NOAA Technical Memorandum NMFS

FEBRUARY 2011

COMPARISON OF REAL-TIME AND POST- CRUISE ACOUSTIC SPECIES IDENTIFICATION OF DOLPHIN WHISTLES USING ROCCA (REAL-TIME ODONTOCETE CALL CLASSIFICATION ALGORITHM)

Yvonne Barkley
Julie N. Oswald
James V. Carretta
Shannon Rankin
Alexis Rudd
and
Marc O. Lammers

NOAA-TM-NMFS-SWFSC-473

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

The National Oceanic and Atmospheric Administration (NOAA), organized in 1970, has evolved into an agency that establishes national policies and manages and conserves our oceanic, coastal, and atmospheric resources. An organizational element within NOAA, the Office of Fisheries is responsible for fisheries policy and the direction of the National Marine Fisheries Service (NMFS).

In addition to its formal publications, the NMFS uses the NOAA Technical Memorandum series to issue informal scientific and technical publications when complete formal review and editorial processing are not appropriate or feasible. Documents within this series, however, reflect sound professional work and may be referenced in the formal scientific and technical literature.

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.

FEBRUARY 2011

COMPARISON OF REAL-TIME AND POST- CRUISE ACOUSTIC SPECIES IDENTIFICATION OF DOLPHIN WHISTLES USING ROCCA (REAL-TIME ODONTOCETE CALL CLASSIFICATION ALGORITHM)

Yvonne Barkley[1], Julie N. Oswald[2], James V. Carretta[1],
Shannon Rankin[1], Alexis Rudd[3], and Marc O. Lammers[4]

[1]National Oceanic & Atmospheric Administration
National Marine Fisheries Service
Southwest Fisheries Science Center, Protected Resources Division
8604 La Jolla Shores Drive, La Jolla, California 92037

[2]Bio-Waves, Inc., 517 Cornish Drive, Encinitas, California 92024

[3]University of Hawaii, Department of Zoology, 2538 McCarthy Mall,
Edmondson 152, Honolulu, Hawaii 96839

[4]Oceanwide Science Institute,
P.O. Box 61692, Honolulu, Hawaii 96839

NOAA-TM-NMFS-SWFSC-473

U.S. DEPARTMENT OF COMMERCE
Gary F. Locke, Secretary
National Oceanic and Atmospheric Administration
Jane Lubchenco, Undersecretary for Oceans and Atmosphere
National Marine Fisheries Service
Eric C. Schwaab, Assistant Administrator for Fisheries

COMPARISON OF REAL-TIME AND POST-CRUISE

ACOUSTIC SPECIES IDENTIFICATION OF DELPHINID WHISTLES USING ROCCA

(*REAL-TIME ODONTOCETE CALL CLASSIFICATION ALGORITHM*)

Cover photograph: Spinner dolphin (*Stenella longirostris*). Photograph by: Shannon Rankin, Southwest Fisheries Science Center

INTRODUCTION

Southwest Fisheries Science Center (SWFSC) has been conducting shipboard visual line-transect surveys of marine mammals since the 1970s (Kinzey et al., 2000). Passive acoustic detection methods were incorporated in a systematic manner in 2000 (Rankin et al., 2008a). Passive acoustic detection of marine mammals is not limited by light, sea, and weather conditions and in some conditions, such as low light and inclement weather, can be more effective than visual methods (Thomas et al., 1986). Passive acoustic detection methods have been shown to improve the accuracy of cetacean abundance estimation for some species when coupled with visual detection (Barlow and Taylor, 2005; Barlow and Rankin, 2007). In addition, passive acoustic detection provides valuable information about marine mammals missed by visual observers during surveys (Rankin et al., 2008a). The ability of passive acoustic methods to aid in population surveys depends on the ability to accurately determine species identity from acoustic detections.

Several species of cetaceans produce calls that are distinct, easily detected with passive acoustics methods, and identifiable to species. These include vocalizations from baleen whales and sperm whales (Rankin and Barlow, 2005; Goold and Jones, 1995). For example, North Pacific minke whales are known for their unusual and unique ‘boing’ vocalization (Rankin and Barlow, 2005) and sperm whales produce distinctive clicks (Watkins, 1980).

Bio-acoustic signals produced by dolphins are more difficult to identify to species compared to those produced by most whales (Oswald, 2007). Their sounds are categorized into three types (echolocation clicks, burst pulses, and whistles), each of which can be used for species classification. Echolocation clicks of some species contain consistent spectral features that are species-specific and therefore can be used for species classification (Soldevilla et al., 2008). Burst pulses are broadband click trains with short interclick intervals. These call types have not yet been used in species identification (Roch et al., 2006). Whistles are frequency-modulated tonal calls that are highly variable within and among species (Lammers et al., 2003). In general, most delphinid whistles have energy in lower frequencies than echolocation clicks, and therefore propagate greater distances. This characteristic makes them useful for species identification during vessel- based marine mammal surveys.

Real-time Odontocete Call Classification Algorithm (ROCCA) is a Matlab-based tool that classifies dolphin species based on analysis of dolphin whistles (Oswald et al., 2007). Whistles are manually selected by an operator, and ROCCA uses the algorithm TRIA (written by M.O. Lammers) to extract the fundamental whistle contour. The contour extraction and measurement processes are semi-automated. ROCCA automatically measures 55 variables from the whistle contour. Whistles are classified to species using a suite of statistical analyses of measured variables (Fig. 1; Appendix A). Previous versions of ROCCA used discriminant function analysis (DFA) and classification and regression trees (CART) analysis to classify whistles (Oswald et al., 2007). Currently, random-forest classifiers (Brieman, 2001) are being developed for use in ROCCA. Breiman (2001) developed random-forest classifiers by creating multiple classification trees (= a forest) modeled from specific training data. The trees are used to provide a consensus prediction for novel input data. Correct classification scores vary by species and analysis methods; however, overall correct classification scores are significantly greater than chance for all classification algorithms that have been tested (Oswald et al., 2007; J. Carretta, pers. comm.; Appendix B). To date, most analysis of acoustic species identity using ROCCA has been post-cruise processed from acoustic recordings.

The SWFSC has used ROCCA in real-time during shipboard field surveys to a limited extent. ROCCA was used in real-time to determine the species identity of false killer whales, *Pseudorca crassidens,* during two SWFSC surveys in 2005 (Barlow and Rankin, 2007) and 2010. ROCCA was implemented in the acoustics protocol in a 2006 survey solely for whistle data collection. ROCCA is currently being restructured to improve accuracy and ease of use in real-time applications by automating several of the features. The value of ROCCA in real-time applications depends on its accuracy when utilized in the field. Correct classification results provided by Oswald et al. (2007) were based on post-cruise processing of whistles selected using standardized random methods. It is unclear if these classification results can be applied to real-time applications. There are at least two differences between real-time (RT) and post-cruise (PC) application of ROCCA that may affect correct classification: whistle selection methods and whistle contour extraction accuracy.

Standardized post-cruise analysis methods allow for careful whistle selection and contour extraction to enable accurate whistle measurements and species classification. In real-time applications, whistle selection is performed by a field technician, and the number and type of whistles selected may vary depending on any number of factors. There is an inherent bias in the whistle selection, and this bias may vary by field technician and by encounter. In fact, for this study, whistles selected by one field technician were specifically chosen if they were considered difficult for ROCCA to correctly classify (S. Rankin, pers. comm.). There is growing evidence that some whistles may be easier to classify to species (J. Oswald, pers. comm.), and therefore the correct classification of a group may vary based on the specific whistles chosen for analysis. Whistle variables used as input to ROCCA's classification algorithms are measured from contours extracted using TRIA. Changes in the accuracy of whistle contour extraction may affect the whistle measurements, and therefore the species classification. In real-time applications, an accurate whistle contour extraction may not be possible due to certain limitations, including time, animal behavior, and equipment issues.

