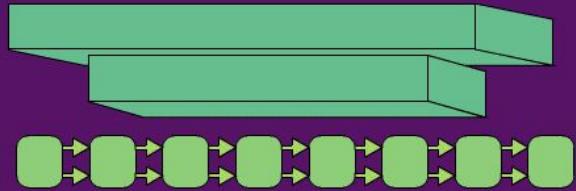




A neural network
framework for
researchers studying
acoustic behavior



nicholdav.info



NickleDave



@nicholdav



@nicholdav@fosstodon.org

David Nicholson, Yarden Cohen
Scipy 2023

Acknowledgements (part 1)

Dr. Yarden Cohen

Dept of Brain Sciences

Weizmann Institute of Science, Israel

<https://www.weizmann.ac.il/brain-sciences/labs/cohen/>



@YardenJCohen



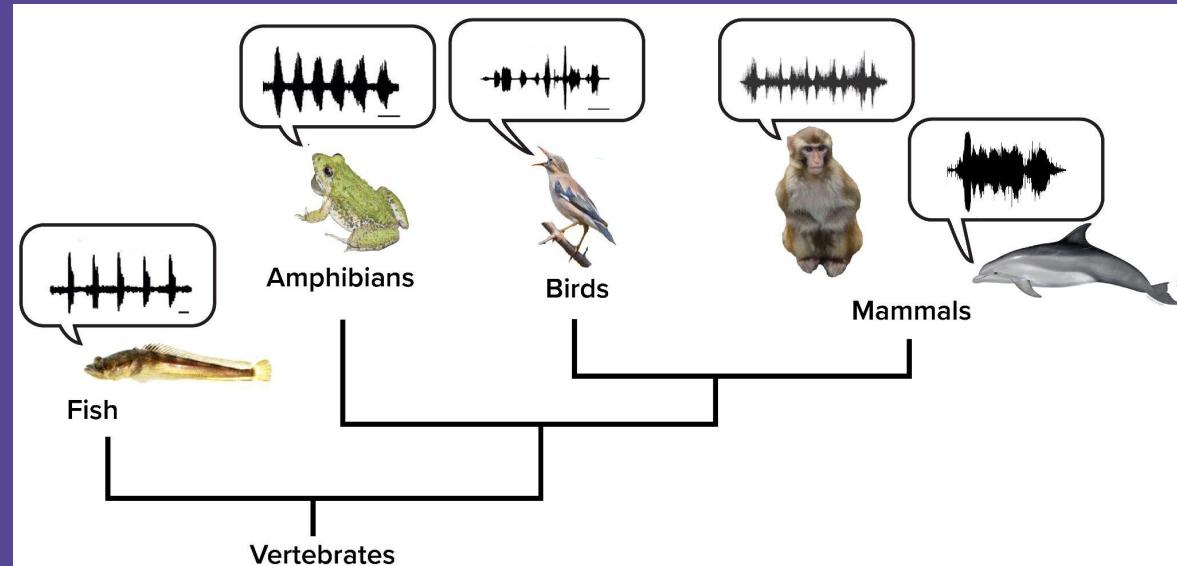
Introduction

Acoustic behavior and animal communication

What makes us human?

Language and speech

How is speech like birdsong
and bat calls?



Adapted from: Bass Chagnaud 2012, Branstetter et al. 2016,
Chen et al. 2020

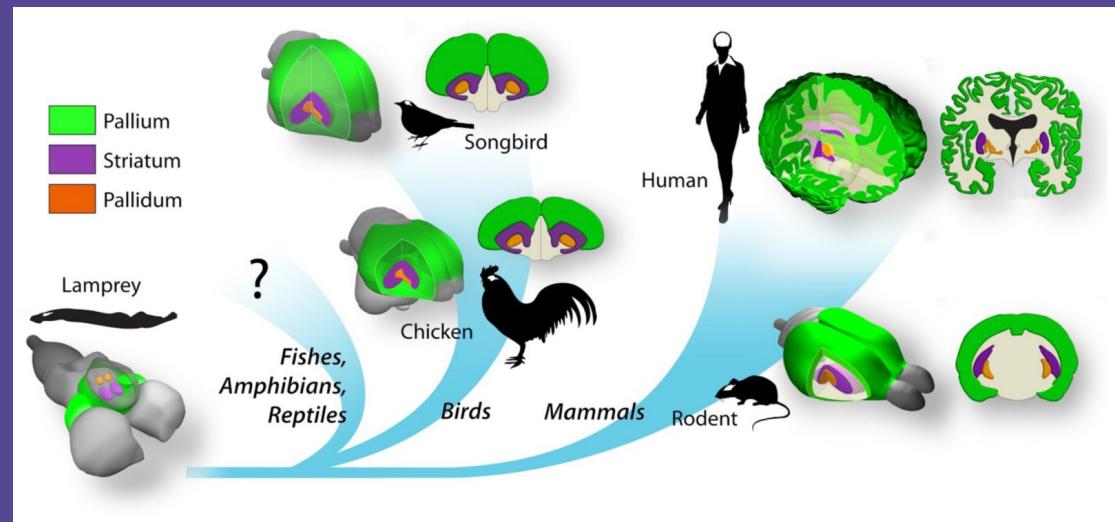
Acoustic behavior and animal communication

What makes us human?

Language and speech

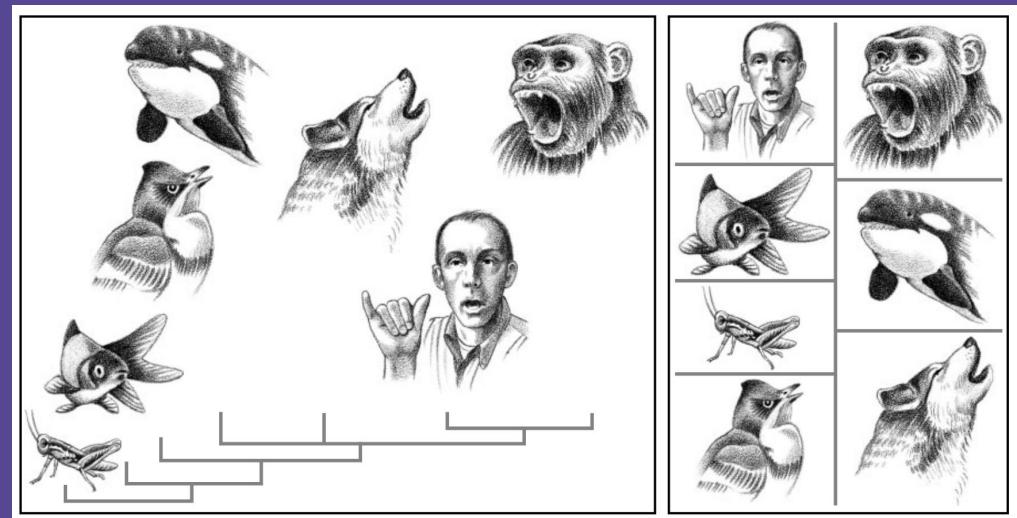
How is speech like birdsong
and bat calls?

How are the brain areas that
produce these behaviors
related?



Acoustic behavior and animal communication

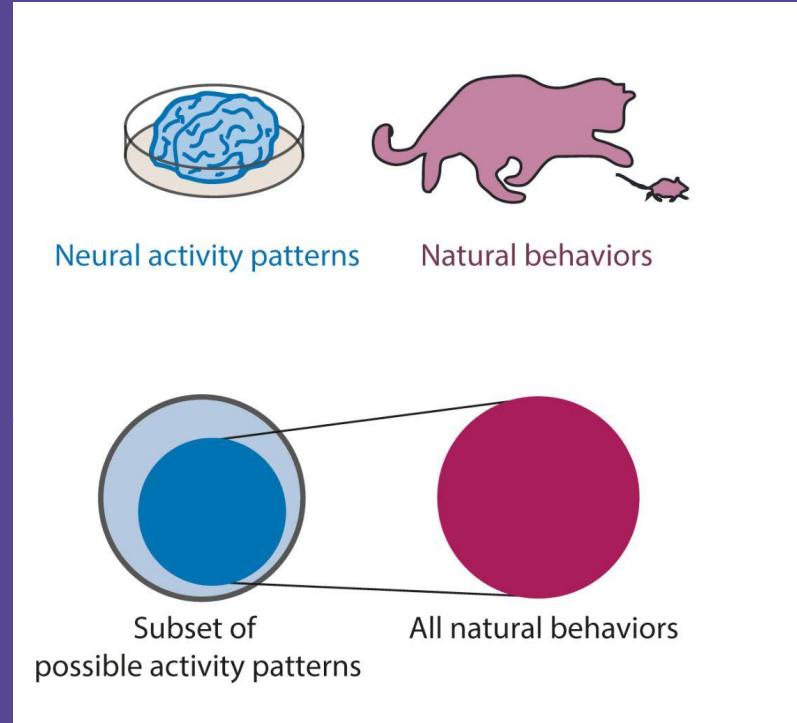
- Interdisciplinary (Wirthlin et al. 2019)
- Big team science (Hauser et al. 2002)
- Cutting edge computational methods
 - **Deep learning, AKA neural network models (Sainburg et al 2021, Stowell 2022)**



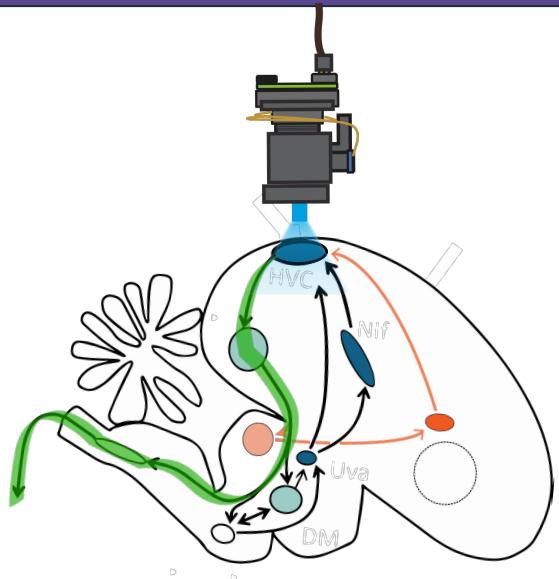
Hauser et al. 2002

Why "acoustic behavior"?

- We want to understand **behavior** and how the **brain produces it**
- First we must be able to **quantify** and **model** real world behavior



Why "acoustic behavior"?



The premotor song circuit



Why "acoustic behavior"?

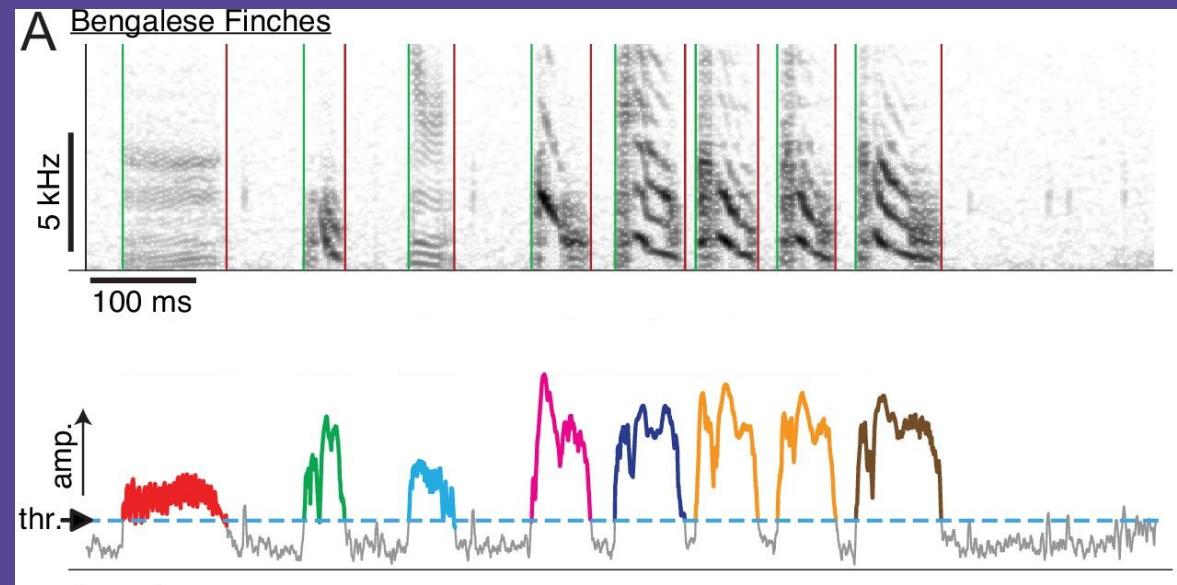
Quantifying and **modeling** real world
behavior requires

- Big data
- Automated analyses

Analyses of acoustic behavior

Many analyses consists of a workflow focused on sequences of *units*

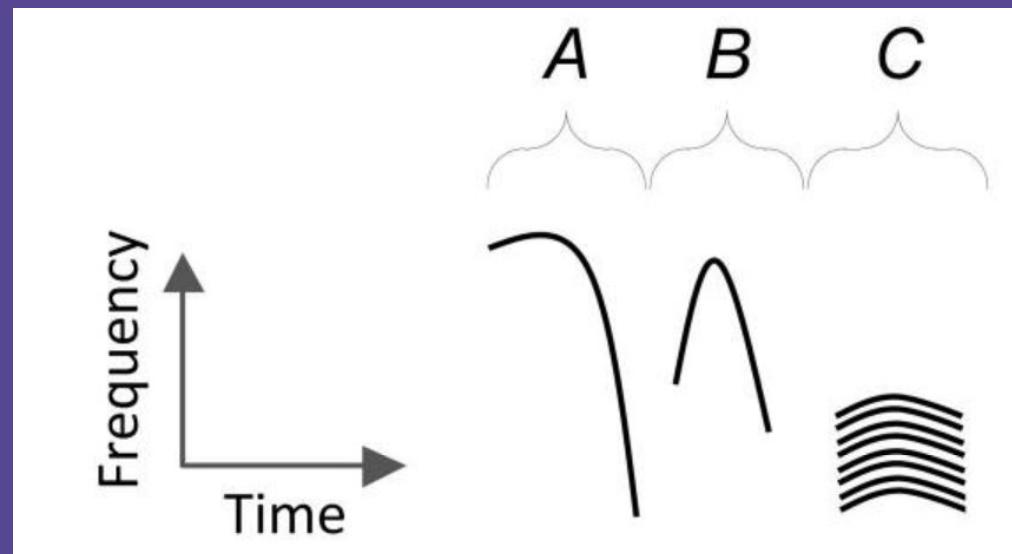
- By *segmenting*, e.g. using *silent gaps*



Analyses of acoustic behavior

Is this the right way to think about acoustic behavior?

