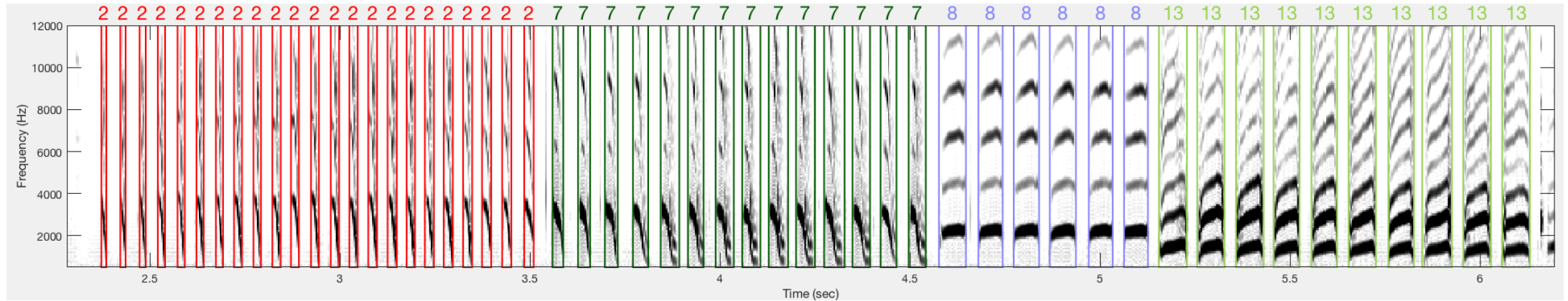
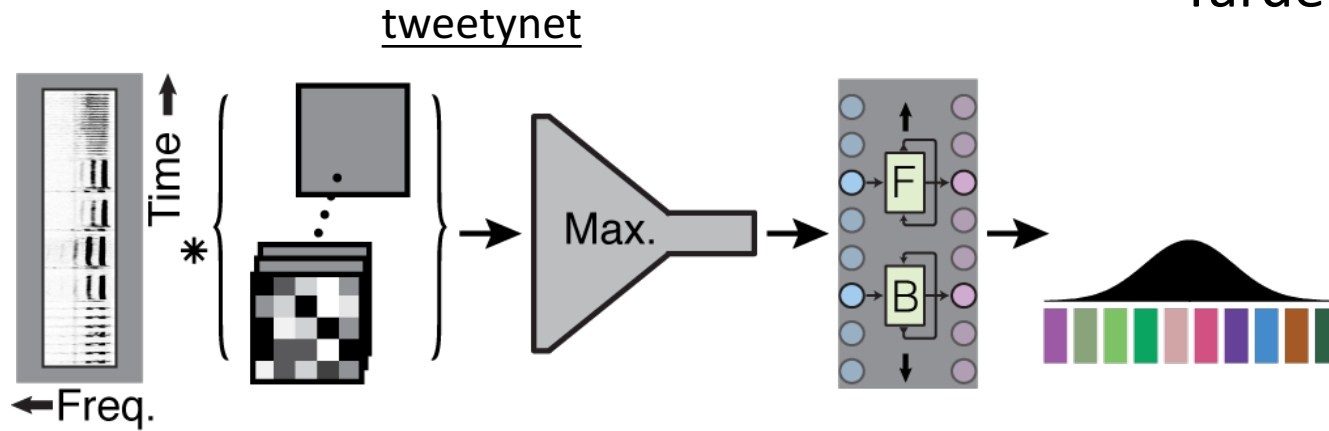


Segmentation and annotation of birdsong with a hybrid recurrent-convolutional neural network

Yarden Cohen, Gardner lab, Boston University



SFN Poster: RR11, Nov 5th, 1-5pm

Repo: <https://github.com/yardencsGitHub/tweetynet>

Segmentation and annotation of birdsong with a hybrid recurrent-convolutional neural network