In this report, we examine the effects of whistle selection methods and whistle contour and measurement accuracy on the classification of whistles collected during a survey in the eastern tropical Pacific Ocean (ETP). During this shipboard survey, whistles were selected in real-time for ROCCA without regard to the accuracy of the whistle contour extraction and then re-analyzed using post-cruise methods to allow for precise contour extractions. A comparison of the results is presented. Whistles from these same dolphin encounters were also selected based on the standardized post-cruise methods. A comparison of the classification results for the whistles selected in real-time and using the standardized post-cruise methods were examined to determine the effect of whistle selection on correct classification. The purpose of this study is to identify how whistle selection, whistle measurements, and overall analysis methods influence ROCCA's species classification results.

METHODS

Shipboard Survey

The Stenella Abundance Research survey (STAR) 2006 was a combined visual and acoustic survey of marine mammals in the ETP, running from 28 July to 7 December 2006 on the NOAA research vessel R/V *McArthur II*. Visual observations followed standard SWFSC protocol (Kinzey et al., 2000), using two visual observers on 25 X 150 'big eye' binoculars and one visual observer as a data-recorder and scanning the field with naked eye and hand-held binoculars. All dolphin groups presented in this report were identified to species using visual methods.

Acoustic monitoring followed standard SWFSC protocol (Rankin et al., 2008). A 2-element hydrophone array (500 Hz – 40 kHz ± 5 dB at -150 dB *re* 1 V/mPa) was towed at a distance of 200-300 m behind the ship at a depth of 8-11 m while traveling at approximately 10 knots during daylight hours. The array was monitored by the primary acoustics team aurally via headphones and visually using a scrolling real-time spectrograph display (ISHMAEL, Mellinger, 2001). Whistles from dolphin schools were selected for classification using ROCCA as time allowed.

Acoustic Species Classification (ROCCA)

Dolphin whistles were selected from a scrolling spectrograph (ISHMAEL, Mellinger, 2001, Fig. 1a) using two methods: real-time whistle selection and a standardized post-cruise whistle selection method. Whistles selected in real-time were chosen by field personnel from all whistling delphinid groups without concern for whistle quality. In an effort to provide a cross-section of a variety of whistle types, poor-quality whistles were at times purposefully selected (S. Rankin, pers. comm.). Poor-quality whistles are defined as having a low signal to noise ratio (SNR) or as overlapping in time and frequency with other vocalizations (Fig. 2c). Standardized post-cruise whistle selection methods included only single-species delphinid groups that were visually detected and were at least one nautical mile from other groups. Standardized whistle selection methods were based on methods used in Oswald et al. (2007), but more restrictive standards were placed on each selected whistle (A. Rudd, pers. comm.). Only good-quality whistles were selected, which are defined as having high SNR, where no part of the whistle appeared to be masked by background noise and there were few or no overlapping vocalizations (Fig. 2a).

The selected whistles were saved as individual wave files (Fig. 1b). ROCCA was then used to classify the signals using a Matlab window opened through ISHMAEL. ROCCA opened the saved wave file and automatically extracted a contour of the selected whistle by stepping through the saved wave file, one FFT (fast Fourier transform, Charif et al., 1995) at a time, using TRIA. The fundamental frequency of the whistle contour was extracted based on the peak frequency in each window (Fig. 1c). A routine within ROCCA ensured that random transient peaks in the spectrum were not mistaken for the fundamental peak frequency. In ROCCA, manual adjustments affecting the extracting algorithms sensitivity to noise *may* be made by the user to ensure that the contour is accurately extracted (Fig. 1d). The contour extractions of whistles selected in real-time were not manually adjusted in real-time, and the accuracy of the whistle contour measurements was not assessed. Post-cruise analysis of whistles included manual adjustments to provide an accurate extraction of the whistle contour. Analysis of whistles required the accurate selection of each whistle's starting point and careful scrutiny of the whistle contour (Fig. 1d; Fig. 2b), including the accurate measurement of steps and inflection points (Fig. 1e). Whistles selected in both real-time and post-cruise were carefully measured using ROCCA during post-cruise analysis to maximize accuracy in the whistle contour extraction and the measurement of whistle variables. If a whistle's contour was poorly extracted and inaccurate, even after manual adjustments (Fig. 2d), the whistle was discarded.

If a good contour trace was achieved (Fig. 1d; Fig. 2b), 55 variables were automatically measured, including slopes, frequencies, steps, and positions and numbers of inflection points (Fig. 1e; Appendix A). These variables were then classified using a random-forest algorithm, which resulted in a species identification for the whistle. ROCCA maintains a running tally of species predictions for a single school of dolphins. Once all whistles in a school have been analyzed, the overall school is classified based on the species that received the highest number of tree votes in the random-forest analysis.

Random-forest Analysis

A random-forest model was created to classify whistles for each trial. A training dataset was created from past SWFSC surveys, including cruises in the ETP and off the west coast of the

United States. This dataset contains dolphin whistle variable measurements obtained using standardized post-cruise methods in ROCCA. This training dataset included 1,997 whistles from 135 detections. All whistles were identified to species via visual sighting confirmation. The training dataset does not include data from the STAR 2006 survey. Results from the training dataset are provided in Appendix B.

Random-forest is an extension of the classification and regression tree (CART) method of Breiman et al. (1984) and creates multiple classification trees (= a "forest"), which are used to provide a consensus prediction for novel input data. The training dataset was used to construct the random-forest model containing 500 classification trees in the predictive forest. Each tree used a randomly-selected two-thirds of the whistles in the training dataset. The remaining one-third of whistles not used to build the forest were then used for classification. Each tree produced its own error rate, and an average error rate for all 500 trees in the model was obtained (Appendix B).

The training dataset was adjusted to provide the most accurate results. Due to the similarities in whistle characteristics, we combined the two common dolphin species, *Delphinus delphinus* and *Delphinus capensis*, into a single generic category called "Common dolphin" or "Cd". Certain outlier species were removed, such as *Peponocephala electra* and *Lagenodelphis hosei*, due to insufficient whistle samples necessary to create each classifier. Two additional hybrid variables, 'ratio.step.duration' and 'ratio.inflection.duration' (Appendix A), were also created for the analysis, providing 57 total variables used for classification.

Three separate datasets of whistle variable measurements were analyzed using the random-forest model:

Trial A: whistles *selected* in real-time and *measured* in real-time

Trial B: whistles *selected* in real-time and *measured* post-cruise

Trial C: whistles *selected* post-cruise and *measured* post-cruise

Only whistles from acoustic detections existing in both the real-time and post-cruise datasets were used in the random-forest analysis to maintain consistency. Each whistle was classified 500 times, once by each predictive tree in the forest. Each tree classified every whistle as a certain species. For each group, the species with the highest percentage of classifications was the predicted species. A confusion matrix of classification percentages was produced for each trial after the random-forest analysis.

Variable importance was assessed within random-forest through a routine that randomized (swapped) variable values between records. Variables were randomized one at a time, trees were built from the randomized data, and out-of-bag error rates (the error rate associated with the 1/3 of data not used to build individual trees) were generated for the forest model. Variables were then 'ranked' by importance, with the 'most important' variables represented by the greatest decline in predictive performance under the condition of randomization.

RESULTS

The acoustics team monitored 9,241 km of trackline over 100 days of effort during the 2006 STAR survey, with a total of 774 acoustic detections of cetaceans (Fig. 3, Rankin et al., 2008a). One hundred fifty-five dolphin groups, producing 971 whistles, were classified using whistles selected in real-time; 19 dolphin groups, producing 243 whistles, were classified using whistles selected in post-cruise analysis. Only acoustic detections of dolphin groups that occurred in both the real-time and post-cruise whistle datasets were included in the analysis (n=12 dolphin groups, Table 1; Fig. 3).

Classification results for the random-forest training dataset are presented in Appendix B. The training data resulted in an overall correct classification score of 65.0% for all 500 trees. This is significantly higher than the 12.5% expected based on chance alone for eight species. Whistles from false killer whales had the highest correct classification score (93.4%) and whistles from common dolphin species (73.4%) had the second highest score. Whistles from rough-toothed dolphins, *Steno bredanensis,* (63.9%) and spotted dolphins, *Stenella attenuata,* (63.7%) had similar scores. Whistles from bottlenose dolphins, *Tursiops truncatus,* (43.2%) and spinner dolphins, *Stenella longirostris,* (37.7%) produced the lowest correct classification scores. A high percentage of striped dolphin whistles, *Stenella coeruleoalba,* (30.1%) and spinner dolphin whistles (17.9%) were misclassified as common dolphins.