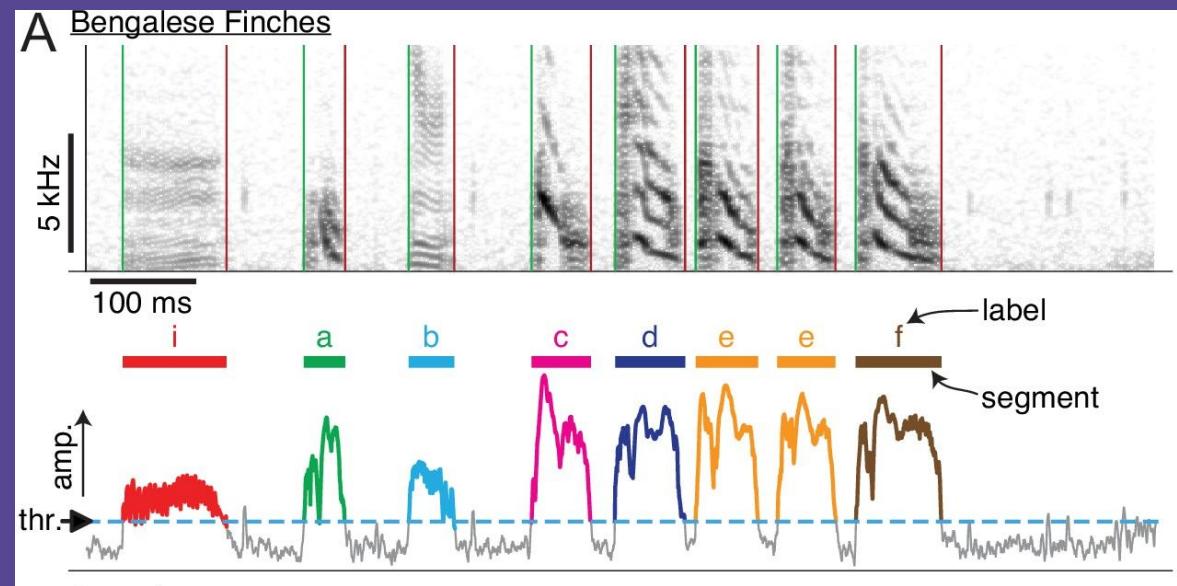
Unclear, in part because doing these analyses has been so hard



Kershenbaum et al. 2015

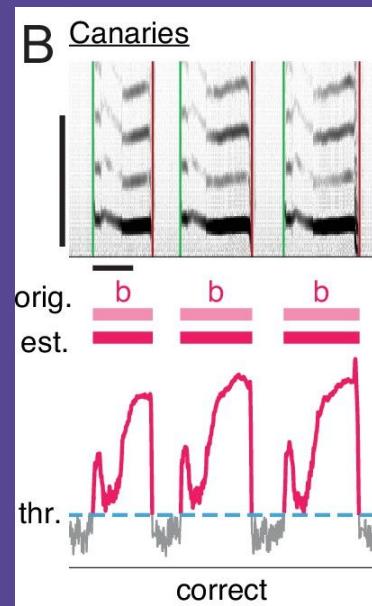
Analyses of acoustic behavior

- Get sequence of units by *segmenting*
- Some analyses require *annotating* the units



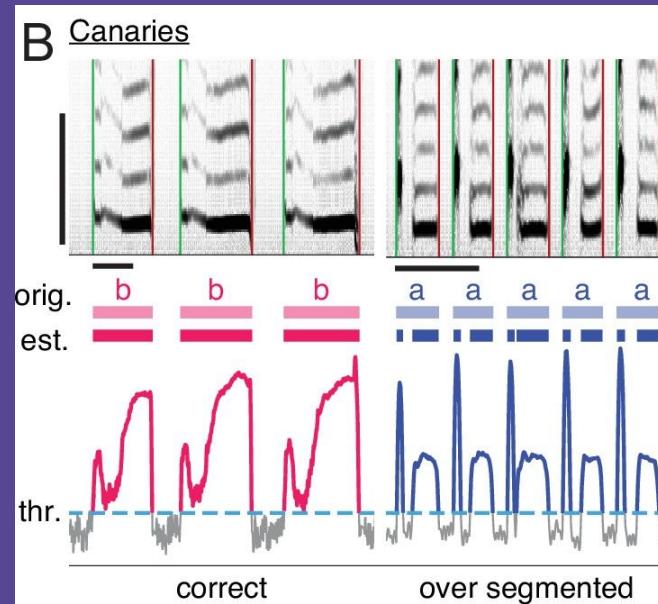
Analyses of acoustic behavior

Off-the-shelf signal processing algorithms for segmenting often don't work



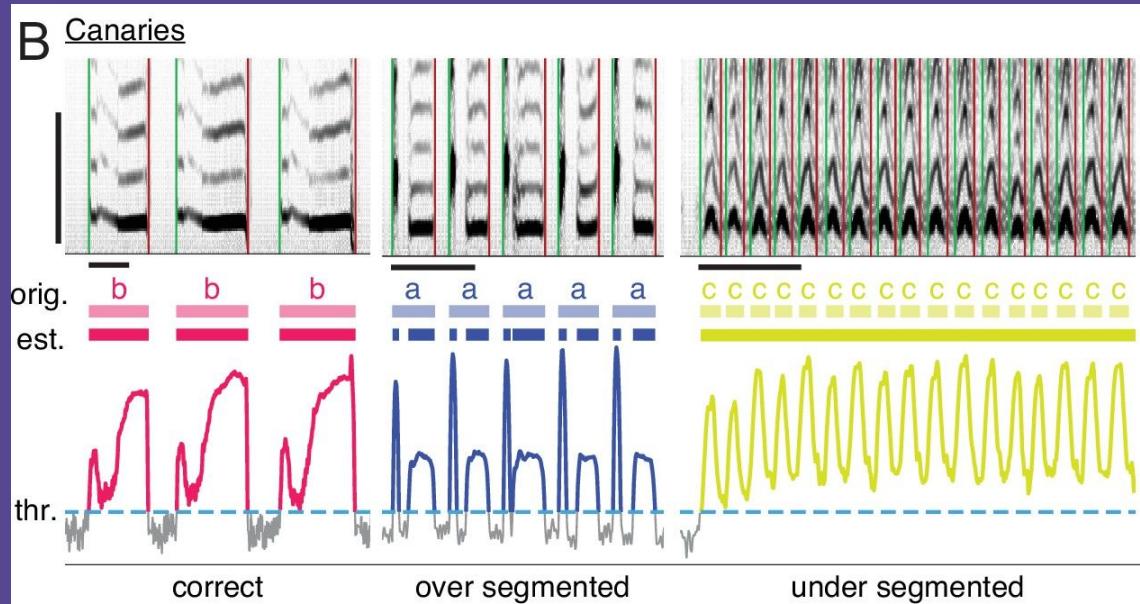
Analyses of acoustic behavior

Off-the-shelf signal processing algorithms for segmenting often don't work



Analyses of acoustic behavior

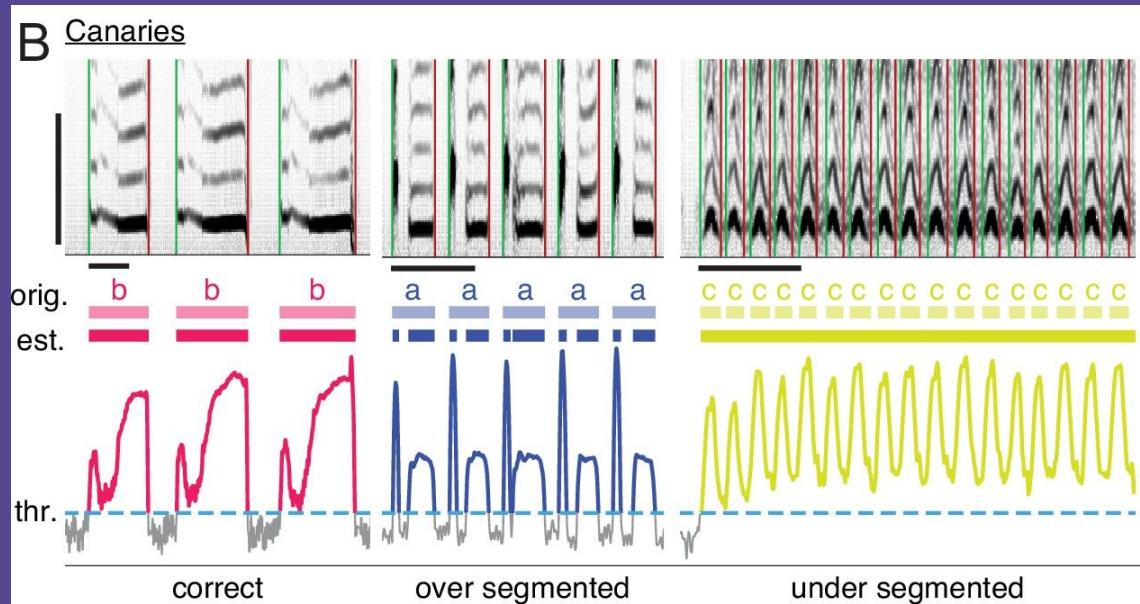
Off-the-shelf signal processing algorithms for segmenting often don't work



Analyses of acoustic behavior

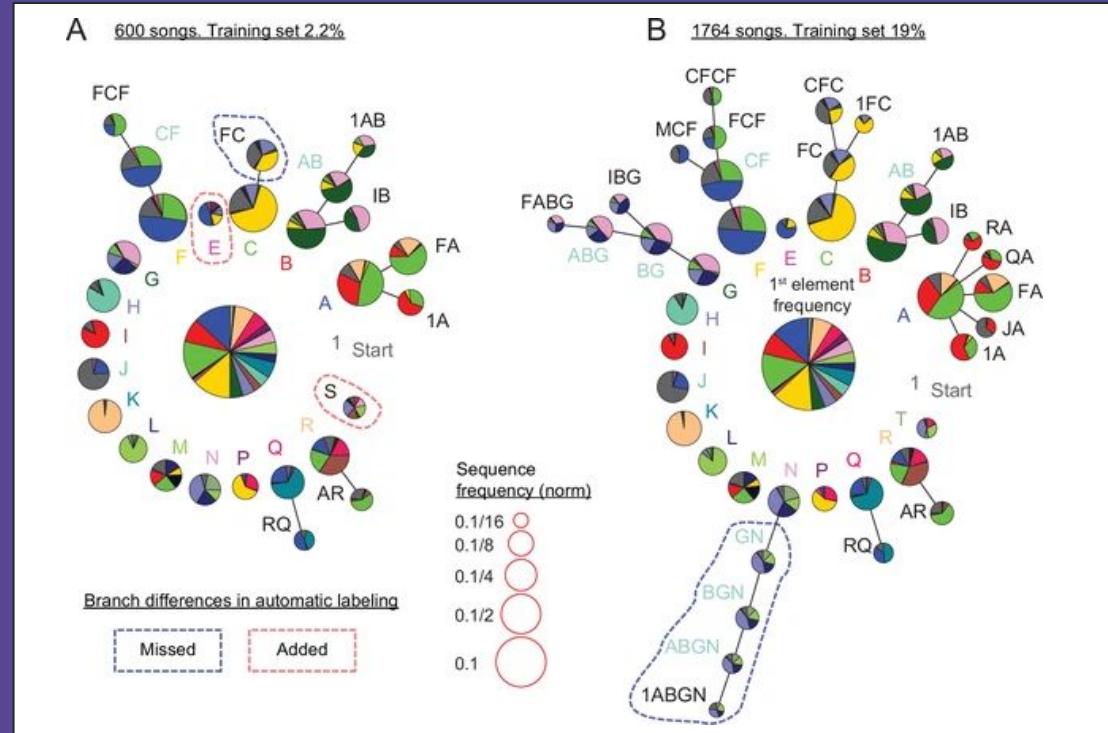
Off-the-shelf signal processing algorithms for segmenting often don't work