Yarden Cohen, Gardner lab, Boston University

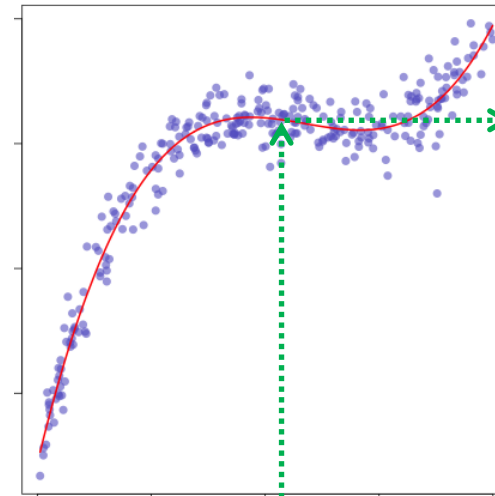
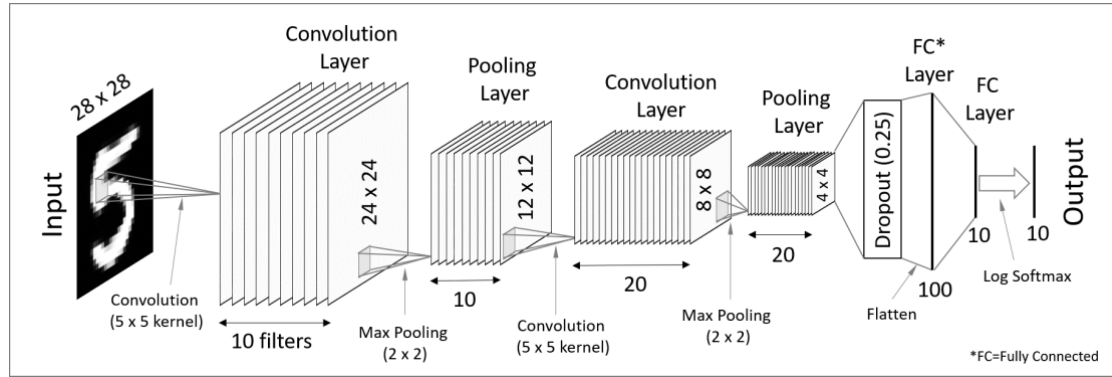


(arXiv:1703.00848, NIPS 2017)

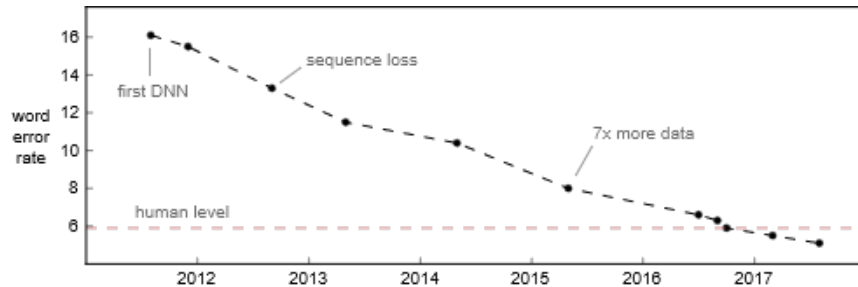
SFN Poster: RR11, Nov 5th, 1-5pm

Repo: <https://github.com/yardencsGitHub/tweetynet>

Deep learning is curve fitting



It is extremely (surprisingly) helpful in replicating human performance.

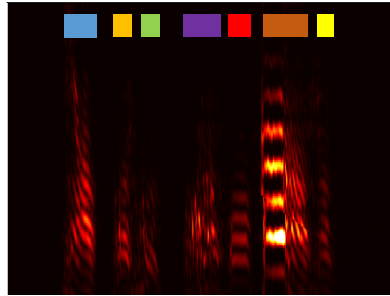


(<https://awni.github.io/speech-recognition/>)



