Cornell University

K. Lisa Yang Center for Conservation Bioacoustics

Identifying Individual Right Whales through Passive Acoustic Upcall Recordings

Irina Tolkova[1], Holger Klinck[1,2], Dana A. Cusano[3], Anke Kügler[3], Susan E. Parks[3]

[1] K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University, Ithaca, NY USA
[2] Marine Mammal Institute, Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Newport, OR, USA
[3] Department of Biology, Syracuse University, Syracuse, NY, USA

NARW Consortium 2024

CornellLab

Individual identification of North Atlantic right whales through passive acoustics would enable streamlined **abundance estimation**.

In prior studies, upcall vocalizations were shown to contain information on both **individual identity** and **age class**.

But this task was approached with **manual features**, and has not yet been addressed with automated **machine learning** techniques.

Vol. 30: 157–169, 2016
doi: 10.3354/esr00735

ENDANGERED SPECIES RESEARCH
Endang Species Res

Published May 31

Calls of North Atlantic right whales *Eubalaena glacialis* contain information on individual identity and age class

Jessica A. McCordic*, Holly Root-Gutteridge, Dana A. Cusano, Samuel L. Denes, Susan E. Parks

Department of Biology, Syracuse University, Syracuse, NY 13244, USA

Table 4. Classification results of discriminant function analysis for North Atlantic right whale *Eubalaena glacialis* individual and age class based on measurements of upcalls

Grouping	No. of whales	No. of calls	Adjusted chance classification	% correctly classified	Cross-validated % correctly classified
Individual	14	212	22.25	76.9	72.6
Age					
Adult	9	157			
Juveniles	4	45			
Total	13	202	65.37	87.6	86.1

Some Key Challenges…

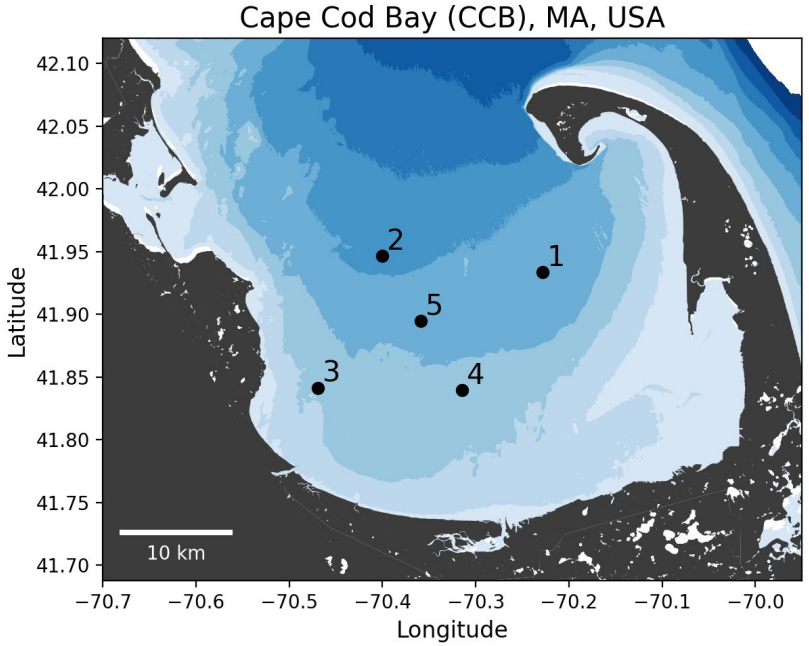
Dataset Construction: how to evaluate individual identification in passive acoustic monitoring data?

Sample Size Limitation: how to leverage deep learning with small datasets?

Ambient Noise: how to ensure that classifiers are learning characteristics of upcalls, not of noise?

Some Key Challenges…

Dataset Construction: how to evaluate individual identification in passive acoustic monitoring data?



Focal upcall recordings were collected with non-invasive suction cup archival acoustic tags. This work uses 180 total upcalls across 9 individuals.

Image: WHOI

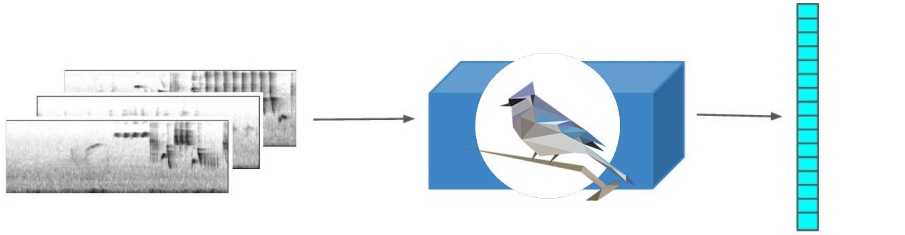
Passive acoustic monitoring data was collected with a bottom-moored five-hydrophone array in Cape Cod Bay between February and June of 2019.