→ Neural network model



Annotation

For behavioral experiments
• e.g., models of syntax



Annotation

Key Problem:

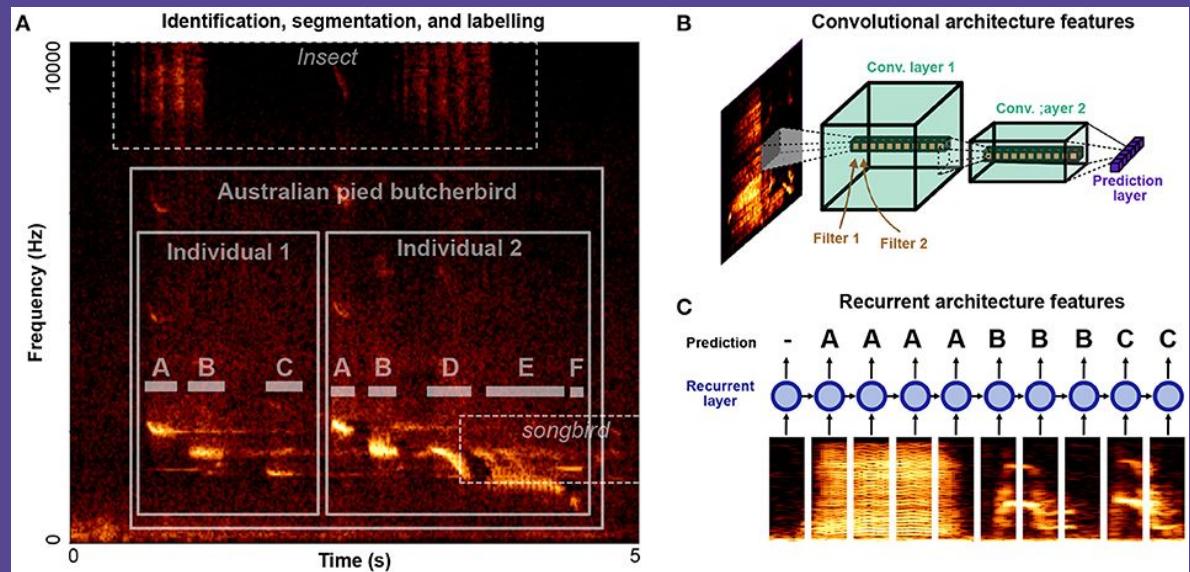
- Animals produce many more sounds than can be annotated by hand
 - e.g., birds sing 100s-1000s of songs a day



Neural network models for acoustic behavior

Multiple models proposed
for related tasks (Sainburg
Gentner 2021, Stowell
2021)

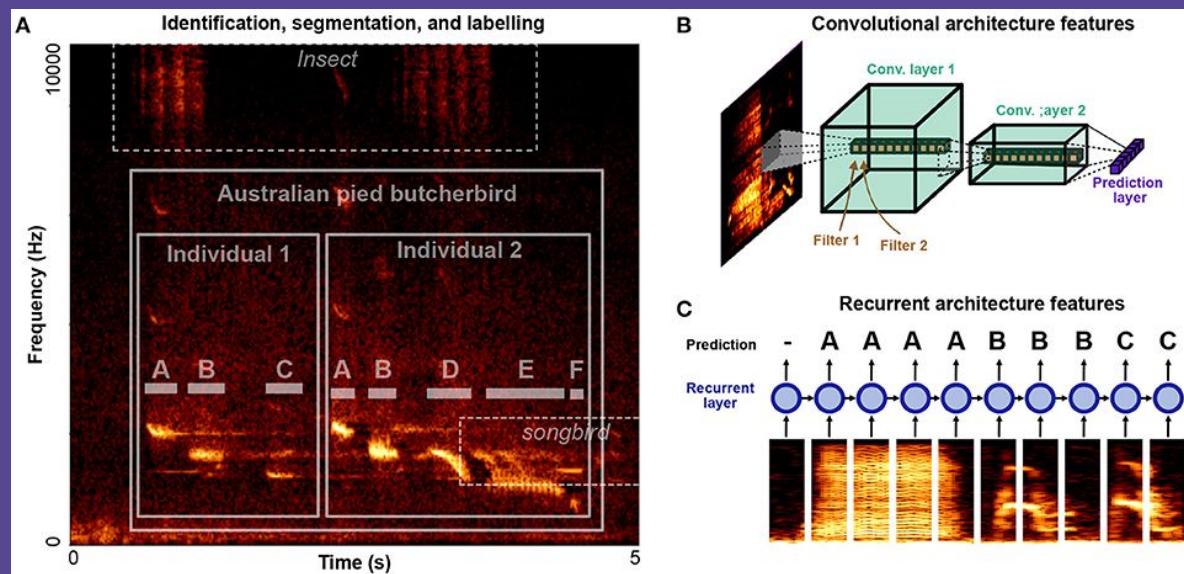
Often: familiar **computer
vision** models



Neural network models for acoustic behavior

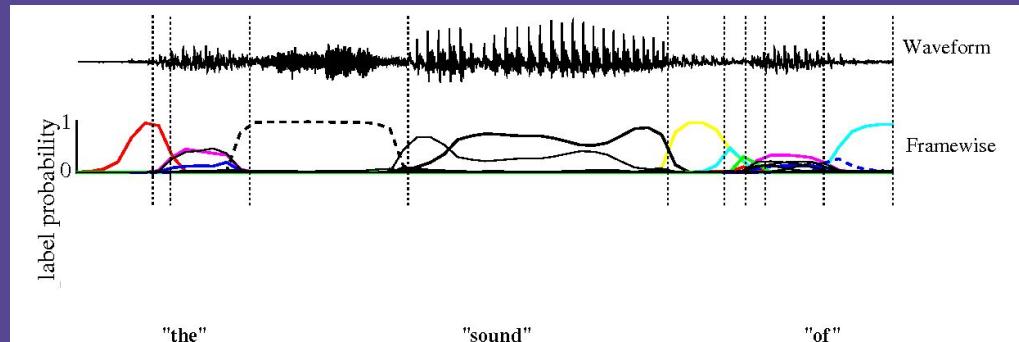
Papers describing these models have software written **with different frameworks**

- keras, tensorflow, pytorch, etc.



Neural network models for acoustic behavior

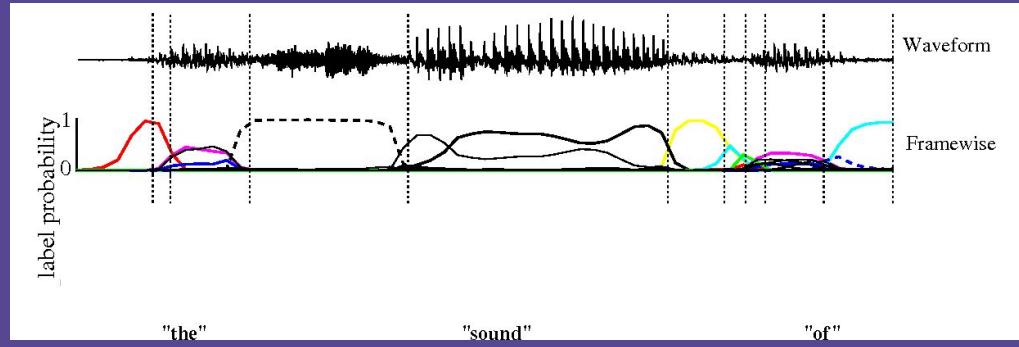
- Annotation as *frame classification* (Graves 2012)
 - a. Classify each frame
 - b. Recover sequences of labeled units by "segmenting" the classifications



Graves et al 2005

Neural network models for acoustic behavior

- Annotation as *frame classification* (Graves 2012)
- Multiple models proposed for this task
 - a. TweetyNet (Cohen 2021, 2022) -- Tensorflow, later PyTorch
 - b. DAS (Steinfath 2021, 2022) -- Keras
 - c. Koumura 2016 -- Java, custom

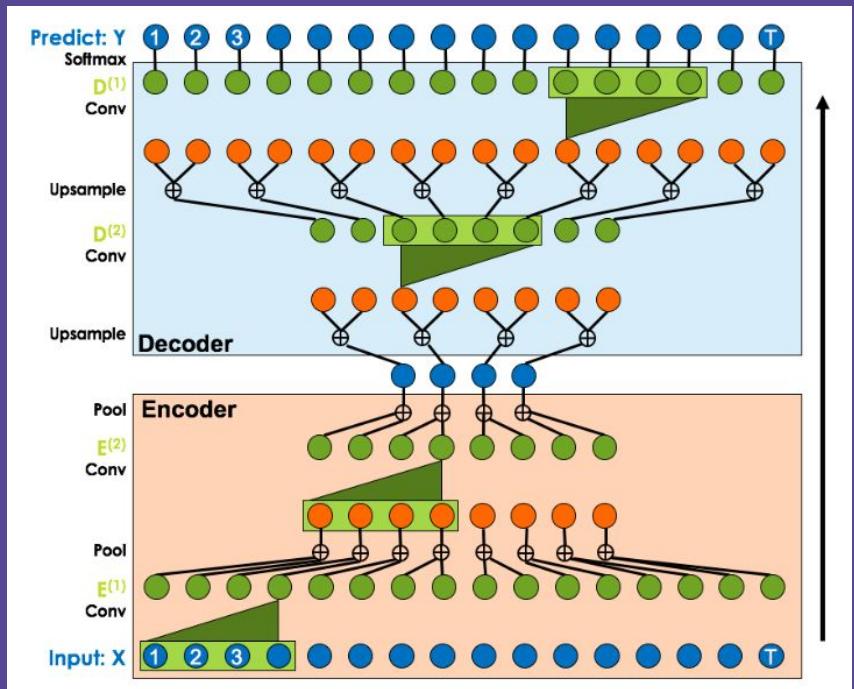


Graves et al 2005

Neural network models for acoustic behavior

Models from other literature could apply here as well:

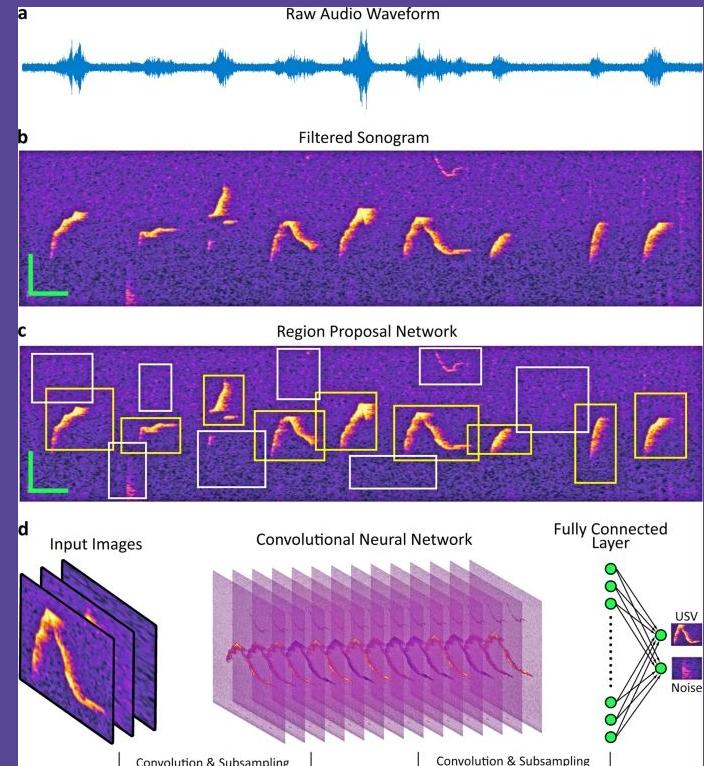
- *Audio event detection*
- *Action segmentation*
 - e.g. Encoder-Decoder Temporal Convolutional Network of Lea et al. 2017