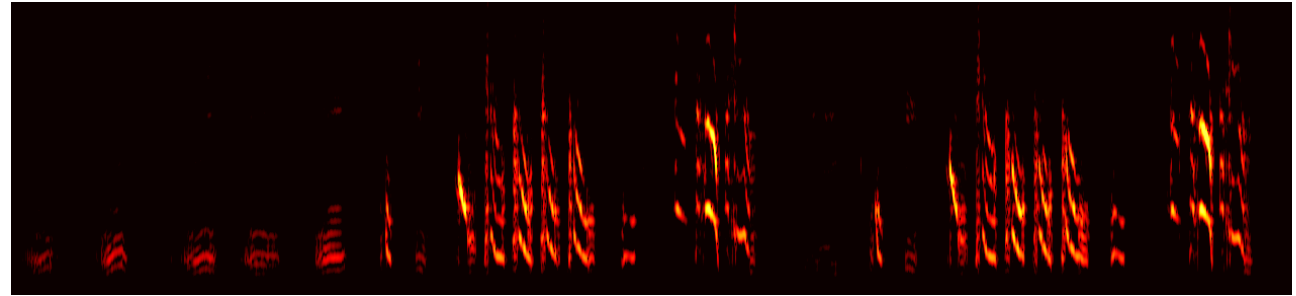
Variable birdsong is challenging to annotate

Zebra finch:



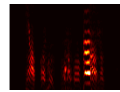
1 Sec

Bengalese finch:

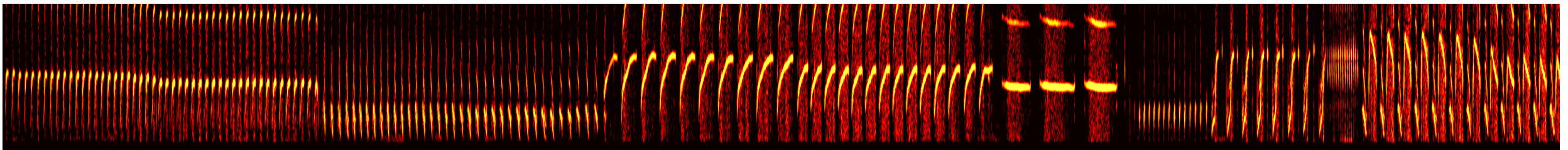


1 Sec

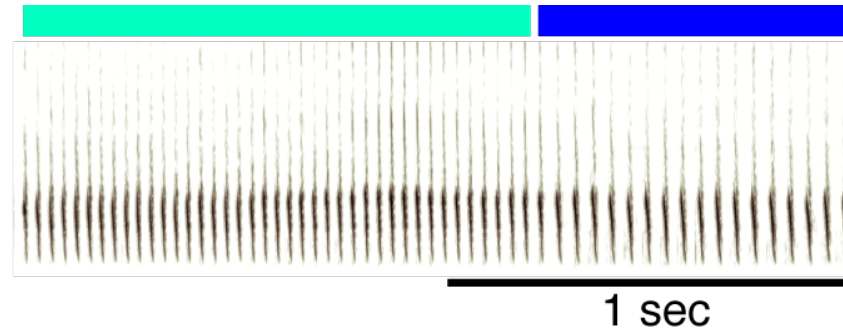
Canaries:



