Toward Automated Passive Acoustic Monitoring of Urban Rats

Nicholas Jourjine[1], Ralph Peterson[1], Sonja Johnson-Yu[2] and Emily Mackevicius[1]

1. Basis Research Institute, Cambridge, MA and New York, NY, USA
2. Harvard University, Cambridge, MA, USA

nick@basis.ai nickjourjine.bsky.social



Urban rats are invasive pests that are also intelligent, social, **highly vocal** mammals

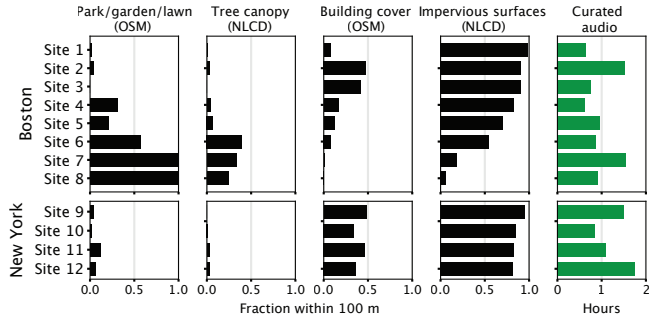
Urban rat management needs automated, **large-scale sensing** of rodent populations

Passive monitoring of rat vocalization is a promosing solution, but tools are lacking!

rats@basis research goals and motivations

- Advance methods for privacy-aware urban bioacoustics
- Create diverse, curated rodent vocalization datasets
- Use these datasets to:
  1. Gain insight into the vocal repertoires of wild rats
  2. Benchmark ML models for automated rat detection

Raw data: urban rats on a habitat gradient





More “urban”



Intermediate



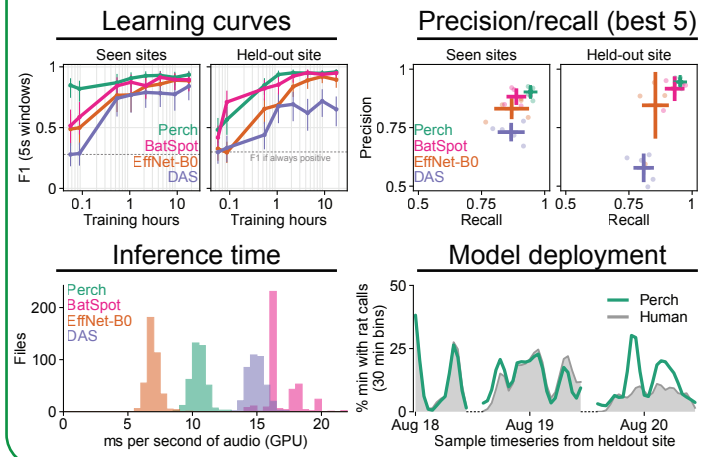
More “natural”

Benchmarking models on curated data

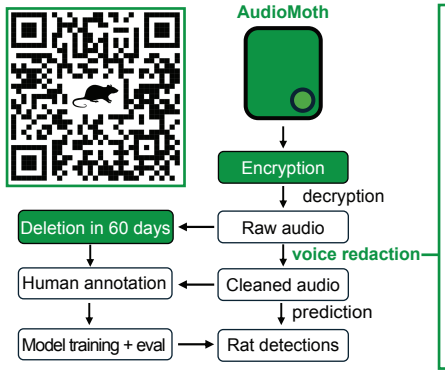
Split	Sites	Files	USVs	Hours
Train	18	1,124	23,394 (81%)	17.6
Validation	18	198	3,844 (13%)	3.1
Seen-site test	3	60	531 (2%)	0.91
Held-out site	1	100	1,270 (4%)	1.53

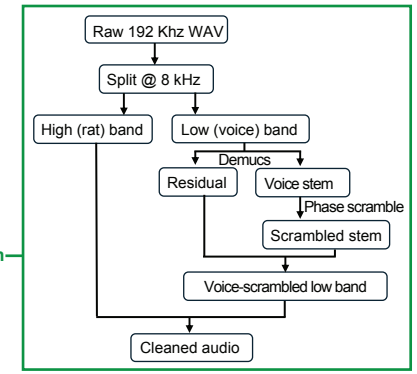
Model	Input	Model fit
Deep Audio Segmenter (DAS) Steinfath et al. eLife 2021	43 ms waveform (192 kHz)	TCN, no pre-training
BatSpot Bergler et al. Scientific Reports 2022	0.5 s spectrogram (15–85 kHz)	ResNet-18, pre-trained
EfficientNet-B0 Tan et al. ICML 2019	1 s spectrogram (15–85 kHz)	Full model, pre-trained
Perch 2.0 van Merriënboer et al. arXiv 2025	830 ms waveform (as if 32 kHz)	Logistic regression

Rat-PAM model training and evaluation



Methods for privacy-aware urban bioacoustics



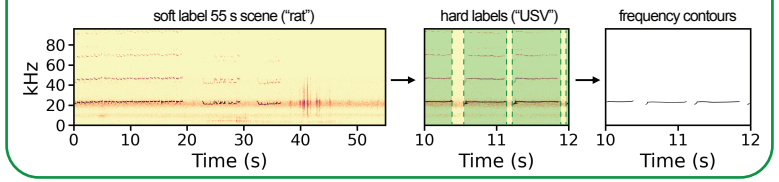


Curated data: acoustic labels and contours

Soft label 55 second-long scenes from spectrogram images
178,364 soft labeled scenes: positive if >1 rat vocalization

Hard-label rat vocalizations within acoustic scenes
30,069 labeled vocalizations: squeak (5%) and USV (95%)

Curate quality frequency contours from USV hard-labels
13,669 high-quality frequency contours: validated by hand



Conclusions and next steps

- Pre-trained embedding models perform well with some generalization to an unseen site
- One hour of labeled audio is sufficient for near-best performance on test annotations
- **Next**: Squeak detection, expand to more sites and cities, frequency contour analysis