Lea et al. 2017

Neural network models for acoustic behavior

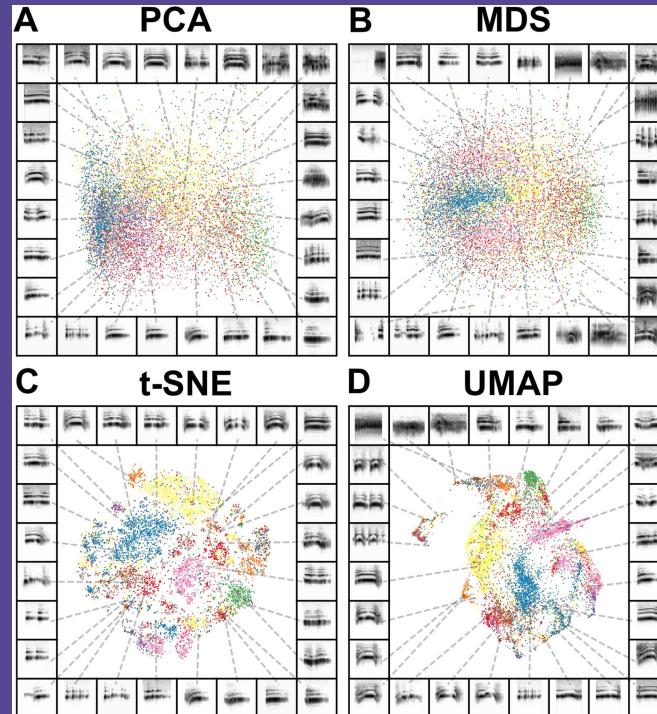
- Annotation as *object detection*
 - Detect units as bounding boxes
 - Often used with mouse ultrasonic vocalizations
- Examples
 - Deepsqueak (Matlab)



Coffey et al 2019

Neural network models for acoustic behavior

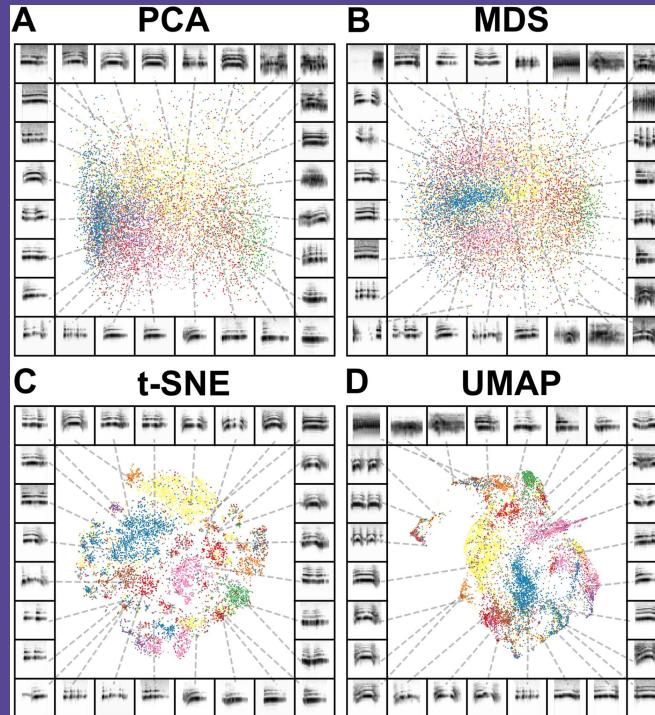
- Unsupervised models
 - For *dimensionality reduction*
 - to *cluster*
 - or measure *similarity*



Sainburg et al. 2020

Neural network models for acoustic behavior

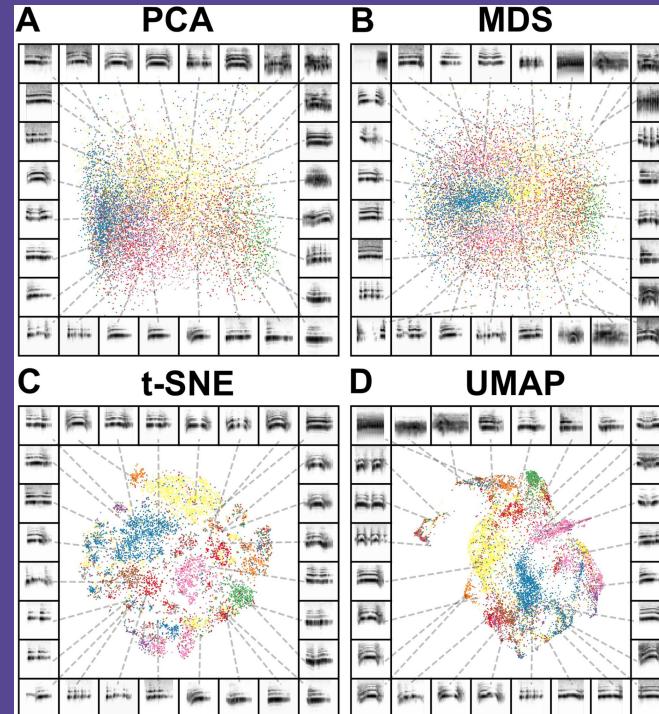
- Unsupervised models
 - Examples
 - Parametric UMAP (Sainburg, Gentner, McInnes)
 - Tensorflow, Pytorch
 - AVA (Goffinet et al. 2021)
 - Pytorch



Sainburg et al. 2020

Neural network models for acoustic behavior

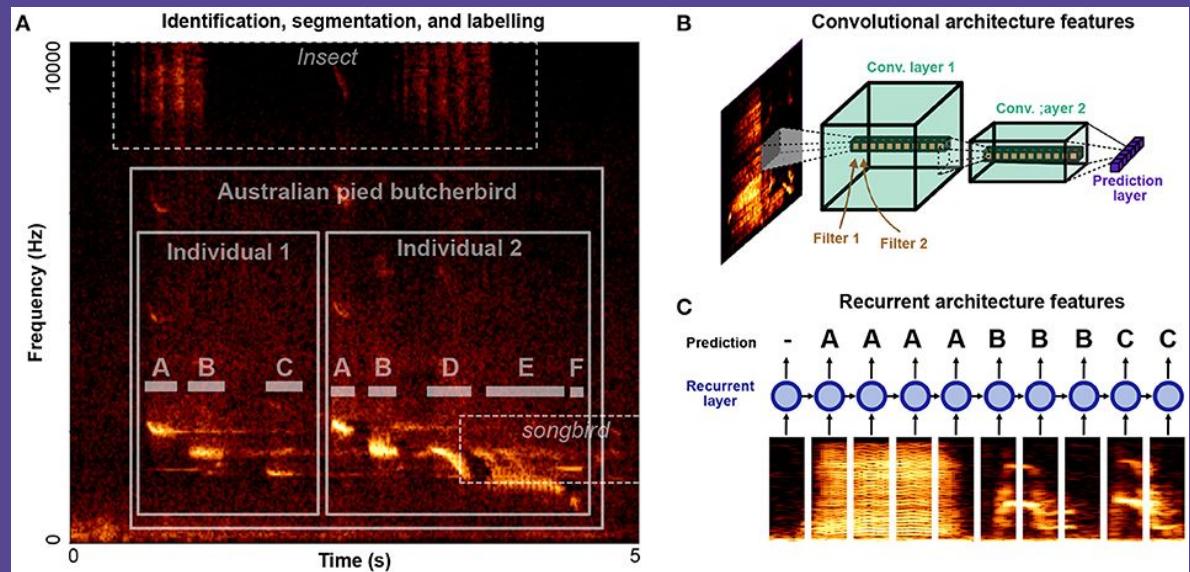
- Unsupervised models
 - Examples
 - UMAP (Sainburg, Gentner, McInnes)
 - Tensorflow, Pytorch
 - AVA (Goffinet et al. 2021)
 - Pytorch
 - Some of these methods **require** segmented audio
 - → Need to measure how results depend on segmentation



Sainburg et al. 2020

Neural network models for acoustic behavior

Need: single framework developed and maintained by research software engineers



Neural network models for acoustic behavior

Need: single framework developed and maintained by research software engineers



<https://us-rse.org/usrse23/>
#USRSECON2023

A neural network framework for acoustic behavior

- Single framework, multiple models
- Easily experiment with new models
- Low/no code
- Works with a wide range of data formats
- Easily understand how much (expensive, human annotated) data you need for your task



<https://github.com/vocalpy/vak>

A neural network framework for acoustic behavior

Single framework, multiple
models

```
@model('MyModelFamily')
class MyModelDefinition:
    network = MyNetwork
    loss = MyLoss
    optimizer = torch.optim.Adam
    metrics = {'acc': metrics.Accuracy}
    default_config = {
        'optimizer':
            {'lr': 1e-3}
    }
```

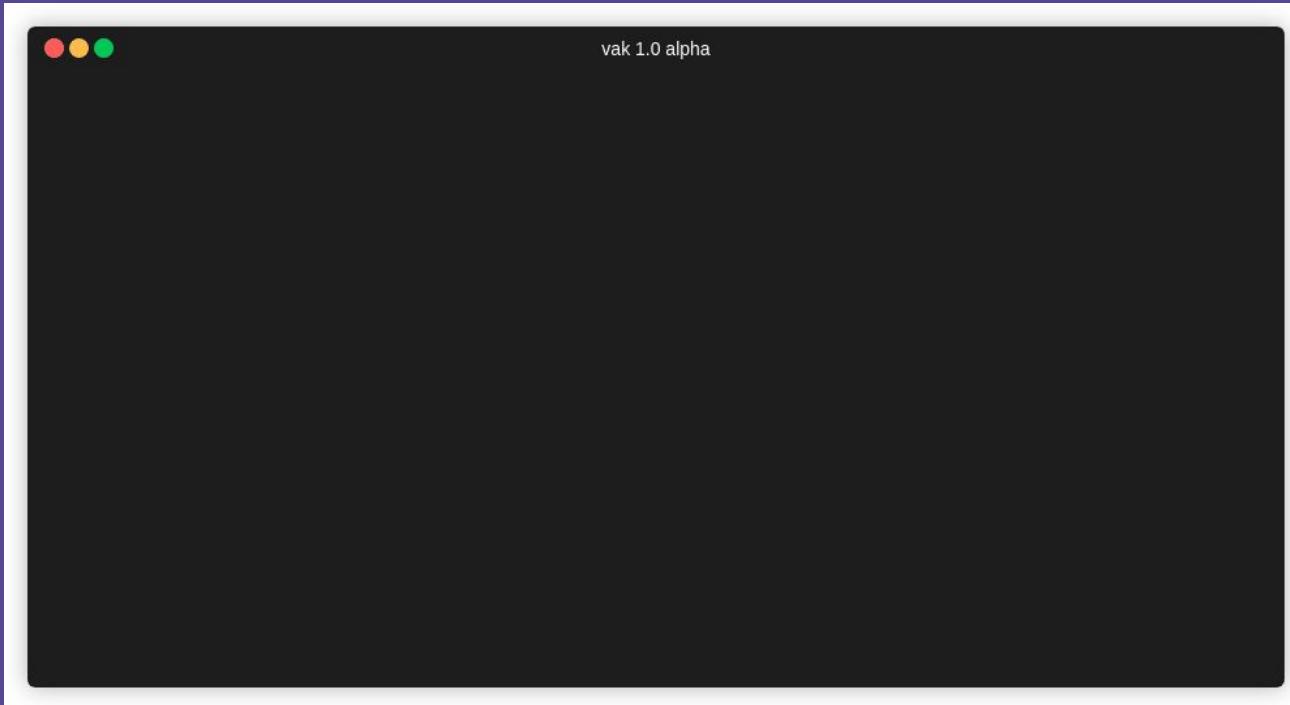
A neural network framework for acoustic behavior

Easily experiment with new
models

```
@model(family=FrameClassificationModel)
class TweetyNoLSTMNet:
    """TweetyNet model without LSTM layer"""
    network = TweetyNetNoLSTM
    loss = torch.nn.CrossEntropyLoss
    optimizer = torch.optim.Adam
    metrics = {'acc': metrics.Accuracy,
               'levenshtein': metrics.Levenshtein,*  
               'segment_error_rate': metrics.SegmentErrorRate,  
               'loss': torch.nn.CrossEntropyLoss}
    default_config = {
        'optimizer':
            {'lr': 0.003}
    }
```

