Few-shot Detection of Animal Sounds at the Planetary Scale

Vincent Lostanlen

LS2N, Centrale Nantes, CNRS

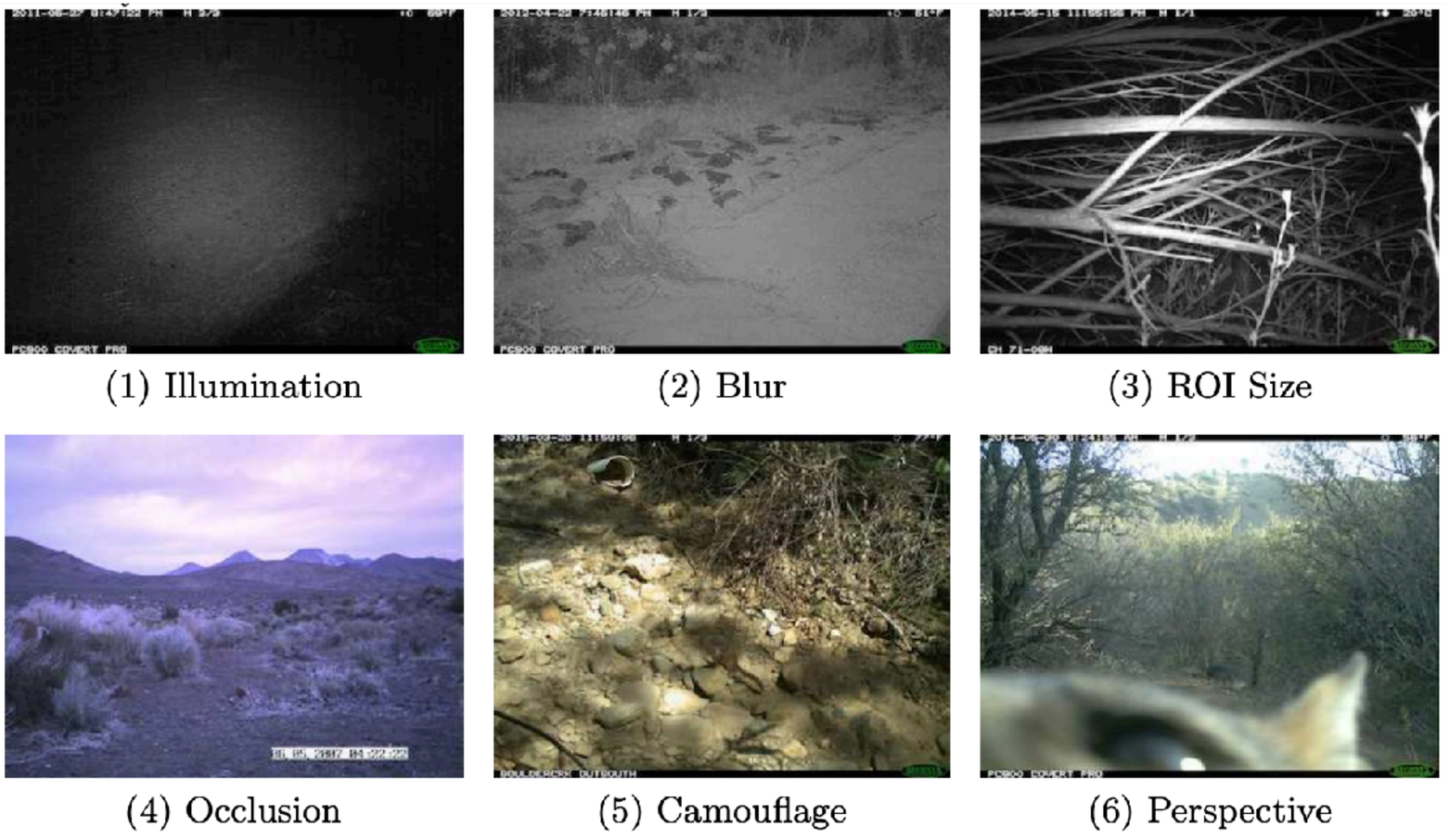
Paris, FR, January 2023. My research is funded by the French government.

Machine learning meets sensor networks

(Tuia *et al.* 2021)

Recognition in terra incognita

(Beery *et al.* 2018)



(1) Illumination (2) Blur (3) ROI Size

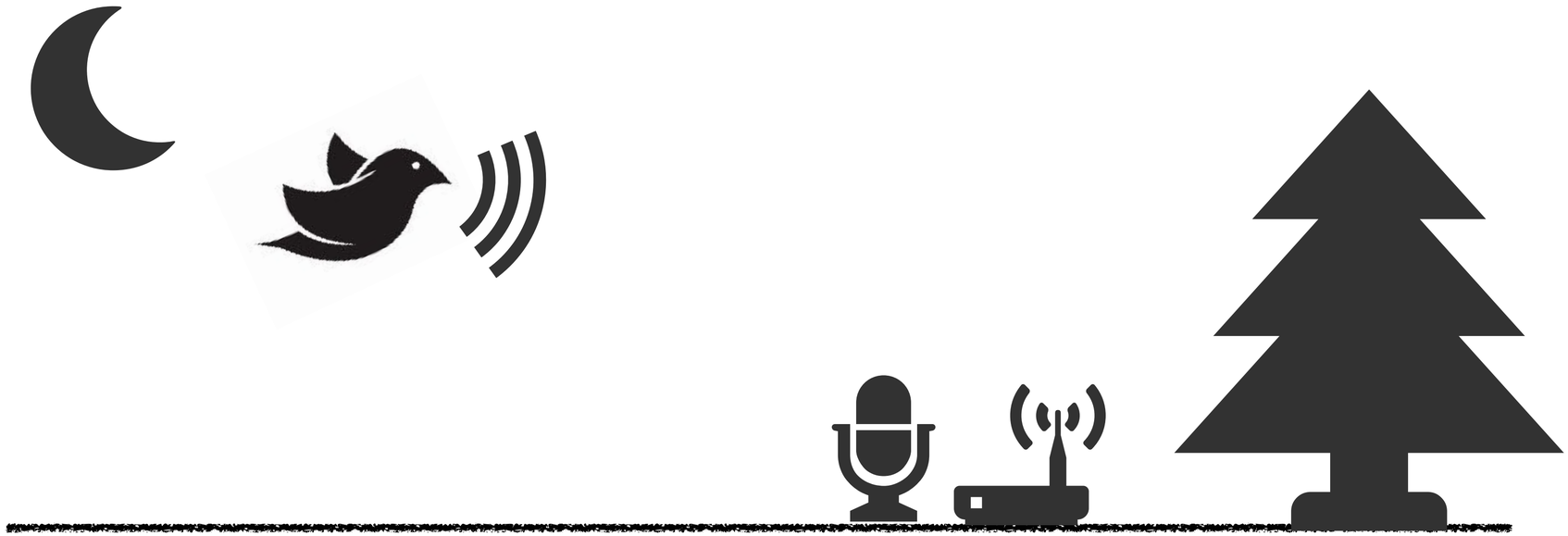
(4) Occlusion (5) Camouflage (6) Perspective

