Dataset Comparison

Several workshop participants have indicated that they are working on the right whale dataset. In order to make a meaningful comparison between different detectors and classifiers we have now uploaded a second dataset to the workshop website. Again, this contains files with many upsweep calls, files with gunshots and files with no right whales present.

We ask you to process these additional datasets with algorithms trained with the other datasets and to submit your results to the workshop organisers at least a week before we meet in June. Your results will then be compared with the log files, which we're not sending round in advance (and are not looking at ourselves - promise !).

For each detected call we will use the following information:

The start and end time of the call.

The lower and upper frequency bounds of the call.

An optional score value which can be any numerical value, a high value indicating a higher degree of confidence that the call is a right whale.

Your logs will be compared with the "ground truth" logs created by human analysts in the following way:

Right whale calls within several sets of SNR ranges will be examined. For each range, we will create precision/recall curves where each point is based on thresholds that cover the range of your detection scores. For each threshold, your algorithm's detections will be said to match a call from the ground truth set if the call overlaps in time and frequency and the score is greater than the current threshold. Based on the set of matches for a given threshold, we will measure **recall** as the fraction of truth calls within that SNR range which had a match with one or more detections with a score >= the current threshold. We will measure **precision** as the number of detected truth calls (as for recall) divided by the number of your calls that had scores >= the minimum threshold (see Roch et al., J. Acoust. Soc. Am., Vol. 130, No. 4, 2011). If you don't provide score values, then we will plot a single point for each SNR range.

It is possible (indeed likely) that the ground truth logs have missed some genuine right whale calls in the data. For this reason, the second IFAW data set is included, in which we are sure, based on a human analysis of the data and on aerial surveys which found no right whales in the vicinity of the recorders, that there are no right whales present. From these data, we will make a second estimate of **precision**, which will be the number of detected truth calls in the NOPP data (as above) divided by the number of detected truth calls in the NOPP data + the total number of detected calls in the IFAW data (again, selected based on your threshold values).

Data format:

Detection logs can be sent in one of two formats. The Silbido format is preferred since it incorporates score values.

Format 1 – XBat.

XBat log files. These can be three files (one per day) or one for the entire dataset. I've a Matlab routine which will convert these to the preferred Silbido format.

Format 2 - Silbido

This is a log format developed by Marie Roch and colleagues for use with the Silbido tonal detector (Roch et al. 2011, *J. Acoust. Soc. Am.* 130(4)). A description of the file format and Matlab and Java code for writing Silbido format data into annotation (.ann) files is available on the workshop website for download. The Silbido format allows you to write the full time-frequency contour of a detected call. Note however that the comparison with the human "truth" logs will only use the minimum and maximum time and frequency values. For the three days of test data, you can do any of the following:

Submit one file for all three days, with times (in seconds) referenced to the start of the first days data

Submit one file per day with times (in seconds) referenced to the start time of each day.

Submit one file per .wav file (i.e. 100's of files) with the times in each log file referenced to the start time of the corresponding wav file.

If you are submitting multiple files, I care little about file names, so long as the files sort sequentially and there is a file for each day / wav file. For ease of book keeping though, please use file names that contain the start time of the data in that file. Our software will read the files in order, add a time offset to each detection based on the files position in the directory and the duration of the files and then concatenate all calls into a single structure.

Any questions, please contact Doug Gillespie dg50@st-andrews.ac.uk