Through a manual examination, it was possible to distinguish individual right whales **calling simultaneously across different channels**.

Some Key Challenges…

Sample Size Limitation: how to leverage deep learning with small datasets?

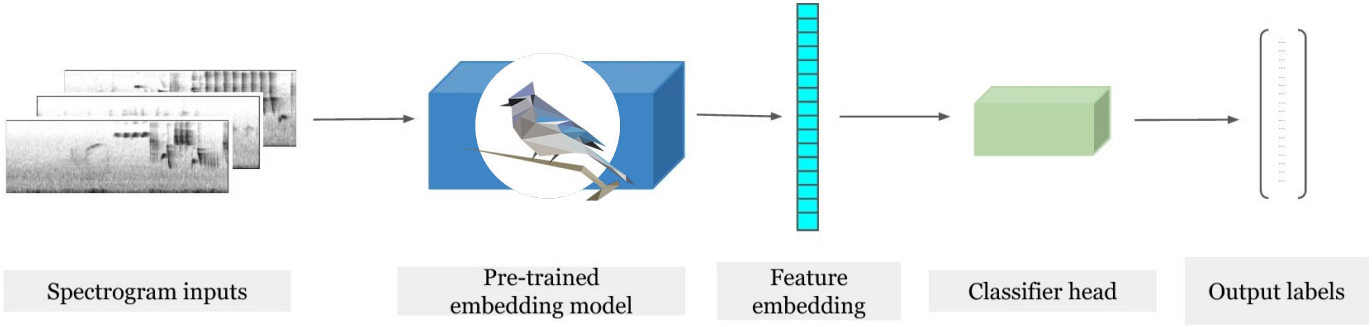
This challenge can be addressed through **transfer learning.** Specifically, we use the **BirdNET** model.



Some Key Challenges…

Sample Size Limitation: how to leverage deep learning with small datasets?

This challenge can be addressed through **transfer learning.** Specifically, we use the **BirdNET** model.

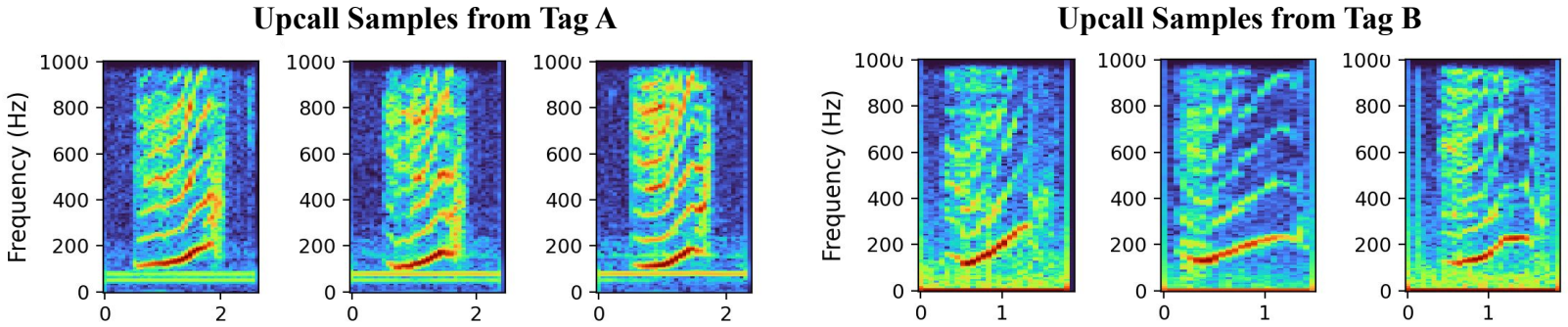


Despite being trained on birdsong, feature embeddings from BirdNET have been **successfully used to classify calls of other taxonomic groups**: including marine mammals, bats, and frogs.

Ghani, Burooj, Tom Denton, Stefan Kahl, and Holger Klinck. "Global birdsong embeddings enable superior transfer learning for bioacoustic classification." Scientific Reports (2023).

