Acoustic Metadata Management and Transparent Access to Networked Oceanographic Data Sets

Marie A. Roch
Dept. of Computer Science, San Diego State University
5500 Campanile Drive, San Diego, CA 92182-7720
Phone: (619) 594-5830 FAX: (619) 594-6746 E-mail: marie.roch@sdsu.edu

Simone Baumann-Pickering
Scripps Institution of Oceanography
University of California at San Diego
9500 Gilman Dr., La Jolla, CA 92093-0205
Phone: (858) 534-7280 FAX: (858) 534-6849 E-mail: sbaumann@ucsd.edu

Catherine L. Berchok
Alaska Fisheries Science Center, National Marine Fisheries Service
National Oceanic and Atmospheric Administration, Department of Commerce
7600 Sand Point Way N.E., Building 4
Seattle, Washington 98115-6349
Phone: (206) 526-6331 FAX: (206) 526-6615 E-mail: Catherine.Berchok@noaa.gov

John A. Hildebrand
Scripps Institution of Oceanography
University of California at San Diego
9500 Gilman Dr., La Jolla, CA 92093-0205
Phone: (858) 534-4069 FAX: (858) 534-6849 E-mail: jhildebrand@ucsd.edu

Erin M. Oleson
Pacific Islands Fisheries Science Center, National Marine Fisheries Service
National Oceanic and Atmospheric Administration, Department of Commerce
1601 Kapiolani Blvd, Ste. 1110
Honolulu, Hawaii 996814
Phone: (808) 944-2172 FAX: (808) 941-0307 E-mail: Erin.Oleson@noaa.gov

Sofie M. Van Parijs
Northeast Fisheries Science Center, National Marine Fisheries Service
National Oceanic and Atmospheric Administration, Department of Commerce
166 Water Street, Woods Hole, MA 02543
Phone: (508) 495-2119 FAX: (508) 495-2258 E-mail: sofie.vanparijs@noaa.gov

Award Number: N00014-11-1-0697
http://cetus.ucsd.edu

LONG-TERM GOALS

The long-term goals of this effort are to produce software capable of organizing and archiving metadata associated with the detection of marine mammals. The software provides a workbench for researchers to integrate queries for both biological and physical data by providing mediation services to support queries that cover networked data sets in manners similar to those used for querying the acoustic data.

The data is to be accessible from a variety of languages used by the scientific community for analysis and modeling. Preservation of acoustic metadata can prevent the loss of detailed information that is not captured in scientific publications and will permit the analysis of decadal-scale data sets. The Tethys server at Scripps now has data sets that cover over five years, and we are beginning to see the first fruits of the preservation of detailed long time-series data.

OBJECTIVES

The objectives of this effort are to produce:

1. A database which can flexibly store multiple types of metadata derived from a variety of acoustic platforms, both stationary and mobile.
2. Standardization of methods to make the data repositories useful to the passive acoustic monitoring community.
3. Secure access on network platforms using industry standard security protocols.
4. Visualization primitives in selected analysis and modeling languages (e.g. Matlab, R).
5. Access methods for the above languages.
6. Primitives to query data both spatially and temporally in an efficient manner.
7. Demonstration projects to show the value of the database as a scientific workbench component.

APPROACH

1) Technical approach

The acoustic metadata database enables researchers to organize, store and most importantly query information derived from passive acoustic monitoring (PAM). Due to the large number of acquisition platforms, types of detection effort, etc., structuring these data is a complicated semi-structured task and traditional databases do not meet the needs of PAM users. By compiling a large team of PAM users who work on a global scale, we are defining data standards that are likely to meet the needs of the PAM community in general. Networking capabilities provide the ability to share data and eventually export summary data to OBIS-SEAMAP. In addition, this effort provides users with access to online physical oceanography databases using a single interface. The project also provides access methods for a variety of computer languages used for analysis by the scientific community.

2) Key Personnel

Dr. Marie A. Roch (San Diego State University) is the project manager and administrator for this project. She also takes the lead for software development.

Dr. John A Hildebrand (Scripps Institution of Oceanography (SIO)) is the project manager for the subaward to SIO.

