Using machine learning to listen out for birds

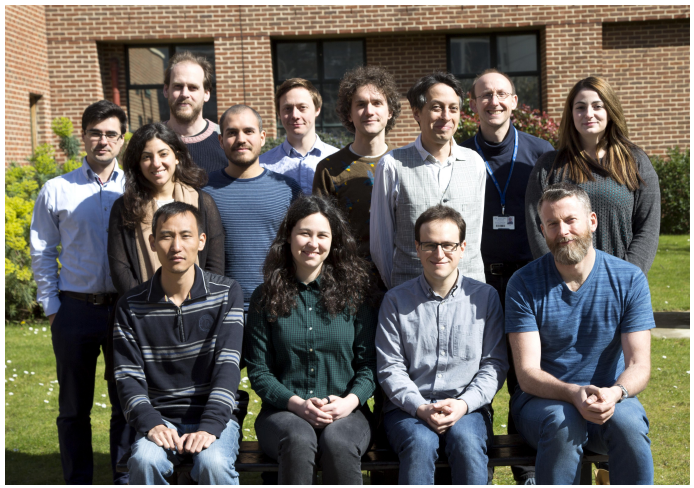
Dan Stowell

Machine Listening Lab
Centre for Digital Music
School of Elec Eng & Computer Science
Queen Mary University of London

The Machine Listening Lab

machine-listening.eecs.qmul.ac.uk

centre for digital music | Queen Mary University of London | CIS centre for intelligent sensing

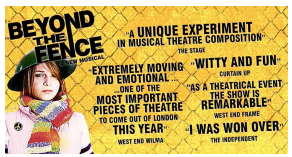


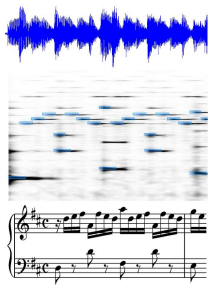
In the Machine Listening Lab we develop methods for making sense of natural sounds, everyday sounds, and recorded music. ***Machine listening*** is the use of signal processing and machine learning to extract useful information from sound.

Applications

Warblr - Automatic bird species recognition in the palm of your hand

The first ever computer-generated musical





Automatic music transcription systems

Lead academics

Dan Stowell

Emmanouil Benetos

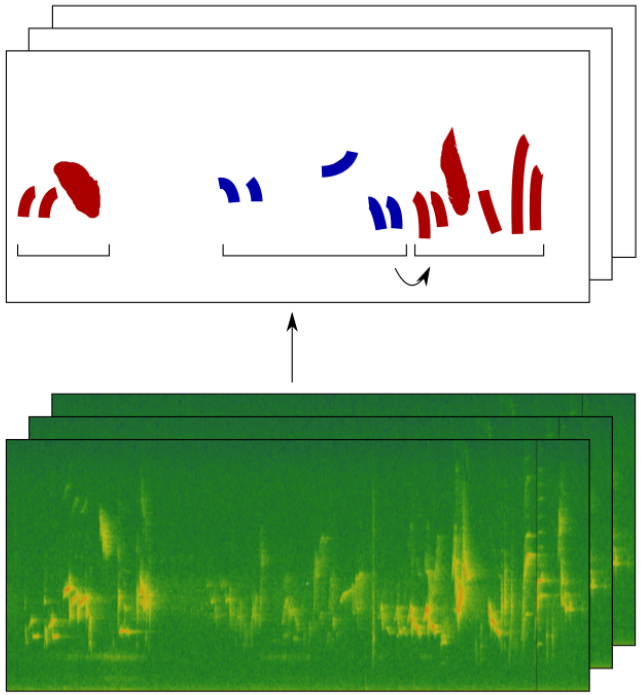
Bob L. Sturm

Sebastian Ewert

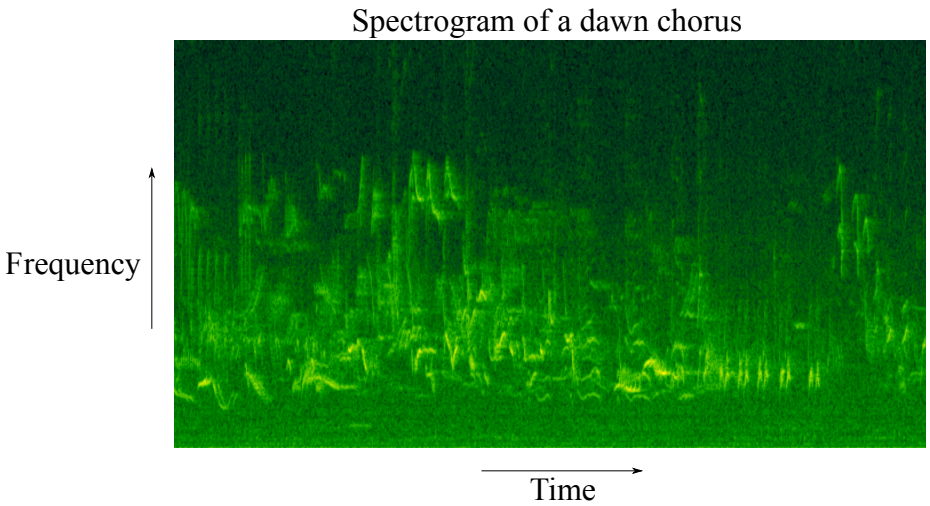
My work: bird sounds

5-year EPSRC Fellowship project:

"Structured machine listening for soundscapes with multiple birds"



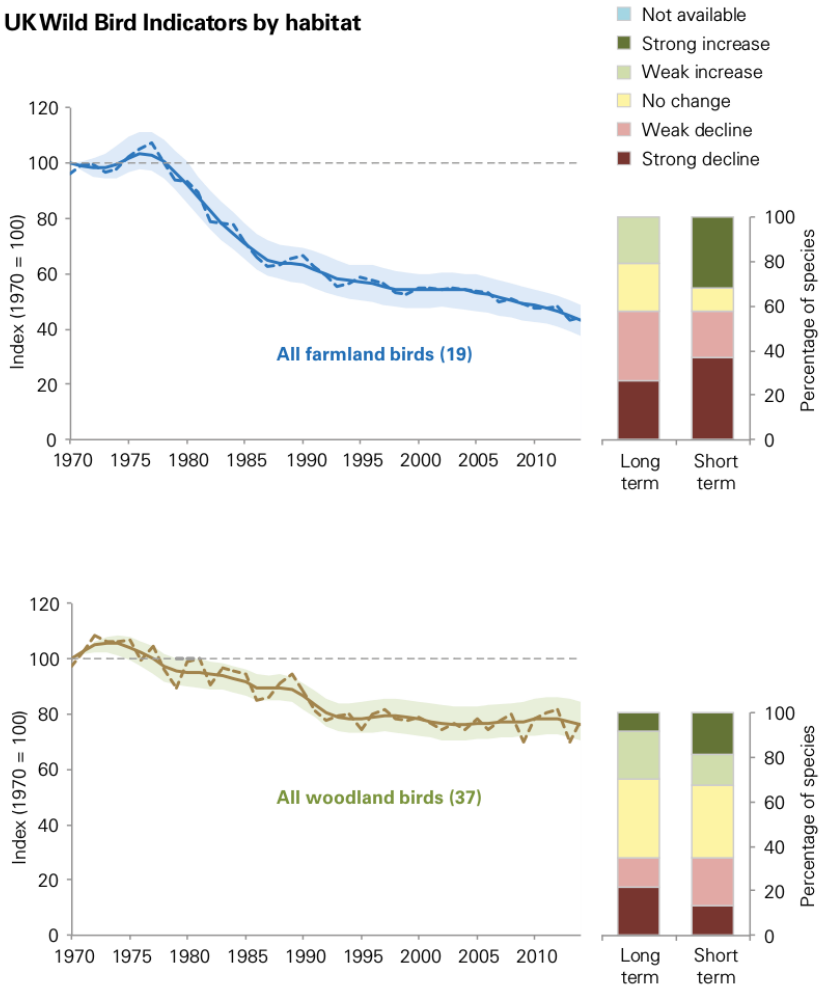
Can we decode the dawn chorus?



- Which species?
- How many birds?
- Singing in response to neighbours?
- Warning about predators?
- Defending a territory, or newly arrived?

Many questions could be answered with the help of modern **machine learning** and **signal processing**

Machine listening and bird sounds - why?



Machine listening and bird sounds - why?

Support The Guardian

The Guardian

News | Opinion | Sport | Culture | Lifestyle

World ▸ Europe US Americas Asia Australia

'Catastrophe' as France's bird population collapses due to pesticides

Dozens of species have seen their numbers decline, in some cases by two-thirds, because insects they feed on have disappeared

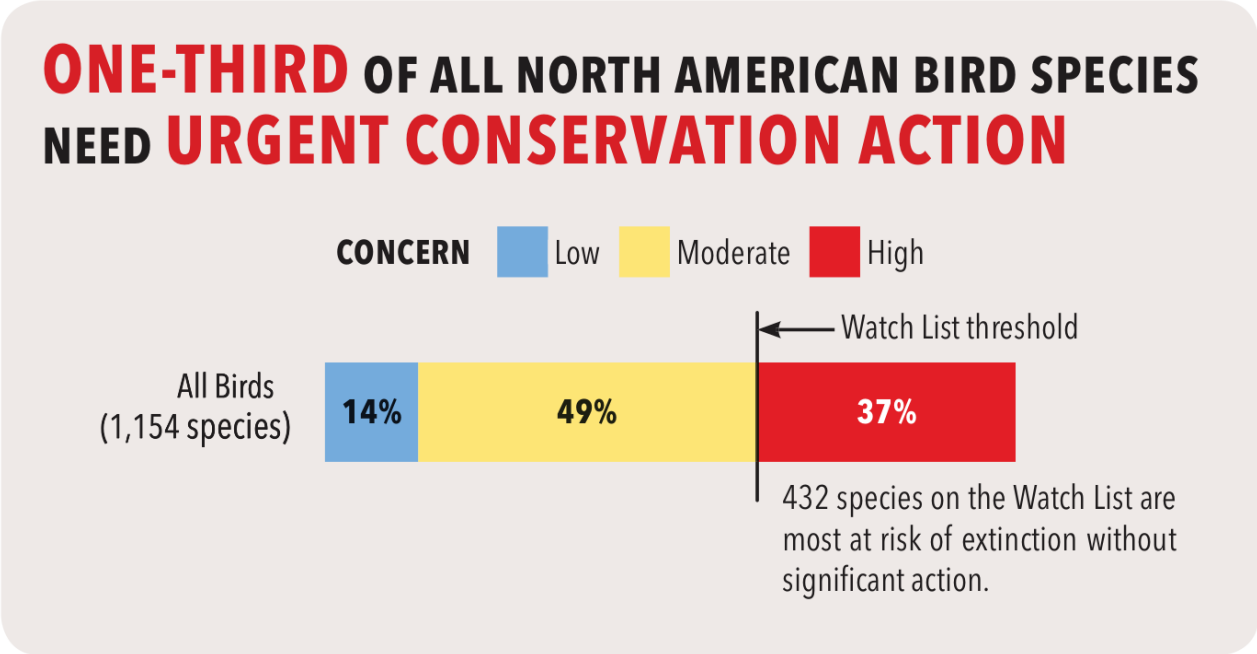
Agence France-Presse

Wed 21 Mar 2018 00.50 GMT

centre for digital music

Queen Mary University of London

Machine listening and bird sounds - why?

