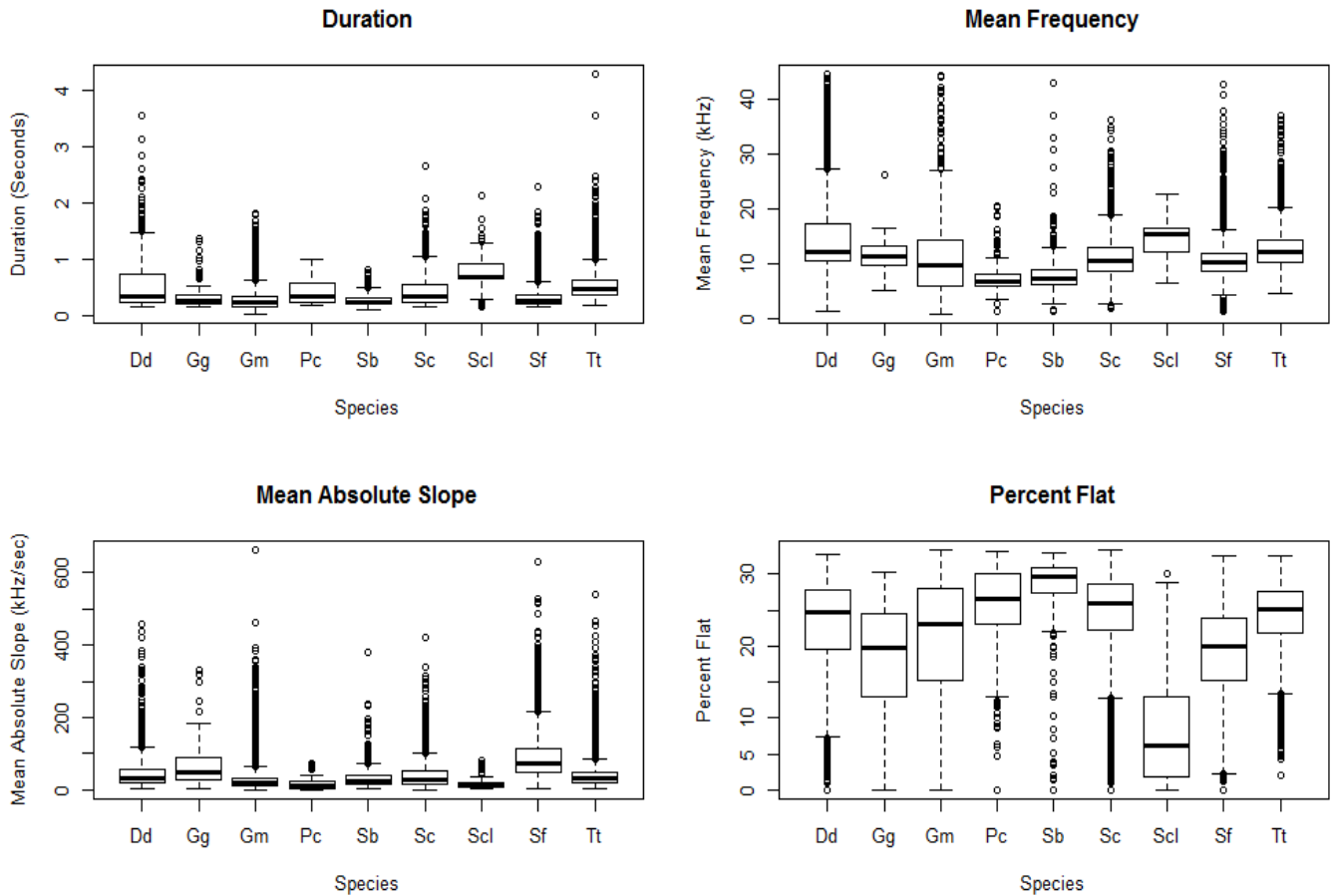


Figure 4. A subset of the time, frequency and slope variables measured from whistles recorded in the northwest Atlantic ocean. All variables are shown by species as box plots. Mean absolute slope refers to the mean of the absolute values of slope measurements taken at many points along the whistle contour. Percent flat refers to the percentage of the whistle contour with zero slope. The boxes represent the upper 75% quartile and the lower 25% quartile, with the solid black line within each box indicating the median. The hinges show the maximum and minimum values excluding outliers and the stars represent outliers that are more than or less than 3/2 times the quartile ranges, respectively. Species listed on the x-axis are: *Delphinus delphis* (Dd), *Grampus griseus* (Gg), *Globecephala macrorhynchus* (Gm), *Pseudorca crassidens* (Pc), *Steno bredanensis* (Sb), *Stenella coeruleoalba* (Sc), *Stenella clymene* (Scl), *Stenella frontalis* (Sf), and *Tursiops truncatus* (Tt).