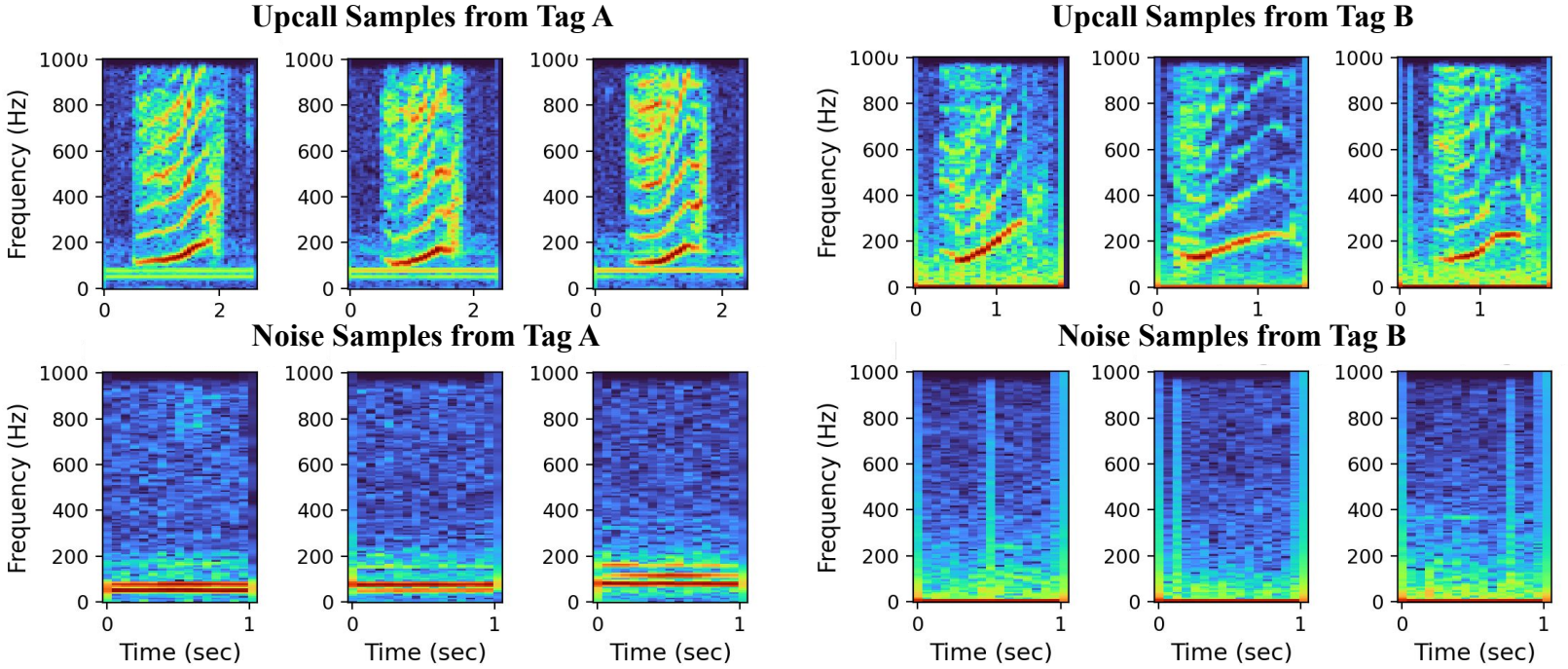
Some Key Challenges…

Ambient Noise: how to ensure that classifiers are learning characteristics of upcall, not of noise?



Some Key Challenges…

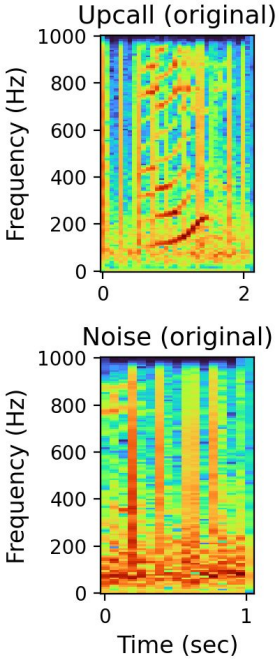
Ambient Noise: how to ensure that classifiers are learning characteristics of upcall, not of noise?



To address this challenge, we apply **rigorous de-noising** and perform **parallel analysis** of ambient noise samples.

Some Key Challenges…

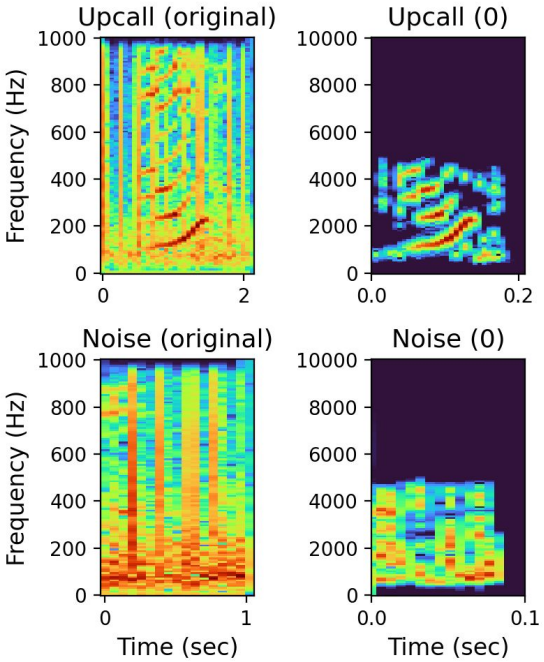
Ambient Noise: how to ensure that classifiers are learning characteristics of upcall, not of noise?