1 Sec

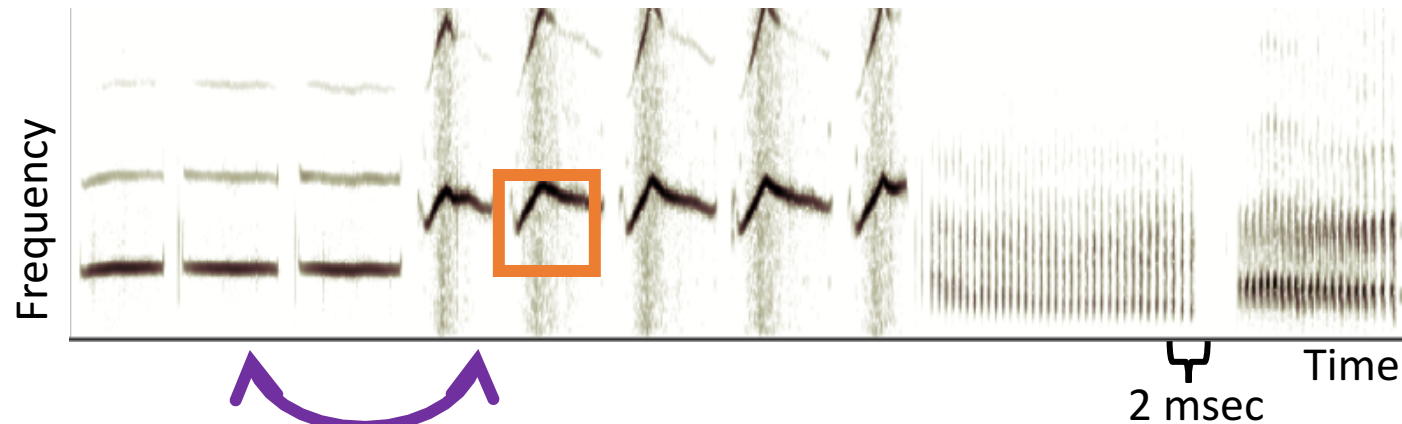


- Manual 20-30 minutes per song (I had >5000)
- Syllables:
 - Variable durations 10-500 msec
 - Variable acoustics
- Variable sequence
 - # syllables/phrase, gaps, order



Clark, Marler, Beeman, 1987
Anderson, Dave, Margoliash 1996,
Glaze and Troyer, 2006
Pearre et al PloS One, 2017

The model captures spectrogram features and outputs the syllable likelihoods in every time bin

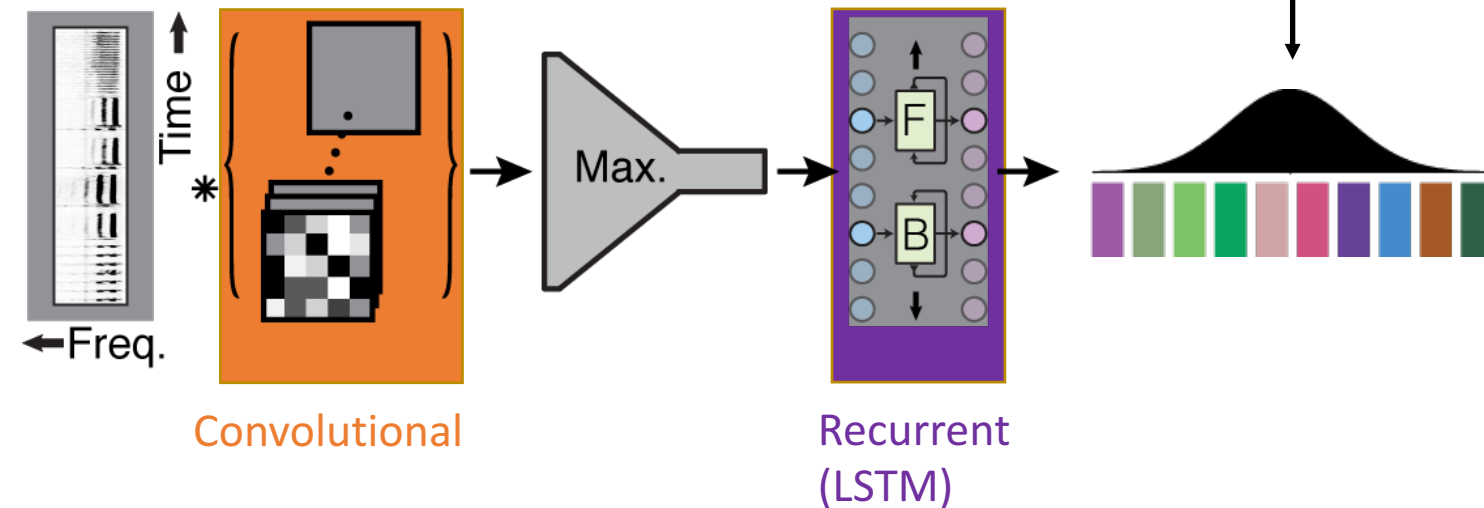


Training:

Audio files + syllable annotation

Output for new data:

- The most likely syllable in each time bin
- (Potentially) A similarity metric to all syllables



Clark, Marler, Beeman, 1987
Anderson, Dave, Margoliash 1996,
Glaze and Troyer, 2006
Pearre et al PloS One, 2017
Koumura Okanoya, 2016
S. Böck and M. Schedl, 2012
Parascandolo, Huttunen, and Virtanen, 2016

Data formats and model definition

Training data

- Audio files:
 - .wav
 - .cbin
- Labels .csv file with columns:
 - File name
 - Syllable tag
 - Onset time
 - Offset time
- Alternatively, Matlab files with spectrograms

Configuration .ini file

```
[NETWORK]
Parameters (learning rate, etc')

[SPECTROGRAM]
Parameters (FFT size, etc')

[DATA]
Where the data is stored

[OUTPUT]
Path of output folder
```

Data formats and model definition

Train a model

```
tweetynet-cli --train --config config_file.ini
```

Evaluate model training

```
tweetynet-cli --config config_file.ini
```

Annotate new data

```
tweetynet-cli --predict --config config_file.ini
```

Configuration .ini file

[NETWORK]

Parameters (learning rate, etc')

[SPECTROGRAM]

Parameters (FFT size, etc')

[DATA]

Where the data is stored

[OUTPUT]

Path of output folder

[TRAIN]

A file path to a .csv file with training audio file names and labels

Training set durations and other definitions

[PREDICT]