Drs. Simone Baumann-Pickering (Scripps Institution of Oceanography), Catherine L. Berchok (NOAA Alaska Fisheries Science Center (AFSC)), Erin M. Oleson (NOAA Pacific Island Fisheries

Science Center (PIFSC)), Melissa Soldevilla (NOAA Southeast Fisheries Science Center (SEFSC)), and Sofie Van Parijs (NOAA Northeast Fisheries Science Center (NEFSC)), all represent data providers who will be using the database and are integrally involved in the operational specification, requirements, and testing.

Dr. Simone Baumann-Pickering is providing the lead on habitat modeling, and Dr. Sofie Van Parijs is the project manager for NOAA as well as the lead on data standardization.

3) Work plans for the upcoming year

Documentation of the application programming interface (API) for generating detection annotations in Tethys format will be written. A similar API will be developed to permit localization software to generate Tethys ready annotations. Final testing and deployment of a new transport layer (see below) will occur.

A new client for the R language will be developed. Continued work on indices to speed queries should continue to progress and be deployed in the final project year. Work on the security model has been delayed due to feedback from project participants that they preferred other functionality to be added first, but the security model will be considered in this next year. In addition, work has begun on a version of Tethys for 64 bit architectures which will address limitations in the processing of some environmental data queries that exceed the capacity of 32 bit address space. There has been an interest in improving our moving platform support for gliders, floats, etc. and better integration of track-lines will be a focus this year.

We will continue to advocate for Tethys use at conferences (e.g. San Francisco meeting of the Acoust. Soc. of Am.) and through mailing lists (e.g. MARMAM, BIOACOUSTICS-L). Tethys's web presence will be improved, incorporating elements from presentations and reports that show how and why Tethys can be beneficial.

WORK COMPLETED

Tethys had its first publicized public release this year. While improvements to the software are ongoing, we have begun advocating its use within the community. A preliminary barebones website has been established (http://tethys.sdsu.edu) that allows access to the software and documentation. An 84 page user manual describes installation, administration, and use of Tethys, and a separate 22 page "cookbook" manual describes how to perform common Tethys tasks within Matlab. A sample database was assembled from a subset of detections on high frequency acoustic recording packages (HARPs) to allow users to learn Tethys without having to use their own data.

Tethys was promoted at the June Detection, Classification, Localization, and Density Estimation of Marine Mammals Using Passive Acoustic Monitoring Workshop (St. Andews, Scotland), during a round table discussion on data archiving and through related projects at the September Intl. Bioacoustics Congress (Pirenópolis, Brazil), and at a presentation at the September IEEE Oceans Conference (San Diego, CA). An IEEE Oceans conference proceeding is in press, and a paper on the spatial-temporal distribution of beaked whales is in review.

Data import and export are critical to the success of Tethys. Discussions with Ei Fujioka have resulted in a data format for export of Tethys data to OBIS-SEAMAP and implementation is beginning. Additional work on data import has been completed, and any open database connectivity (ODBC)

compliant data source for which drivers are available (e.g. MySQL, Oracle database, Postgres) can now be imported.

Implementation of indices to speed queries is ongoing. After discussion with ocean observatory community members, the data transport layer that communicates between the the data service and clients is being modified to support a RESTful interface that can support web services. The REST service is currently undergoing final development and beta testing and should be available in the fall public release. The transition to REST offers many advantages such as the possibility of speeding data transmission through compression (implemented) or the potential to use alternative data formats such as Java script object notation (JSON) which carry lower transmission overhead.

An application programming interface for automated detectors to produce data that can be directly imported into Tethys has been completed. The interface has been tested with a beaked (*Hyperoodontidae spp.*), fin (*Balaenoptera physalus*), and blue (*Balaenoptera musculus*) whale detectors and is currently being integrated into other detectors.

The spring 2013 workshop with NOAA partners focused on importing data into Tethys and how to use it. All participants were able to successfully import samples of their own acoustic detections and instrument deployments into their database. Additional functionality such as the ability to retrieve arbitrary detection details have been implemented as a result of user feedfack.