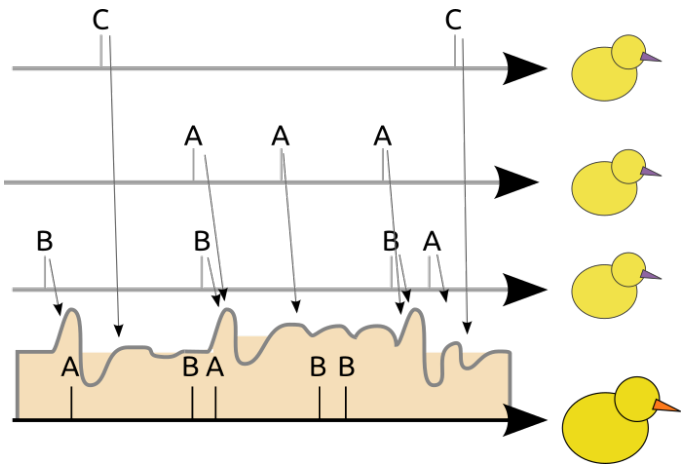


Opportunities for us:

- **Audio** is the best medium for bird monitoring
- **Bioacoustic** monitoring has truly entered its **big data** era

And there's much more we can do, beyond monitoring...

Postcard 1: bird communication networks



Collab with Clayton lab (QMUL) and MPIO
Stowell et al. (2016) **Royal Soc Interface**

centre for digital music
Queen Mary University of London

Postcard 2: "on-bird" sound recording

Understanding daily behaviour of birds in the wild

Collab with Lisa Gill, MPIO
Stowell et al. (2017) **IEEE Trans ASLP**

Postcard 2: "on-bird" sound recording

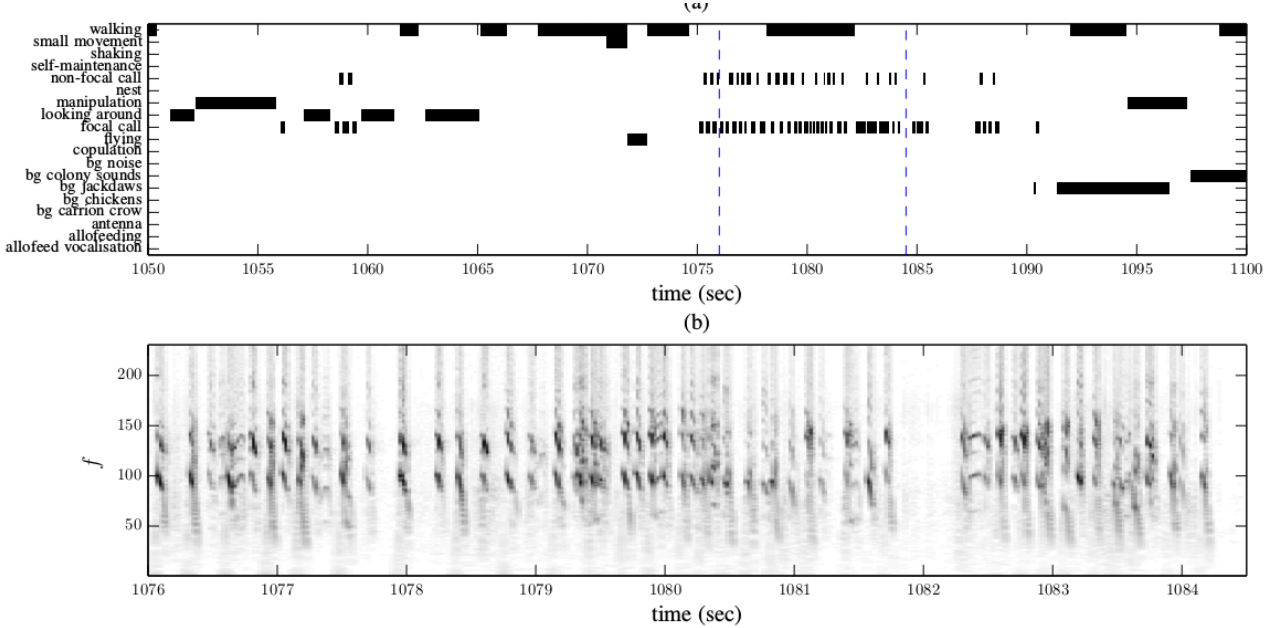
Understanding daily behaviour of birds in the wild

Collab with Lisa Gill, MPIO
Stowell et al. (2017) **IEEE Trans ASLP**

centre for digital music
Queen Mary University of London

Postcard 2: "on-bird" sound recording

Understanding daily behaviour of birds in the wild



Collab with Lisa Gill, MPIO
Stowell et al. (2017) **IEEE Trans ASLP**

centre for digital music
Queen Mary University of London

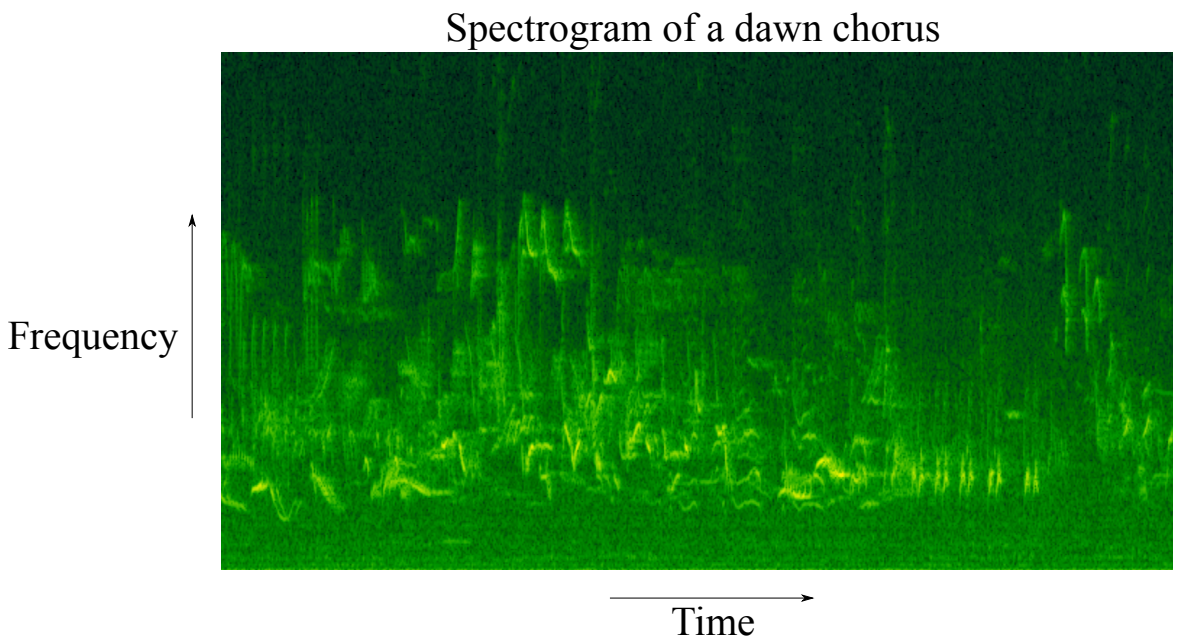
Postcard 2: "on-bird" sound recording

[video]

In this talk...

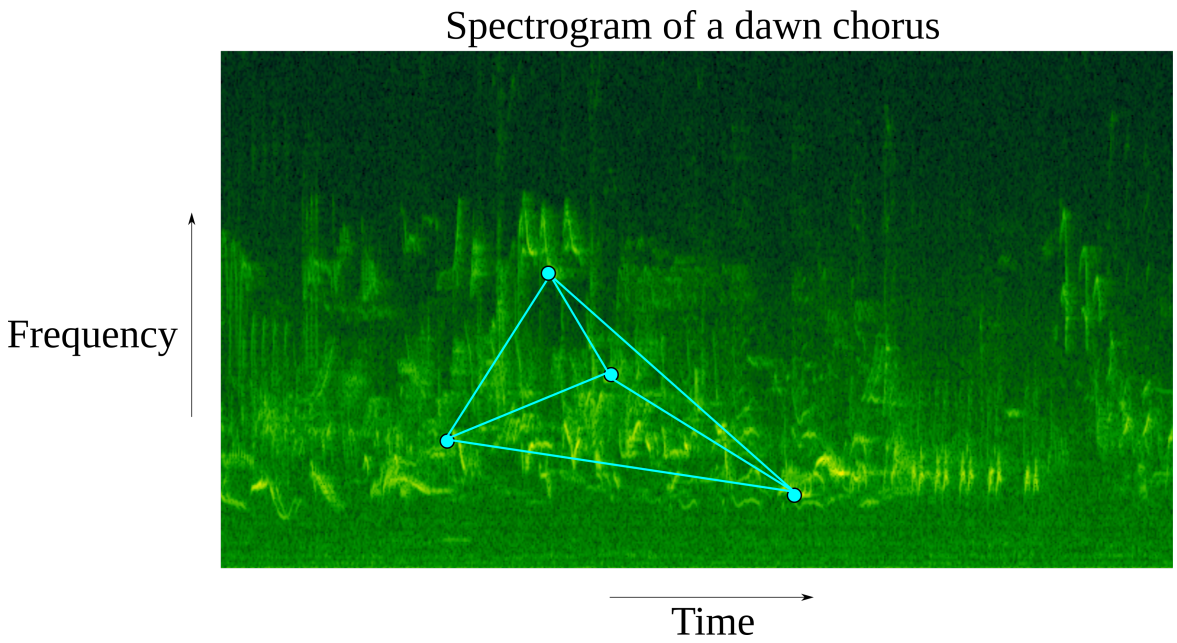
- Applying machine learning to sound signals – **state of the art**
- **Customising** the methods for bird sound characteristics
- Machine learning that **generalises** well
- Working with **limited datasets**
- Pictures of birds

