Decoding gray wolf (*Canis lupus*) communication at scale in Yellowstone National Park

Amir Dehbi[1], Daniel R. Stahler[2], Matthew C. Metz[2], Kira A. Cassidy[2], Jeffrey T. Reed[3], Jeremy D. SunderRaj[2], Taylor L. Rabe[2], Brett A. Melbourne[4], Joanna E. Lambert[1,4]

1 Department of Environmental Studies, University of Colorado Boulder, Boulder, CO, USA
2 Yellowstone Center for Resources, Yellowstone National Park, Mammoth Hot Springs, WY, USA
3 Grizzly Systems & Cry Wolf Project, Emigrant, MT, USA
4 Department of Ecology and Evolutionary Biology, University of Colorado Boulder, Boulder, CO, USA

NEURAL INFORMATION PROCESSING SYSTEMS

Wolf Communication Ecology & Conservation

Gray wolves (*Canis lupus*) are an ideal species for bioacoustics and machine learning research due to their complex, socially rich, and ecologically meaningful vocal communication system. These highly social carnivores use long-distance acoustic signals for territory defense, social cohesion, and mate attraction.[1-2] They occupy only a small portion of their historic U.S. range and remain a conservation priority.[3] Wolf vocalizations encode individual identity, group membership, and emotional state, but their vocal repertoire remains poorly understood.[4-7] Although recent studies demonstrate the feasibility of automated howl detection using deep learning, ML use in wolf bioacoustics remains nascent, offering opportunities to advance automated detection, unsupervised repertoire discovery, and semantic exploration.[8-9] Yellowstone's gray wolf monitoring program is uniquely positioned for this work because its decades-long tracking of individuals, packs, and social dynamics provides an unparalleled ecological and behavioral foundation for interpreting and validating bioacoustics and machine learning analyses.

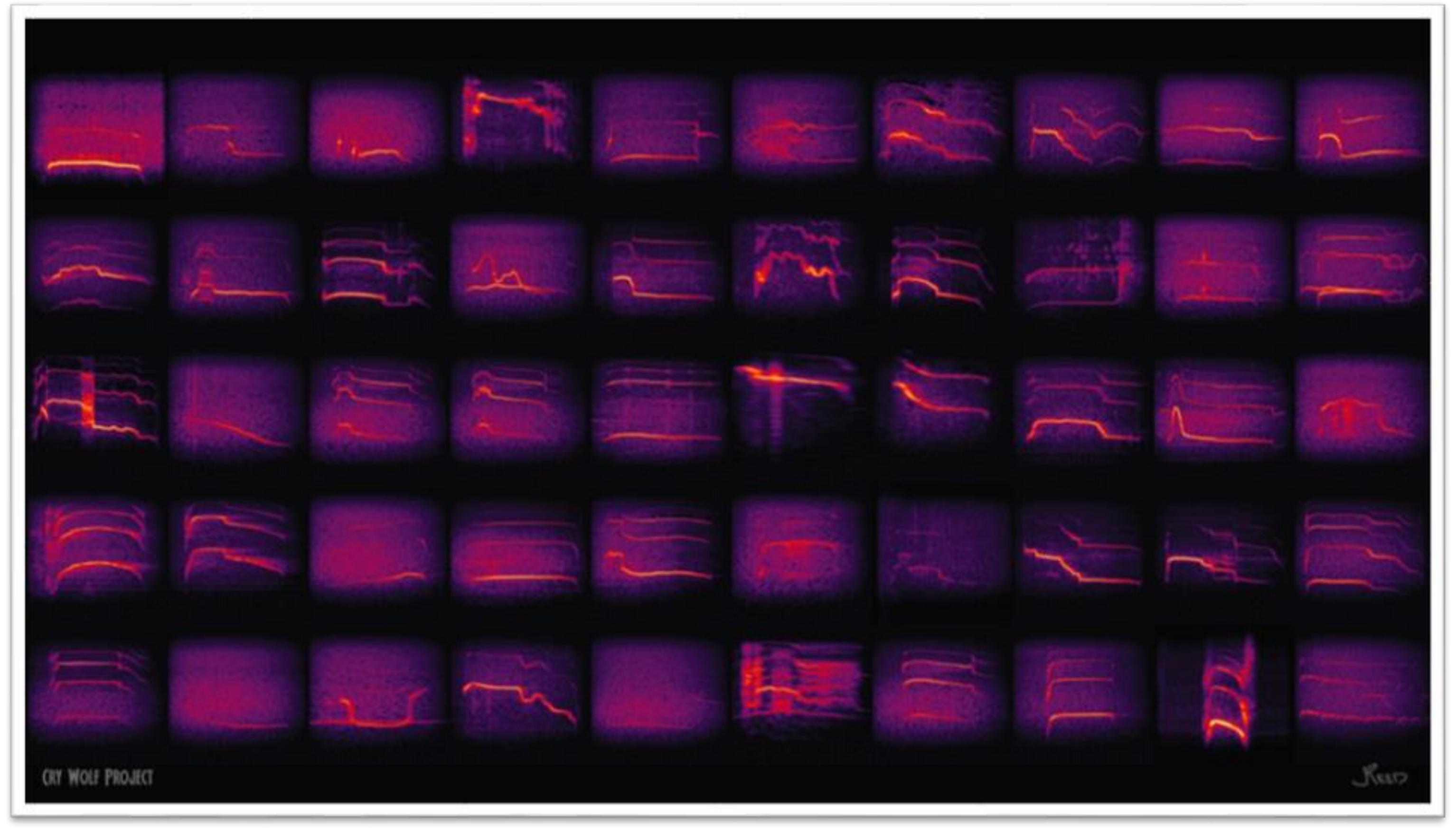
Yellowstone Wolf Bioacoustics Project Aims:

- Assemble a large, representative, in-situ gray wolf bioacoustics dataset.
- Pilot on-collar bioacoustics autonomous recording units (ARUs).
- Develop automated wolf detection models using existing state-of-the-art (SOTA) convolutional neural network (CNN) and transformer models.
- Describe wolf vocal repertoire and structure via unsupervised learning.
- Evaluate data scarcity strategies for rare call types.
- Develop workflows for occupancy and species distribution models.

The one-eyed matriarch of the Junction Butte pack, 907F, performing a solo howl. Photo courtesy of Evan Stout, Yellowstone Wildlife Guide Company.

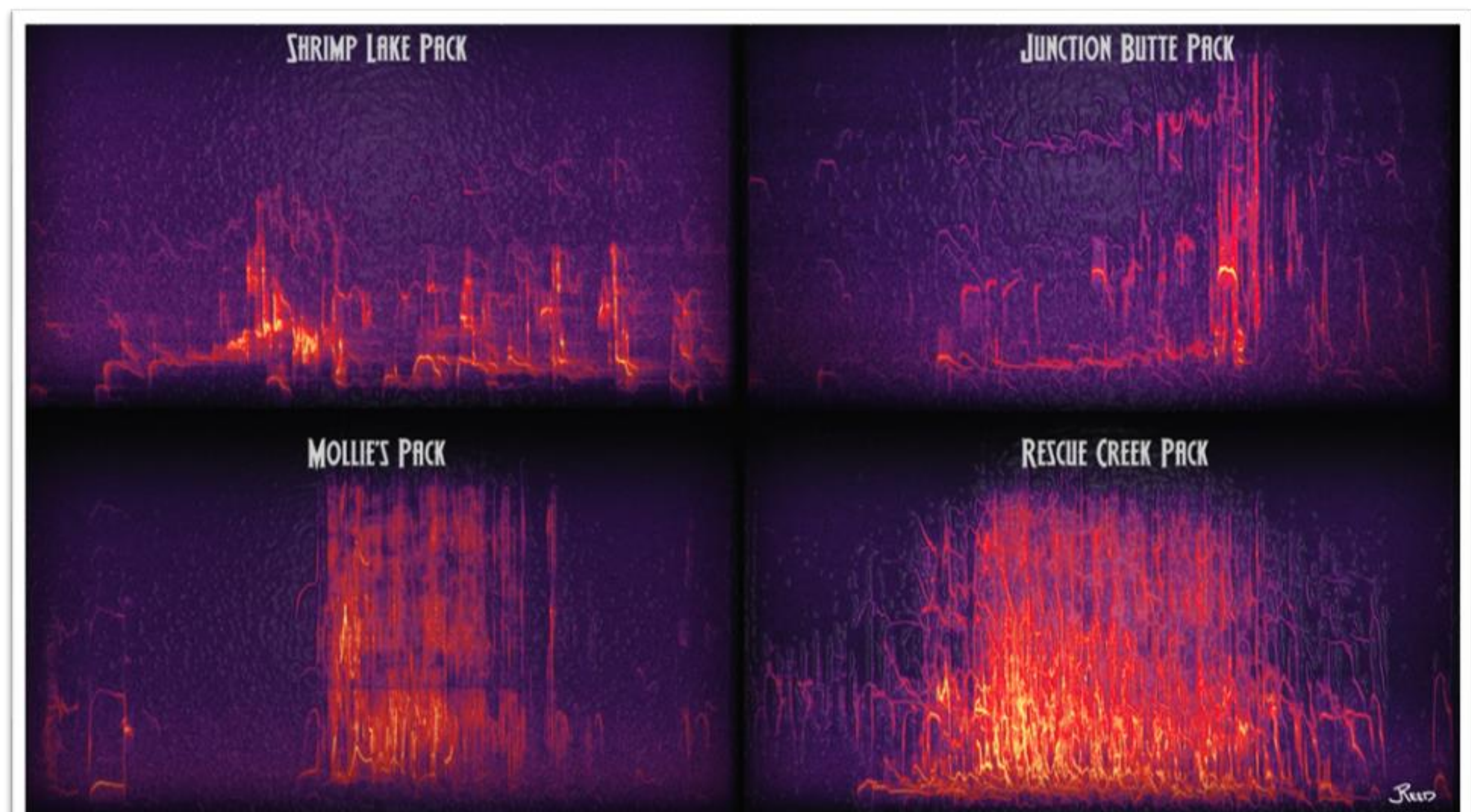
Wolf Vocal Complexity & Repertoire

Wolf vocal signals exhibit complex temporal–spectral structures with individually distinctive features, nonlinear phenomena, and overlapping acoustic spaces between call categories – all of which challenge traditional categorical classification approaches.[1-2,4-5] Their vocalizations are multi-scale: with short-range and long-range call types as well as individual and collective signal production mechanisms.[2]



Spectrogram images showing spectral variety in 50 solo wild gray wolf (*Canis lupus*) howls recorded in Yellowstone National Park in 2023. Image prepared by Jeffrey T. Reed.

Existing wolf vocal repertoire descriptions are typically based on human auditory perception and describe 4-8 call types. Recent machine learning analyses reveal additional structure within these categories, suggesting that wolf vocal communication may be better represented as a continuous, high-dimensional acoustic space rather than a fixed set of discrete classes.[1-2,5,7]



Spectrogram images showing 4 distinct chorus howls produced by 4 different wild gray wolf (*Canis lupus*) packs recorded Yellowstone National Park in 2023. Image prepared by Jeffrey T. Reed.

Vocalization Detection & Repertoire Discovery

Yellowstone National Park Bioacoustics Dataset

With over 4,500 hours of field work, the biologists at Yellowstone have maintained a 50-unit ARU network across the northern range, yielding more than 200,000 hours of continuous audio and over 7,000 wolf vocalization events (~800 hours).