Specifically, we subtract the **time-averaged ambient noise spectrum**, subtract **mean spectrogram row and column values** to mitigate impulsive or narrowband noise, and then **threshold** time-frequency bins below a certain threshold to 0.

Some Key Challenges…

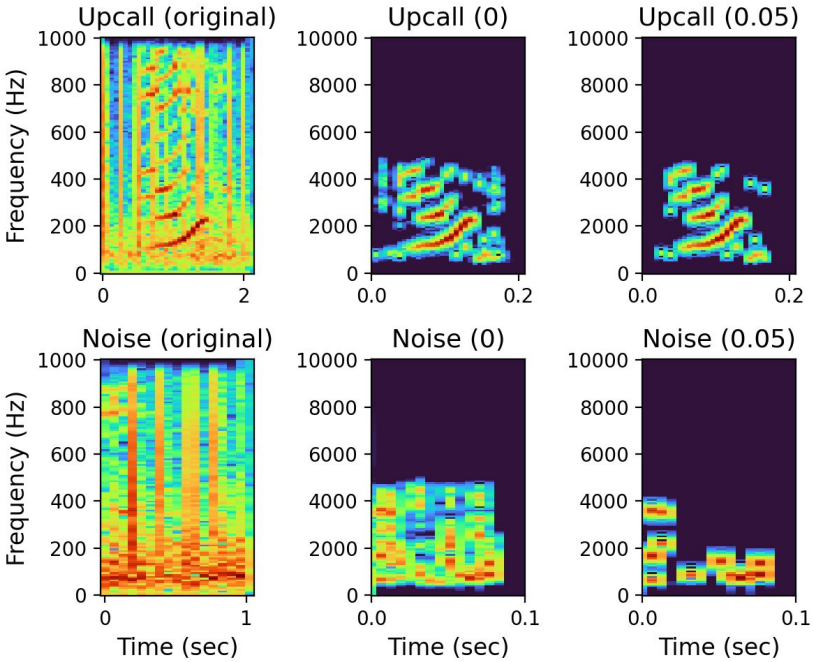
Ambient Noise: how to ensure that classifiers are learning characteristics of upcall, not of noise?



Specifically, we subtract the **time-averaged ambient noise spectrum**, subtract **mean spectrogram row and column values** to mitigate impulsive or narrowband noise, and then **threshold** time-frequency bins below a certain threshold to 0.

Some Key Challenges…

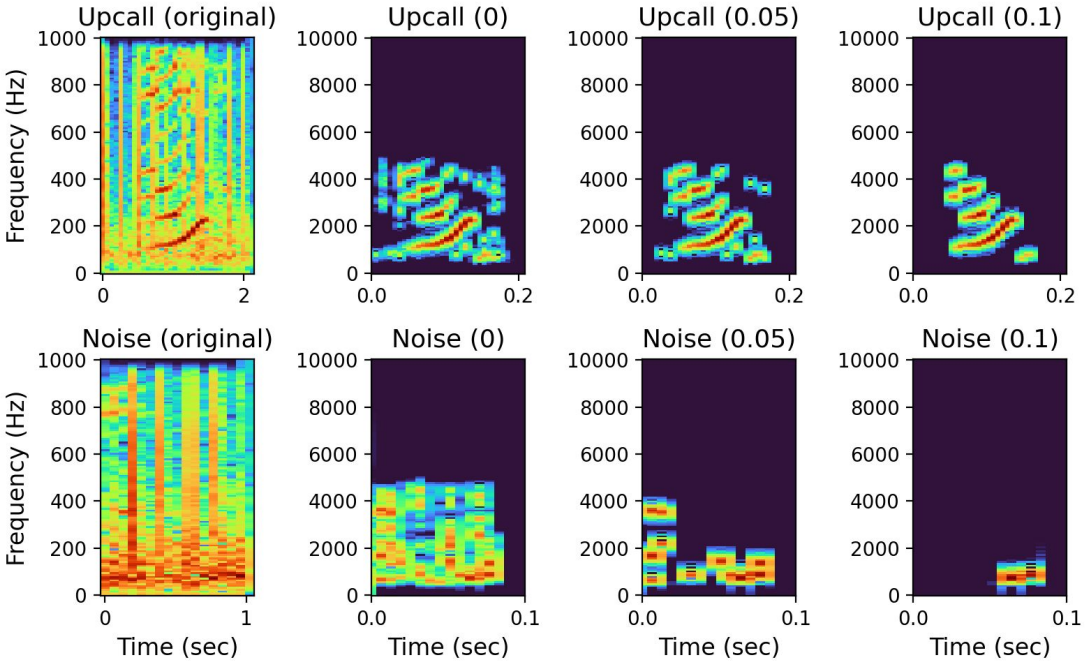
Ambient Noise: how to ensure that classifiers are learning characteristics of upcall, not of noise?