Fig. 3. Common data challenges: (1) **Illumination**: Animals are not always salient. (2) **Motion blur**: common with poor illumination at night. (3) **Size of the region of interest** (ROI): Animals can be small or far from the camera. (4) **Occlusion**: e.g. by bushes or rocks. (5) **Camouflage**: decreases saliency in animals' natural habitat. (6) **Perspective**: Animals can be close to the camera, resulting in partial views of the body.

Advantages of computational bioacoustics

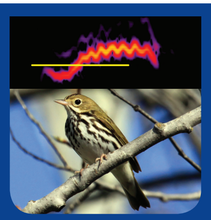
- (Quasi) omnidirectional
- Operates at night
- (Quasi) independent of weather
- Able to localize sources

- Many species are more easily heard than seen
- Potentially long-distance
- Potentially polyphonic
- Gives insight on animal behavior



Rosetta Stone to the Warblers

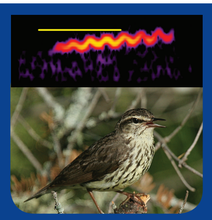
The Cornell Lab of Ornithology



Ovenbird
(Seiurus aurocapilla)

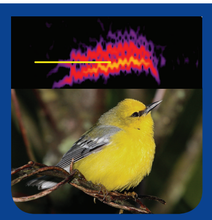
Worm-eating Warbler
(Helmitheros vermivorum)

Louisiana Waterthrush
(Parkesia motacilla)



Northern Waterthrush
(Parkesia noveboracensis)

Golden-winged Warbler
(Vermivora chrysoptera)



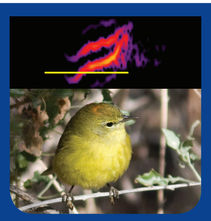
Blue-winged Warbler
(Vermivora cyanoptera)

Black-and-white Warbler
(Mniotilta varia)

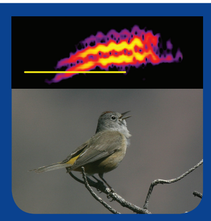
Prothonotary Warbler
(Protonotaria citrea)

Swainson's Warbler
(Limnothlypis swainsonii)

Tennessee Warbler
(Oreothlypis peregrina)

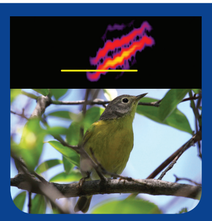


Orange-crowned Warbler
(Oreothlypis celata)



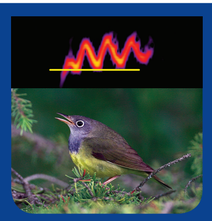
Colima Warbler
(Oreothlypis crissalis)

Lucy's Warbler
(Oreothlypis luciae)



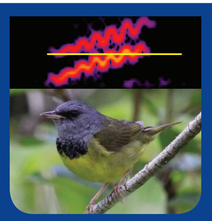
Nashville Warbler
(Oreothlypis ruficapilla)

Virginia's Warbler
(Oreothlypis virginiae)

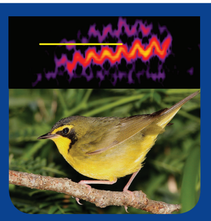


Connecticut Warbler
(Oporornis agilis)

MacGillivray's Warbler
(Geothlypis tolmiei)



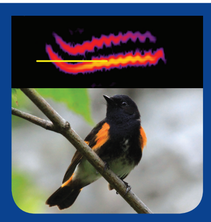
Mourning Warbler
(Geothlypis philadelphia)



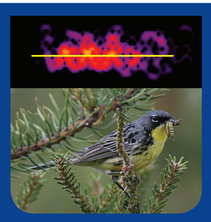
Kentucky Warbler
(Geothlypis formosa)

Common Yellowthroat
(Geothlypis trichas)

Hooded Warbler
(Setophaga citrina)



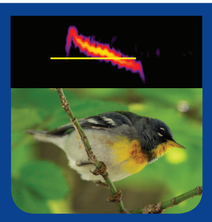
American Redstart
(Setophaga ruticilla)



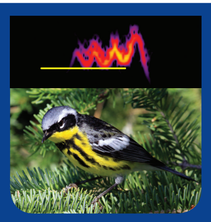
Kirtland's Warbler
(Setophaga kirtlandii)

Cape May Warbler
(Setophaga tigrina)

Cerulean Warbler
(Setophaga cerulea)



Northern Parula
(Setophaga americana)



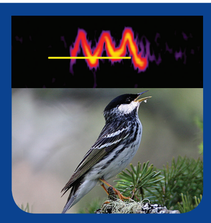
Magnolia Warbler
(Setophaga magnolia)

Bay-breasted Warbler
(Setophaga castanea)

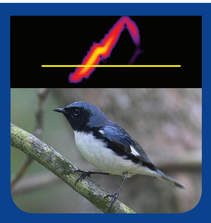
Blackburnian Warbler
(Setophaga fusca)

Yellow Warbler
(Setophaga petechia)

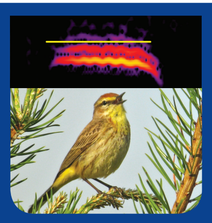
Chestnut-sided Warbler
(Setophaga pensylvanica)



Blackpoll Warbler
(Setophaga striata)

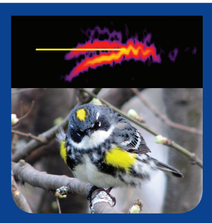


Black-throated Blue Warbler
(Setophaga caerulescens)

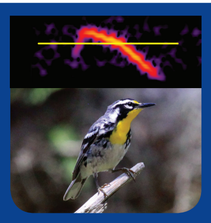


Palm Warbler
(Setophaga palmarum)