For Trial A (RT selection, RT measurement), an overall correct classification score of 33.9% was obtained (Table 2). Whistles from false killer whales, (72.2%) received the highest classification score, followed by rough-toothed dolphins (55.6%). Whistles from striped dolphins (25.4%), spotted dolphins (25.0%), and spinner dolphins (20.8%), produced the lowest correct classification scores. A high percentage of whistles from spotted dolphins (25.0%) and striped dolphins (18.6%) were misclassified as common dolphin.

Trial B (RT selection, PC measurement), resulted in an overall correct classification score of 36.4% (Table 3). Whistles from false killer whales (72.2%) had the highest correct classification score and spotted dolphins (62.5%) had the second highest correct classification score, followed by rough-toothed dolphins (55.6%). Whistles from spinner dolphins (25.0%) and striped dolphins (23.7%) received the lowest classification scores. A high percentage of whistles from spotted dolphins (25.0%) and spinner dolphins (23.7%) were misclassified as common dolphin.

For Trial C (PC selection, PC measurement), an overall correct classification score of 17.7% was obtained (Table 4). Whistles from false killer whales (86.5%) received the highest correct classification score, and rough-toothed dolphins (20.0%) received the second highest score. Whistles from spotted dolphins (7.7%) and striped dolphin (1.7%) received the lowest scores overall. The majority of whistles from spinner dolphins (100.0%), striped dolphins (98.3%), and spotted dolphins (92.3%) were misclassified as common dolphins in Trial C.

Species classification for each group in Trial C consisted mostly of common dolphin. Every tie that occurred between species classifications in Trials A and B included common dolphin as well (Table 1).

The importance of each variable was assessed and plotted (Fig. 4). The 30 most important variables used in ROCCA were determined after the randomization of variable values. The top three most important variables included maximum frequency, mean frequency, and center frequency.

DISCUSSION

The random-forest analysis, consisting of five hundred trees built with the training dataset, resulted in an overall correct classification score of 65%. The correct classification score expected by chance is 12.5% (for eight species given in Appendix B). The random-forest (Appendix B) produced more accurate classification results for individual dolphin whistles than the DFA and CART methods previously used in Oswald et al. 2007. Correct classification scores were greatly improved using the random-forest method for several species, including false killer whales, short-finned pilot whales, common dolphins, spinner dolphins, spotted dolphins, and striped dolphins. This random-forest training dataset was used to evaluate the effects of whistle measurement accuracy and whistle selection methods for this study.

The effect of whistle measurement accuracy on acoustic species identification can be seen by comparing Trials A and B (Table 5). Correct classification scores were very similar in both trials for several species, including false killer whales, rough-toothed dolphins, and striped dolphins. These results suggest that the additional effort applied during the post-cruise measurement of whistles in Trial B may not improve the correct classification scores for most species. One noticeable difference between real-time and post-cruise analysis methods for whistles selected in real-time occurred for spotted dolphins. Correct classification scores spotted dolphins increased from Trial A (25.0%) to Trial B (62.5%). The spotted dolphin whistles measured in Trials A and B were good-quality whistles, ruling out whistle quality as a factor of poor classification scores using real-time methods. For this species, it is possible that important whistle characteristics may be inaccurately measured in real-time due to what appear to be small differences in the initial whistle trace. Cursory examination of these whistles suggests that the end and maximum frequency, and possibly the number of steps, may be critical whistle characteristics that must be accurately measured for correctly classifying spotted dolphins (Fig. 5). Studying specific characteristics of dolphin whistles may also clarify why many individual whistles and dolphin groups were misclassified as common dolphin.

The effect of the whistle selection method on acoustic species identification can be seen by comparing Trials B and C (Table 5). Overall correct classification scores decreased between whistles selected in real-time (Trial B; 36.4%) and whistles selected in post-cruise (Trial C;

17.7%). Only false killer whale whistles showed improved scores between Trial B (72.2%) and Trial C (86.5%). Post-cruise selection of whistles for the 2006 survey included only whistles with a high SNR for *all* parts of the whistle. These data suggest that whistle selection methods may influence our ability to accurately identify most species using dolphin whistles.

During SWFSC surveys, after the initial visual detection of a dolphin school, the course and speed of the survey vessel are often altered to approach the animals for confirmation of species identity and group size estimation. This change in the vessels track to a ‘chase’ mode may lead to changes in behavior of dolphins nearby. For example, in the ETP, dolphins of the genus *Stenella* (spotted, spinner, and striped dolphins) typically exhibit evasive behaviors, often swimming rapidly away from the ship. Changes could also affect vocal behavior. For example, whistles with a high SNR are often recorded during the final close approach of the ship to the group of animals. It is possible that, for many species, these whistles are not representative of their normal whistle repertoire or behavior. Whistles used in the training dataset were chosen from a wider range of SNRs and may contain a more representative sample of normal whistle repertoires than the whistle dataset selected for post-cruise analysis. Our data suggest that only selecting ‘good quality’ whistles with a high SNR may create a bias that negatively affects ROCCA’s ability to correctly classify many species. Future studies should include all whistles that have a sufficiently high SNR for whistle contour extraction in ROCCA.

In general, the accuracy of ROCCA whistle measurements made in real-time may be sufficient for the identification of most species (Trial A). However, spotted dolphin classification scores improved with post-cruise whistle measurements (Trial B). Unfortunately, the small sample size in this study is insufficient to explain these improved scores, which may be due to inaccurate measurement of specific variables. Of the 57 variables measured, only a subset of variables is important in classifying dolphin whistles, and variable importance is likely to vary by species (Fig. 4). Analyzing a larger sample of dolphin whistles would allow us to identify variables that are more likely to impact classification results.

The comparison of whistle selection methods (Trial B versus Trial C) suggests that it may be possible to relax the strict whistle selection methods applied during post-cruise analysis. In fact, the post-cruise methods for the 2006 dataset (Trial C) were more restrictive than the methods used to create the training dataset presented in Appendix B, where the overall correct classification score was 65%. Future studies should examine the effects of chase on vocal behaviors of delphinids. A larger dataset needs to be examined to study any changes in whistle characteristics due to chase and the effects that these changes have on ROCCA species identification. In addition, not all of the species included in ROCCA were present in the STAR 2006 dataset. As a result, nothing is known of the effects of whistle selection or whistle extraction methods on classification of whistles produced by common dolphins, short-finned pilot whales, and bottlenose dolphins. It would be beneficial to analyze a dataset of whistles that includes all eight of the species present in the training dataset.

The results of this study suggest that selecting and measuring whistles using real-time methods may result in correct classification scores that are comparable to, if not better than, pure post-cruise methods. This study was based on a small selection of whistles from 12 dolphin schools and a larger dataset should be examined to confirm these results. Although it is possible that the small sample sizes for the STAR 2006 cruise data may be responsible for the extremely poor classification scores obtained during Trial C, we are optimistic that the results obtained using real-time selection and measurement methods are reasonable. Examination of a large dataset of whistles measured in real-time and post-cruise will provide information on the critical variables necessary for correct acoustics species identification and will identify situations in which post-cruise analysis of whistles may be necessary. Likewise, a large dataset should be examined to determine if the correct classification of whistles is affected by the approach of the ship, and if so, the point at which whistles should not be considered useable for classification. All analyses should be conducted by species, and possibly by geographic region. Improvements in ROCCA based on the results of these studies will allow ROCCA to be reliably used to determine acoustic species identity of whistling dolphins in real-time applications.

ACKNOWLEDGEMENTS

This study could not have been performed without the help of the officers and crew of the NOAA ship R/V *McArthur II*. Special thanks to the scientists and Lisa Ballance, the principal investigator for the STAR 2006 survey, for their cooperation. Liz Zele provided valuable acoustic assistance during the STAR 2006 survey. This manuscript was improved by helpful reviews by Tom Norris and Susannah Calderan. Funding for this work was provided by the Navy (N45) and the Joint Industry Program.

