Identifying Delphinid Clicks

Melissa Grier, Kristen Forte, Jarod Teague, Dylan Bush

Introduction

Delphinids produce directional, broadband clicks for echolocation in a sequence. There is currently no catalogue of known recorded sounds produced by delphinids. Using passive acoustic methods provided by the Challenge Data [1], delphinid species may be identified by the frequency and inter-click interval of their clicks. The benefit of this type of data analysis may best be seen by possible communications. Once the frequencies were examined and assigned to their respective species it could be further analyzed to show the various musical notes so-called "songs" delphinids produce. This leads to the possibility of future computer programs being designed using musical notes to create a dolphin lexicon allowing human and dolphin communications.

Methodology

FREQUENCY CLUSTER PREDICTIONS

The main task was delphinid classification using a given set of data that recorded various click frequencies (kHz). Clicks are very short (typically <1 ms) with broadband pulses of peak energy at high frequencies (typically tens of kHz).

Extracting random sample sets, the clicks were analyzed by frequency (kHz) and assigned to the various delphinid species.

K-mean clustering was used on the random data samples to locate frequency clusters in the data sets.

Delphinid Species Identification by Musical Notes

Delphinid	Music Notes
Risso's Dolphin	A#0, B0, B1, C1, D#1, F#0m G#0
Gervais beaked whale	A0, A#0, C1, C#1, D1, D#1, E1, F1, F#1
Sperm whale	B-2, D-2, G-2, G-3, G-4
Stenella genus	D1, F1
Sowerby's beaked whale	A#1, B1, C2
Atlantic white-sided dolphin	A#0, A0, D0, D#0, F0, F#0, G0
Short-fined pilot whale	A-1, B-1, C0, D0, G-1
Cuvier's beaked whale	A-1, B-1, C0, D0

EXPLORATION

We used numerical and musical frequencies to identify various delphinid species.

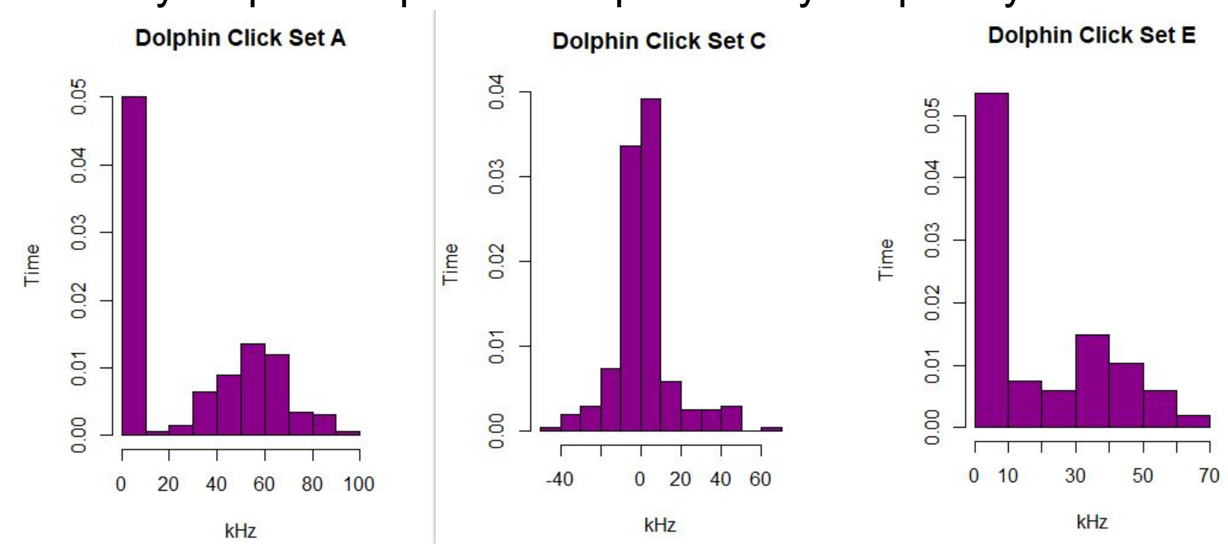
Resources

1. https://challengedata.ens.fr/participants/challenges/32/
2. https://www.intmath.com/trigonometric-graphs/music.php
3. https://www.sciencedirect.com/science/article/abs/pii/S1043661805800295

Results

EXPLORATION

How many delphinid species are present by frequency?