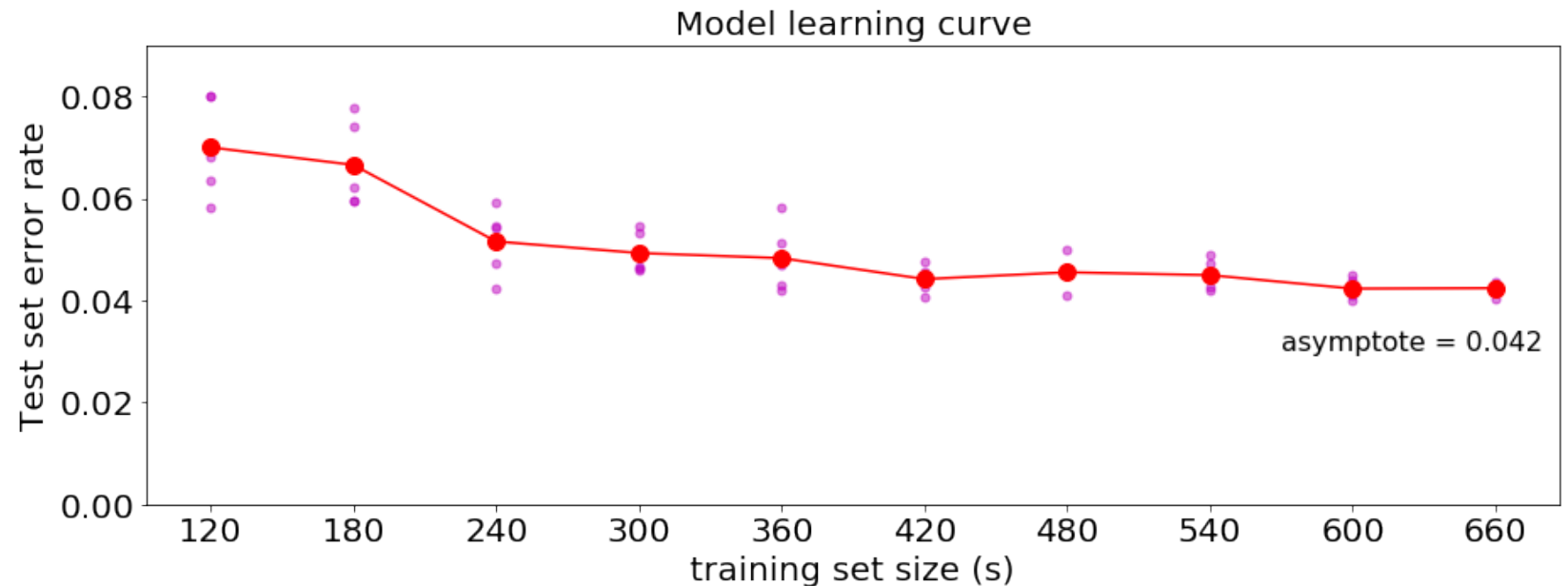
Path to trained model

Path to new data folder

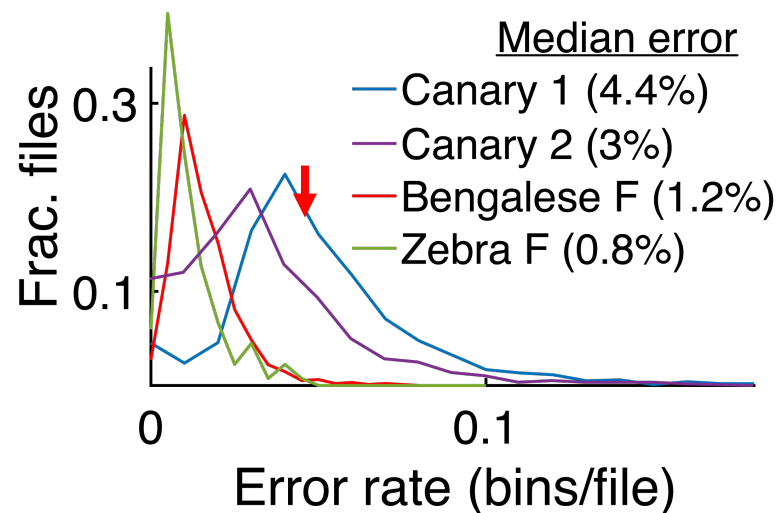
Model training evaluating (learning curves)

Now run the experiment:

```
activate tweetynet  
tweetynet-cli --config ./configs/config_bird0.ini
```



The model performs well in finches and does >95% of the work on canaries

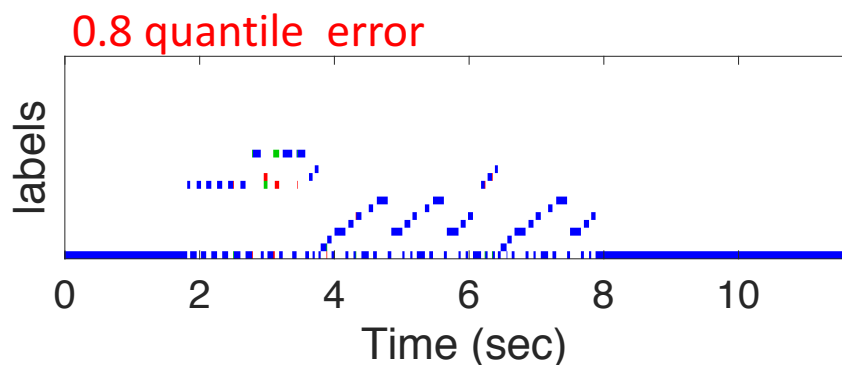
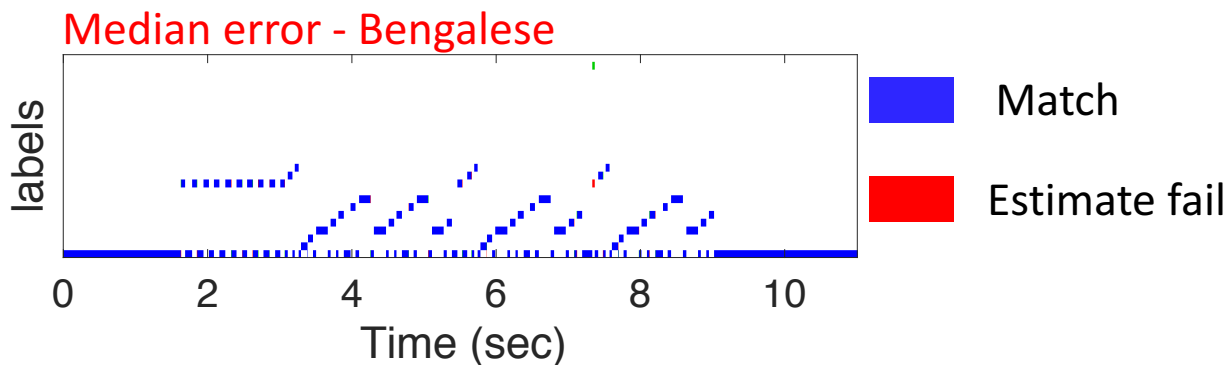