LITERATURE CITED

Barlow, J., and B. Taylor. 2005. Estimates of sperm whale abundance in the northeastern temperate Pacific from a combine acoustic and visual survey. Marine Mammal Science. 21(3): 429-445.

Barlow, J., and S. Rankin. 2007. False killer whale abundance and density: Preliminary estimates for the PICEAS study area south of Hawaii and new estimates for the US EEZ around Hawaii. NOAA Administrative Report LJ-07-02, available from Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037, USA.

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

Breiman, L., J. Friedman, R. Olshen, and C. Stone. 1984. Classification and regression trees. Wadsworth, Pacific Grove, CA. 358p.

Caldwell, M. C., and D. K. Caldwell. 1967. Intraspecific transfer of information via the pulsed sound in captive odontocete cetaceans. In R. G. Busnel (Ed.), Animal sonar systems, biology and bionics. Jouyen- Josas, France: Laborative de Physiologie: 879-936.

Charif, R. A., Mitchell, S. & Clark, C. W. 1995: Canary 1.2 User's Manual. Appendix B: A Biologist’s Introduction to Spectrum Analysis. Cornell Laboratory of Ornithology, Ithaca, NY.

Goold, J. C., and S. E. Jones. 1995. Time and frequency domain characteristics of sperm whale clicks. J. Acoust. Soc. Am. 98(3): 1279-1291.

Jackson, A., T. Gerrodette, S. Chivers, M. Lynn, S. Rankin, S. Mesnick. 2008. Marine mammal data collected during a survey in the Eastern Tropical Pacific Ocean aboard NOAA ships David Starr Jordan and McArcthur II, July 28 – December 7, 2008. NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-421. 54p.

Kinzey D., P. Olson, and T. Gerrodette. 2000. Marine mammal data collection procedures on research ship line-transect surveys by the Southwest Fisheries Science Center. NOAA Administrative Report LJ-00-07C, available from Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037, USA.

Kohavi, R., F. Provost. 1998. "*Glossary of Terms.*" In: Editorial for the Special Issue on Applications of Machine Learning and the Knowledge Discovery Process, Machine Learning 30: 2-3.

Lammers, M.O., W.W. L. Au, and D.L. Herzing. 2003. The broadband social acoustic signaling behavior of spinner and spotted dolphins. J. Acoust. Soc. Am. 114(3): 1629-1639.

Mellinger, D.K. 2001. ISHMAEL 1.0 User's Guide: Integrated System for Holistic Multi-channel Acoustic Exploration and Localization. NOAA Technical Memorandum OAR-PMEL-120.

Oswald, J.N., S. Rankin, J. Barlow, and T.F. Norris. 2003. Acoustic identification of nine delphinid species in the Eastern Tropical Pacific Ocean. Mar. Mam. Sci. 19(1):20-37.

Oswald, J.N., S. Rankin, J. Barlow, and M.O. Lammers. 2007. A tool for real-time acoustic species identification of delphinid whistles. J. Acoust. Soc. Am. 122(1):587-595.

Oswald, J.N., T. Norris, and R. Sousa-Lima. In review. A review of methods for the automated detection, extraction, and classification of marine mammal calls. Aquatic Mammals.

Rankin, S., and J. Barlow. 2005. Source of the North Pacific "boing" sound attributed to minke whales. Acoust. Soc. Am. 118(5): 3346-3351.

Rankin, S., J. Barlow, J.N. Oswald, and L. Ballance. 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. NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-429. 58p.

Roch, M., M.S. Soldevilla, J.C. Burtenshaw, E.E. Henderson, and J. A. Hildebrand. 2006. Gaussian mixture model classification of odontocetes in the southern California bight and the gulf of California. J. Acoust. Soc. Am. 121(3):1737-1748.

Soldevilla, M.S., E.E. Henderson, G.S. Campbell, S.M. Wiggins, J.A. Hildebrand, and M.A. Roch. 2008. Classification of Risso's and Pacific white-sided dolphins using spectral properties of echolocation clicks. J. Acoust. Soc. Am. 124(1): 609-624.

Watkins, W. A. 1980. Acoustics and the behavior of sperm whales. In R.-G. Busnel and J. F. Fish, eds. Animal sonar systems. Plenum Publishing Corp., New York, NY. 283-290.

TABLES

Table 1. Summary of the 12 dolphin groups included in the analysis and their group species identification based on individual whistles measured and selected in real-time (RT) and post-cruise (PC). Groups identified as multiple species did not obtain a single majority of classified whistles. Species codes are as follows: *Stenella attenuata* (Sa), *Stenella longirostris* (Sl), *Stenella coeruleoalba* (Sc), *Steno bredanensis* (Sb), *Tursiops truncatus* (Tt), *Pseudorca crassidens* (Pc), and combined common dolphin species *Delphinus delphis* and *Delphinus capensis* (Cd).

						Acoustic Species ID		
Date	Latitude	Longitude	# RT Whistles	# PC Whistles	Actual Species	Trial A	Trial B	Trial C
6-Sep-06	8.82	-148.16	2	3	Sl	Tt	Tt	Cd
6-Sep-06	8.81	-147.98	20	49	Sc	Tt	Tt	Cd
6-Sep-06	8.75	-147.86	3	11	Sb	Sb	Cd/Pc/Sb	Cd
7-Sep-06	8.01	-145.55	10	20	Sc	Sb	Sb	Cd
10-Sep-06	4.75	-135.49	18	37	Pc	Pc	Pc	Pc
14-Sep-06	1.08	-122.86	13	34	Sc	Sc	Sc	Cd
16-Oct-06	-2.82	-96.65	6	8	Sc	Cd/Pc/Sc	Cd	Cd
22-Oct-06	7.71	-97.87	10	9	Sc	Cd/Sc	Sa	Cd
28-Oct-06	4.48	-112.85	10	1	Sl	Sc	Cd	Cd
28-Oct-06	4.59	-113.03	12	5	Sl	Sl	Sl	Cd
31-Oct-06	13.13	-108.68	6	14	Sb	Sb	Sb	Cd/Pc/Sb
17-Nov-06	16.08	-98.69	8	52	Sa	Cd/Sa/Sc/Tt	Sa	Cd

Table 2. Confusion matrix of classification percentages for Trial A - individual whistles selected and measured in real-time. Predicted species (1st row) and observed species (1st column) are displayed. The number of whistles included in the analysis for each species (*n*) is given in the last column. 'NA' represents species that were not present in the novel data. Correct classifications are shown in bold. The overall correct classification was 33.9%.

	% Classified as								
Actual Species	False killer whale	Common dolphin	Rough-toothed dolphin	Spotted dolphin	Short-finned pilot whale	Striped dolphin	Bottlenose dolphin	Spinner dolphin	*n*
False killer whale	**72.2**	0.0	16.7	0.0	11.1	0.0	0.0	0.0	*18*
Common dolphin	NA	NA	NA	NA	NA	NA	NA	NA	*0*
Rough-toothed dolphin	22.2	0.0	**55.6**	0.0	11.1	11.1	0.0	0.0	*9*
Spotted dolphin	0.0	25.0	0.0	**25.0**	0.0	25.0	25.0	0.0	*8*
Short-finned pilot whale	NA	NA	NA	NA	NA	NA	NA	NA	*0*
Striped dolphin	6.8	18.6	25.4	3.4	1.7	**25.4**	15.3	3.4	*59*
Bottlenose dolphin	NA	NA	NA	NA	NA	NA	NA	NA	*0*
Spinner dolphin	0.0	8.3	8.3	8.3	0.0	29.2	25.0	**20.8**	*24*

Table 3. Confusion matrix of classification percentages for Trial B - individual whistles selected in real-time and measured in post-cruise analysis. Predicted species (1st row) and observed species (1st column) are displayed. The number of whistles included in the analysis for each species (*n*) is given in the last column. 'NA' represents species that were not observed in the novel data. Correct classifications are shown in bold. The overall correct classification was 36.4%.