Pine Warbler
(Setophaga pinus)



Yellow-rumped Warbler
(Setophaga coronata)



Yellow-throated Warbler
(Setophaga dominica)

Prairie Warbler
(Setophaga discolor)

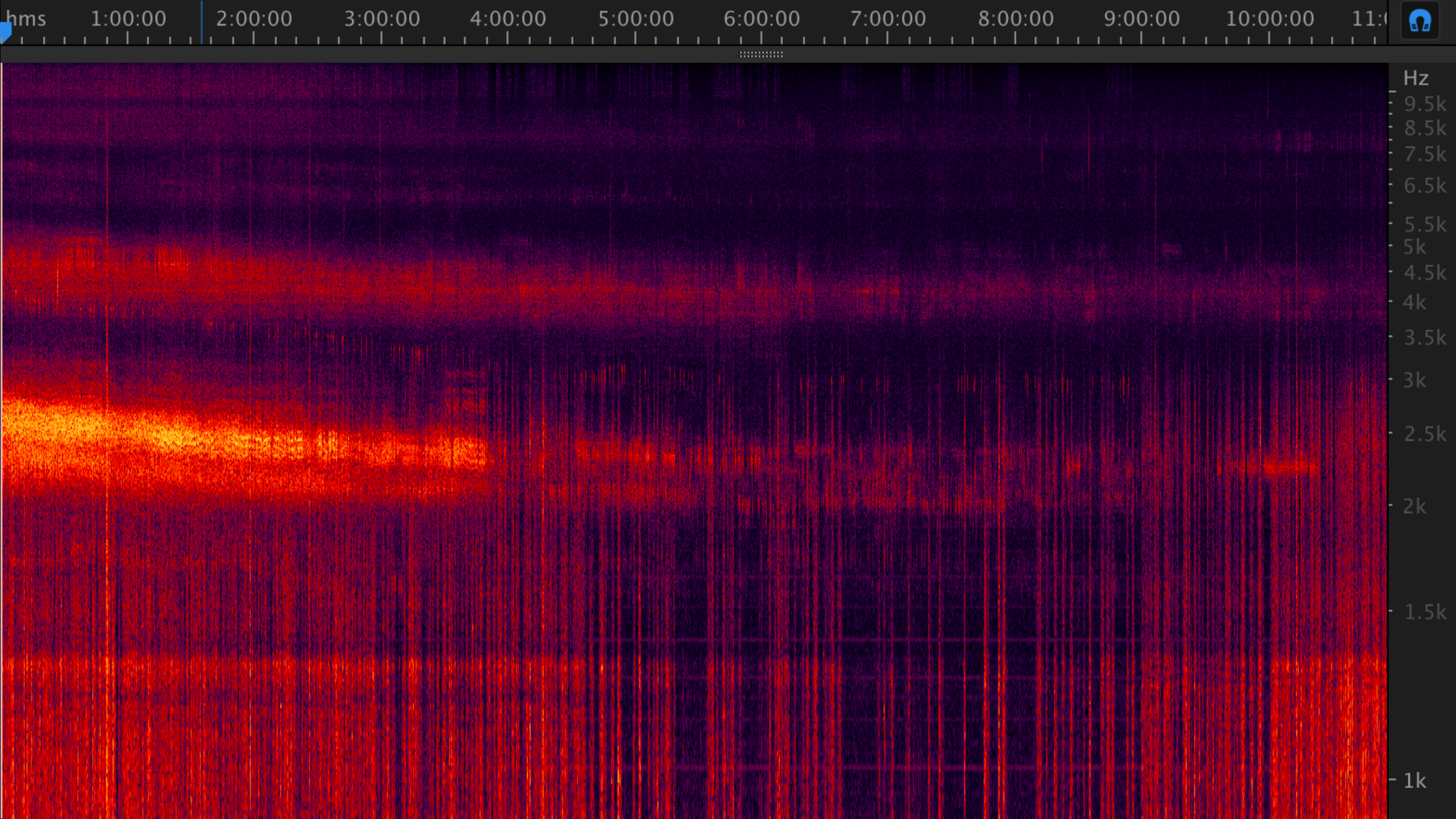
Grace's Warbler
(Setophaga graciae)

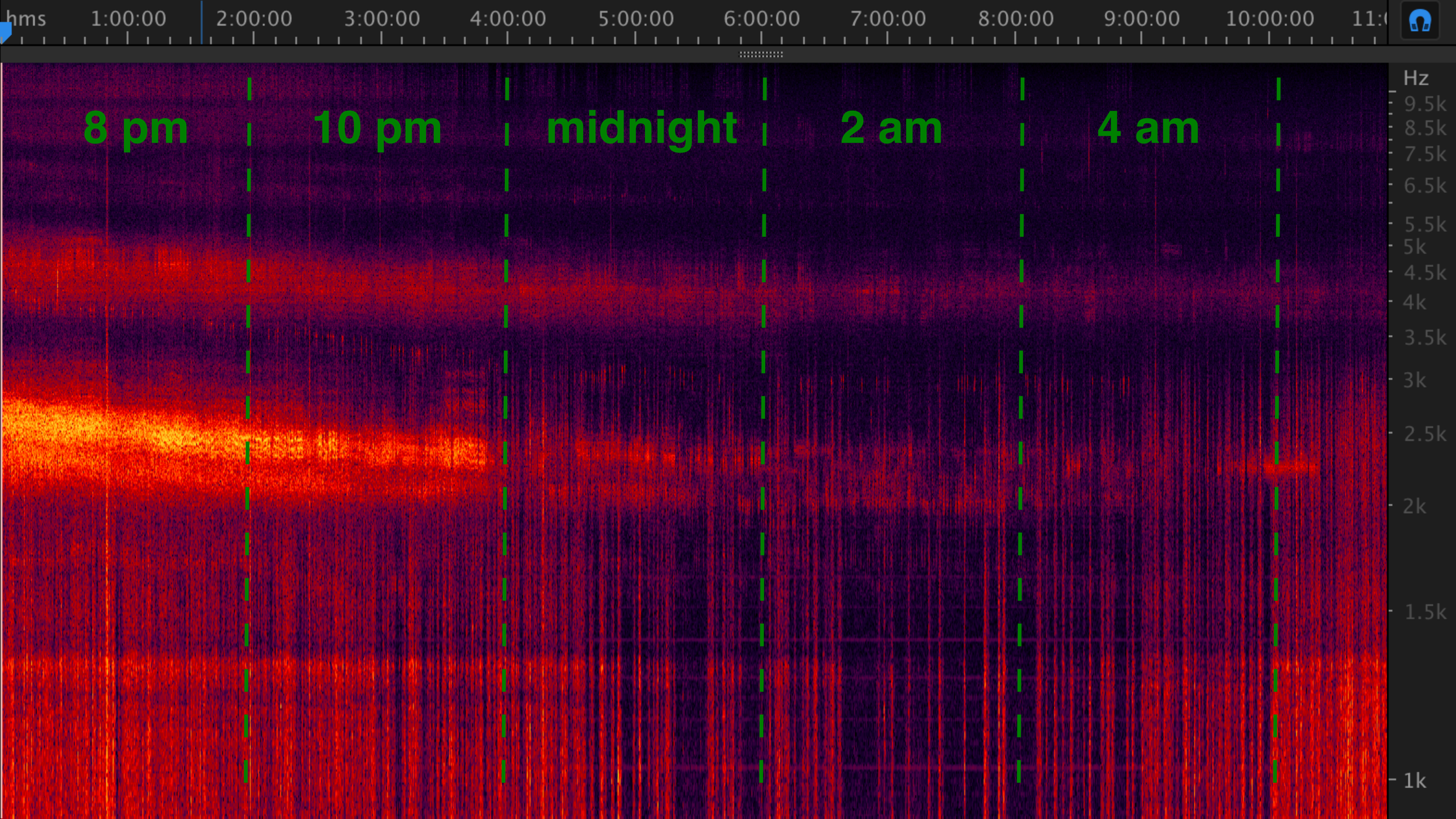
Black-throated Gray Warbler
(Setophaga nigrescens)