Analysing sound using machine learning



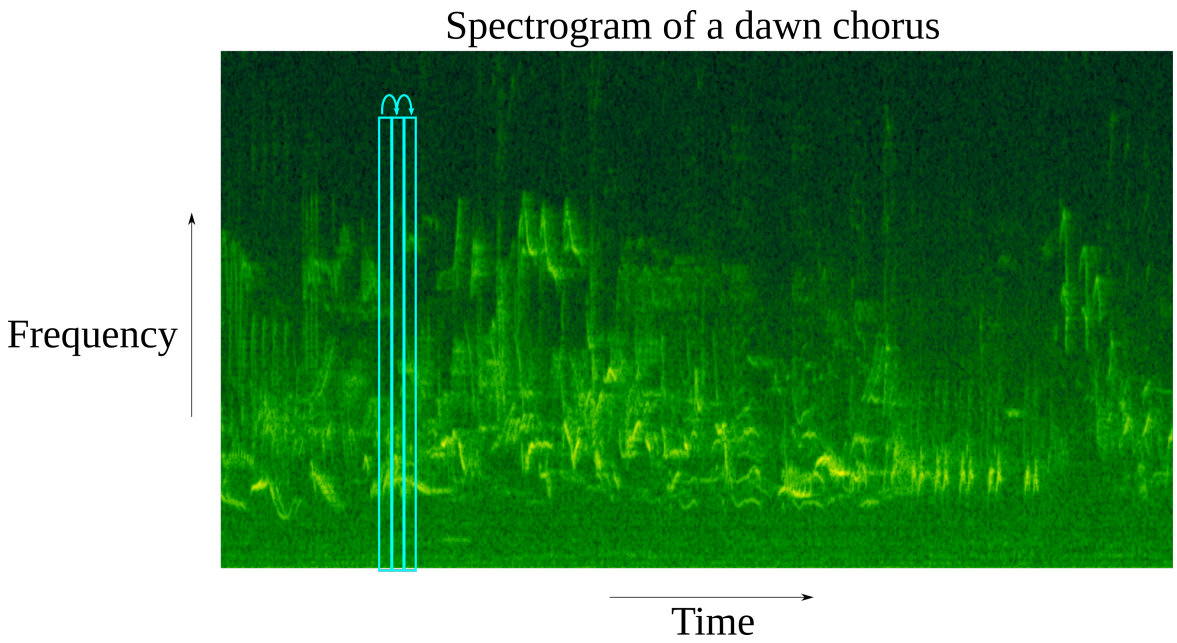
Analysing sound using machine learning

Fingerprinting? ('Shazam')



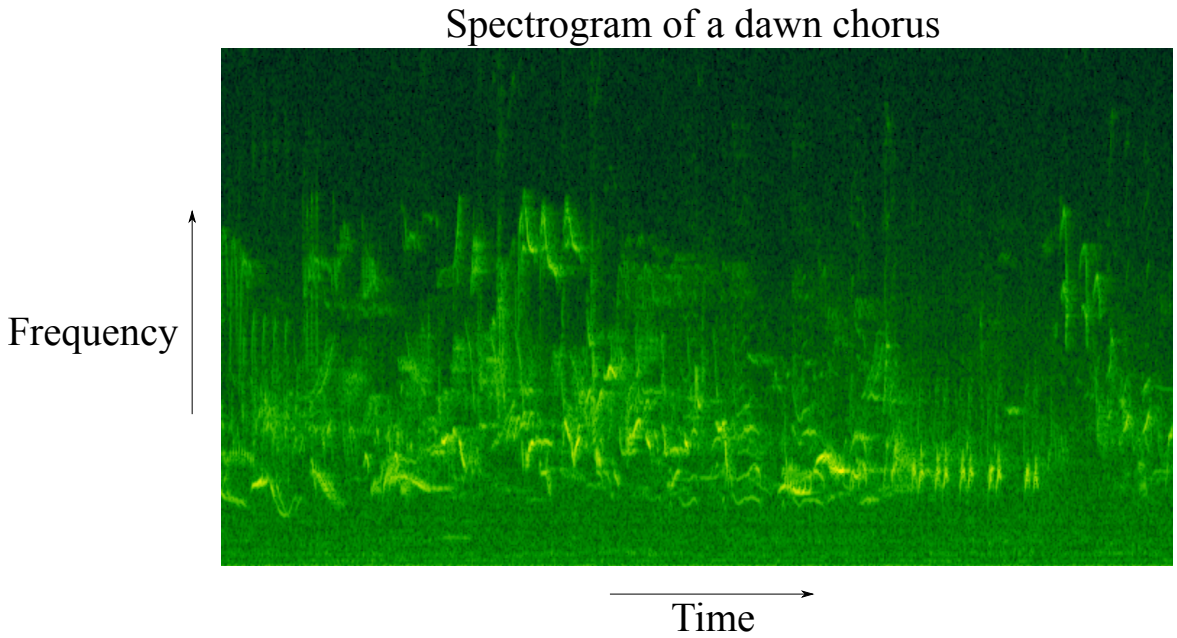
Analysing sound using machine learning

Speech recognition methods? Hidden Markov model (HMM)

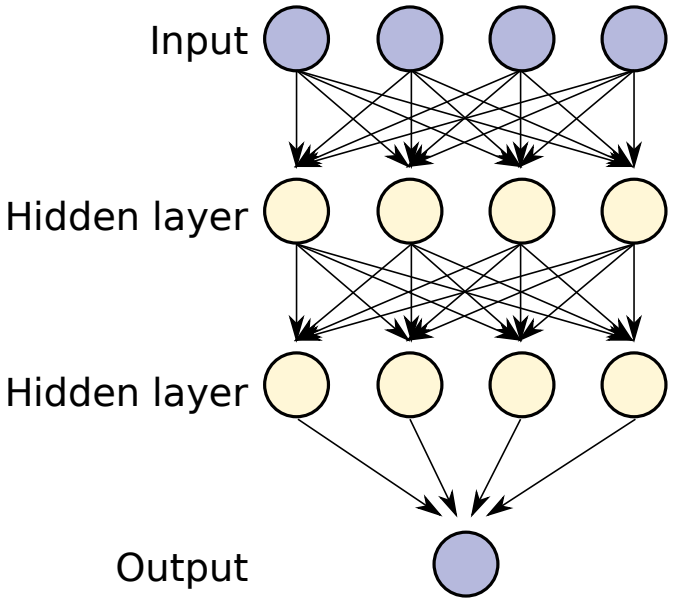


Analysing sound using machine learning

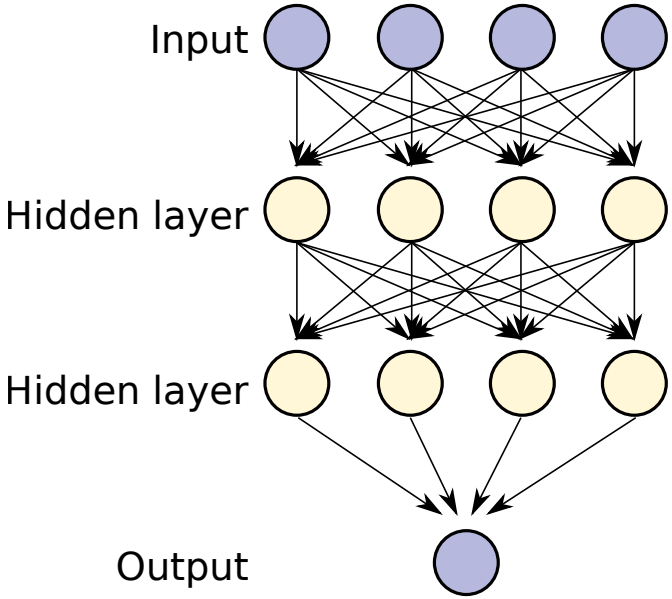
Deep learning?

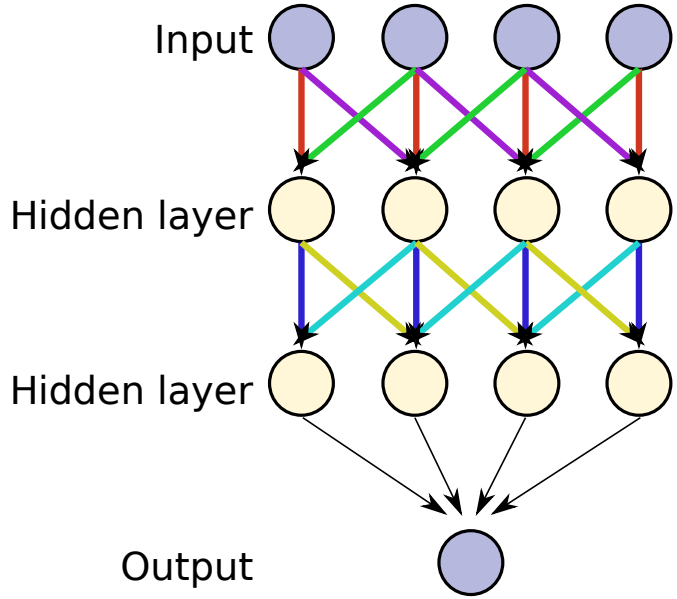


Neural nets, versus convolutional neural nets (CNN)



Neural nets, versus convolutional neural nets (CNN)

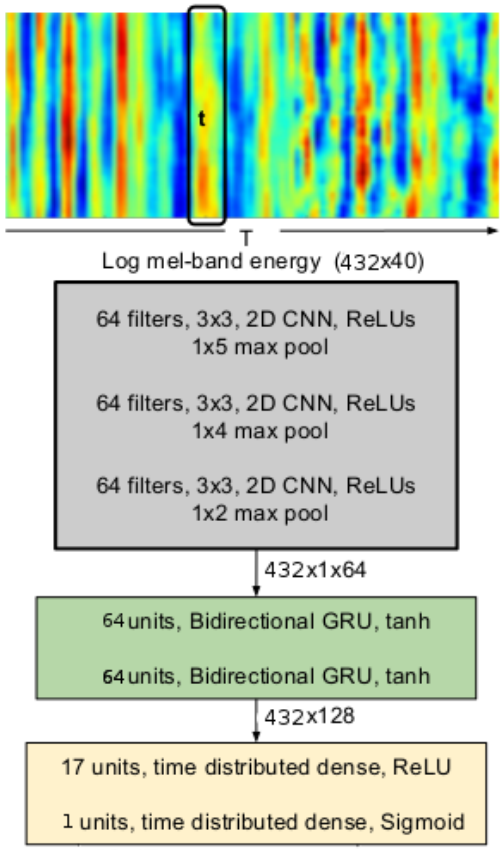




Shared weights among neurons

- Dramatically fewer free parameters → easier to train
- Shift-invariance built in
- Locality built in