A neural network framework for acoustic behavior

Low / no code, command-line interface, with just a few commands



<https://github.com/vocalpy/vak>

A neural network framework for acoustic behavior

Low / no code

- command-line interface
 - with just a few commands
 - TOML config files capture options
 - **reproducibility built in by default**

```
[PREP]
data_dir = "~/Documents/data/vocal/bird1/week2"
output_dir = "./data/prep/bird1"
audio_format = "wav"
annot_format = "simple-seq"
train_dur = 900
val_dur = 80
test_dur = 400

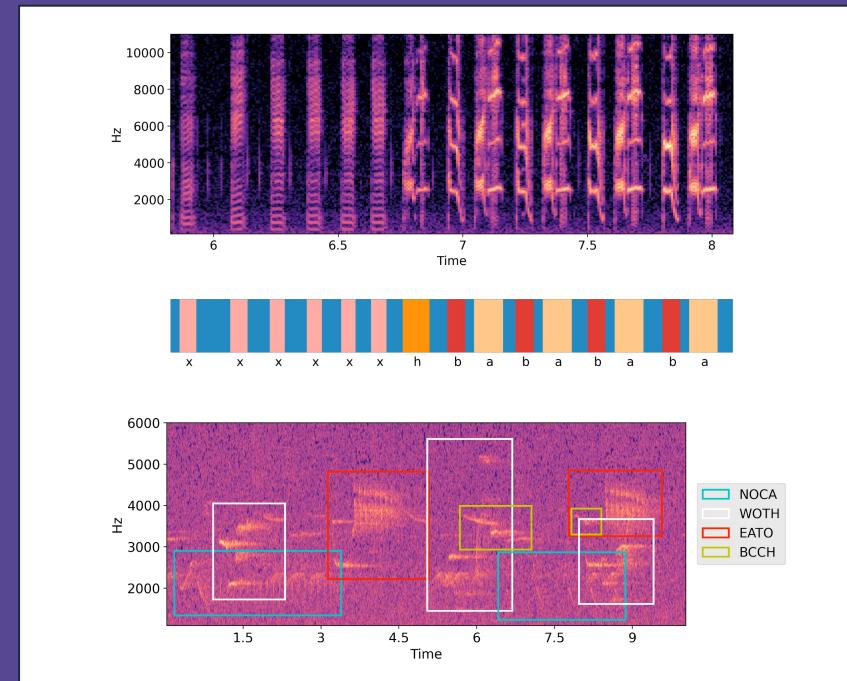
[SPECT_PARAMS]
fft_size = 512
step_size = 64
freq_cutoffs = [ 500, 10000, ]

[LEARNCURVE]
model = "TweetyNet"
train_set_durs = [ 30, 45, 75, 120, 180, 480, 600, ]
```

A neural network framework for acoustic behavior

—
Works with a wide array of data formats

- vak parses annotation formats with crowsetta
- now a pyOpenSci package
 - <https://www.pyopensci.org>
 - published in Journal of Open Source Software
- PyCon 2023 lightning talk:
https://youtu.be/54q_cPCNNS8?list=PL2_Uw4_HvXqvY2zhJ9AMUa_Z6dtMGF3gtb&t=1082



A neural network framework for acoustic behavior

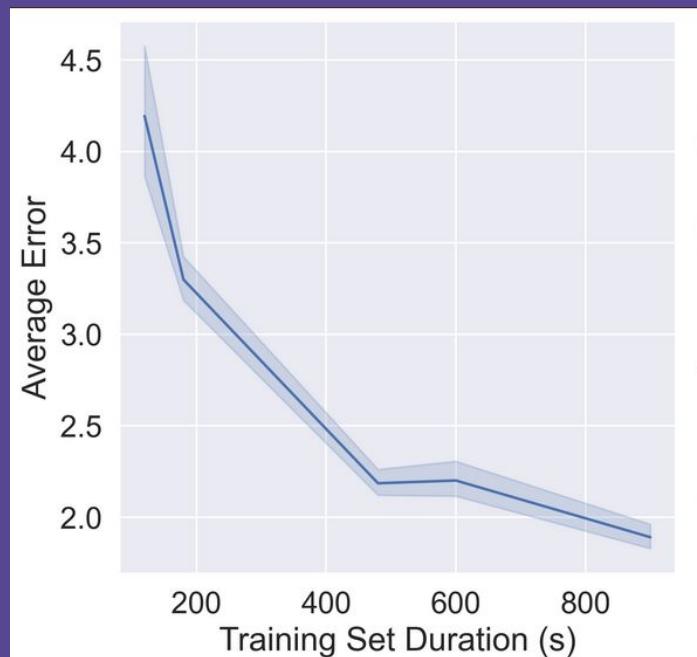
Works with a wide array of data formats

- audio formats
 - pysoundfile
- array formats
 - npz files for spectrograms
 - scipy.io.loadmat

A neural network framework for acoustic behavior

Easily understand how much training data you need for your task

```
[LEARNCURVE]
models = "TweetyNet"
train_set_durs = [ 30, 45, 75, 120, 180, 480, 600]
num_replicates = 10
```



A neural network framework for acoustic behavior

- Built on the scientific Python stack
 - numpy
 - scipy
 - matplotlib
 - pandas
 - dask
- Version 1.0
 - Lightning backend
 - Better abstractions for models, tasks, and datasets



A neural network framework for acoustic behavior

Design / API

- like other torch libraries for domain models
 - `models` module
 - `family`
 - `transforms` module
 - `frame labels`
 - `nn` module
 - `loss` module

Publications using vak

- vak has been used by at least ~20 labs we have worked with directly, and are basis for several papers since.
- <https://vak.readthedocs.io/en/latest/reference/about.html#publications-using-vak>

Low-dimensional learned feature spaces
quantify individual and group differences in
vocal repertoires

Jack Goffinet, Samuel Brudner, Richard Mooney, John Pearson 

Shared mechanisms of auditory and non-auditory vocal learning in the songbird brain

James N McGregor , Abigail L Grassler, Paul I Jaffe, Amanda Louise Jacob, Michael S Brainard, Samuel J Sober

The impacts of fine-tuning, phylogenetic distance, and sample size on big-data bioacoustics

Kaiya L. Provost , Jiaying Yang, Bryan C. Carstens



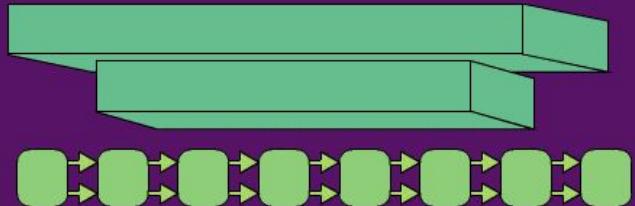
<https://github.com/vocalpy/vak>

Results: vak o.x

TweetyNet

vak was developed for: "Automated annotation of birdsong with a neural network that segments spectrograms", Cohen et al., 2021

- <https://elifesciences.org/articles/63853>
- <https://github.com/yardencsGitHub/tweetynet>



How to automate annotation?

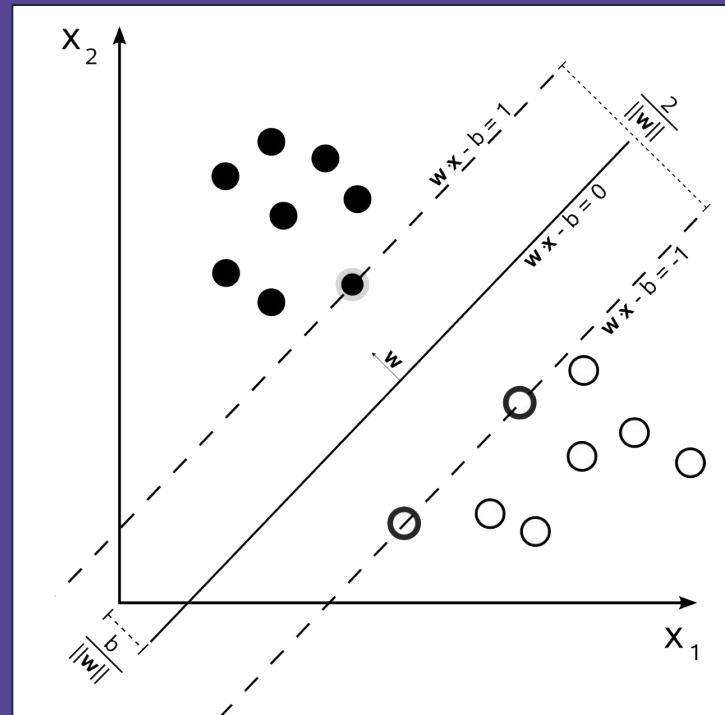
- Machine learning methods with engineered features?
 - e.g. Support Vector Machine (SVM), as in Tachibana et al. 2014
- Problem:
 - **Methods that use features extracted from segments can fail when segmentation is noisy**



<https://bit.ly/43fSFzT>



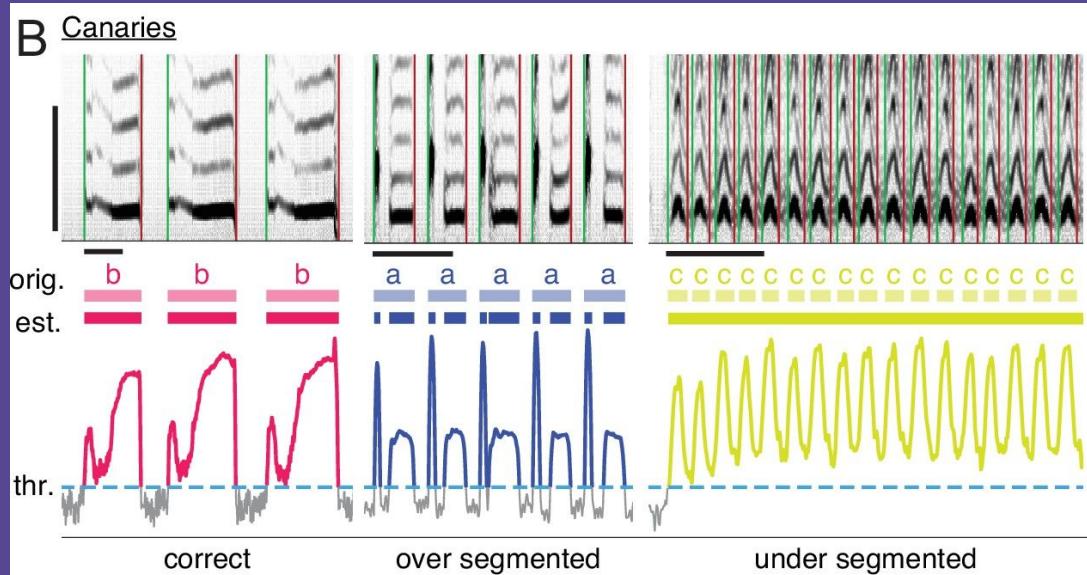
<https://youtu.be/BwNeVNou9-s>



How to automate annotation?

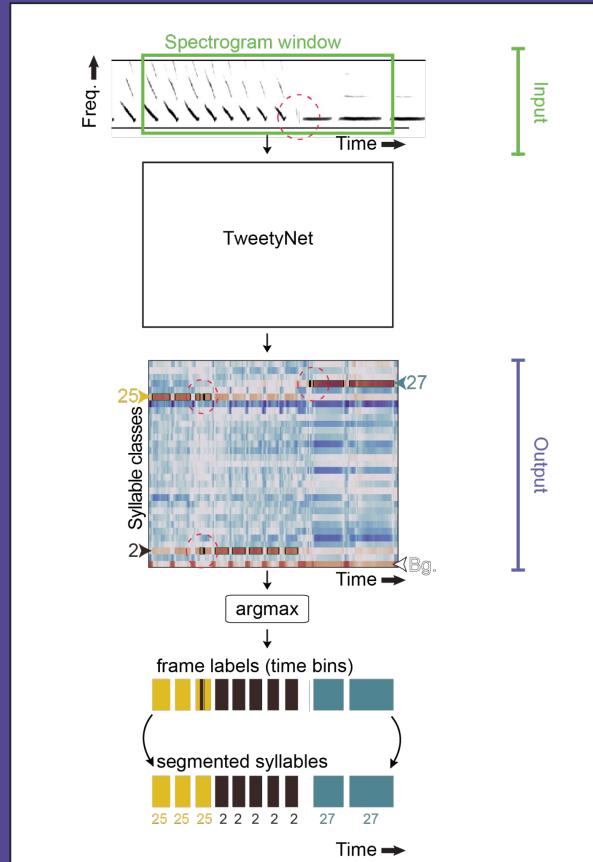
Problem:

- Methods that use features extracted from segments can fail when segmentation is noisy