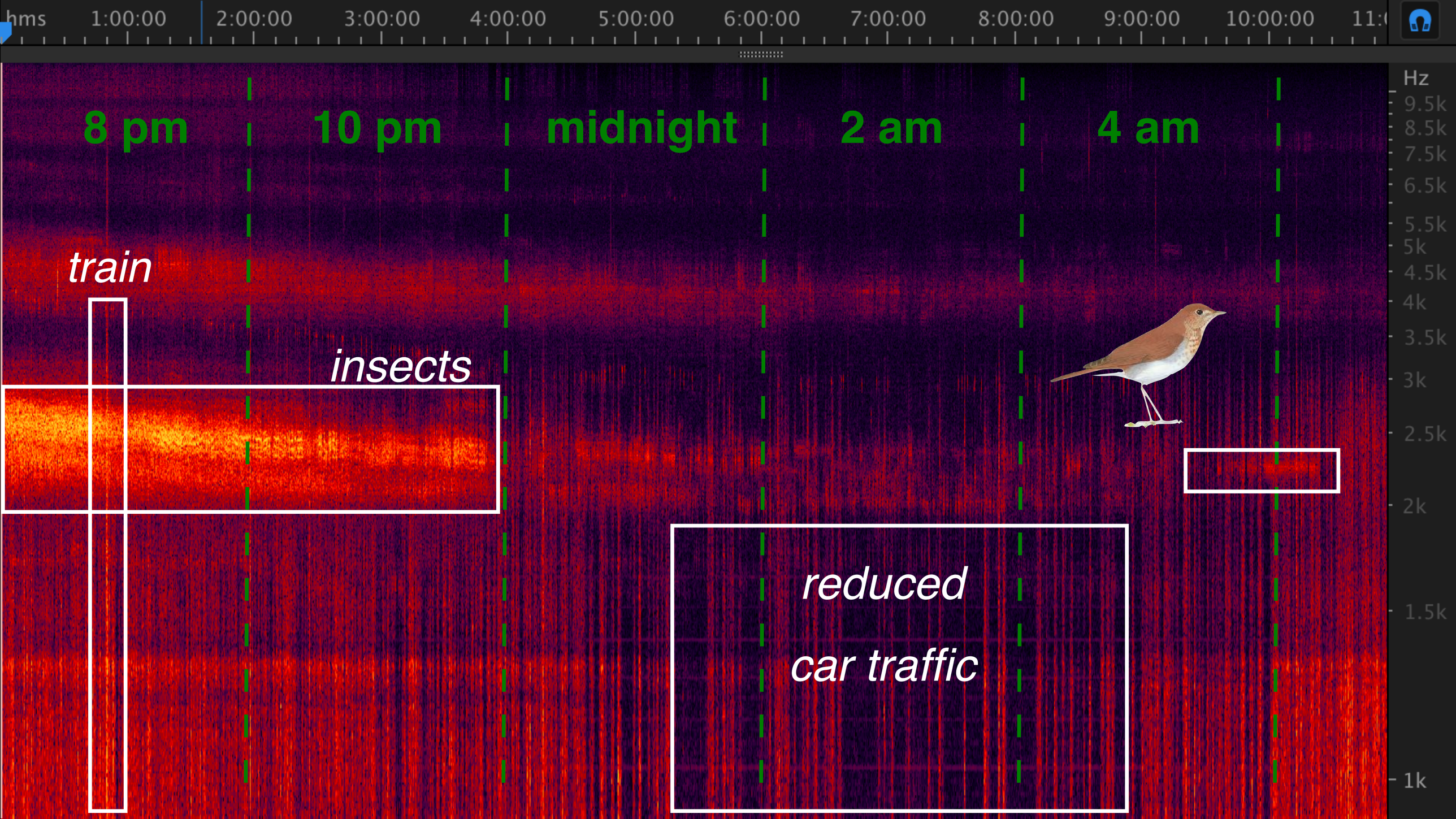
Bill Evans. www.oldbird.org

Bill Evans. www.oldbird.org

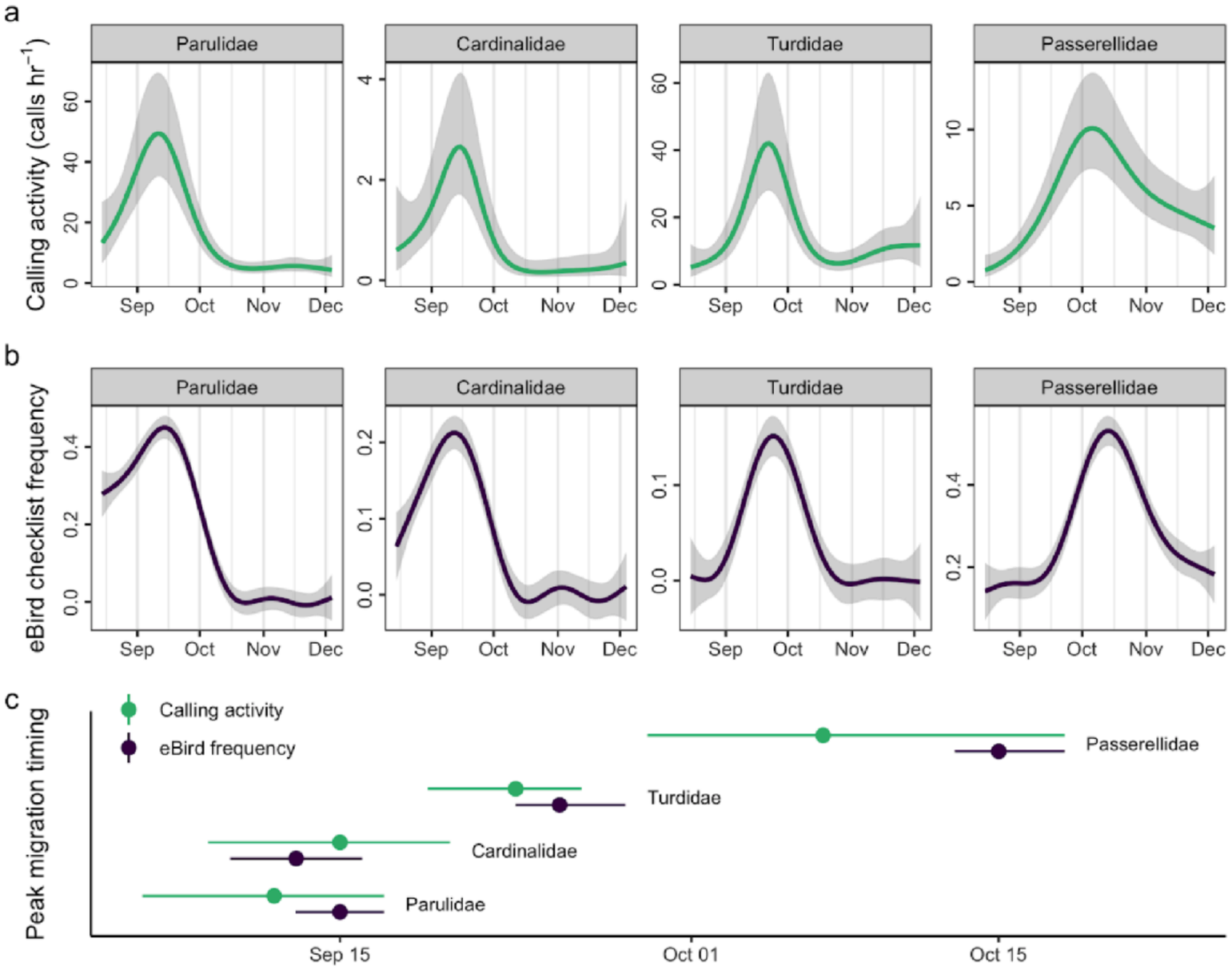
autonomous recording units


hms
1:00:00
2:00:00
3:00:00
4:00:00
5:00:00
6:00:00
7:00:00
8:00:00
9:00:00
10:00:00
Hz
9.5k
8.5k
7.5k
6.5k
5.5k
5k
4.5k
4k
3.5k
3k
2.5k
2k
1.5k
1k


hms
1:00:00
2:00:00
3:00:00
4:00:00
5:00:00
6:00:00
7:00:00
8:00:00
9:00:00
10:00:00
8 pm
10 pm
midnight
2 am
4 am
Hz
9.5k
8.5k
7.5k
6.5k
5.5k
5k
4.5k
4k
3.5k
3k
2.5k
2k
1.5k
1k


hms
1:00:00
2:00:00
3:00:00
4:00:00
5:00:00
6:00:00
7:00:00
8:00:00
9:00:00
10:00:00
8 pm
10 pm
midnight
2 am
4 am
Hz
9.5k
8.5k
7.5k
6.5k
5.5k
5k
4.5k
4k
3.5k
3k
2.5k
2k
1.5k
1k
train
insects
reduced
car traffic

How it began: bird migration monitoring



with

B. Van Doren

A. Dokter

A. Cramer

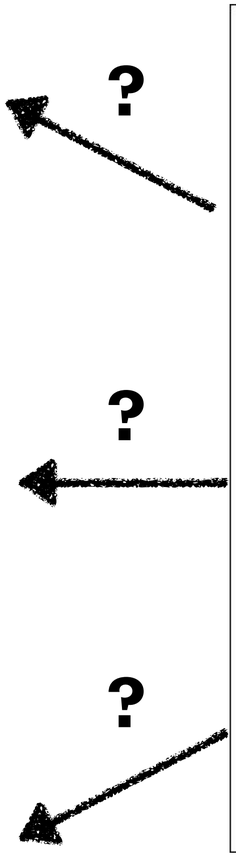
J. Salamon

A. Farnsworth

J. P. Bello

Generalizing to new species?

?



?

?

?

?

?

Idea: a new benchmark and open challenge

https://dcase.community/challenge2023

Coordinators

Ines Nolasco
Queen Mary University of London

Shubhr Singh
Queen Mary University of London

Vincent Lostanlen
Centre National de la Recherche Scientifique(CNRS)
Laboratoire des Sciences du Numérique de Nantes (LS2N)

Ariana Strandburg-Peshkin
University of Konstanz
Max Planck Institute of Animal Behavior

This task focuses on sound event detection in a few-shot learning setting for animal (mammal and bird) vocalisations. Participants will be expected to create a method that can extract information from five exemplar vocalisations (shots) of mammals or birds and detect and classify sounds in field recordings.

Challenge has ended. Full results for this task can be found in the Results page.