Convolutional **and** recurrent neural nets ('CRNN')



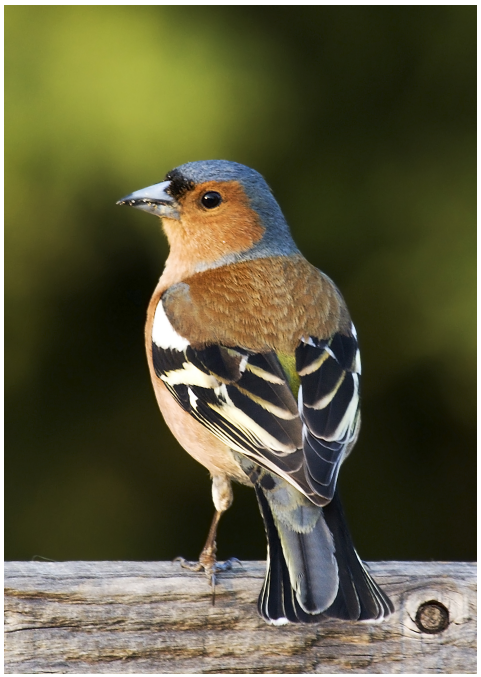
e.g. Cakir et al (2018); Morfi & Stowell (2018)

centre for digital music
Queen Mary University of London

Bird species classification

Automatic classification of bird species

- Useful for monitoring, research, archive management
  - —but only if it works in realistic conditions: noisy, and **many** possible species ($>$ 200 singing birds in UK, $>$ 400 in Brazil)

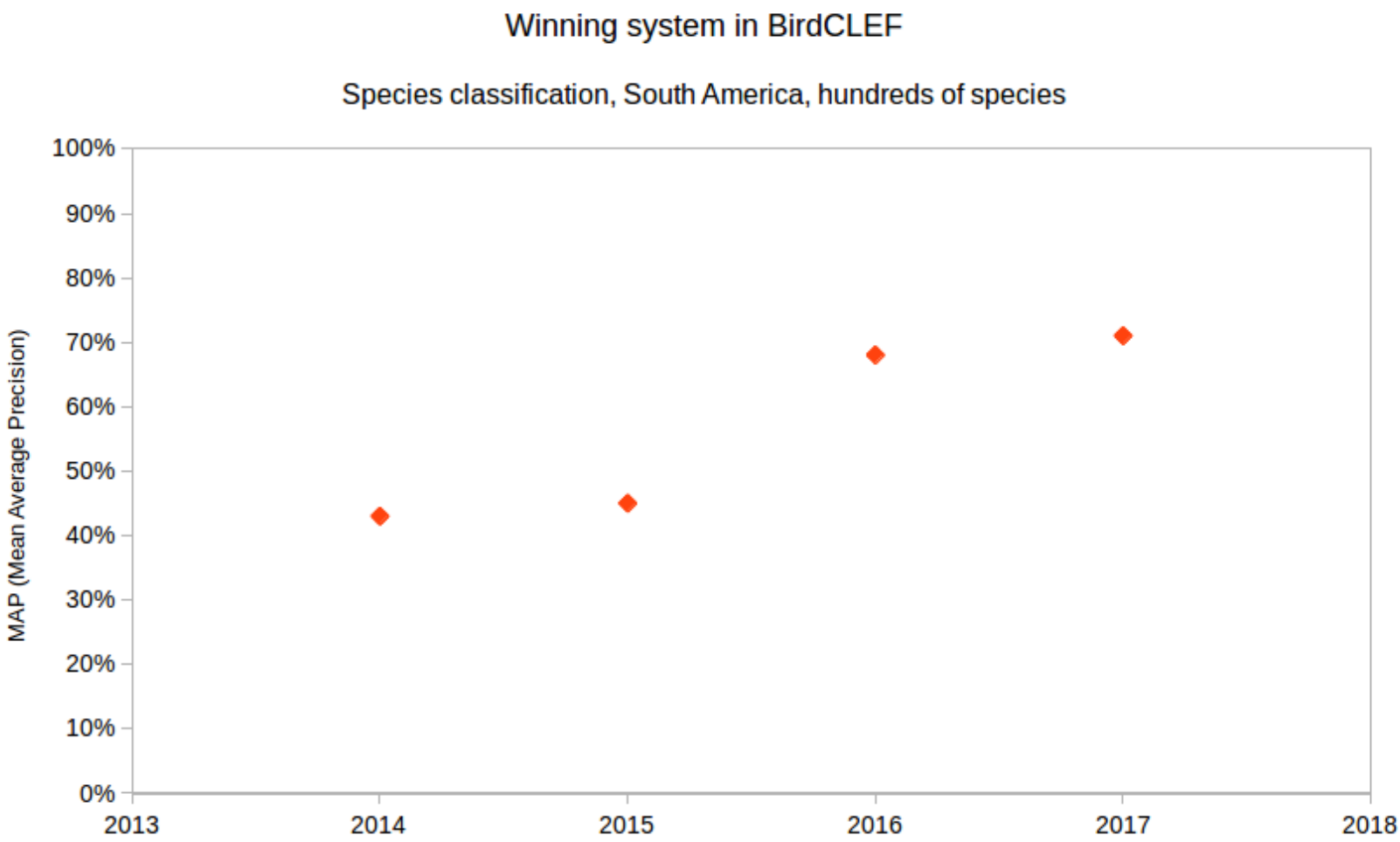


(Photo credit: Michael Maggs CC-by-sa)

centre for digital music

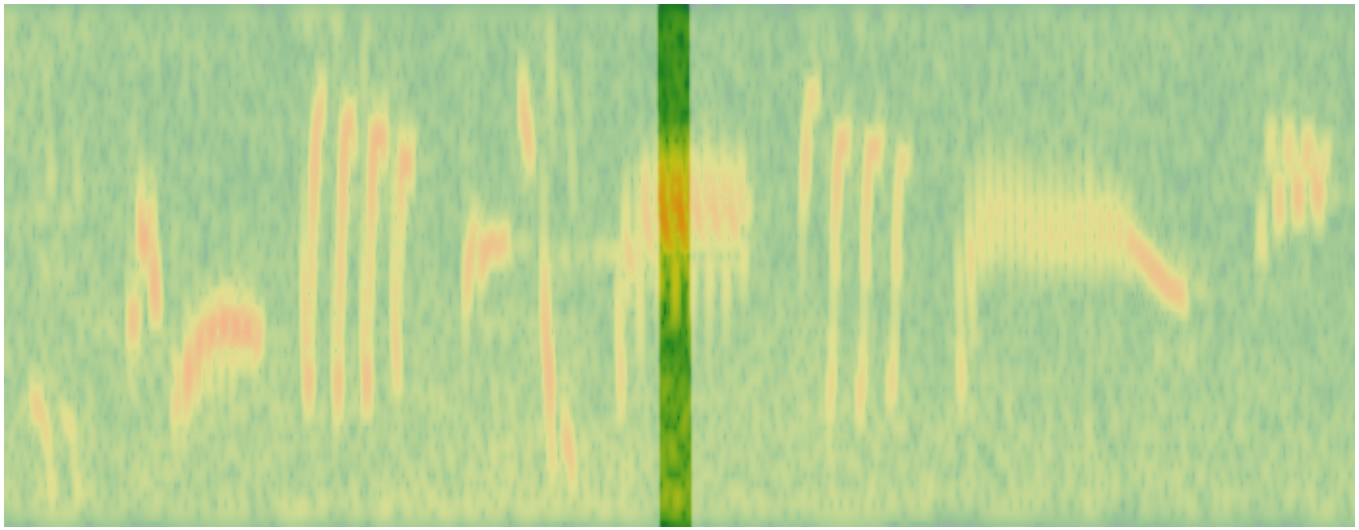
Queen Mary University of London

Bird species classification



Feature choice

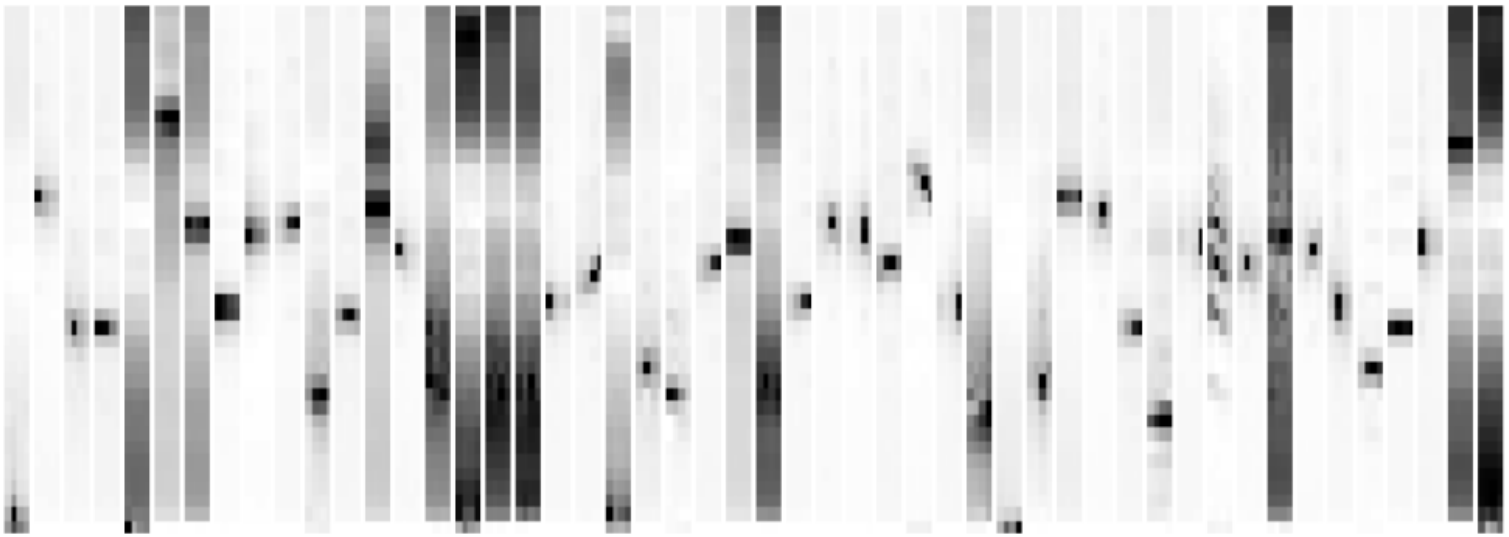
Our input data: "slices" from spectrograms
(e.g. 4 frames, approx 50 ms)