Specifically, we subtract the **time-averaged ambient noise spectrum**, subtract **mean spectrogram row and column values** to mitigate impulsive or narrowband noise, and then **threshold** time-frequency bins below a certain threshold to 0.

Some Key Challenges…

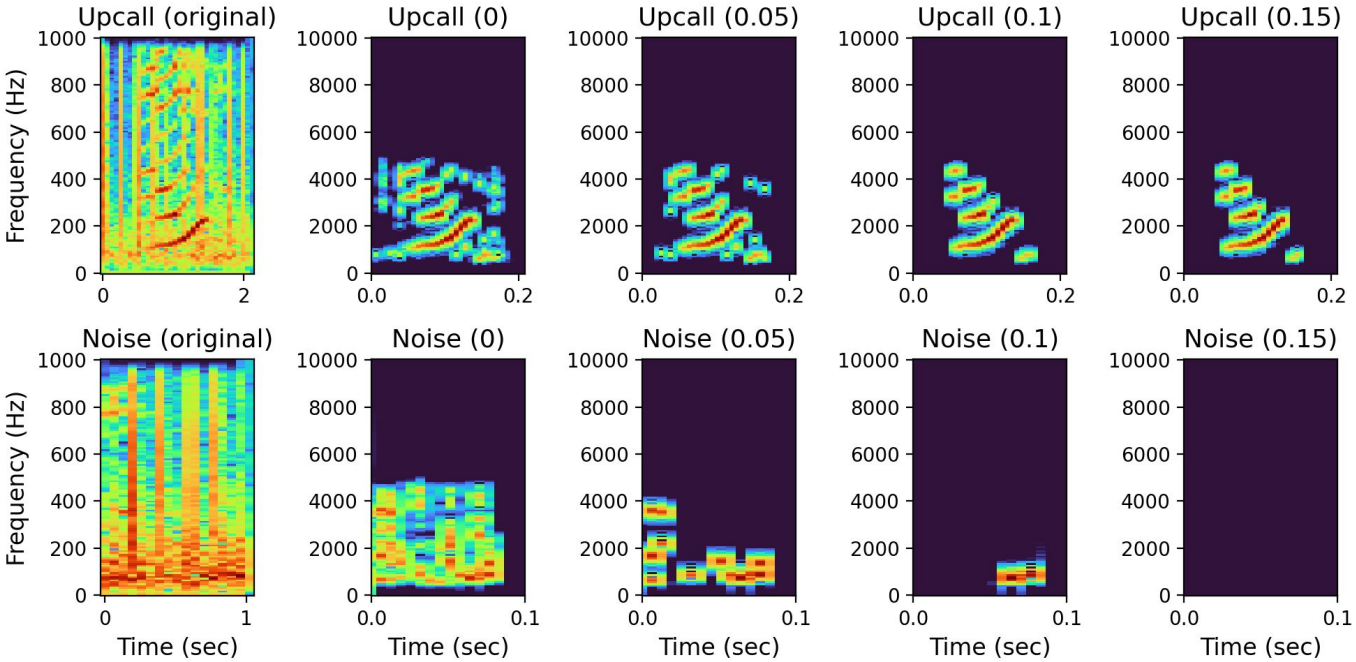
Ambient Noise: how to ensure that classifiers are learning characteristics of upcall, not of noise?



Specifically, we subtract the **time-averaged ambient noise spectrum**, subtract **mean spectrogram row and column values** to mitigate impulsive or narrowband noise, and then **threshold** time-frequency bins below a certain threshold to 0.

Some Key Challenges…

Ambient Noise: how to ensure that classifiers are learning characteristics of upcall, not of noise?



Specifically, we subtract the **time-averaged ambient noise spectrum**, subtract **mean spectrogram row and column values** to mitigate impulsive or narrowband noise, and then **threshold** time-frequency bins below a certain threshold to 0.