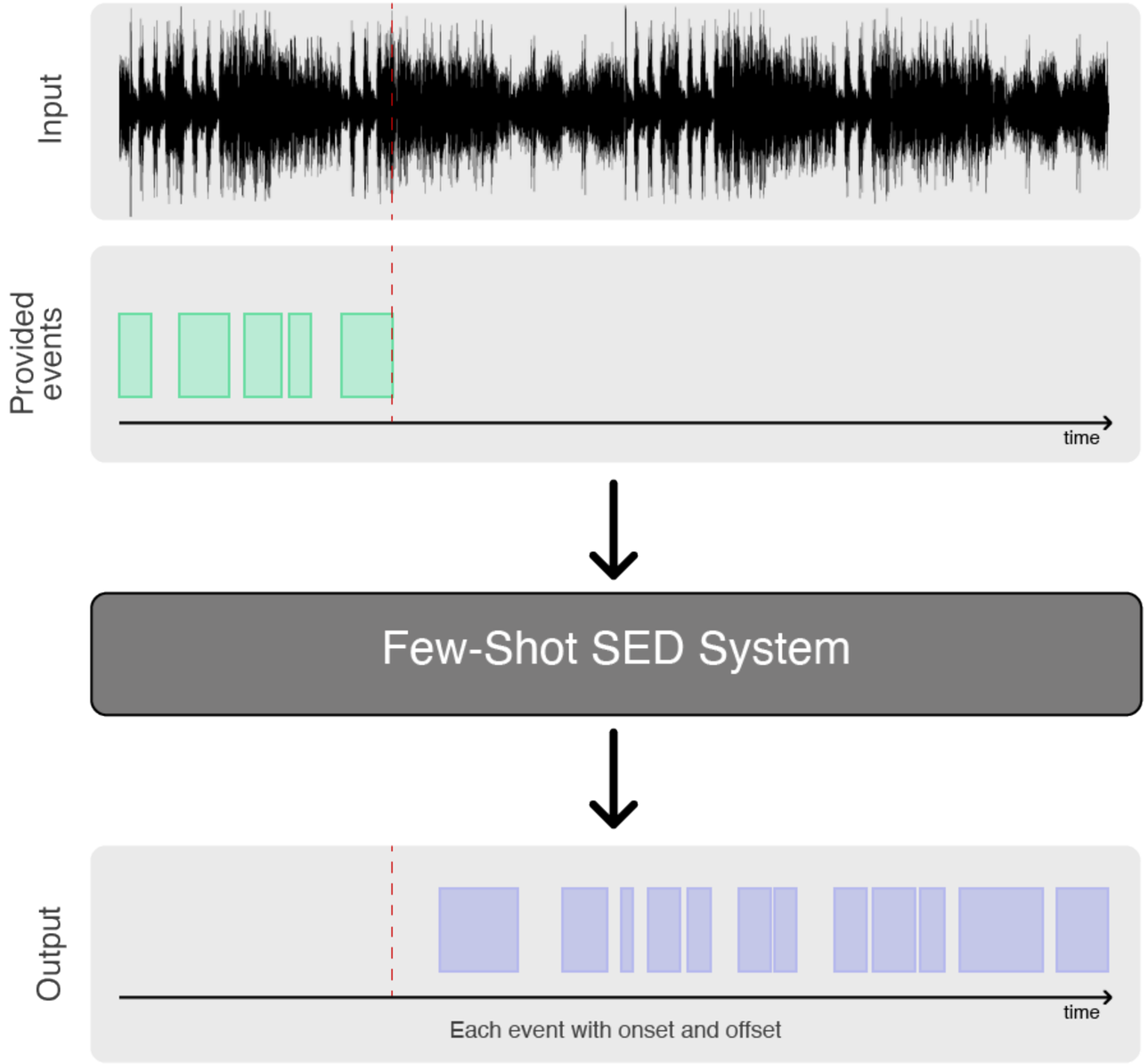
The development dataset has been changed. on 25th of April. Please download the new version from https://doi.org/10.5281/zenodo.6012309

If you are interested in the task, you can join us on **dedicated slack : task-fewshot-bio-sed**.