	% Classified as								
Actual Species	False killer whale	Common dolphin	Rough-toothed dolphin	Spotted dolphin	Short-finned pilot whale	Striped dolphin	Bottlenose dolphin	Spinner dolphin	*n*
False killer whale	**72.2**	0.0	11.1	0.0	16.7	0.0	0.0	0.0	*18*
Common dolphin	NA	NA	NA	NA	NA	NA	NA	NA	*0*
Rough-toothed dolphin	11.1	22.2	**55.6**	0.0	11.1	0.0	0.0	0.0	*9*
Spotted dolphin	0.0	25.0	0.0	**62.5**	0.0	12.5	0.0	0.0	*8*
Short-finned pilot whale	NA	NA	NA	NA	NA	NA	NA	NA	*0*
Striped dolphin	3.4	16.7	15.3	15.3	0.0	**23.7**	16.9	1.7	*59*
Bottlenose dolphin	NA	NA	NA	NA	NA	NA	NA	NA	*0*
Spinner dolphin	0.0	23.7	0.0	16.7	0.0	16.7	25.0	**25.0**	*24*

Table 4. Confusion matrix of classification percentages for Trial C - individual whistles selected and measured in post-cruise analysis. Predicted species (1st row) and observed species (1st column) are displayed. The number of whistles included in the analysis for each species (*n*) is given in the last column. 'NA' represents species that were not observed in the novel data. Correct classifications are shown in bold. The overall correct classification was 17.7%.

	% Classified as								
Actual Species	False killer whale	Common dolphin	Rough-toothed dolphin	Spotted dolphin	Short-finned pilot whale	Striped dolphin	Bottlenose dolphin	Spinner dolphin	*n*
False killer whale	**86.5**	8.1	0.0	0.0	2.7	2.7	NA	NA	*37*
Common dolphin	NA	NA	NA	NA	NA	NA	NA	NA	*0*
Rough-toothed dolphin	16.0	44.0	**20.0**	0.0	0.0	20.0	NA	NA	*25*
Spotted dolphin	0.0	92.3	0.0	**7.7**	0.0	0.0	NA	NA	*52*
Striped dolphin	0.0	98.3	0.0	0.0	0.0	**1.7**	NA	NA	*120*
Short-finned pilot whale	NA	NA	NA	NA	NA	NA	NA	NA	*0*
Bottlenose dolphin	NA	NA	NA	NA	NA	NA	NA	NA	*0*
Spinner dolphin	0.0	100.0	0.0	0.0	0.0	0.0	NA	NA	*9*

Table 5. Correct classification scores for individual whistles measured in real-time and post-cruise for each species. 'NA' represents species that were not observed in the novel data.

	% Correct Classification Scores		
Species	Trial A	Trial B	Trial C
False killer whale	72.2 (*18*)	72.2 *(18)*	86.5 *(37)*
Common dolphin	NA	NA	NA
Rough-toothed dolphin	55.6 *(9)*	55.6 *(9)*	20.0 *(25)*
Spotted dolphin	25.0 *(8)*	62.5 *(8)*	7.7 *(52)*
Short-finned pilot whale	NA	NA	NA
Striped dolphin	25.4 *(59)*	23.7 *(59)*	1.7 *(120)*
Bottlenose dolphin	NA	NA	NA
Spinner dolphin	20.8 *(24)*	25.0 *(24)*	0.0 *(9)*
Overall Correct Classification	33.9 *(118)*	36.4 *(118)*	17.7 *(243)*

FIGURES

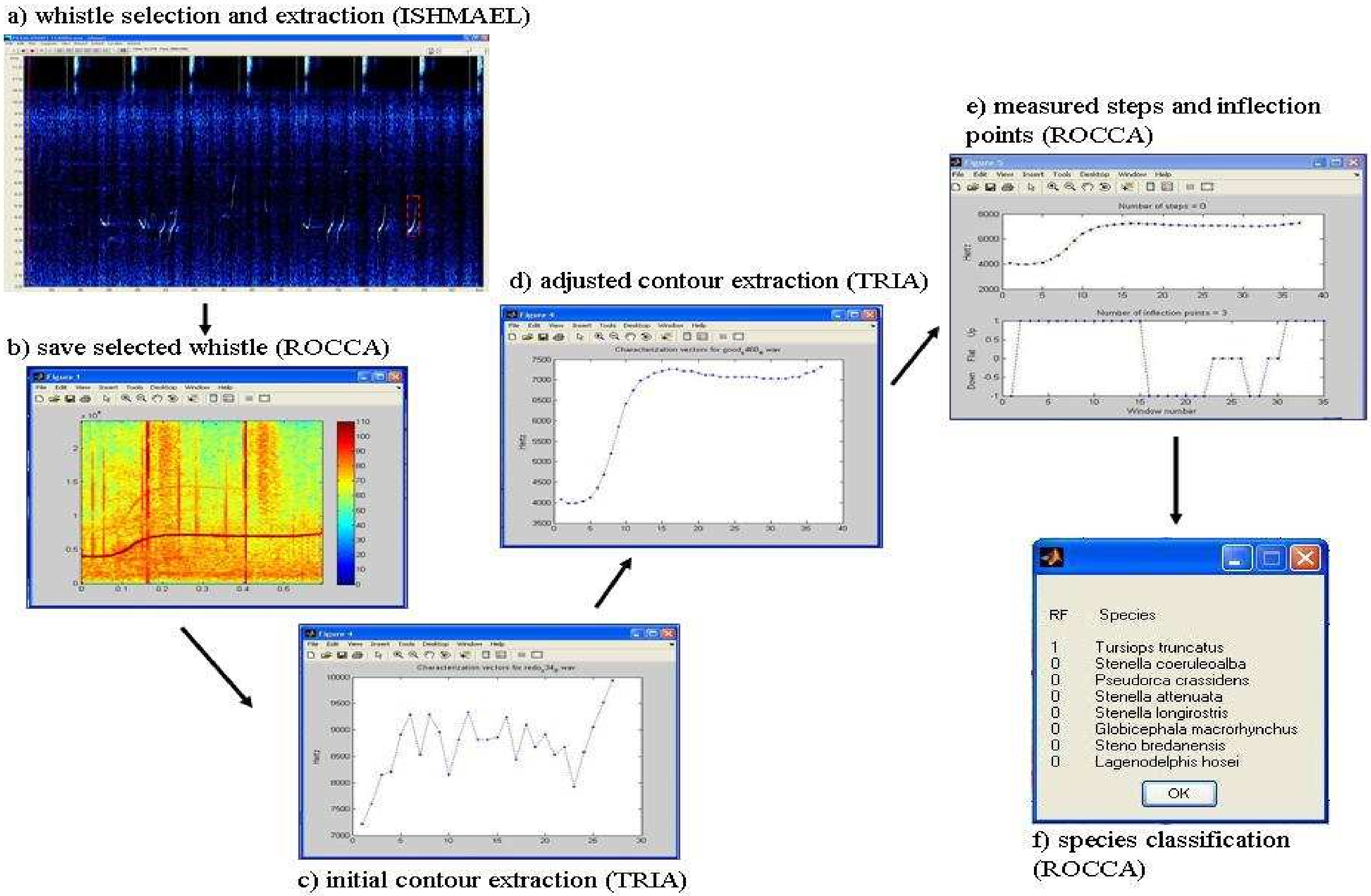

Figure 1. The flow chart depicts how a dolphin whistle is identified to species using ROCCA. The whistle is selected from the real-time spectrogram (ISHMAEL, Mellinger 2001) (a), saved as a wave file and sent to ROCCA in Matlab (b). The fundamental time-frequency whistle contour is extracted and does not always match the actual whistle (c). The contour can be adjusted and extracted multiple times until it closely resembles the original whistle (d). 55 variables are measured from the whistle contour, including the number of steps and inflection points, (e), and the whistle is classified to species and added to the running tally of species identifications for that school (f). In real-time, the initial whistle extraction was not adjusted to improve its accuracy (d).