RESULTS

Tethys is beginning to transition towards being a production tool. This year has produced improvements to Tethys that are both visible (enhanced documentation and capabilities, improved user interface and error reporting) and internal (improvements in speed and reliability).

We show an example of analysis done by Širović (2013) from the Marine Bioacoustics Lab at Scripps Institution of Oceanography. We helped Dr. Širović incorporate the new application programming interface into her blue and fin whale detectors. Tethys enabled her to quickly pull together information to analyze five and a half years worth of data. Dr. Širović was kind enough to let us showcase this work as an example of analysis enabled by Tethys.

In this work, Širović et al. examined the results of analysis of five years of data. Blue whale B calls were detected via spectrogram correlation (Mellinger and Clark, 2000) with seasonal and annual estimation of kernels to adjust for seasonal and annual variation in B call production. Variation in precision (percentage of detections that were correct) and recall (percentage of expected detections that were found) were shown over seasonal periods (Fig. 1). Tethys enables users to track their detections and then query by varying criteria making analysis for such studies easier. Similar analyses were made for interannual variation of blue whale B call detection rates and the impact of shipping and instrument noise on detector performance.

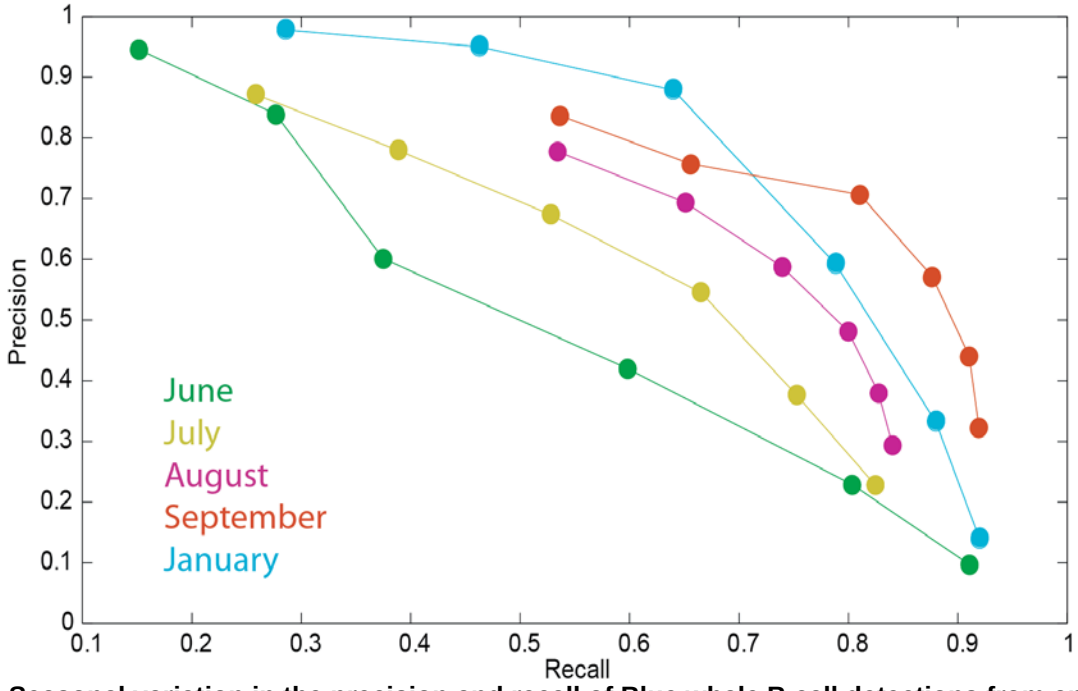

Figure 1 - Seasonal variation in the precision and recall of Blue whale B call detections from an instrument in the Southern California Bight (118.57˚ W, 32.37˚ N) during the 2011 field season.

Fin whales were detected using the acoustic power index method (Širović et al., 2004), an energy ratio comparison that tests a signal band in relation to noise bands on either side of the signal band. This is useful for fin whales whose calls can be so dense that individual calls cannot be distinguished. An example of multiyear analysis (Fig. 2) clearly shows seasonal trends in the detection of fin whales at a Southern California Bight site over a five and a half year period. Spatial-temporal trends were also analyzed (Fig. 3).