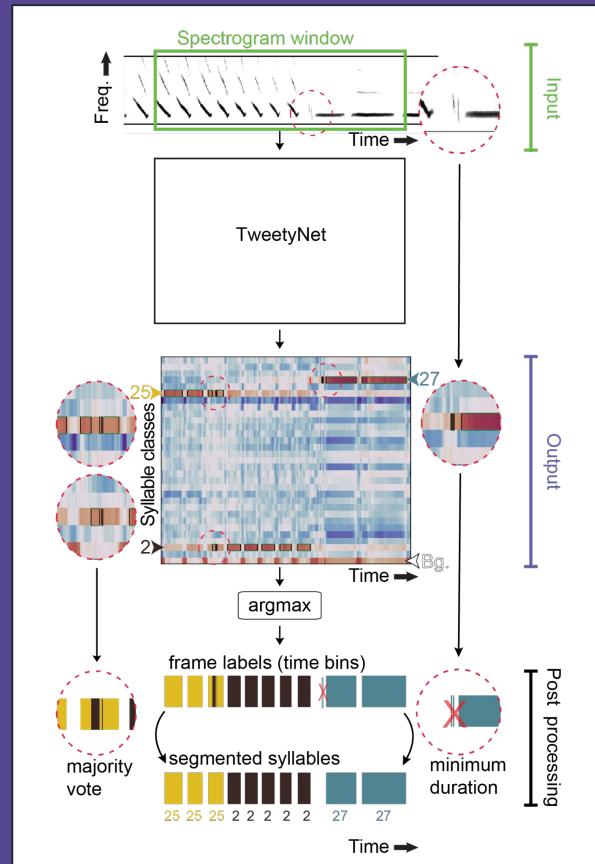
TweetyNet

A way to automate
annotation without
needing cleanly
segmented audio



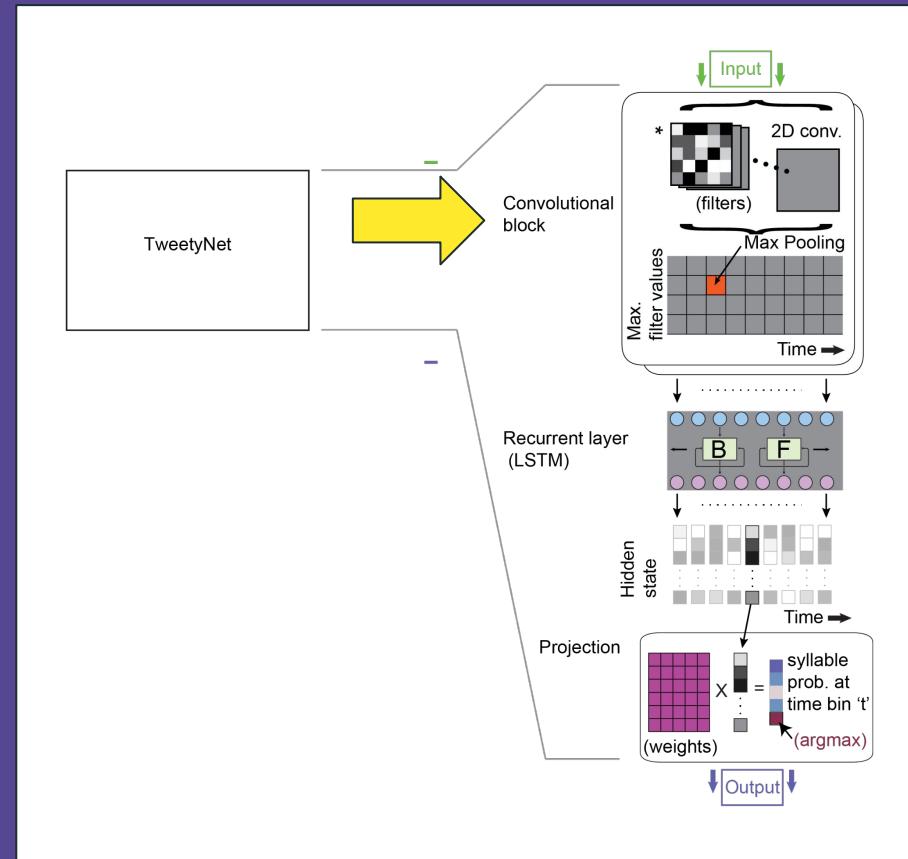
TweetyNet

A way to automate
annotation without
needing cleanly
segmented audio



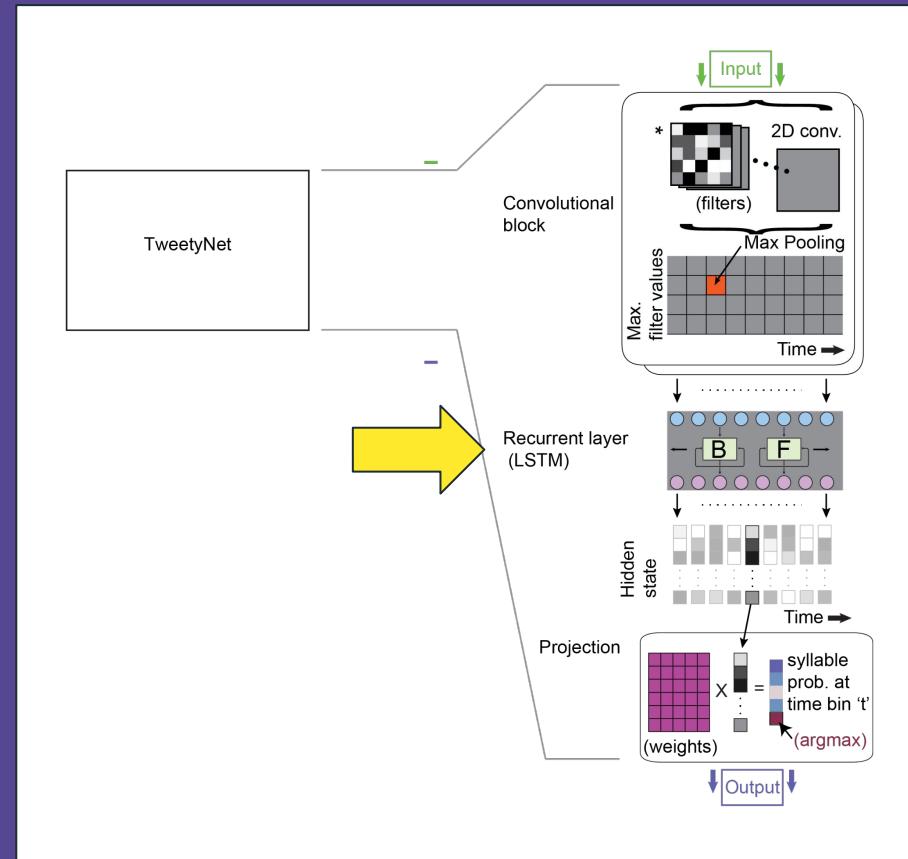
TweetyNet

A way to automate
annotation without
needing cleanly
segmented audio



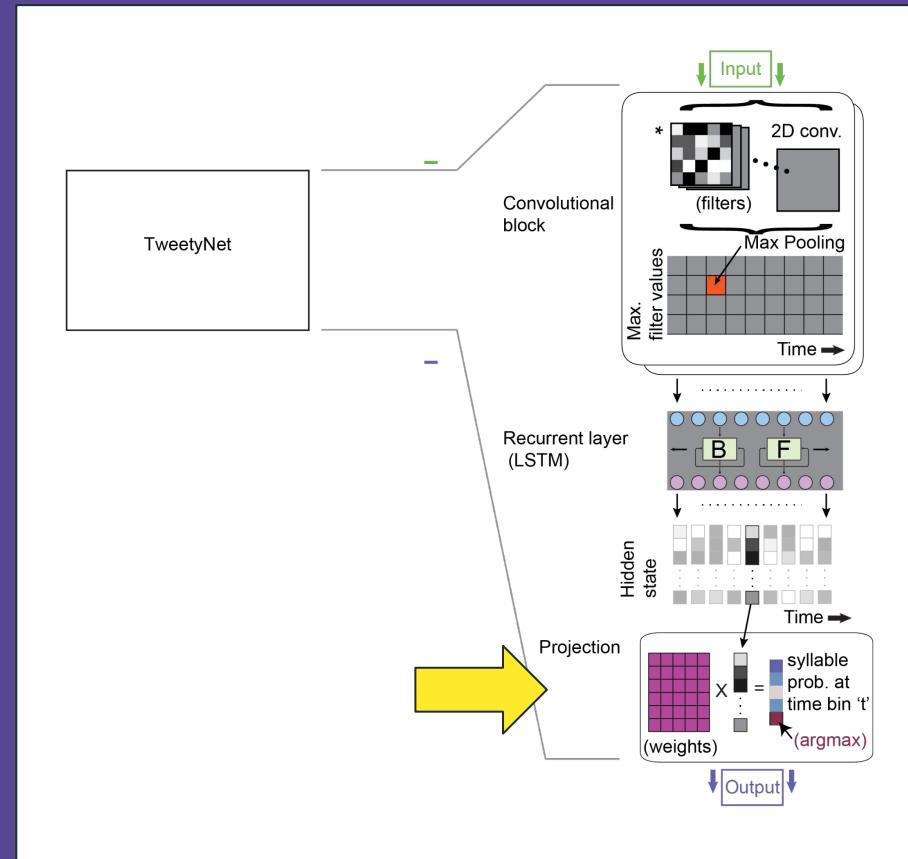
TweetyNet

A way to automate
annotation without
needing cleanly
segmented audio



TweetyNet

A way to automate
annotation without
needing cleanly
segmented audio



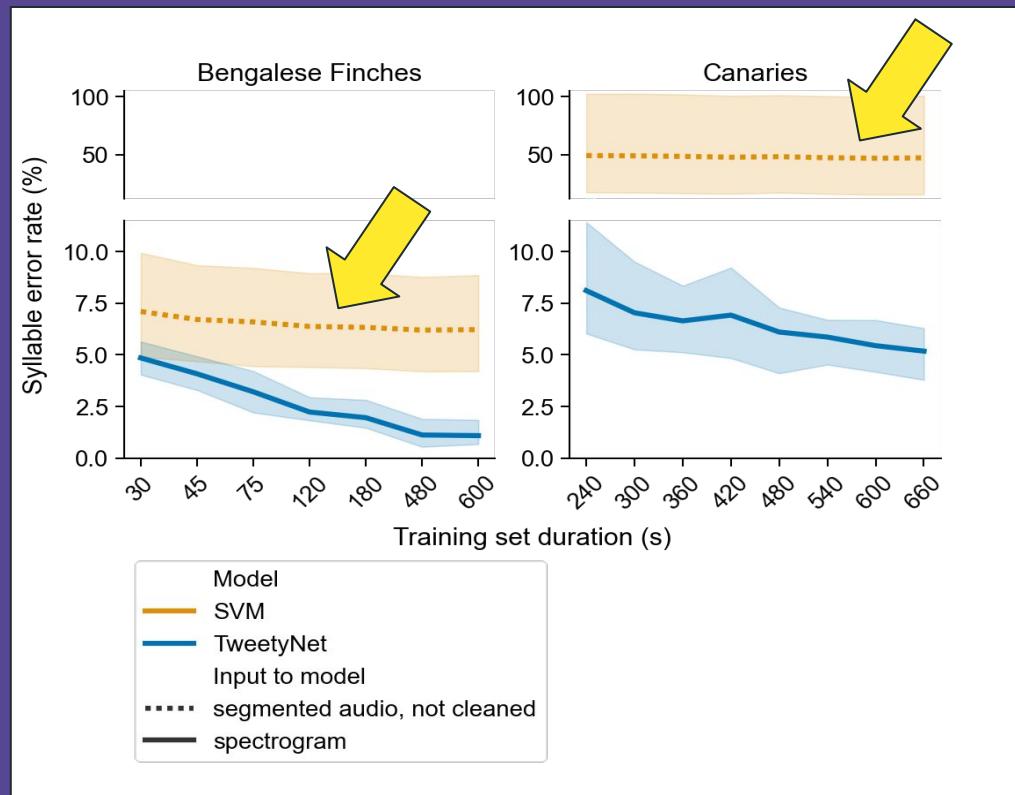
Metrics

- Frame error
- Syllable error rate
 - an edit distance
 - normalized

$$\frac{\text{Syllable Word Error Rate}}{=} \frac{\text{Insertions} + \text{Deletions} + \text{Substitutions}}{\text{Number of Words in Reference Transcript}}$$