Results: Tag Data

upcalls and noise samples from **animal-borne tag recordings**

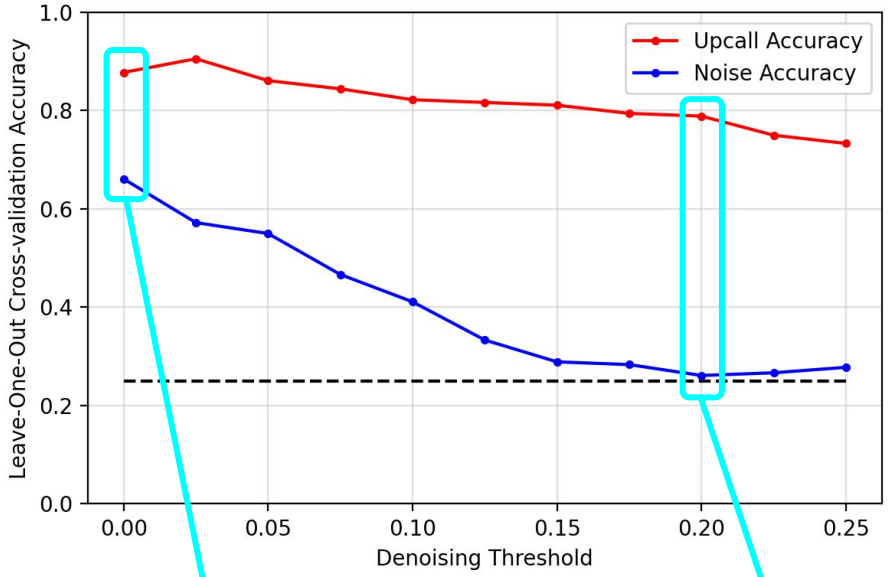
denoising (time-averaged and frequency-averaged spectral subtraction, thresholding) & **upsampling**

BirdNET embeddings (feature vector of size 1024)

evaluate **leave-one-out cross-validation** accuracy on both upcalls and noise samples

dimensionality reduction and visualization with **UMAP**

Classification accuracy across thresholds:



After de-noising without thresholding, we have an upcall accuracy of **88%** and noise classification accuracy of **66%**.

At a threshold of **0.2**, we observe accuracy of **79%** on upcalls, and near-random performance on noise with an accuracy of **26%**.