Stowell & Plumbley, **PeerJ** 2014

Feature learning

A selection of items from the 'dictionary' (learnt from Brazil birdsong, 77.8 hours):

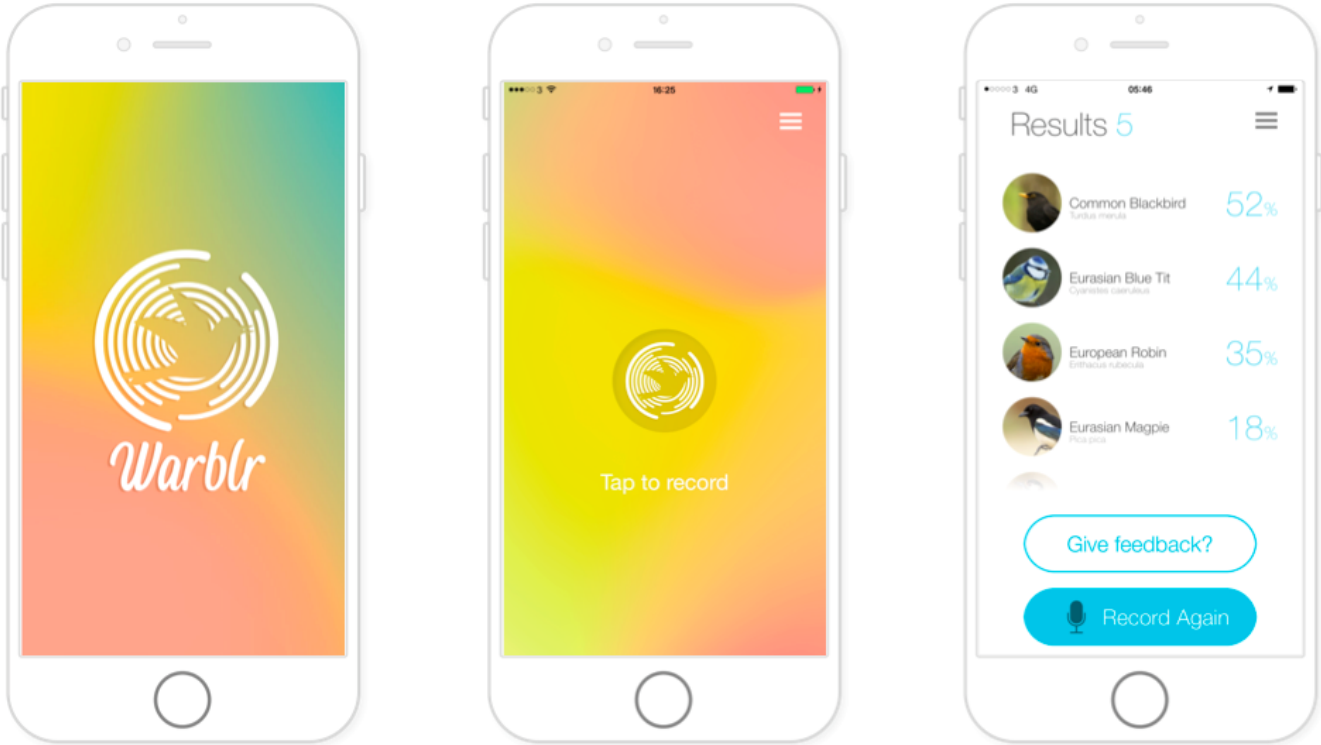


Stowell & Plumbley, **PeerJ** 2014

Bird species classification: Warblr

'Warblr' app – for Android and iOS

Bird species classification: Warblr

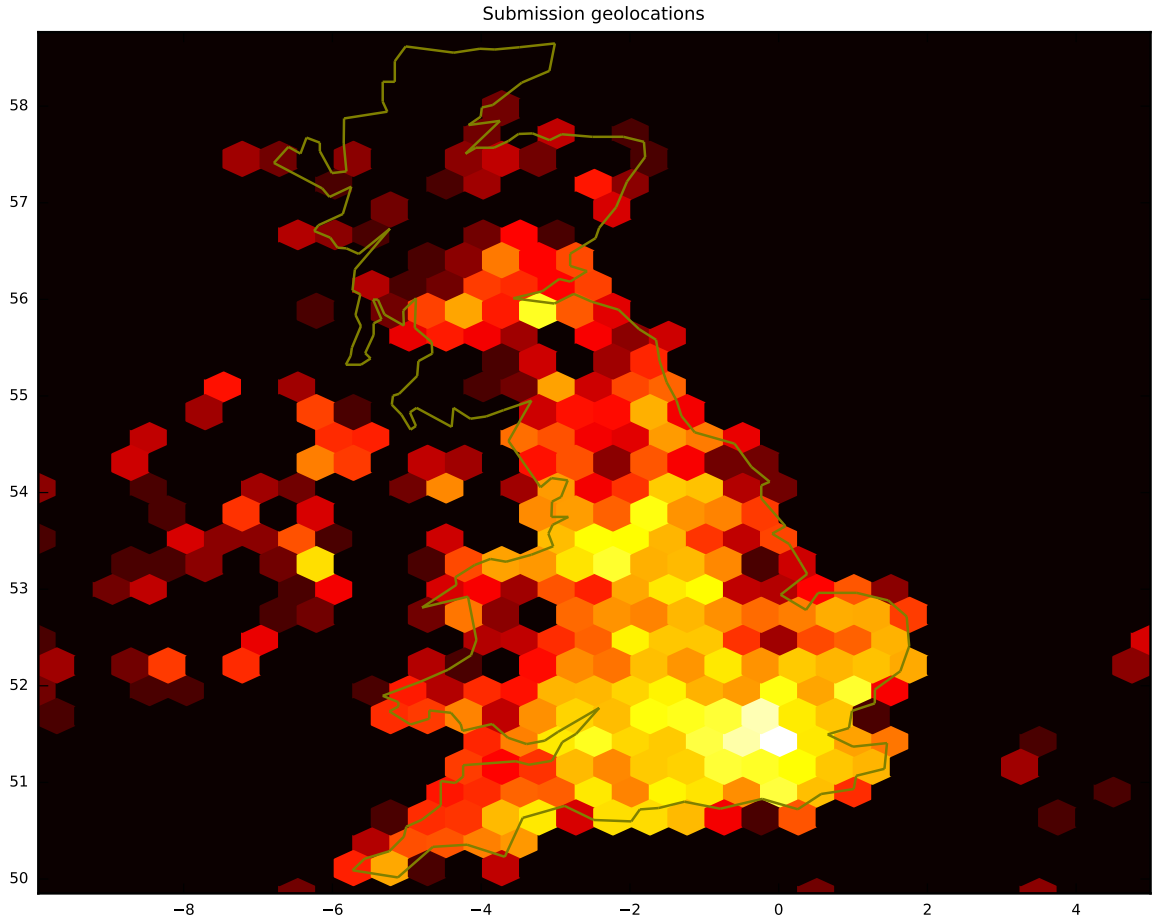


'Warblr' app – for Android and iOS

Warblr: who are our users?

5,000+ paying users since August 2015

Over 45,000 recordings submitted to our database (≈ 80/day)



Queen Mary University of London

Some of our users...

Some of our users...

Detection – a missing link in the chain?

Many projects need reliable *detection* of bird sounds
e.g. in long unattended recordings

Detection of bird sounds not the same as **classification**

...easier? harder?

Bird Audio Detection challenge

We designed the **Bird Audio Detection challenge**

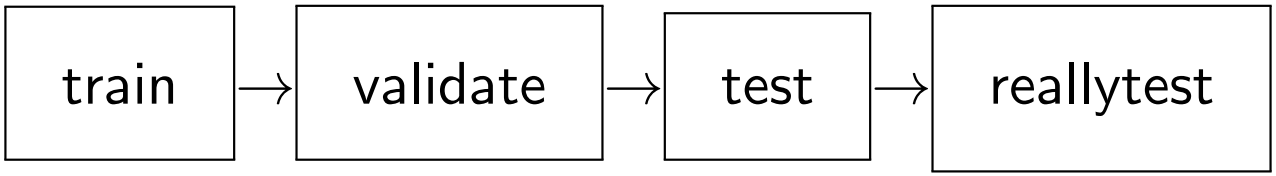
Dev set 1: 10k items, crowdsourced audio from around the UK (Warblr phone app)

Dev set 2: 7k items, crowdsourced audio from misc field recordings

Testing set: 10k items, remote monitoring, Chernobyl Exclusion Zone

Machine learning workflow

Machine learning workflow



Bird Audio Detection challenge: outcomes

- 30 teams submitted
- Strong results (up to 89% AUC)

User	Preview Score	Final Score	Classifier	Domain adaptation	Ensembling
bulbul	88.9 %	88.7 %	CNN	Pseudo-labelling	Model averaging
cakir	88.3 %	88.5 %	CRNN	no	no (for strongest submission); Model avera
topel	88.8 %	88.2 %	CNN-DenseNet	Pseudo-labelling	Multi-epoch, Model averaging(geom)
MarioElias	88.5 %	88.1 %	CNN, ExtraTreesRegressor	no	Model averaging (over 2 diverse methods)
adavanne	88.2 %	88.1 %	CRNN	Test mixing	no

- CNN vs CRNN
- *Domain adaptation* strategies
  - Pseudo-labelling, test mixing
  - Though not always needed

Stowell et al (accepted) **Meth Ecol and Evol**

centre for digital music — Queen Mary University of London

Low-resource tasks

"Bird detection" addressable as a big data situation
(2 classes, 10000s of items for each)

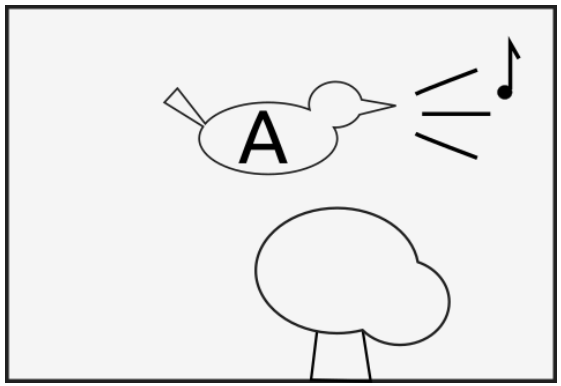
– what about tasks with much less available data?