Multidisciplinary consortium

https://dcase.community/challenge2023

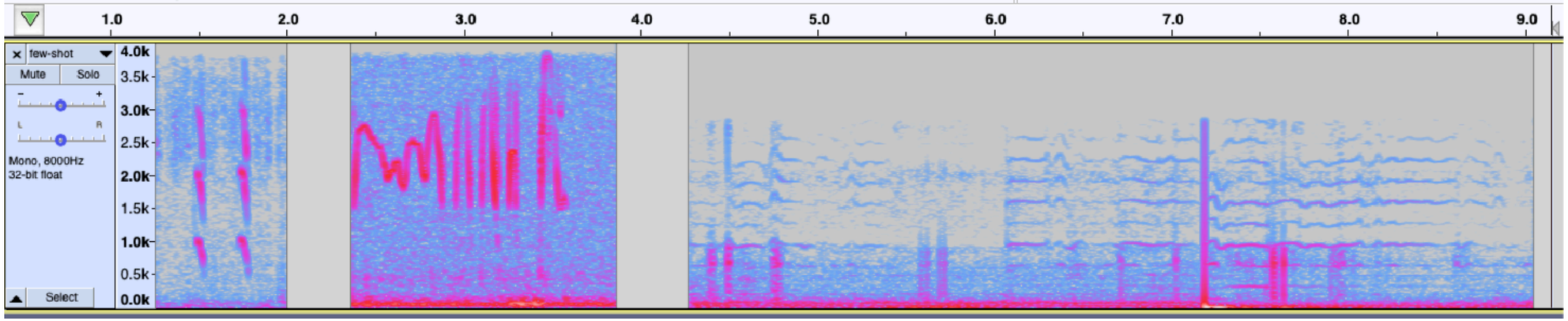
Task formulation

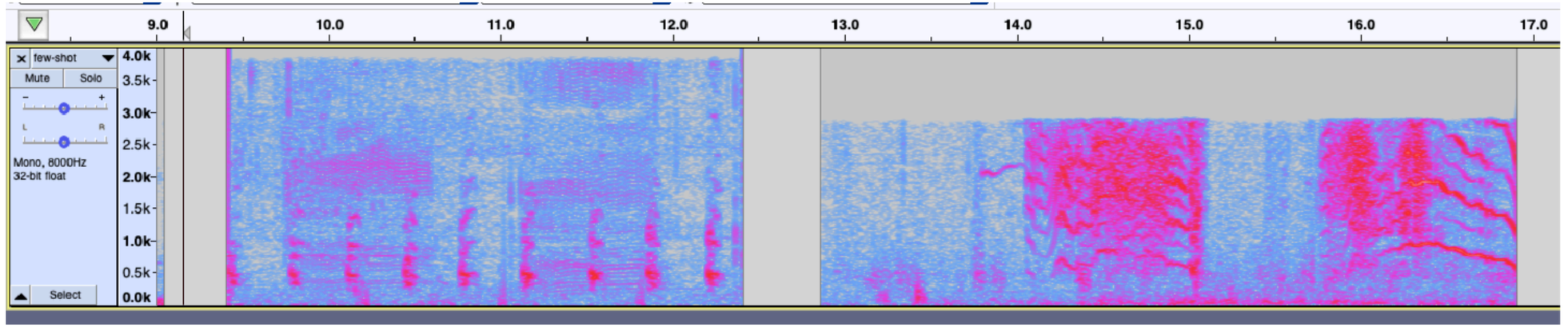


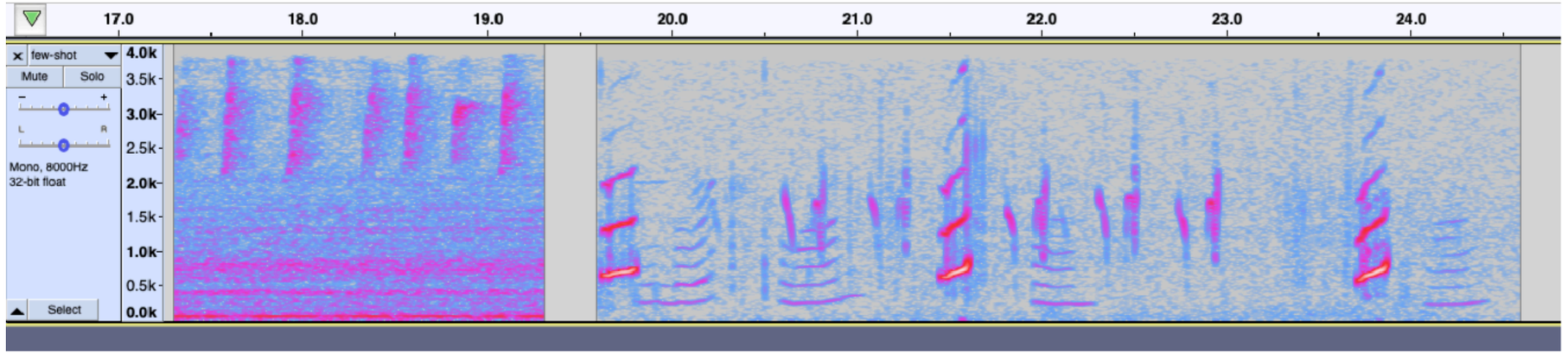
Onset-offset prompt

"Like tab completion for sound!"

Samples


1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
few-shot
Mute
Solo
Mono, 8000Hz
32-bit float
Select
4.0k
3.5k
3.0k
2.5k
2.0k
1.5k
1.0k
0.5k
0.0k


9.0
10.0
11.0
12.0
13.0
14.0
15.0
16.0
17.0
few-shot
Mute
Solo
Mono, 8000Hz
32-bit float
Select
4.0k
3.5k
3.0k
2.5k
2.0k
1.5k
1.0k
0.5k
0.0k


17.0
18.0
19.0
20.0
21.0
22.0
23.0
24.0
few-shot
Mute
Solo
Mono, 8000Hz
32-bit float
Select
4.0k
3.5k
3.0k
2.5k
2.0k
1.5k
1.0k
0.5k
0.0k