(70 training files, 600s,
~2000 test files)

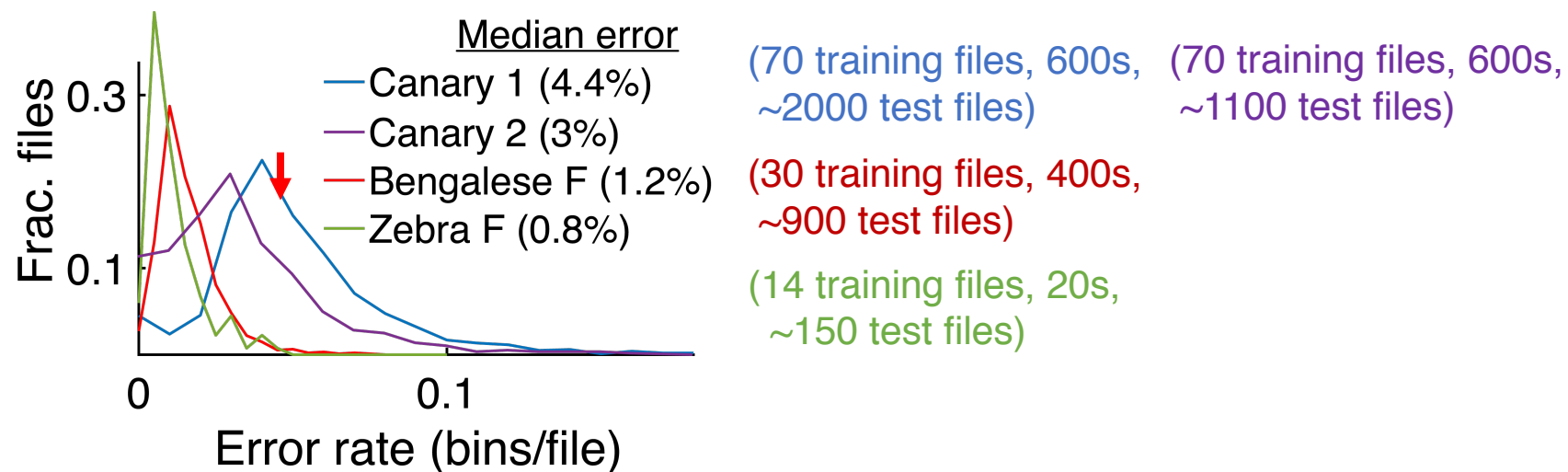
(70 training files, 600s,
~1100 test files)

(30 training files, 400s,
~900 test files)

(14 training files, 20s,
~150 test files)