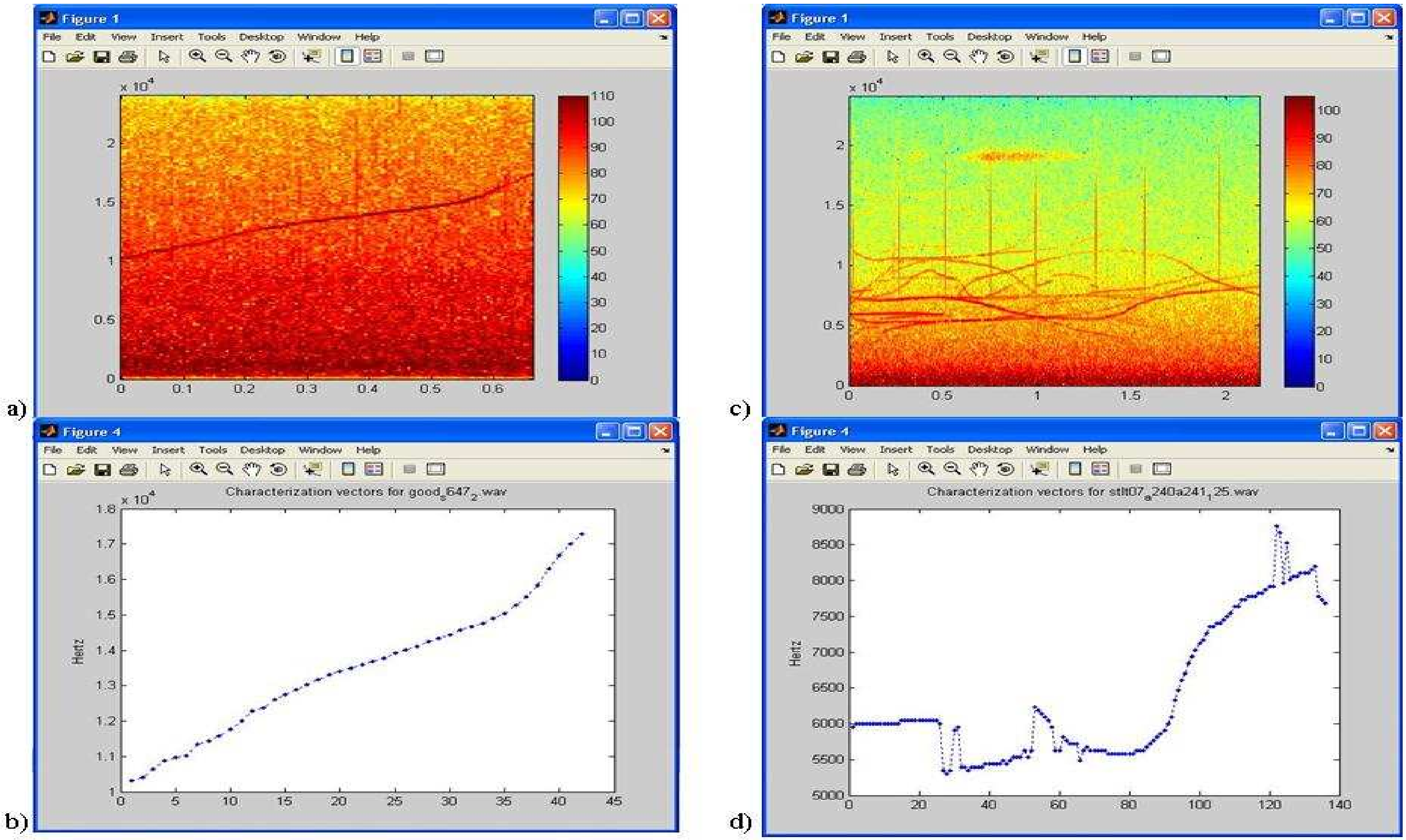

Figure 2. A good quality whistle is selected for ROCCA (a) and a good contour trace is obtained (b). When a poor quality whistle is selected (c), it is more likely that a poor contour trace will be produced that does not fit the selected whistle (d).

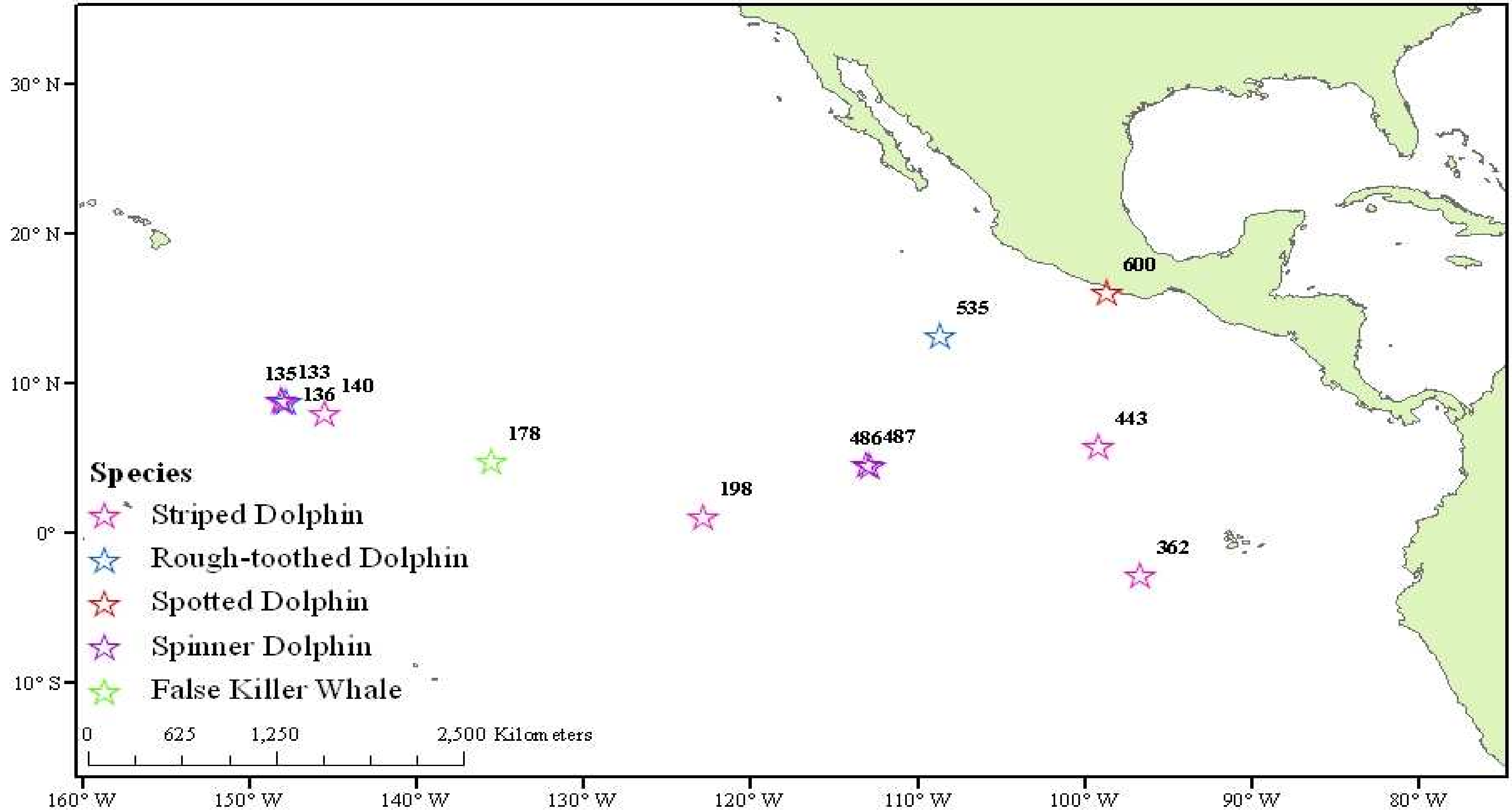

Figure 3. Map of the acoustic detections included in this study from STAR 2006. Sighting number is provided next to the sighting location on the map.