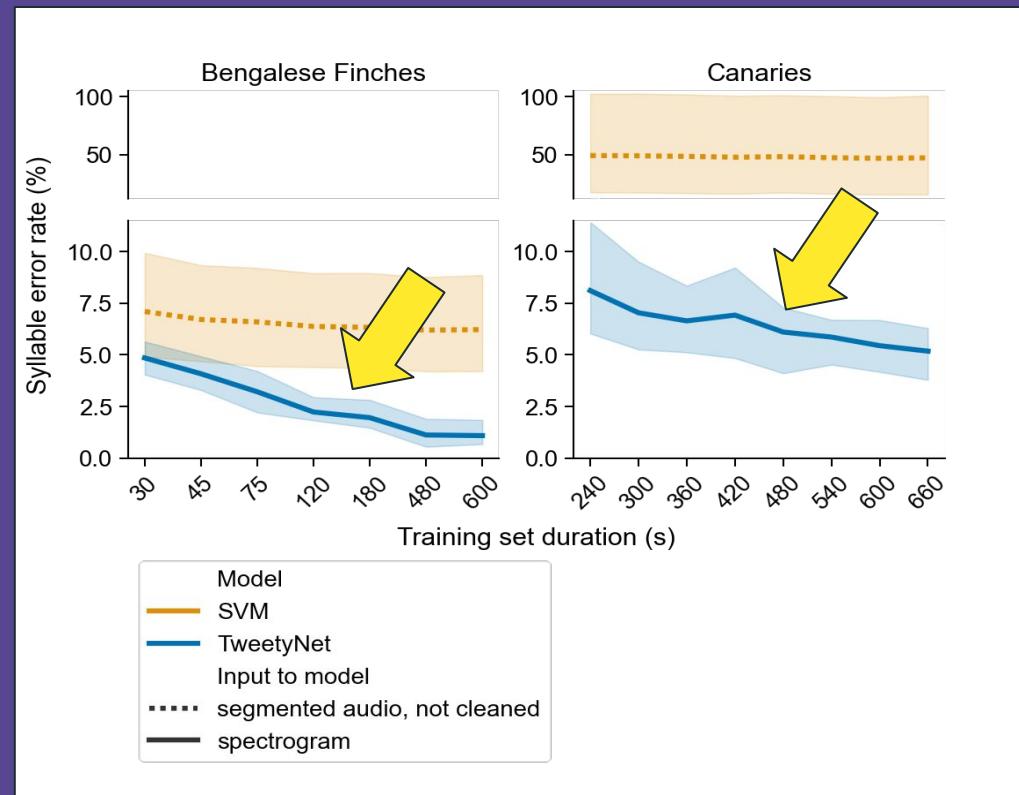
TweetyNet v. Support Vector Machine

Support Vector Machine
applied to "raw" segmented
audio = very high syllable
error rate



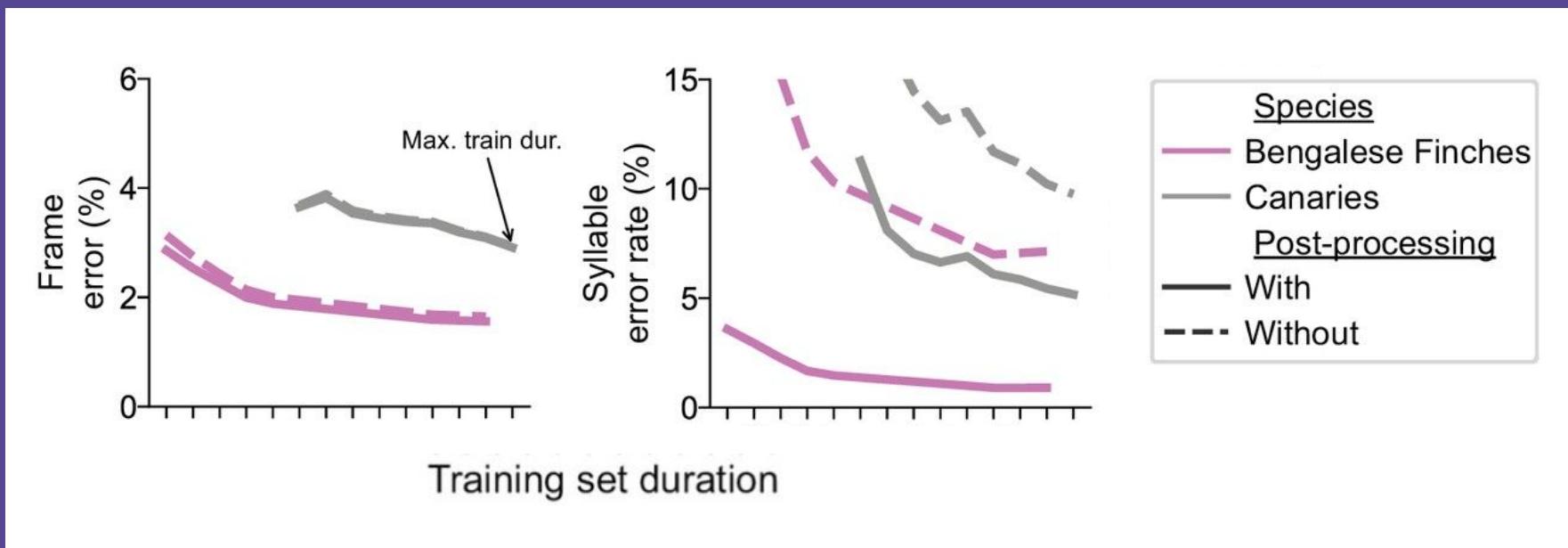
TweetyNet v. Support Vector Machine

TweetyNet applied to spectrogram, without needing to segment audio: much lower syllable error rate



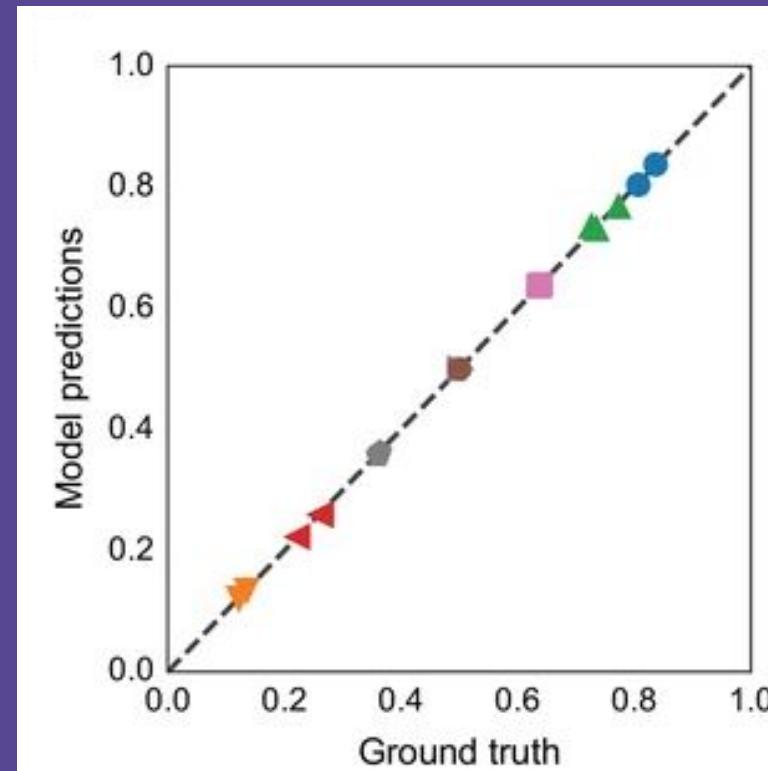
Post-processing

Reduces syllable error rate, not frame error



TweetyNet performance

TweetyNet predictions replicate key results about stability of *branch points* in Bengalese finch song



Results: Vak 1.0

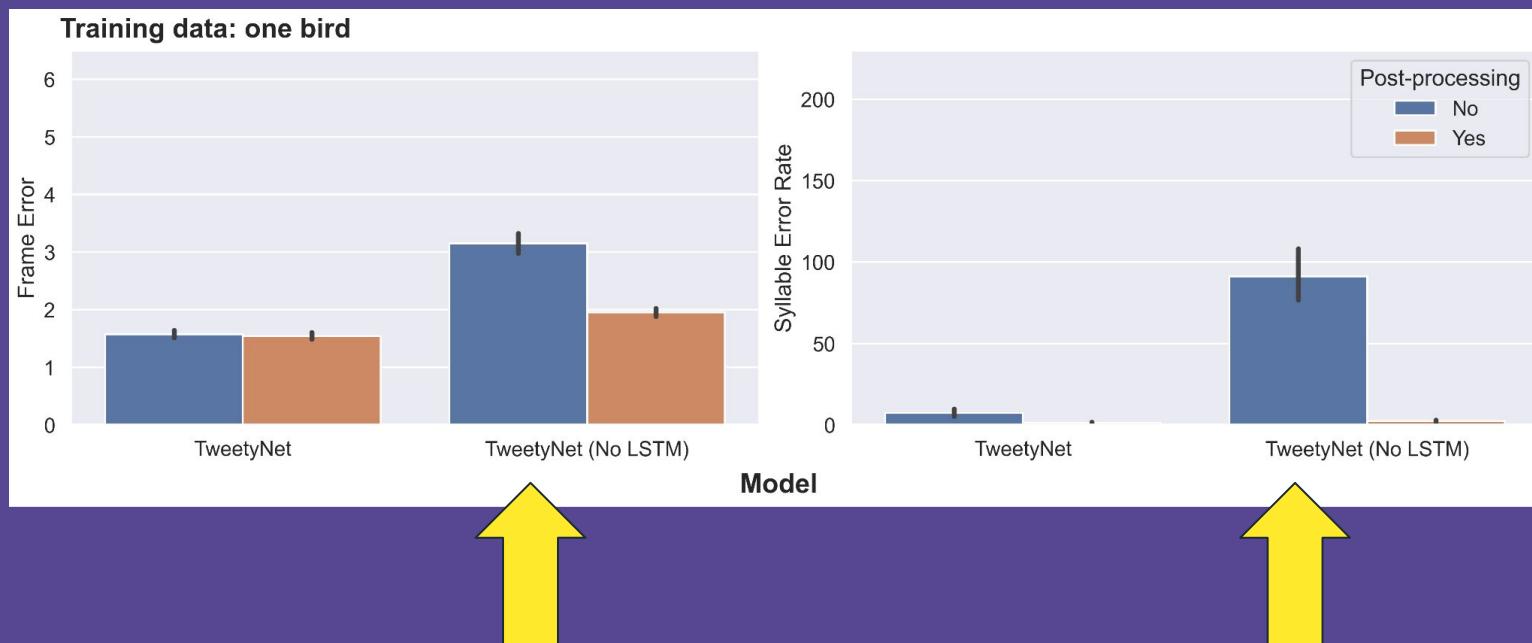
Ablation experiment

Ablation = remove recurrent
Long Short Term Memory
(LSTM) layer

```
@model(family=FrameClassificationModel)
class TweetyNoLSTMNet:
    """TweetyNet model without LSTM layer"""
    network = TweetyNetNoLSTM
    loss = torch.nn.CrossEntropyLoss
    optimizer = torch.optim.Adam
    metrics = {'acc': metrics.Accuracy,
               'levenshtein': metrics.Levenshtein,*  
               'segment_error_rate': metrics.SegmentErrorRate,  
               'loss': torch.nn.CrossEntropyLoss}
    default_config = {
        'optimizer':
            {'lr': 0.003}
    }
```

Ablation experiment

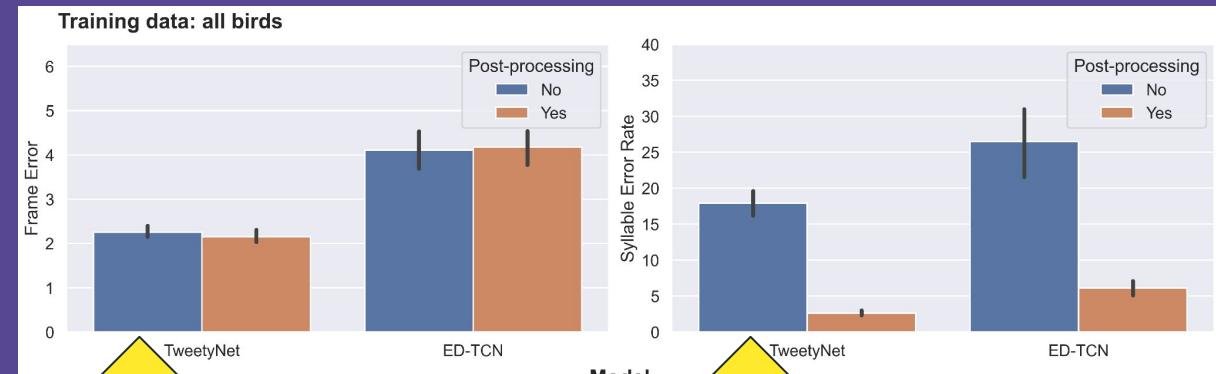
TweetyNet without LSTM has higher frame rate and syllable error rate



Comparing architectures

On this task, TweetyNet achieves a lower segment error rate than ED-TCN (Lea et al. 2017)

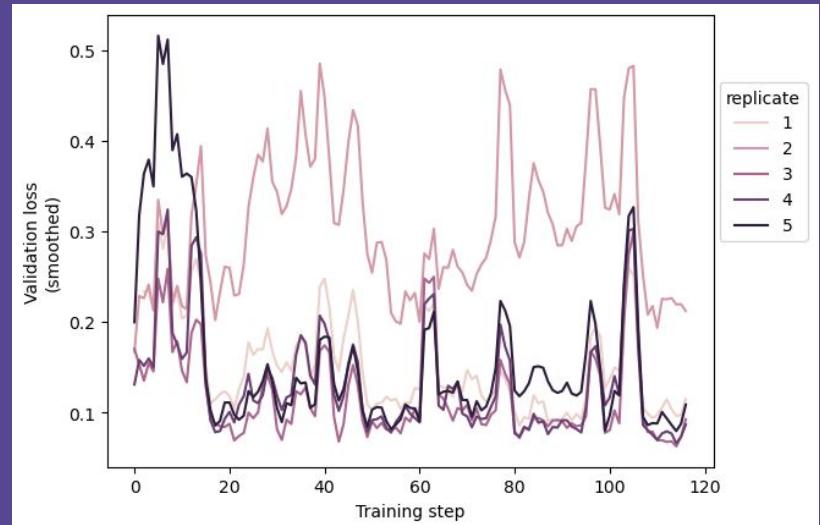
- Lea et al. report that ED-TCN reduces oversegmenting compared to a CNN-LSTM model