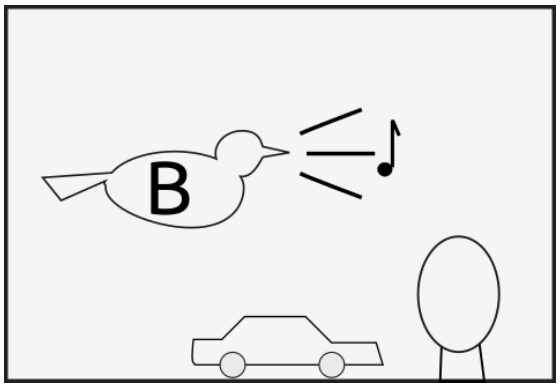
1. Identifying *individual* birds
2. Transcribing sound scenes in detail

Identifying individual bird ID

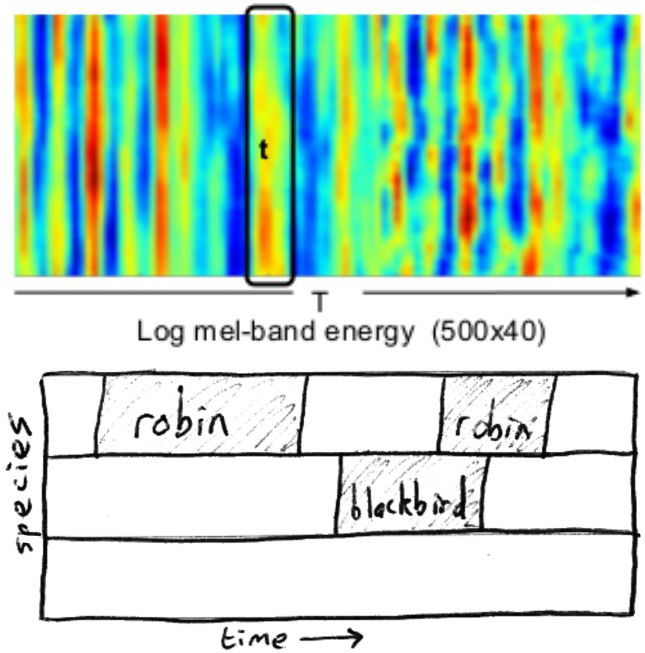
Motivation: reduce intrusive monitoring (capturing/tagging/ringing)

Many birds do have individual signature





"Automatic wildlife transcription"

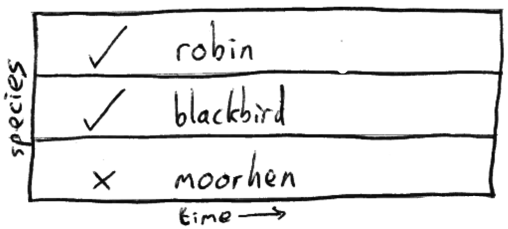


A bit like "automatic music transcription" or "speech diarisation"

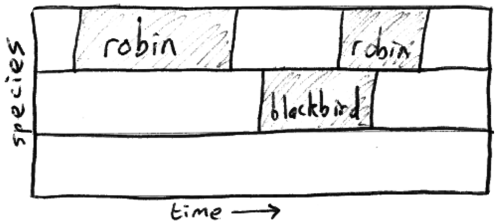
- Much more variability (templates / HMM are of limited use)
- V difficult to collect training annotations covering all sounds
- Many sounds are inherently rare

"Weakly-labelled" learning

Training data:



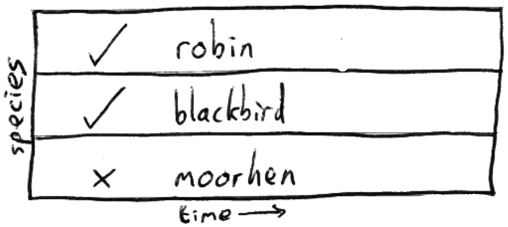
But we want:



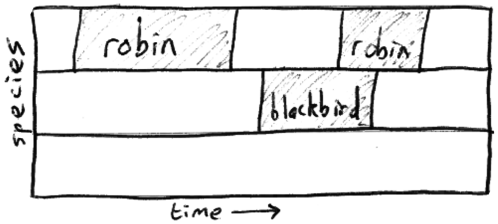
How to train?

"Weakly-labelled" learning

Training data:

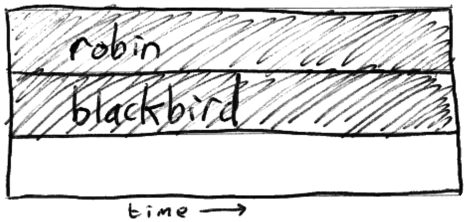


But we want:



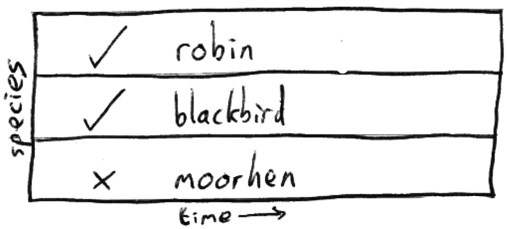
How to train?

- False strong labelling

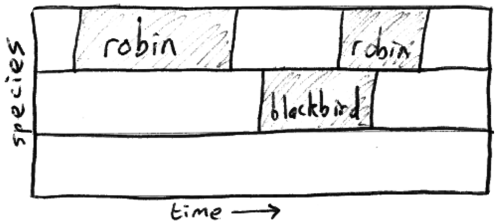


"Weakly-labelled" learning

Training data:

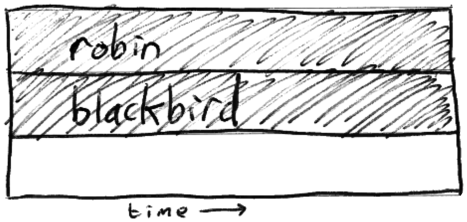


But we want:

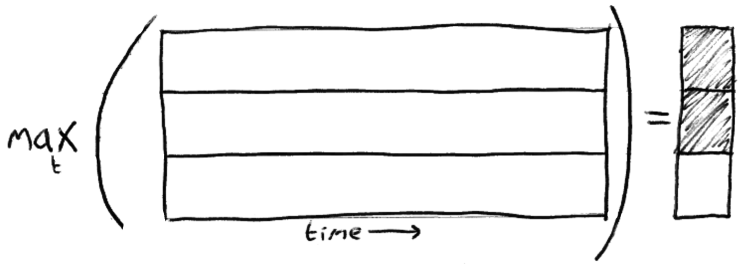


How to train?

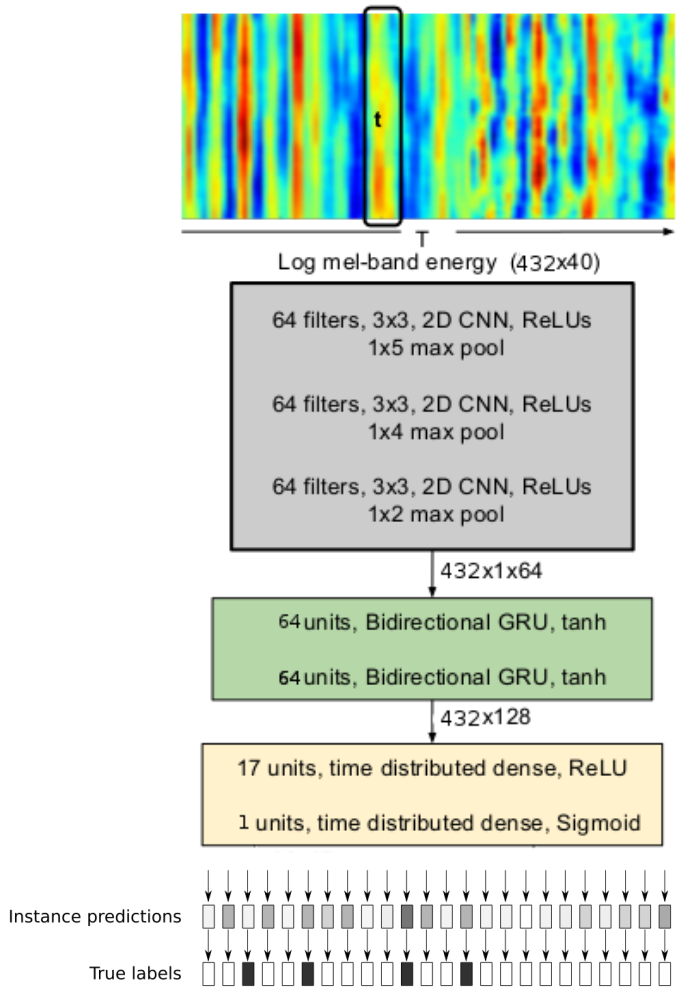
- False strong labelling



- 'Multi-instance learning' target:

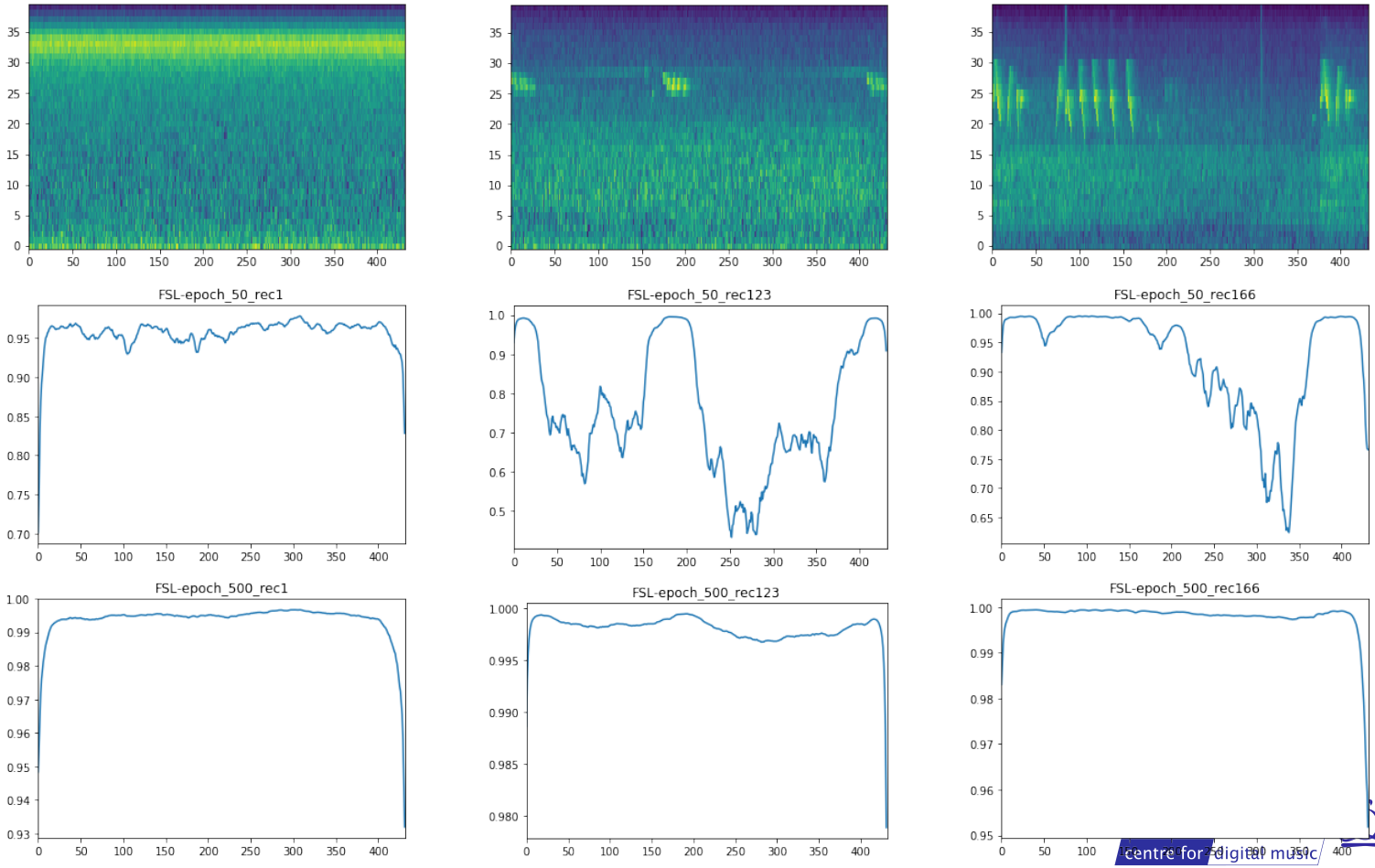


Convolutional recurrent neural network (CRNN)



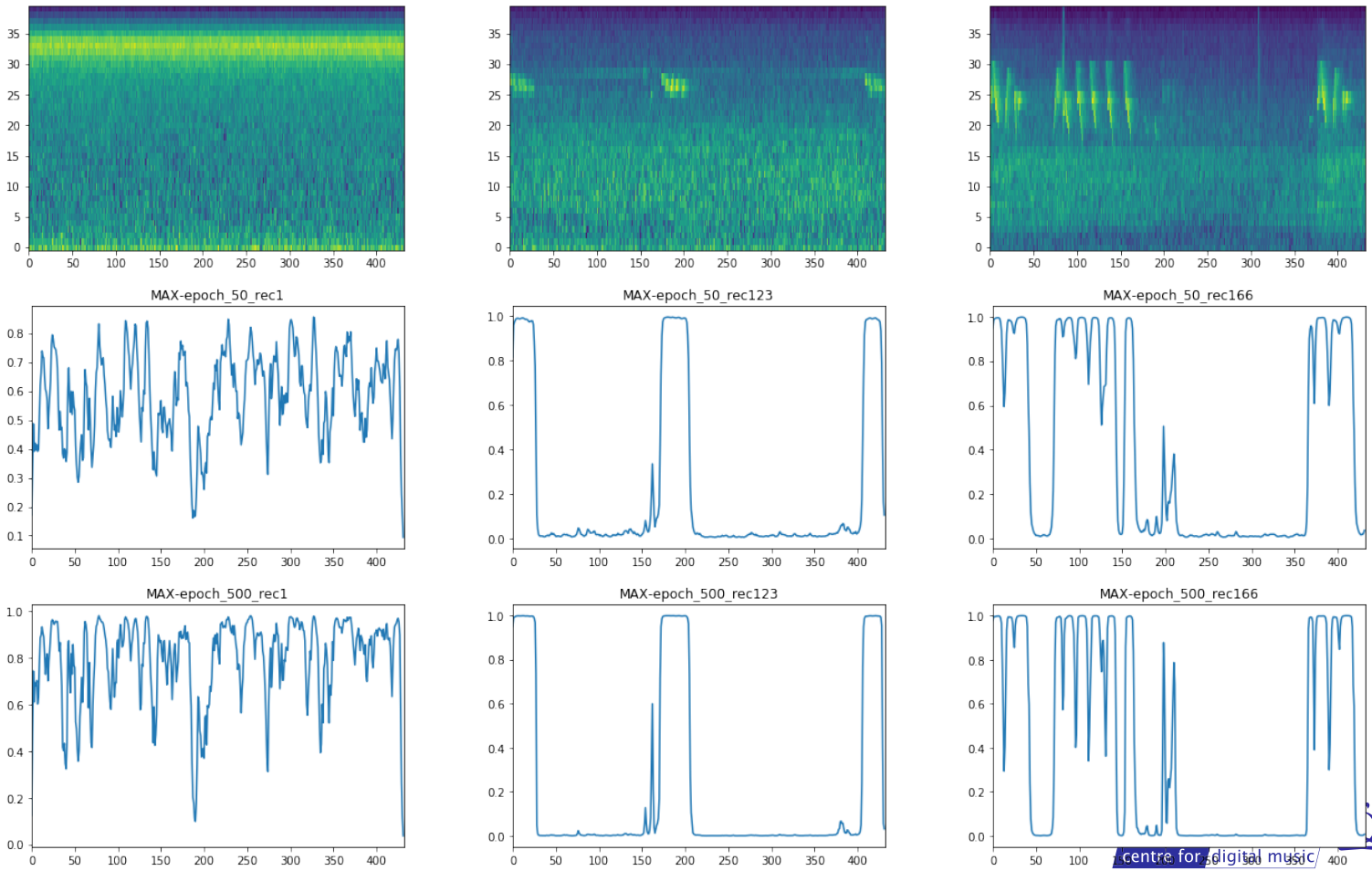
Weakly-labelled learning: results

Trained with "false strong labelling":



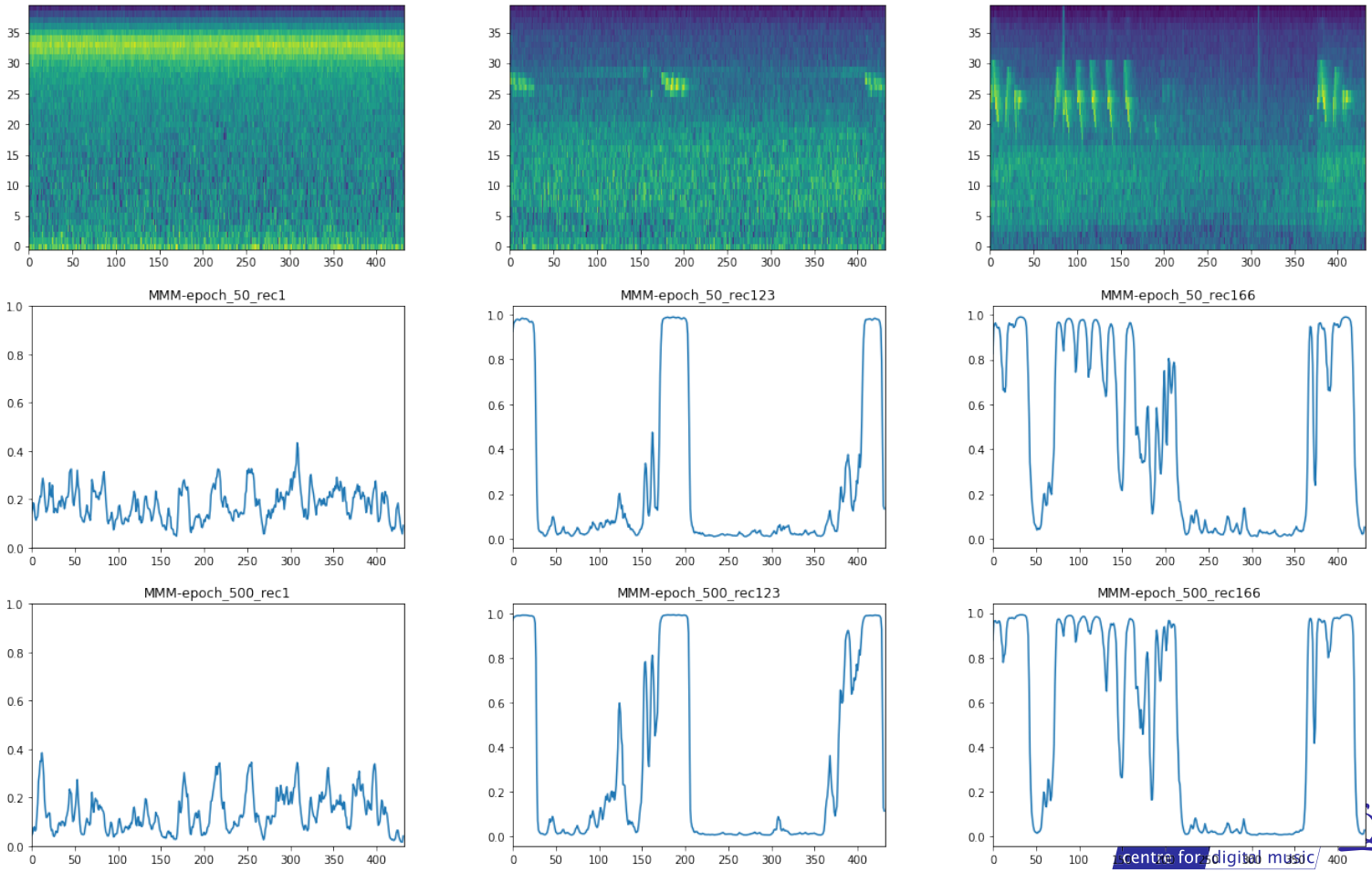
Weakly-labelled learning: results

Trained with multi-instance (max) objective:



Weakly-labelled learning: results

Trained with improved multi-instance objective:



Weakly-labelled learning

"Deep Learning for Audio Event Detection and Tagging on Low-Resource Datasets"
Morfi & Stowell (2018) **Applied Sciences**, 8(8), 1397; https://doi.org/10.3390/app8081397