IMPACT/APPLICATIONS

Passive acoustic monitoring via towed hydrophone arrays, seafloor hydrophones, glider and other recording systems has become an important component of projects designed for surveying and monitoring marine mammals in addition to mitigation of human activities that may impact these federally protected living marine resources. The enormous volumes of data currently being generated during PAM projects require automated methods for efficient and timely processing. Also, the ability to automatically detect, measure and classify vocalizations in real-time reduces the need for highly skilled, costly personnel during at-sea operations and post-processing and thereby reduces operational costs.

Including three broad geographic areas in this research effort will provide insight to the degree of geographic variation in whistles and clicks and how this variation might vary based on separation of populations (i.e. within the same ocean basin vs. widely separated ocean basins). If significant geographic variation in characteristics of whistles and clicks exists within species, then classifiers created using recordings made in one location will not be appropriate for use in other locations. However, if significant geographic variation in characteristics of whistles and clicks does not exist within a species, then vocalizations produced in one location can be used to augment the training dataset for another region.

Combining whistle and click classifiers will allow the classification of a greater proportion of acoustic encounters with greater accuracy than is possible when relying on whistles alone. The ability to classify odontocete vocalizations to species with a high degree of confidence will allow the examination of distribution patterns and habitat use as well as the development of management and mitigation strategies on a species-by-species basis. The three study areas included in this effort all include training areas that are frequently used by the U.S. Navy. Therefore knowledge of when and how species are using these areas is particularly important. PAM is now being used extensively to collect information regarding marine mammal occurrence, distribution and behavior in Naval exercise areas. However, it is not possible with current technology to identify many delphinids to species based on acoustic data alone. It is important to be able to identify delphinids to the species level for multiple reasons. Some species or populations are listed as threatened or endangered (ex. insular false killer whales in Hawaii) and thus it is particularly important to be able to differentiate them from other species in acoustic datasets. Also, different species may react differently to Naval activities. Knowledge of species specific responses will make it possible to plan more effective mitigation measures.

RELATED PROJECTS

This project is part of a larger effort funded jointly by LMR and ONR. The LMR-funded effort began in April, 2014. During this effort, we evaluated the performance of three automated tonal detection and extraction algorithms and the algorithm that produced the most accurate detections and contour extractions (PAMGuards tonal detector) will be used to detect and measure whistles in this ONR-funded effort. During the FY14 LMR-funded effort, we also integrated PAMGuards automated click detector with ROCCA and added automated click measurement capabilities to ROCCA via PAMGuard software. These capabilities will be used to measure clicks in this ONR-funded effort. In FY16, the LMR-funded effort will continue. During this time we will train and test temperate Pacific classifiers (using feature vectors created during ONR FY15) as well as integrate the automated detectors and classifiers into both PAMGuard and Ishmael software to make them easily accessible to end-users.

REFERENCES

Au, W. W. L. 1993. *The sonar of dolphins* (Springer, New York).

Baumann-Pickering, S, Wiggins, S., Hildebrand, J.A., Roch, M.A., Schnitzer, H.U. 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(4), 2212-2224.

- Breiman, L. (2001). Random forests. *Machine Learning* 45, 5-32.
- Gillespie, D.M., Gordon, J., McHugh, R., McLaren, D., Mellinger, D., 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. *Proceedings of the Institute of Acoustics*, 30 Pt 5. United Kingdom.
- 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.
- Roch, M.A., Klinck, H., Baumann-Pickering, S., Mellinger, D.K., Qui, S., et al. 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., Henderson, E.E., Campbell, G.S., Wiggins, S.M., Hildebrand, J.A., and Roch, M.A. 2008. Classification of Risso's and Pacific white-sided dolphins using spectral properties of echolocation clicks. *Journal of the Acoustical Society of America* 124:609–624.