Classification of blue whale D calls and fin whale 40-Hz calls using deep learning

Jeremy Karnowski¹ & Yair Movshovitz-Attias²
¹University of California, San Diego, USA
²Carnegie Mellon University, Pittsburgh, USA
jkarnows@cogsci.ucsd.edu  @mwakanosya
Passive Acoustic Monitoring

- Blue whale and fin whale population sizes are declining.
- Vocalizations found from passive acoustic monitoring can provide massive amounts of data on population sizes and migratory patterns.
“Since the fin whale detectors can be triggered by blue whale calls, a separate detection algorithm for blue whales is being developed to allow for differentiation between the two.“

Weirathmueller, Wilcock, Soule (DCLDE 2011)

“Finally, ambiguity could arise in distinguishing blue whale D calls from fin whale 40-Hz calls in an LTSA even though D calls have a distinctly broader bandwidth (Oleson et al. 2007)”

Širović, Williams, Kerosky, Wiggins, and Hildebrand (2013)
Closely related species, so call production may be similar

Evidence suggests that the fin whale 40-Hz call may be feeding call, similar to the blue whale D call

(Watkins (1981); Sirovic, Williams, Kerosky, Wiggins, and Hildebrand (2013))
Whistle Classification

Feature Selection
(Gannier et al., 2000)

Dynamic Time Warping
(Deecke & Janik, 2006)

Spectrogram Correlation
(Mellinger & Clark, 2000)

Generalized Power-Law
(Helble, Ierley, D'Spain, Roch, Hildebrand, 2012)
GPU and Deep Learning Packages

- **Shallow Network:**
  - Recognizing transient low-frequency whale sounds by spectrogram correlation (Mellinger, Clark 2000)

- **Deep Network:**
  - Practical deep neural nets for detecting marine mammals (Nouri, DCLDE 2013)

[Image of NVIDIA GPU]

[Academic Hardware Grant Request Form]

[Images of Caffe, Theano, and Torch]
Over 1387 hours of audio recorded between 2009-2013 off the coast of Southern California
Dataset Creation

Creating annotation files for each audio file for visual inspection

Creating audio dataset and spectrogram dataset

4796 D calls
415 40-Hz calls
AlexNet + SVM

Input

10 samples

1 sample

256x256

96
55x55

conv max norm

256
27x27

conv max norm

384
13x13

conv

384
13x13

conv

256
13x13

conv max

fc6

fc7

4096

4096

SVM

Class

Extract high level features

Classify each sample

4096-dim feature vectors
fc7 Feature Vectors in 2D (PCA)

blue whale D calls

fin whale 40-Hz calls
AlexNet + SVM

10 samples

Input

1 sample

Extract high level features

Classify each sample

- Linear SVM (C=0.0625)
- 10-fold cross-validation
- For each image, classify each sample as 0 (fin) or 1 (blue)
  - Take average of 10 samples and label as blue if > 0.5
## Results

### Confusion Matrix

<table>
<thead>
<tr>
<th></th>
<th>Blue whale</th>
<th>Fin whale</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Blue whale</td>
<td>4738</td>
<td>58</td>
</tr>
<tr>
<td>True Fin whale</td>
<td>66</td>
<td>349</td>
</tr>
</tbody>
</table>

97.62% Accuracy

For blue whales: 98.63% Precision
98.79% Recall
Results

Precision Recall Curve

Precision Recall Curve
Results

Correctly classified blue whale D calls

Misclassified blue whale D calls

Lower frequency range? Different slope?

Harmonics? Different call?
Future Directions

- Compare this method with other classification methods
- Clean up noise in spectrogram before classification
- Obtain more data from noisier environments to make the detectors more robust
- Add in detection: Use GPL detector (Helble et al. 2012) and then classification on the found calls.
  - Determine the time savings for users
Contributions

- Created more targeted classification datasets
- Used deep learning methods for a novel whistle classification task
- Very good performance in accuracy, precision, and recall
- Easy to modify - researchers can add in additional categories: 50-Hz calls and other false tonal detections
Thanks!
(for all the fish)

Support
Elizabeth Vu
Tyler Helble
John Hildebrand
Edwin Hutchins
Christine Johnson

Funding