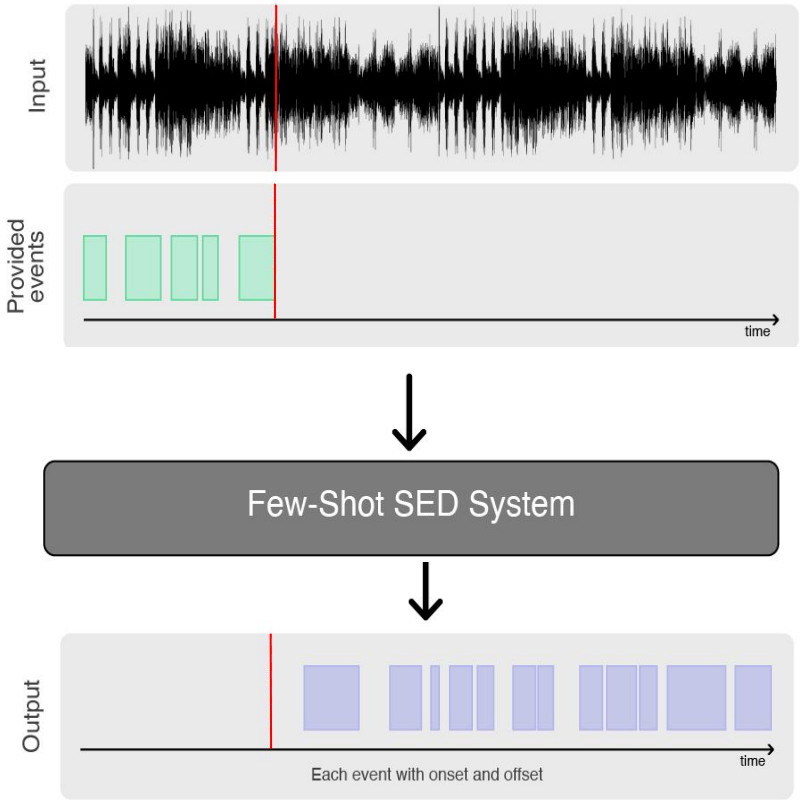
FEW-SHOT BIOACOUSTIC EVENT DETECTION

Ines Nolasco, Shubhr Singh, Ester Vidana Vila, Vincent Lostanlen, Ariana Strandburg-Peshkin, Emily Grout, Lisa Gill, Hanna Pamula, Joe Morford, Michael Emmerson, Frants Jensen, Helen Whitehead, Ivan Kiskin, Veronica Morfi and Dan Stowell.



Datasets








Training Set

- Multispecies flight calls (BV)
- Jackdaw calls (JD)
- Western Mediterranean Wetlands Bird calls (WMW)
- Hyena (HT)
- Meerkat (MT)

Validation Set

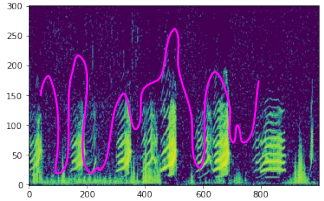
- Poland bird flight calls (PB)
- Meerkat stationary mics (ME)
- HumBug dataset (HB)

Evaluation Set

- Manx Shearwaters (MS)
- Chick calls (MGE)
- Biotopia dawn chorus birds (DC)
- Coati (south American mammal) (CT)
- Chernobyl TREE (chiffchaff, cuckoo) (CHE)
- Dolphin calls (underwater) (QU)

Baselines

- Template matching
- Prototypical networks



Evaluation Metric

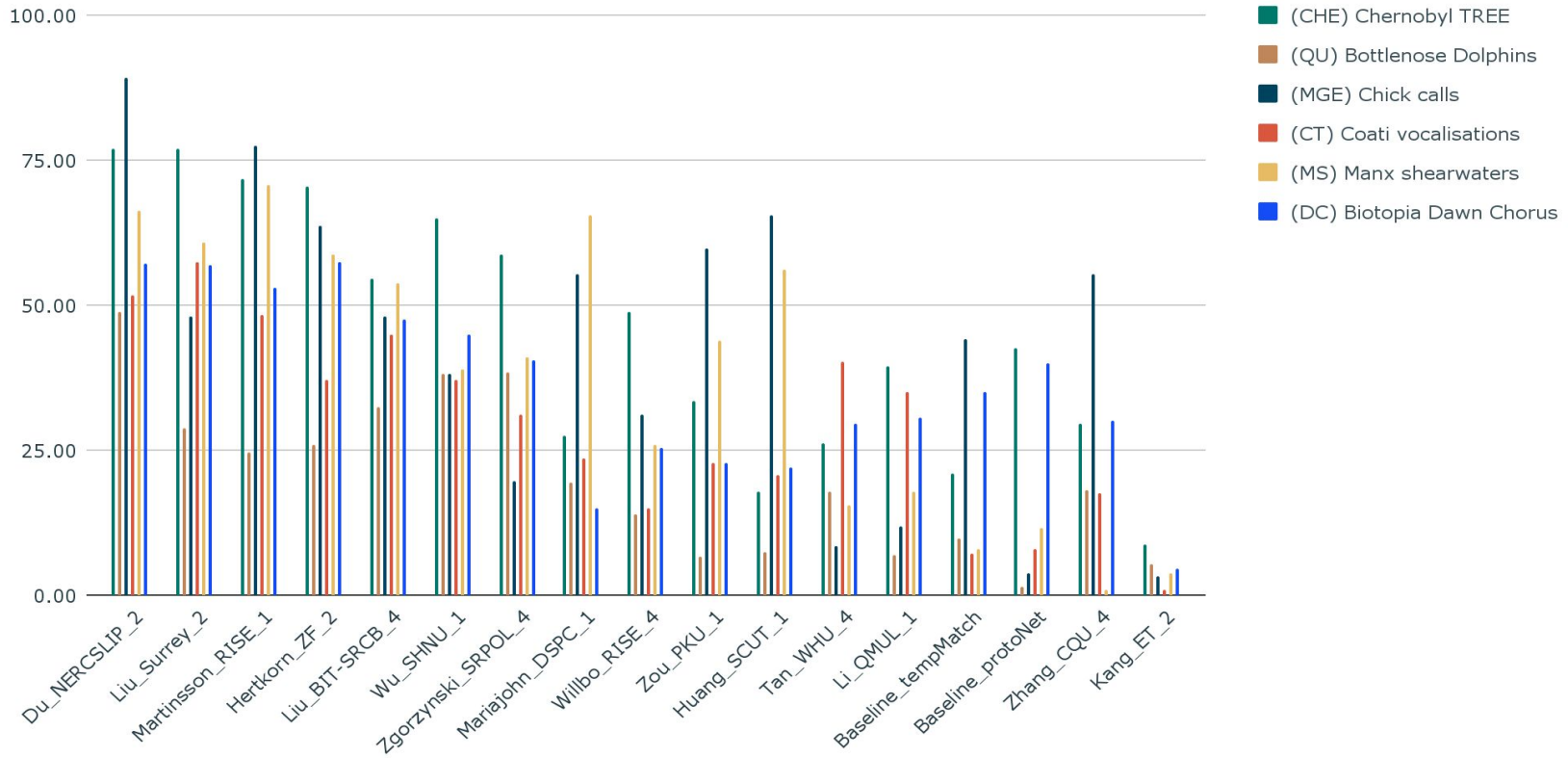
- Intersection over Union
- Bipartite graph matching
- Count TP, FP, FN
- Ranking given by F-score

Results

15 Teams, 46 systems submitted

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

F-score on best system per team and evaluation data subsets



Thank you to all the organizers and participants!