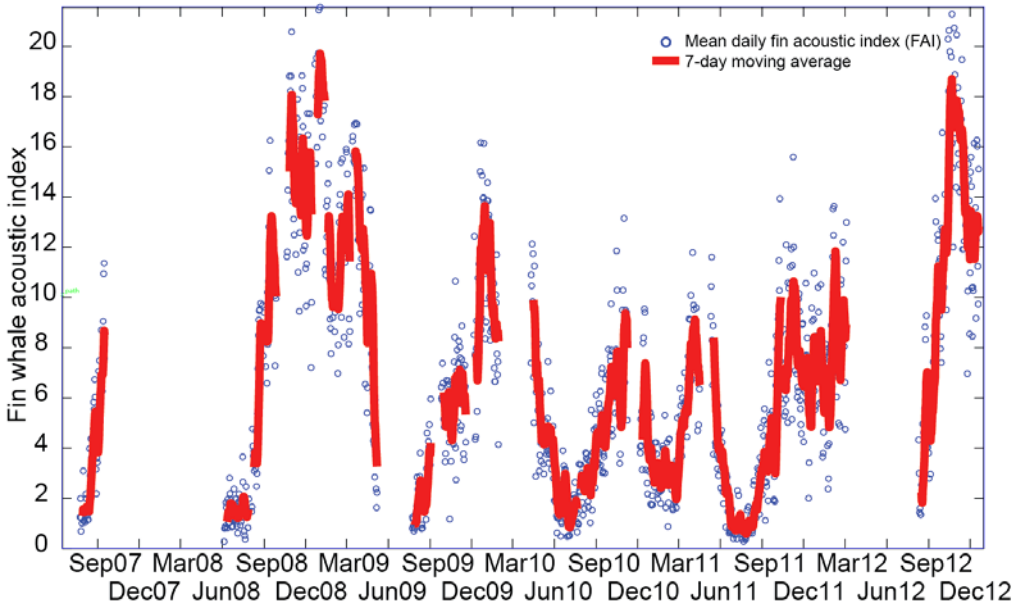

Figure 2 - Acoustic power index denoting presence of calling fin whales over a five year period in the Southern California Bight (119.17˚ W, 32.84˚ N). A seven day moving average is plotted over the daily indices.

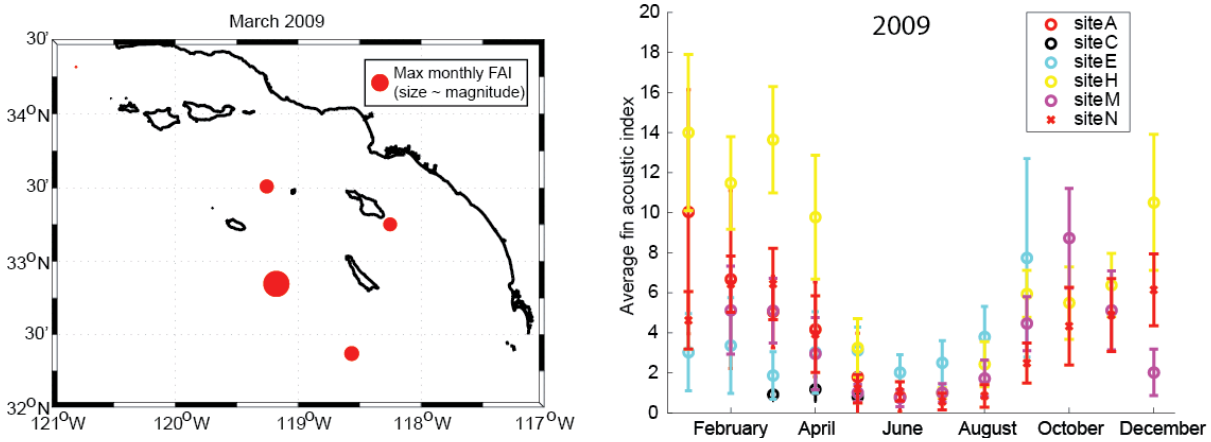

Figure 3 - Spatial-temporal variation of fin whale vocalizations as measured by the fin whale acoustic index. Left panel shows geo-spatial measurement of fin whale index at several sites in the Southern California Bight during March 2009. Right panel summarizes fin whale monthly acoustic power at multiple Southern California Bight sites over 2009.