Click Set A

Bin	Frequency Range	Clusters
1	15 - 20	1
2	20 - 25	1
3	25 - 30	2
4	30 - 35	6
5	35 - 40	7
6	40 - 45	13
7	45 - 50	5
8	50 - 55	16
9	55 - 60	11
10	60 - 65	11

Frequency (kHz) to Delphinid Identification Estimates

Delphinid Species	Peak Frequency (kHz)	Centroid Frequency (kHz)
Risso's Dolphin *(Grampus griseus)*	22, 26, 33	22-47.9
Gervais beaked whale (*Mesoplodon europaeus*)	44	30-50
Sperm whale (*Physeter macrocephalus*)	——	2-15
Stenella genus (*Stenella sp.Stenellid*)	——	6-19
Sowerby's beaked whale (*Mesoplodon bidens*)	65.8	55-90
Atlantic white-sided dolphin (*Lagenorhynchus acutus*)	24	19-29.7
Short-fined pilot whale (*Globicephala macrorhynchus*)	18, 55	12-18 & 50-55
Cuvier's beaked whale (*Ziphius cavirostris*)	——	15-20

Centroid Frequencies were the best estimate to use in classifying the bins to the various delphinid species. In cases of overlap, the peak frequencies (*when available)* were used to further estimate the probability for which delphinid species created the individual click frequency. Despite, the variability and appearance of continuous variables each range typically contained only one to three frequencies regardless of how many datasets were sampled. This further made it easier to classify each delphinid species to a click frequency.

Techniques

K-Mean Cluster

1. Specify the number of cluster
2. Initialize centroids by first shuffling the dataset and then randomly selecting *K* data points for the centroids without replacement
3. Keep iterating until there is no change to the centroids. i.e assignment of data points to clusters isn't changing.

Scientific Pitch Notation

Frequency needs to double every 12 notes

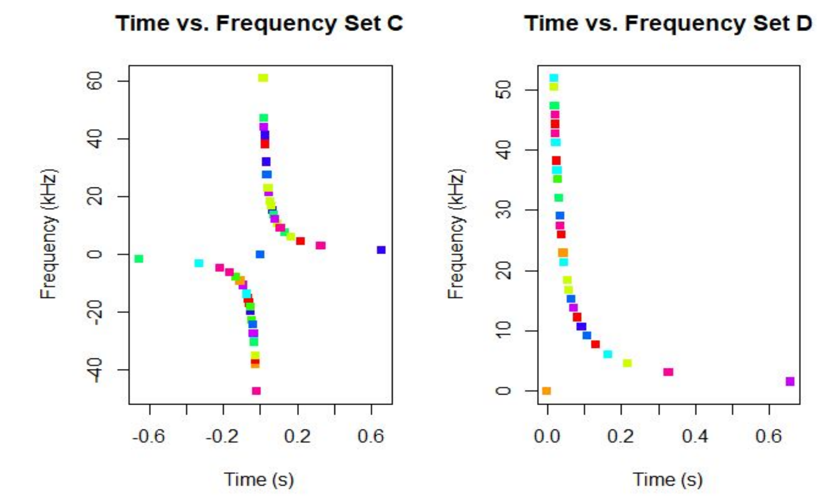
2. $$\text{frequency} = 440 \times 2^{n/12}$$

Mathematically, given the number *n* of semitones of music note A above middle C, the fundamental frequency in hertz is given by the above equation

Further Results

Kernel Density Estimation

The technique was entirely unsuited. It normalized the frequencies but the smooth curve provided no useful data when compared to the histogram method.



K-Means Clustering

K-means clustering algorithm was used for determining the best clustering scheme from the sets using a combination of the number of clusters and distance measure of the frequencies to produce bins for classification estimates.

Maximum mean accuracy (72.21%) is at k=13.

Percentage Prediction Error

3. $$\textit{percentage prediction error} = \frac{\textit{measured value} - \textit{predicted value}}{\textit{measured value}} \; x\, 100$$

The prediction error rate was difficult to calculate due to the little information about the properties of clicks among free-ranging delphinids in offshore habitats. There is no formally recognized catalogue of frequency clicks, so most of the classification data was based on various publications in scientific journals. As such, our error rate was an average of 1.52.

Conclusion

Kernel Density Estimation was redundant for analyzing individual frequency clicks. *K means* was best for identifying clusters and placing them into bins. Scientific Pitch Notation allowed an alternative to viewing frequencies as musical notes.

Future Work

Utilizing python to analyze the data would have produced more results, but the we did not have the time to explore and learn that programming language.

We are grateful for our professor Dr. William Eberle who taught us the fundamentals of data science and encouraged us to participate in this showcase. We are also grateful to the University for the chance to learn and participate in events such as this.