Inspecting loss plots

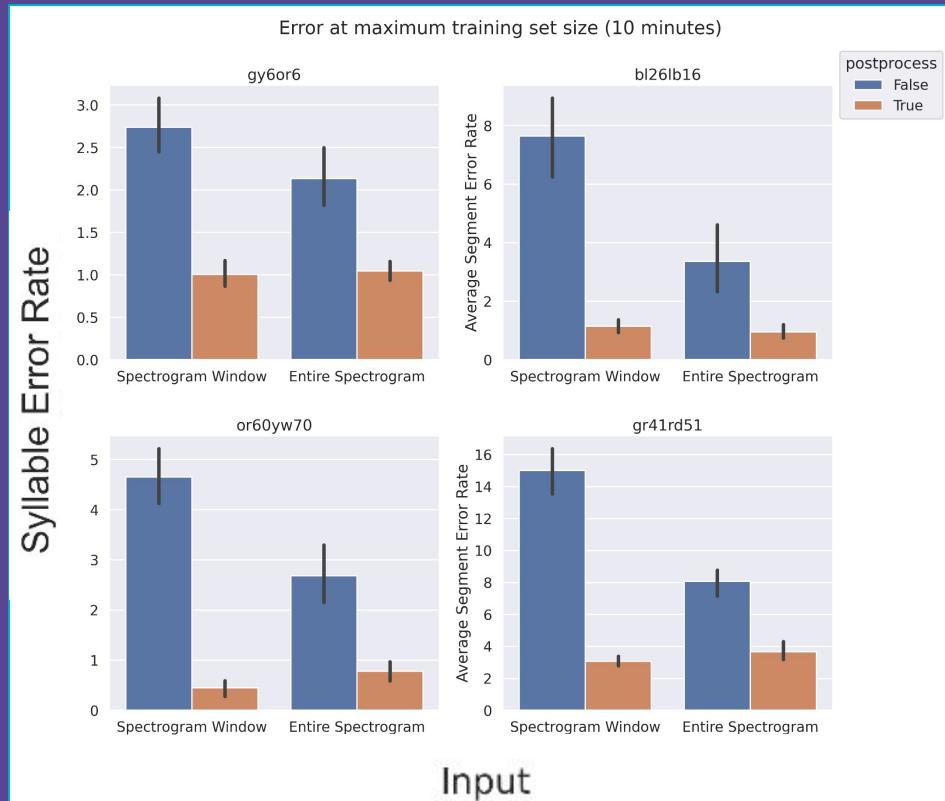
Can convert Tensorboard events files to pandas Dataframes with
`vak.common.tensorboard.events2df`

No clear evidence of overfitting



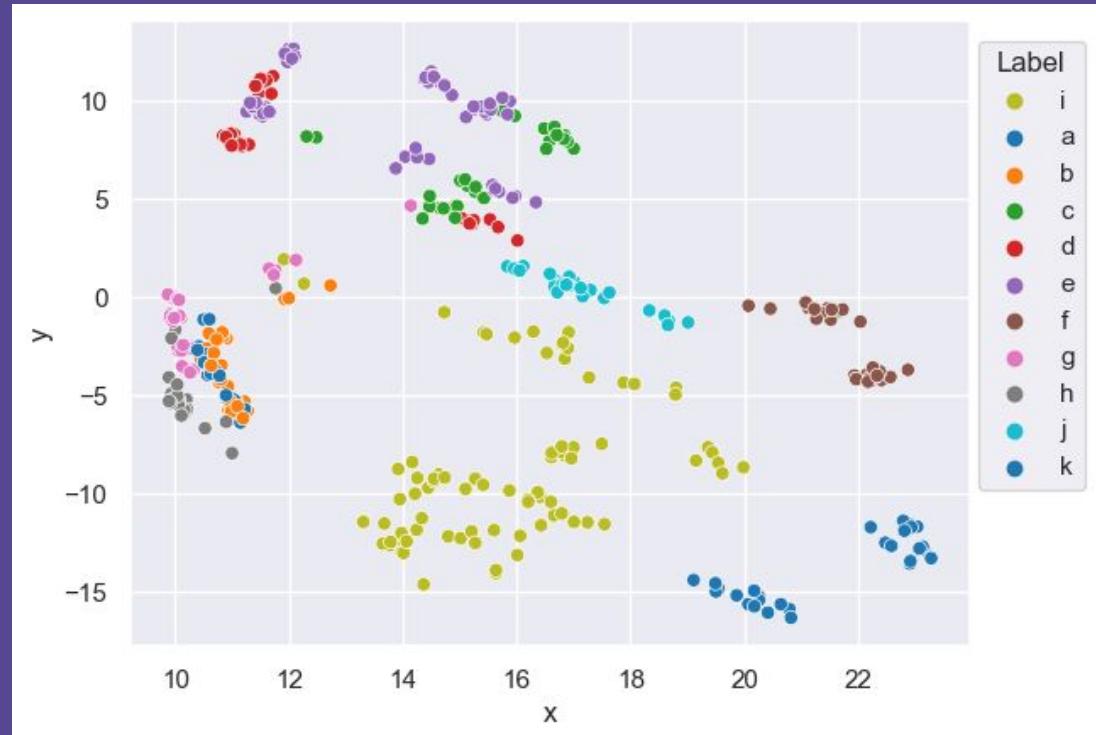
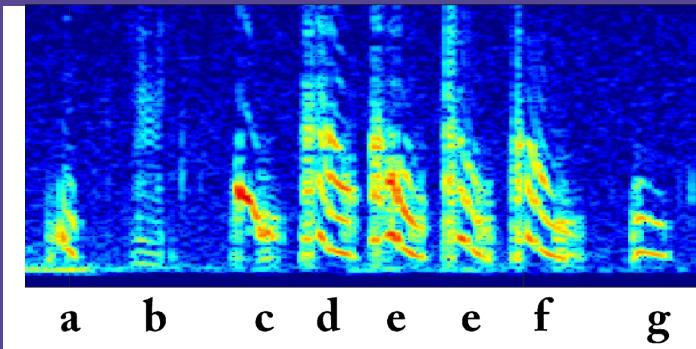
Training TweetyNet on entire song bouts

Improves syllable error rate **without** post-processing



Other model families: Parametric UMAP

Embedding by a fit Parametric UMAP model that uses a convolutional encoder



vak development roadmap

In progress as part of version 1.0: <https://github.com/vocalpy/vak/issues/614>

- Additional model families
 - Object detection
 - Sequence-to-sequence
- Pre-trained models
- Built-in benchmark datasets
- VocalPy as backend for prep step
 - <https://github.com/vocalpy/vocalpy>
 - Proceedings paper in Forum Acusticum 2023

Acknowledgements (part 2)

- Thank you contributors!

Contributors ✨

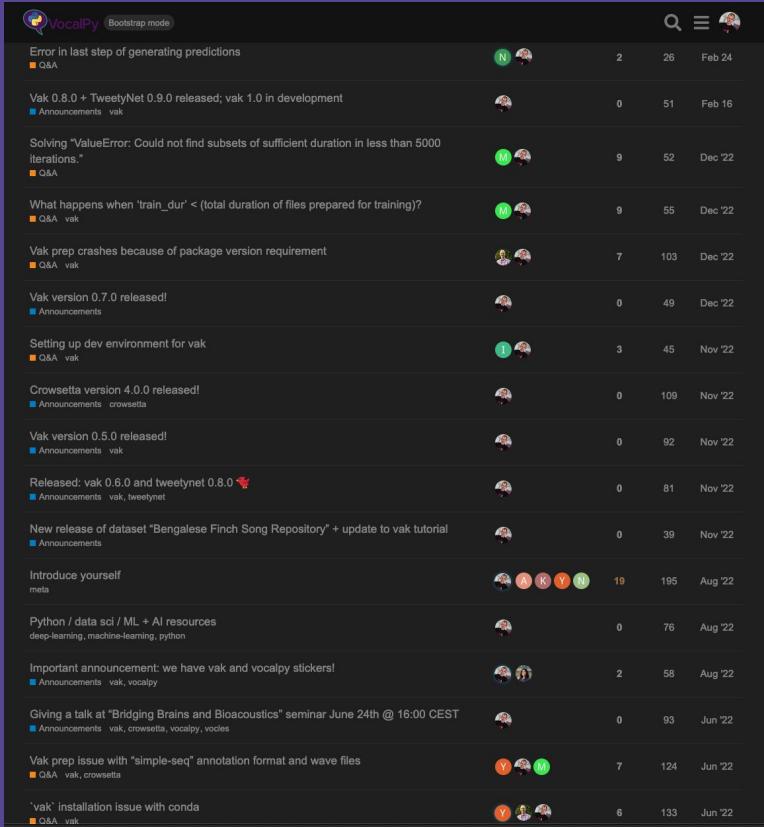
Thanks goes to these wonderful people (emoji key):

 avanikop 	 Luke Poeppl 	 marichard123 	 Therese Koch 	 alyndanoel 		
 adamfishbein 	 vivinastase 	 kaiyaprovost 	 ymk12345 	 neuronalIX 	 Khoa 	 sthaar 
 yangzheng-121 	 Impascual 	 ItamarFruchter 	 Hjalmar K. Turesson 	 nhoglen 		

This project follows the [all-contributors](#) specification. Contributions of any kind welcome!

VocalPy community

- We already have a growing forum:
<https://forum.vocalpy.org/>



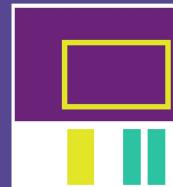
The screenshot shows the VocalPy forum's homepage with a list of recent posts. The posts are organized into categories: Q&A and Announcements. Each post includes the author's profile picture, the number of replies, the number of views, and the date it was posted.

Post Content	Category	Author	Replies	Views	Date
Error in last step of generating predictions	Q&A	N	2	26	Feb 24
Vak 0.8.0 + TweetyNet 0.9.0 released; vak 1.0 in development	Announcements	vak	0	51	Feb 16
Solving "ValueError: Could not find subsets of sufficient duration in less than 5000 iterations."	Q&A	M	9	52	Dec '22
What happens when 'train_dur' < (total duration of files prepared for training)?	Q&A	vak	9	55	Dec '22
Vak prep crashes because of package version requirement	Q&A	vak	7	103	Dec '22
Vak version 0.7.0 released!	Announcements		0	49	Dec '22
Setting up dev environment for vak	Q&A	vak	3	45	Nov '22
Crowsetta version 4.0.0 released!	Announcements	crowsetta	0	109	Nov '22
Vak version 0.5.0 released!	Announcements	vak	0	92	Nov '22
Released: vak 0.6.0 and tweetynet 0.8.0 🎉	Announcements	vak, tweetynet	0	81	Nov '22
New release of dataset "Bengalese Finch Song Repository" + update to vak tutorial	Announcements		0	39	Nov '22
Introduce yourself	meta	A, K, Y, N	19	195	Aug '22
Python / data sci / ML + AI resources	deep-learning, machine-learning, python		0	76	Aug '22
Important announcement: we have vak and vocalpy stickers!	Announcements	vak, vocalpy	2	58	Aug '22
Giving a talk at "Bridging Brains and Bioacoustics" seminar June 24th @ 16:00 CEST	Announcements	vak, crowsetta, vocalpy, voces	0	93	Jun '22
Vak prep issue with "simple-seq" annotation format and wave files	Q&A	vak, crowsetta	7	124	Jun '22
'vak' installation issue with conda	Q&A	vak	6	133	Jun '22

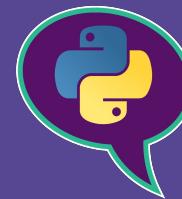
Ways you can contribute

- **Star our repositories on GitHub**
- **Join the forum**
 - **Ask questions**
 - **Answer questions!**
 - **Share examples**
- **Contribute code**

Questions?



crowsetta



VocalPy

- **VocalPy organization on GitHub:** <https://github.com/vocalpy>
 - <https://github.com/vocalpy/vak>
 - <https://github.com/vocalpy/crowsetta>
 - <https://github.com/vocalpy/vocalpy>
- **vak docs:** <https://vak.readthedocs.io/en/latest/>
- **crowsetta docs:** <https://crowsetta.readthedocs.io/en/latest/>
- **VocalPy forum:** <https://forum.vocalpy.org/>