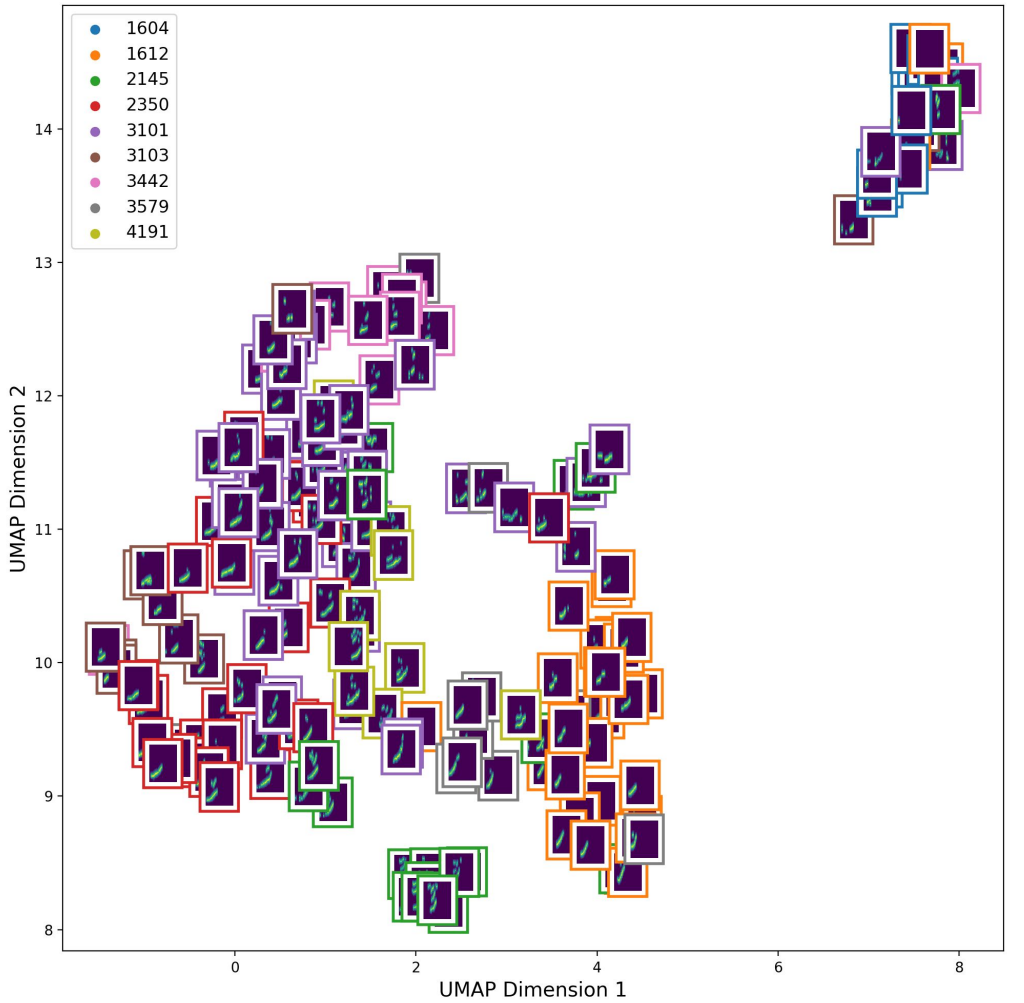
Results: Tag Data

To **visualize lower-dimensional structure**, we further reduce feature dimensionality from 1024 to 2 with **UMAP**, and display all upcalls.

We find that the upcall samples do **cluster by individual**, after accounting for the effects of tag-specific noise.

Altogether, we find that:

- Denoising can successfully mitigate the confounding effect of **tag-specific noise**.
- **BirdNET embeddings** can meaningfully distinguish individual identity within focal upcalls.
- Individual identity is a significant **driver of variability** in upcall structure.



Can we apply this same analysis to passive acoustic recordings?

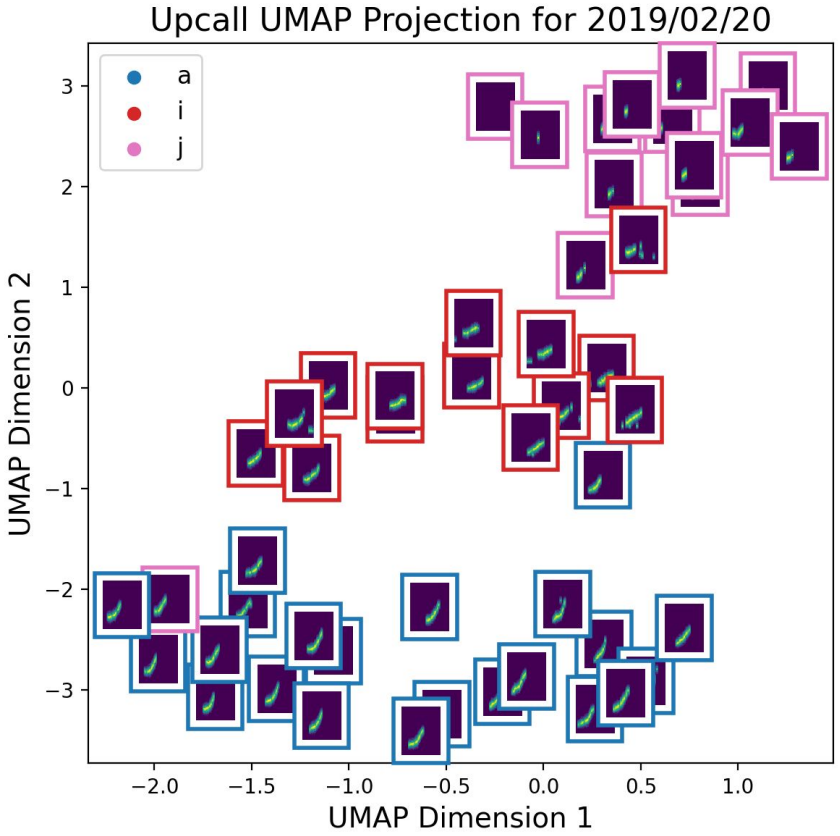
upcalls and noise samples from **passive acoustic recordings** → **denoising & pre-processing** →  → **BirdNET embeddings** → evaluate **leave-one-out cross-validation** accuracies