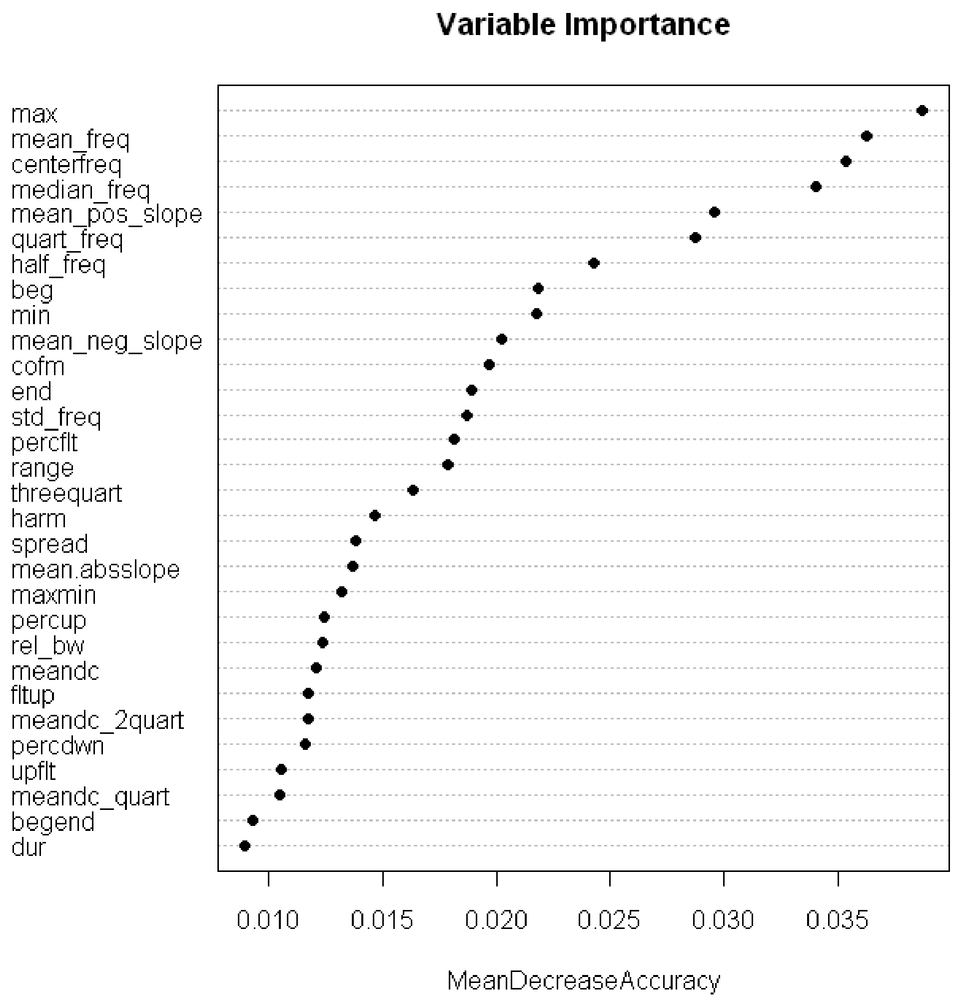

Figure 4. Plot showing the 30 most important variables in ROCCA's random-forest classifier, starting with the most important variable at the top. Variables are given on the y-axis, and the relative importance (shown as a Mean Decrease Accuracy after variable randomization) is given on the x-axis. Full description of the variables are given in Appendix A.

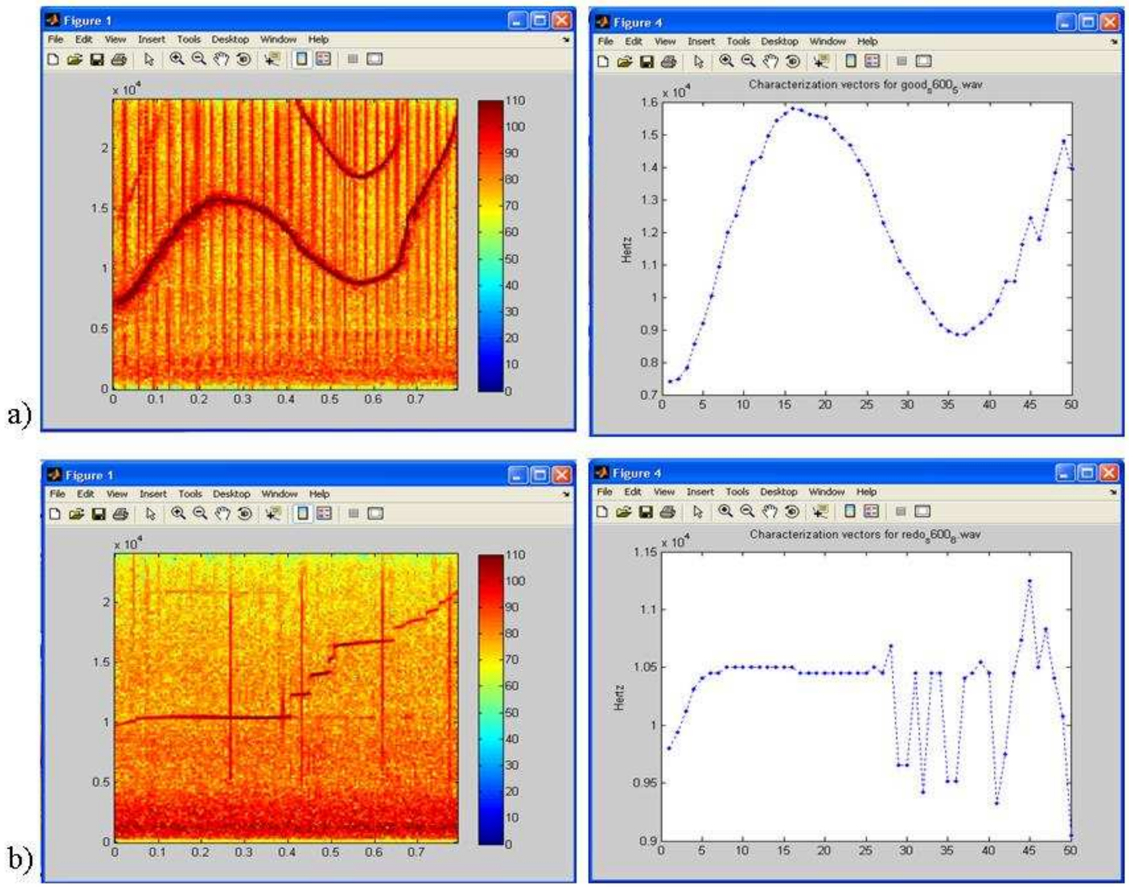

Figure 5. A good quality spotted dolphin whistle chosen for ROCCA (a) produced an initial extraction contour with an inaccurate end frequency. A second good quality spotted dolphin whistle (b) produced an initial extraction contour with inaccurate maximum and end frequencies as well as the number of steps. Despite the high SNR of each whistle, they were misclassified indicating that certain whistle characteristics may prove more critical when classifying spotted dolphins.

APPENDIX A
Table displaying the 57 whistle variables measured in ROCCA. Measurements for each variable were used in the random-forest analysis to classify each dolphin whistle.