The model performs well in finches and does >95% of the work on canaries

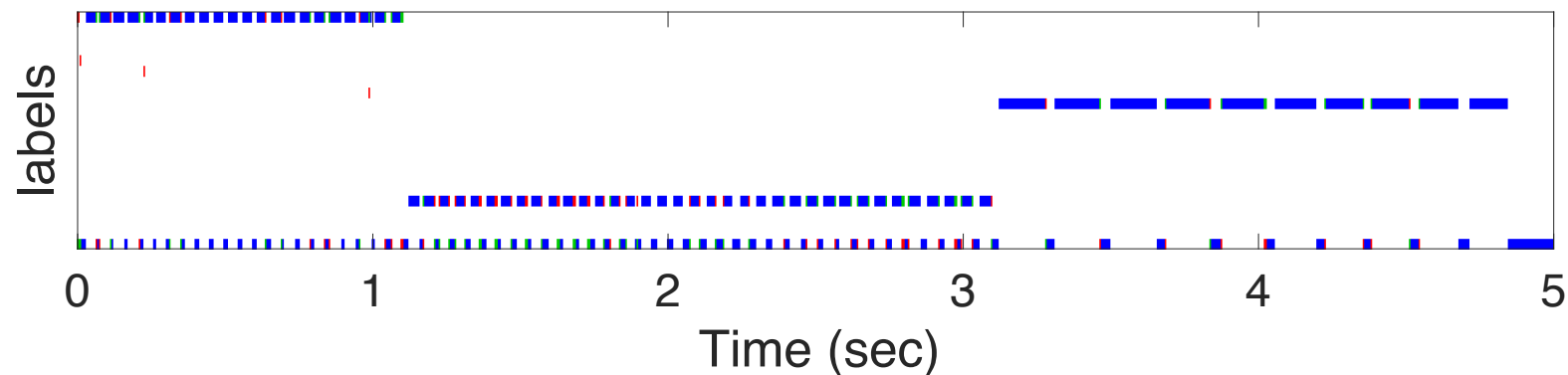


Canary songs:

- Hand annotating ~ 30 minutes
- Proofreading ~ 1 minute

 Match

 Estimate fail



Thanks

Gardner lab (BU):

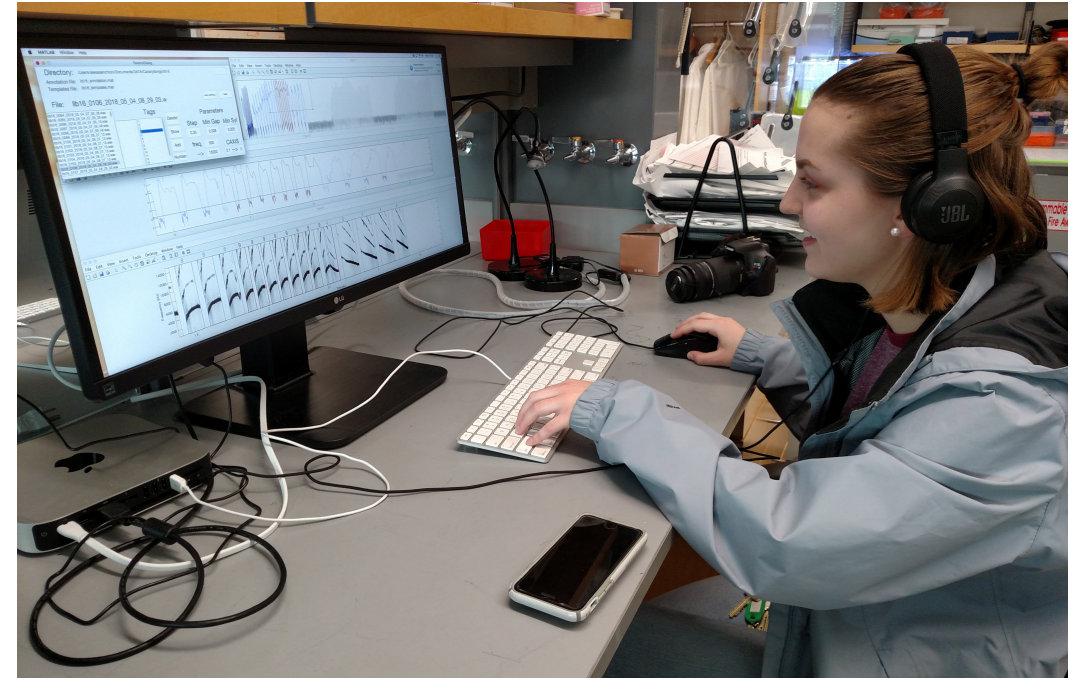
Tim Gardner

Alexa Sanchioni

Emily Mallaber

Sober lab (Emory):

David Nicholson



Both the Gardner and Sober labs acknowledge NVIDIA Investigator grants of GPUs that made this research possible