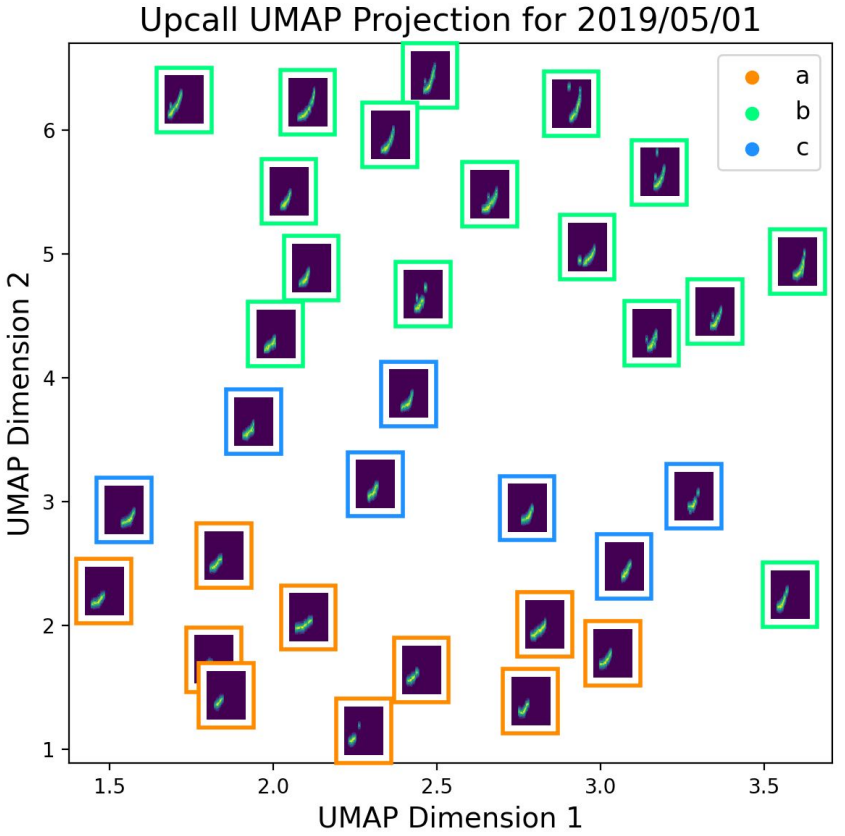
Day	2 / 20	2 / 22	4 / 19	5 / 01
Individuals	3	2	2	3
Total Calls	50	33	34	33
Upcall Accuracy	96%	96%	91%	87%
Noise Accuracy	44%	75%	61%	48%
Random Accuracy	44%	51%	64%	48%

We find that across the different sub-datasets, the BirdNET embeddings allow to **distinguish upcalls with high accuracy**, while obtaining near-random noise classification accuracy.

What does an **unsupervised (UMAP) projection** of the embeddings look like?

Can we apply this same analysis to passive acoustic recordings?

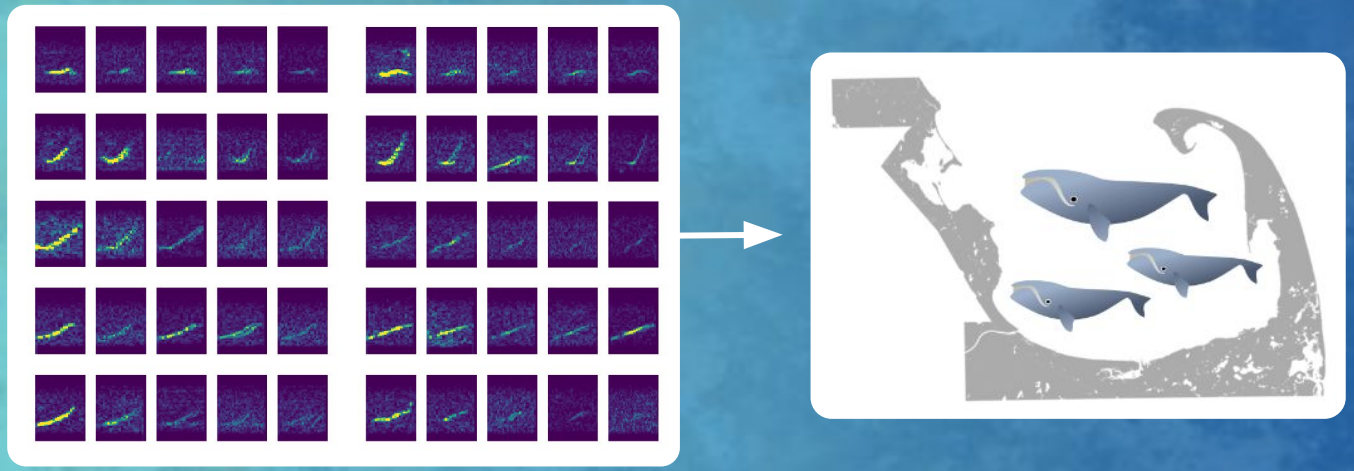




In conclusion…

Feature embeddings can be used to distinguish **individual identity** in **NARW upcalls**, within both **focal and PAM recordings**. In both, individual identity is a significant driver of **variability in vocalization structure.**

Future work: Can we apply this analysis to passive acoustic data to estimate abundance?



Thank you all for listening!

Thank you to my co-authors Holger Klinck, Dana A. Cusano, Anke Kügler, and Susan E. Parks.

Questions?

Irina Tolkova
itolkova@cornell.edu
https://avokloti.github.io/

Crane Beach, MA – August 4th 2024