IMPACT/APPLICATIONS

Currently, the Scripps Institution of Oceanography Tethys node has the most data. Over 800,000 detections of cetaceans, pinnipeds, and anthropogenic activity have been logged. As shown in the results section of this year and previous reports, the growing body of the data will enable longitudinal spatial-temporal studies with significantly reduced effort.

TRANSITIONS

NOAA partners have begun to use Tethys. While we are not currently tracking downloads as the web site is in its infancy, we have attracted interest from other institutions. We have had discussions with Cornell University's Bioacoustics Research Program and met with them during IEEE Oceans.

RELATED PROJECTS

ONR N0001411WX21401 – Advanced Methods for Passive Acoustic Detection, Classification, and Localization of Marine Mammals. PI Jonathan Klay, Dave Mellinger, Dave Moretti, Steve Martin and Marie Roch. Outputs from this project will be in a form that can be easily fed to the Tethys database.

N00014-12-1-0273 – Modeling of Habitat and Foraging Behavior of Beaked Whales in the Southern California Bight, PI John Hildebrand, Simone Baumann-Pickering – The work performed in this grant makes use of Tethys and has overlapping key personnel.

N000141210904 – Blue and fin whale habitat modeling from long-term year-round passive acoustic data from the Southern California Bight, PI John Hildebrand, Ana Širović. – The work performed in this grant makes use of Tethys and has overlapping key personnel.

NSF-OCE-11-38046 – OBIS-SEAMAP, PI Patrick N. Halpin – OBIS-SEAMAP collects visual and acoustic detection information for marine mammals, sea birds, and sea turtles. We have worked with Halpin's team to help them integrate acoustic detections into their platform and are currently working with them to permit transfer of data summaries from Tethys to OBIS-SEAMAP.

NAVFAC Living Marine Resources Applied Research (Bob Gisiner and Frank Stone) – Much of the acoustic data used in the development of this project has been funded by NAVFAC LMR under a variety of grants. Additional data acquisition has been funded through other US Navy fleet operations (Chip Johnson).

REFERENCES

Mellinger, D. K. and Clark, C. W. (2000). Recognizing transient low-frequency whale sounds by spectrogram correlation. *J. Acous. Soc. Am.* 107, 3518-3529.

Širović, A., Hildebrand, J. A., Wiggins, S. M., McDonald, M. A., Moore, S. E. and Thiele, D. (2004). Seasonality of blue and fin whale calls and the influence of sea lee in the Western Antarctic Peninsula. *Deep-Sea Research Part Ii-Topical Studies in Oceanography* 51, 2327-2344, doi:10.1016/j.dsr2.2004.08.005.

Širović, A., Parsell, M., Kerosky, S. M. and Hildebrand, J. A. (2013). Implementation of automatic detectors for blue whale and fin whale calls. In *Intl. Bioacoustics Conference*. Pirenópolis, Brazil.

PUBLICATIONS

M. A. Roch, S. Baumann-Pickering, H. Batchelor, D. Hwang, A. Širović , J. A. Hildebrand, C. L. Berchok, D. Cholewiak, L. Munger, E. M. Oleson, S. Van Parijs, D. Risch, and M. S. Soldevilla, "Tethys: A workbench and database for passive acoustic metadata," *Proc. IEEE Oceans*, 5 pp., San Diego, CA, September 23-26 2013.

S. Baumann-Pickering, A. E. Simonis 1, M. A. Roch, M A. McDonald, A. Solsona-Berga, E. M. Oleson, S. M. Wiggins, R. L. Brownell Jr., J. A. Hildebrand, "Spatio-temporal patterns of beaked whale echolocation signals in the North Pacific", in review.