Full Name of Variable (in order of measurement taken by ROCCA)	Abbreviated Name of Variable	Explanation of Variable
mean duty cycle	meandc	mean duty cycle (proportion of time signal 'on' vs 'off')
mean duty cycle of first quarter	meandc_quart	mean duty cycle of the first quarter of the whistle
mean duty cycle of second quarter	meandc_2quart	mean duty cycle of the second quarter of the whistle
mean duty cycle of third quarter	meandc_3quart	mean duty cycle of the third quarter of the whistle
mean duty cycle of fourth quarter	meandc_4quart	mean duty cycle of the fourth quarter of the whistle
standard deviatio of duty cycle	dc_std	standard deviation of the duty cycle
mean frequency (Hz)	mean freq	mean frequency (Hz)
median frequency (Hz)	median freq	median frequency (Hz)
standard deviation of the frequency (Hz)	std freq	standard deviation of the frequency (Hz)
Spread	spread	difference between the 75th and the 25th percentiles of the frequency
quarter frequency	quart freq	frequency at one quarter of the duration (Hz)
half frequency	half freq	frequency at one half of the duration (Hz)
three-quarter frequency	threequart	frequency at three quarters of the duration (Hz)
center frequency	centerfreq	(minimum frequency+(maximum frequency-minimum frequency))/2
relative bandwidth	rel bw	(max freq - min freq)/center freq
maximum - minimum ratio	maxmin	maximum frequency / minimum frequency
beginning - end ratio	begend	beginning frequency / end frequency
coefficient of frequency modulation	cofm	20 equally spaced frequency measurements are taken, then each frequency value is subtracted from the one before it. The sum of the absolute values of these differences is calculated and divided by 10000.
beginning frequency	beg freq	beginning frequency of whistle (Hz)
end frequency	end freq	ending frequency of whistle (Hz)

Full Name of Variable (in order of measurement taken by ROCCA)	**Abbreviated Name of Variable**	**Explanation of Variable**
minimum frequency	min freq	minimum frequency of whistle (Hz)
maximum frequency	max	maximum frequency of whistle (Hz)
frequency range	range	maximum frequency - minimum frequency of whistle (Hz)
duration	duration	duration of whistle (seconds)
harmonics	harms	binary variable: 1=harmonics are present, 0=harmonics are absent
steps	steps	number of steps in the whistle (a step is defined as having a 10% or greater increase or decrease in frequency over 2 contour points)
inflections	inflect	number of inflection points in the whistle (changes from positive to negative or negative to positive slope)
up down	up dwn	number of inflection points in the whistle that go from positive slope to negative slope
down up	dwn up	number of inflection points in the whistle that go from negative slope to positive slope
up flat	up flat	number of times the slope changes from positive to zero
down flat	dwn flat	number of times the slope changes from negative to zero
flat down	flat dwn	number of times the slope changes from zero to negative
flat up	flat up	number of times the slope changes from zero to positive
step up	step up	number of steps that have increasing frequency
step down	step dwn	number of steps that have decreasing frequency
maximum delta	max delta	maximum time between inflection points
minimum delta	min delta	minimum time between inflection points
maximum - minimum delta ratio	maxmin delta	max delta/min delta
mean delta	mean delta	mean time between inflection points
standard deviation delta	std delta	standard deviation of the time between inflection points
median delta	median delta	median of the time between inflection points
mean slope	mean slope	overall mean slope
mean positive slope	mean pos	mean positive slope

Full Name of Variable (in order of measurement taken by ROCCA)	**Abbreviated Name of Variable**	**Explanation of Variable**
mean negative slope	mean neg	mean negative slope
mean absolute	mean abs	absolute value of the mean of the slope
positive - negative slope ratio	posneg	positive slope/negative slope
percent up	perc up	percent of the whistle that has a positive slope
percent down	perc dwn	percent of the whistle that has a negative slope
percent flat	perc flat	percent of the whistle that has zero slope
beginning sweep	begsw	categorical variable: 1=beginning slope is positive, -1=beginning slope is negative, 0=beginning slope is 0
beginning up	begup	binary variable: 1=beginning slope is positive, 0=beginning slope is negative
beginning down	begdwn	binary variable: 1=beginning slope is negative, 0=beginning slope is positive
end sweep	endsw	categorical variable: 1=ending slope is positive, -1=ending slope is negative, 0=ending slope is 0
end up	endup	binary variable: 1=ending slope is positive, 0=ending slope is negative
end down	enddwn	binary variable: 1=ending slope is negative, 0=ending slope is positive
ratio of steps to duration	ratio.step.dur	number of steps / duration
ratio of inflection points to duration	ratio.inflect.dur	inflection / duration

APPENDIX B

The training dataset was created using dolphin whistles from the following SWFSC surveys: HICEAS 2002, PICEAS 2005, ORCAWALE 2001, STAR 2000, and STAR 2003 (Rankin et al. 2008). Methods for whistle selection and measurement are given in Oswald et al. (2003). A random number generator was used to select 35 whistles from each detection to be used in ROCCA's classifier. If a detection contained fewer than 35 whistles, then all whistles were analyzed. A total of 1,997 whistles from 135 unique detections were included in the training dataset.

Table 6. Confusion matrix of classification percentages from the Random-forest analysis using, with predicted species (1st row) and observed species (1st column). The number of whistles included in the analysis for each species (*n*) is given in the last column. Correct classification scores are shown in bold. The overall out-of-bag correct classification is 65%.

	% Classified as								
Actual species	False killer whale	Common dolphin	Rough-toothed dolphin	Spotted dolphin	Short-finned pilot whale	Striped dolphin	Bottlenose dolphin	Spinner dolphin	*n*
False killer whale	**93.4**	0.7	1.1	0.4	1.1	2.2	0.7	0.4	*272*
Common dolphin	2.4	**73.4**	1.5	6.6	0.0	11.8	1.0	3.4	*594*
Rough-toothed dolphin	18.8	8.3	**63.9**	0.0	0.0	9.0	0.0	0.0	*133*
Spotted dolphin	2.1	12.1	0.3	**63.7**	0.3	11.8	6.2	3.5	*297*
Short-finned pilot whale	31.0	1.4	2.8	4.2	**57.7**	2.8	0.0	0.0	*71*
Striped dolphin	3.3	30.1	4.2	4.5	0.0	**53.0**	1.8	3.0	*332*
Bottlenose dolphin	5.2	15.5	3.2	20.6	0.0	7.7	**43.2**	4.5	*155*
Spinner dolphin	6.0	17.9	6.6	11.9	0.0	13.2	6.6	**37.7**	*151*

RECENT TECHNICAL MEMORANDUMS

SWFSC Technical Memorandums are accessible online at the SWFSC web site (http://swfsc.noaa.gov). Copies are also available form the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161 (http://www.ntis.gov). Recent issues of NOAA Technical Memorandums from the NMFS Southwest Fisheries Science Center are listed below:

NOAA-TM-NMFS-SWFSC-463 Spawning biomass of Pacific sardine *(Sardinops sagax)* off the U.S. in 2010.
N.C.H. LO, B.J. MACEWICZ, and D.A. GRIFFITH
(October 2010)

464 Ecosystem survey of *Delphinus* species cruise report.
S.J. CHIVERS, W.L. PERRYMAN, N.M. KELLAR, J.V. CARRETTA, F.I. ARCHER, J.V. REDFERN, A.E. HENRY, M.S. LYNN, C. HALL A. JACKSON, G. SERRA-VALENTE, T.J. MOORE, C. SURREY-MARSDEN, and L.T. BALLANCE
(October 2010)

465 Oregon, California and Washington line-transect and ecosystem (ORCAWALE) 2008 cruise report.
J. BARLOW, A.E. HENRY, J.V. REDFERN, T. YACK, A. JACKSON, C. HALL, E. ARCHER, and L.T. BALLANCE
(October 2010)

466 A forward-looking scientific frame of reference for steelhead recovery on the south-central and southern California coast.
D.A. BOUGHTON
(October 2010)

467 Some research questions on recovery of steelhead on the south-central and southern California coast.
D.A. BOUGHTON
(October 2010)

468 Is the September 1 river return date approximation appropriate for Klamath River fall Chinook?
M.R. O'FARRELL, M.L. PALMER-ZWAHLEN and J. SIMON
(November 2010)

469 Assessment of the Pacific sardine resource in 2010 for U.S. management in 2011.
K.T. HILL, N.C.H. LO, B.J. MACEWICZ, P.R. CRONE, and R. FELIX-URAGA
(December 2010)

470 AMLR 2009/10 Field Season Report.
A.M. VAN CISE, Editor
(December 2010)

471 Rationale for the 2010 revision of stock boundaries for the Hawai'i insular and pelagic stocks of false killer whales, *Pseudorca crassidens*.
K.A. FORNEY, R.W. BAIRD, and E.M. OLESON
(December 2010)

472 Historical occurrence of coho salmon *(Oncorhynchus kisutch)* in streams of the Santa Cruz Mountain region of California: response to an Endangered Species Act petition to delist coho salmon south of San Francisco Bay.
B.C. SPENCE, W.G. DUFFY, J.C. GARZA, B C. HARVEY, S.M. SOGARD, L.A. WEITKAMP, T.H. WILLIAMS, and D.A. BOUGHTON
(February 2011)