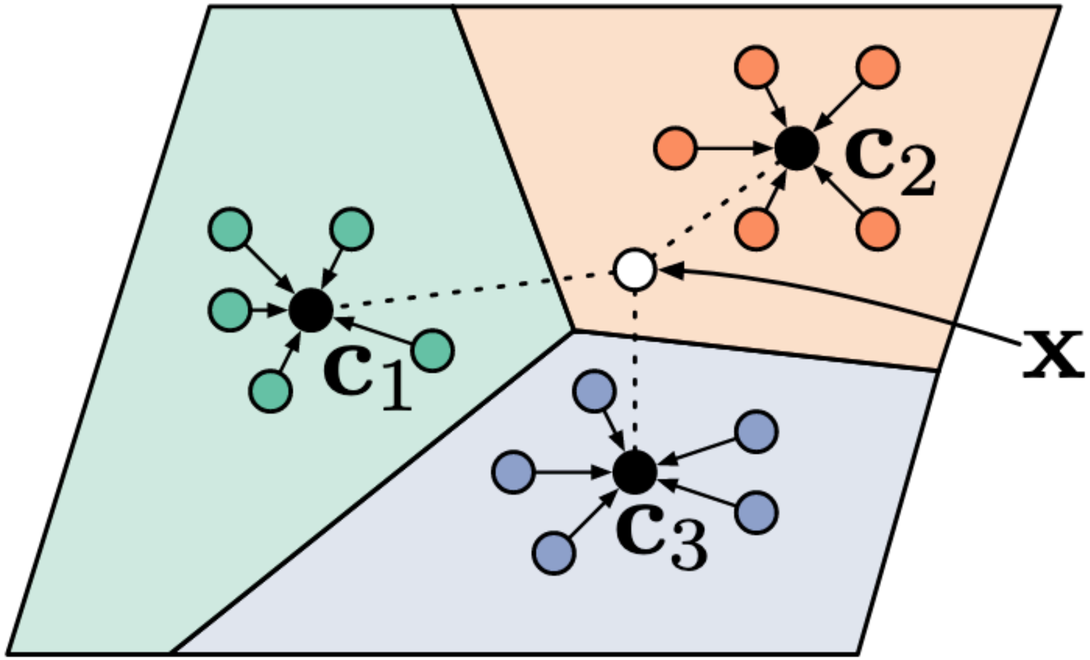
Baseline

Prototypical Networks for Few-shot Learning

Jake Snell
University of Toronto*
Vector Institute

Kevin Swersky
Twitter

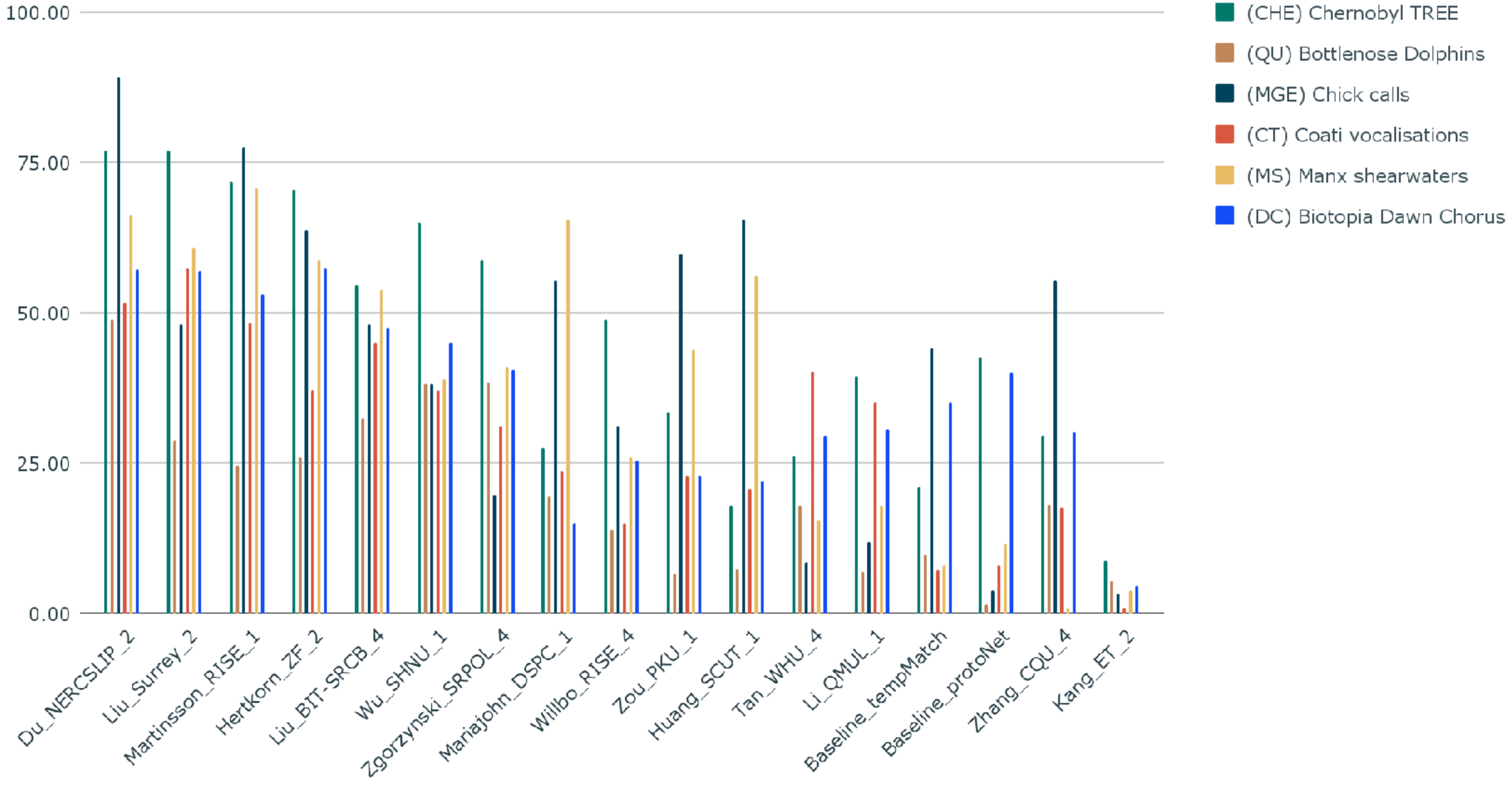
Richard Zemel
University of Toronto
Vector Institute
Canadian Institute for Advanced Research



2022 competitors

Team code	Code	Eval set: F-score % (95% CI)	Val set F-score %	Main characteristics
Du_NERCSLIP	(A)	60.22 (59.66-60.70)	74.4	CNN+ProtoNet; Frame-level embeddings; PCEN;
Liu_Surrey	(B)	48.52 (48.18-48.85)	50.03	CNN+ProtoNet; extra data; PCEN+$\triangle MFCC$; various post-process.
Martinsson_RISE	(C)	47.97 (47.48-48.40)	60	ResNet+ProtoNet; Ensemble(15) based input size; logMel+PCEN
Hertkorn_ZF	(D)	44.98 (44.44-45.42)	61.76	CNN; Frequency resolution preserving pooling; various post-process
Liu_BIT-SRCB	(E)	44.26 (43.85-44.62)	64.77	CNN+ProtoNet; Transductive inference
Wu_SHNU	(F)	40.93 (40.48-41.30)	53.88	ResNet+ProtoNet; Continual-learning; spectrogram
Zgorzynski_SRPOL	(G)	33.24 (32.69-33.69)	57.2	CNN+Siamese Networks; Emsemble (3) average event-length;
Mariajohn_DSPC	(H)	25.66 (25.40-25.91)	43.89	CNN+ProtoNet; logMel; augmentation with time-shifting and mirroring
Wilbo_RISE	(I)	21.67 (21.32-21.97)	47.94	ResNet+ProtoNEt; Semi-supervised; Melspect+PCEN; various post-process
Zou_PKU	(J)	19.20 (18.88-19.51)	51.99	CNN+protoNet; mutual information loss; time frequency masking + mixup
Huang_SCUT	(K)	18.29 (18.01-18.56)	54.63	Transductive inference + Adapted central difference convolution
Tan_WHU	(L)	17.22 (16.82-17.55)	54.53	CNN+ProtoNet pretrained; transductive inference; task adaptive features
Li_QMUL	(M)	15.49 (15.16-15.77)	47.88	CNN+protoNet; PCEN; time, frequency masking + time warping
baseline-TempMatch	[5]	12.35 (11.52-12.75)	3.37	Spectrogram Cross correlation
baseline-ProtoNet	[5]	5.3 (5.1-5.2)	28.45	ResNet+ProtoNet
Zhang_CQU	(N)	4.34 (3.74-4.56)	44.17	CNN+protoNet; Fine tunning with MIMI; PCEN
Kang_ET	(O)	2.82 (2.76-2.87)	-	CNN+ProtoNEt; pretrained ECAPA-TDNN; Fine-tuning; Specaugment

Leaderboard



Conclusion

- Machine listening has untapped potential in the life sciences.
- Spectrogram annotation is tedious and demands expertise.
- We have organized the **first challenge for few-shot bioacoustics**.
- SOTA has went from ~25% to ~75% F-score since 2021.
- We plan to add **new animal taxa in future years**.
- Workshop in Tampere (Finland) in October 2023

https://dcase.community/challenge2023