Classifiers for Passive Acoustic Detection

We are evaluating a suite of SOTA convolutional and transformer-based models across visual and audio representation learning. These architectures span conventional CNN and vision transformer models pretrained and fine-tuned on large benchmark datasets, and bioacoustics domain-specific models developed through self-supervised or pretraining. These models will be used to assess fine-tuning and embedding-based classification strategies applied to wolf vocalizations within a modern deep learning framework.

Model	Architecture	Pretraining	Fine-tuning	Inclusion Justification
EfficientNetV2	CNN	ImageNet-21K	ImageNet-1K	Standard CNN baseline
ConvNext	CNN	ImageNet-1K	None	Modern CNN with transformer-like scaling
VGGish	CNN	YouTube-8M	AudioSet	CNN baseline in recent wolf work[9]
PANNs-CNN14	CNN	AudioSet	-	Bioacoustics reference CNN model
ViT	Vision Transformer	ImageNet-21K	ImageNet-1K	Standard transformer baseline
AST	Vision Transformer	ImageNet-21K	AudioSet	Models use a ViT backbone fine-tuned on AudioSet spectrogram images. HTS-AT offers hierarchical token merging and PaSST offers a patch-out transformer variant.
HTS-AT	Vision Transformer	ImageNet-21K	AudioSet	
PaSST	Vision Transformer	ImageNet-21K	AudioSet	
Perch	CNN + Attention	Several Acoustic	-	General-purpose bioacoustics model
BirdNET	CNN	xeno-canto	-	Established bioacoustics model
DINOv3	Vision Transformer	ImageNet-1K	-	Self-supervised transformer baseline
AVES	Audio Transformer	Several Acoustic	-	Self-supervised based on HuBERT

Unsupervised Vocal Repertoire & Semantic Discovery

Beginning in 2026, we will apply unsupervised learning methods to identify structure in vocalizations to explore acoustic groupings that may represent elements of the vocal repertoire. Using embeddings extracted from pretrained models, we will cluster calls in learned feature spaces with tools such as UMAP for dimensionality reduction and HDBSCAN or Gaussian mixture models for cluster formation. Resulting clusters will be validated using ecological and behavioral data to assess biological relevance.

References

[1] Passilongo et al. (2017). "Singing in a Wolf Chorus..." In: Hystrix
[2] Harrington & Mech (1979). "Wolf Howling and…" In: Behaviour
[3] U.S. Fish & Wildlife Service (2022). "Gray Wolf …". In: fws.gov
[4] Root-Gutteridge et al. (2013). "Improving Individual..." In: Bioacoustics
[5] Palacios et al. (2007). "Iberian Wolf…" In: Journal of Mammalogy
[6] Sadhukan et al. (2019). "Characterizing the..." In: PLoS One
[7] Kershenbaum et al. (2016). "Disentangling..." In: Behavioural…
[8] Makarov et al. (2025). "Automated Detection..." In: Sci. Reports
[9] Campos et al. (2025). "HOWLish: a..." In: Remote Sensing in..

Acknowledgements

Thank you to the field technicians at the Yellowstone Wolf Project who contributed thousands of hours of field work for data collection, and to the American Canid Project, Conservation Nation, Yellowstone Forever, and the Colossal Foundation for providing funds for graduate student research fellowships, field technicians, and research equipment. Additionally, thank you to the Department of Environmental Studies and the Department of Ecology & Evolutionary Biology at the University of Colorado Boulder for your support. This work utilized the Alpine high performance computing resource at the University of Colorado Boulder. Alpine is jointly funded by the University of Colorado Boulder, the University of Colorado Anschutz, Colorado State University, and the National Science Foundation (award 2201538).

NATIONAL PARK SERVICE
YELLOWSTONE FOREVER
University of Colorado Boulder
CONSERVATION NATION
American Canid Project
GRIZZLY SYSTEMS
colossal FOUNDATION