Veronica Morfi (PhD student, 3rd year)

centre for digital music
Queen Mary University of London

Putting it all together

Putting it all together

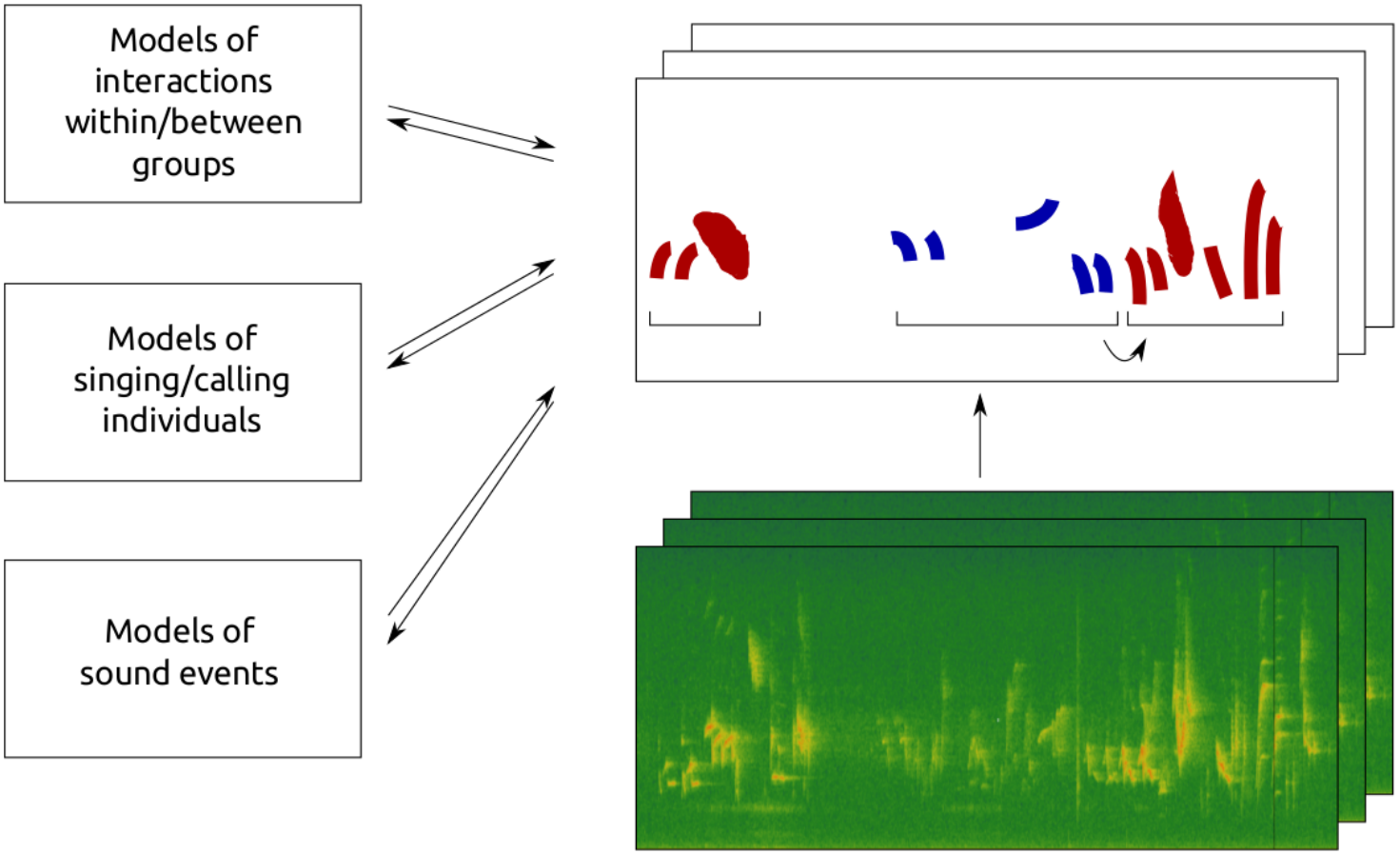
- Large-scale tasks
  - Bird detection
  - Bird species classification

  Achievable, but care needed—generalisation

- They feed into more fine-grained analysis
  - Identifying individuals
  - Transcription of sound scenes

  Care needed—choice of features, constraining the problem

Putting it all together



Conclusions

- Audio analysis
  - Many interesting open problems
  - Good examples of time series for machine learning development
- Bird sound analysis—useful and fascinating
  - Monitoring birds in the wild
  - Understanding animal behaviour
  - Comparative linguistics
  - Challenges to overcome:
    Limited data, unlabelled data, unknown information content

`mcld.co.uk/research`

centre for digital music
Queen Mary University of London

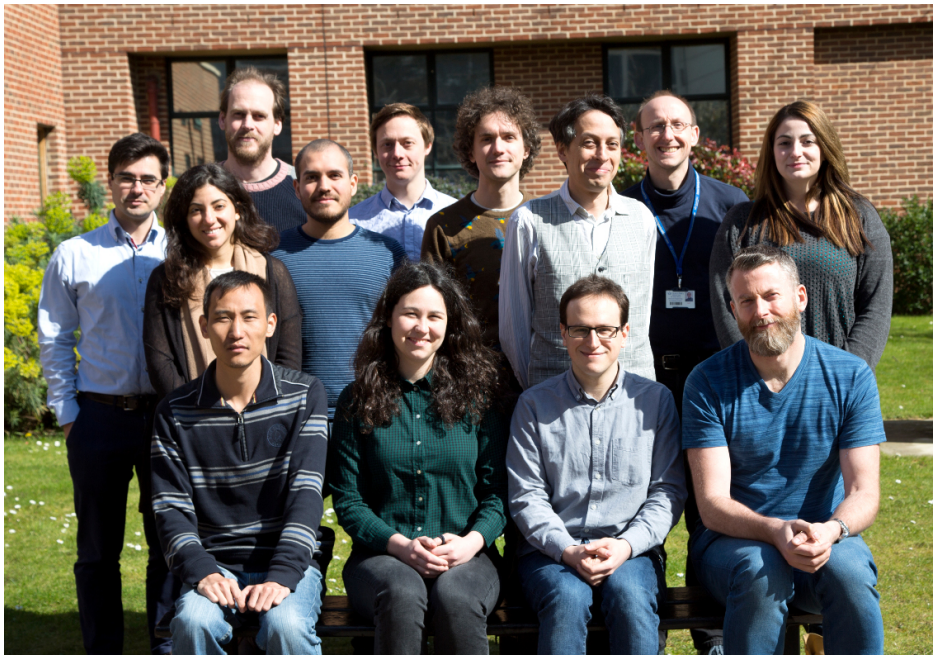
Conclusions

- Audio analysis
  - Many interesting open problems
  - Good examples of time series for machine learning development
- Bird sound analysis—useful and fascinating
  - Monitoring birds in the wild
  - Understanding animal behaviour
  - Comparative linguistics
  - Challenges to overcome:
    Limited data, unlabelled data, unknown information content

`mcld.co.uk/research`

Thanks

Machine Listening Lab



`machine-listening.eecs.qmul.ac.uk`

Thank you to: my collaborators, EPSRC for funding

centre for digital music

Queen Mary University of London

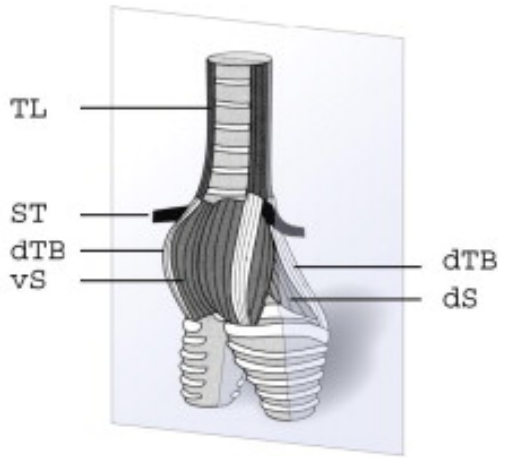
extra slides

Auto-ID: jackdaws

Collab with Lisa Gill, MPIO
Stowell et al. (2016), Proc InterSpeech

Auto-ID: jackdaws

Source-filter model of vocal production

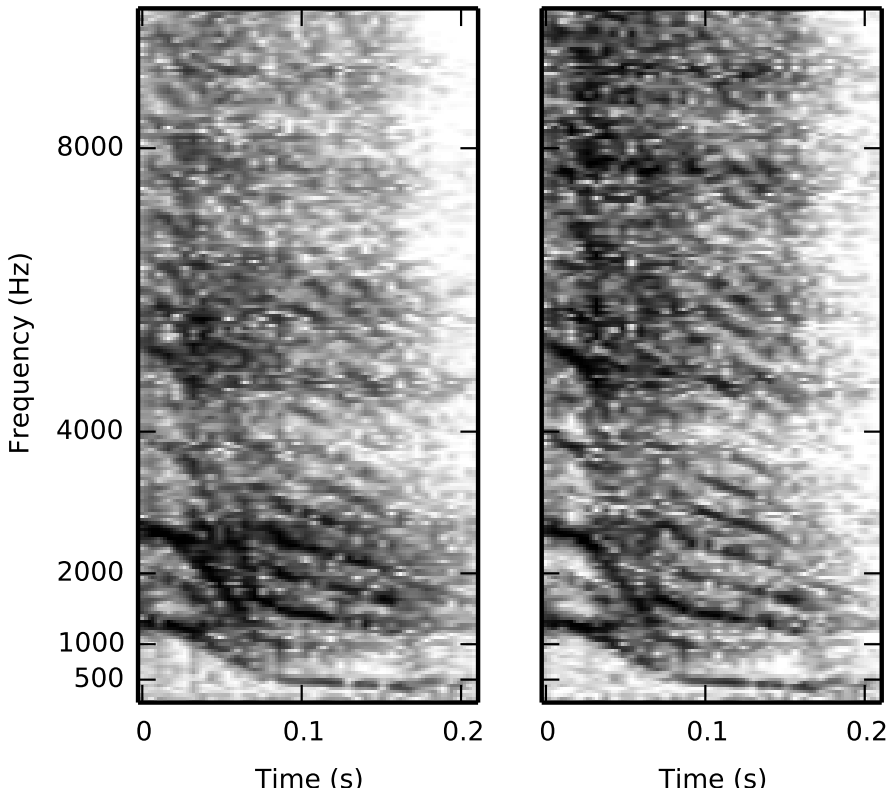


Songbird syrinx (Riede & Goller 2010)

(NB sound propagation through environment)

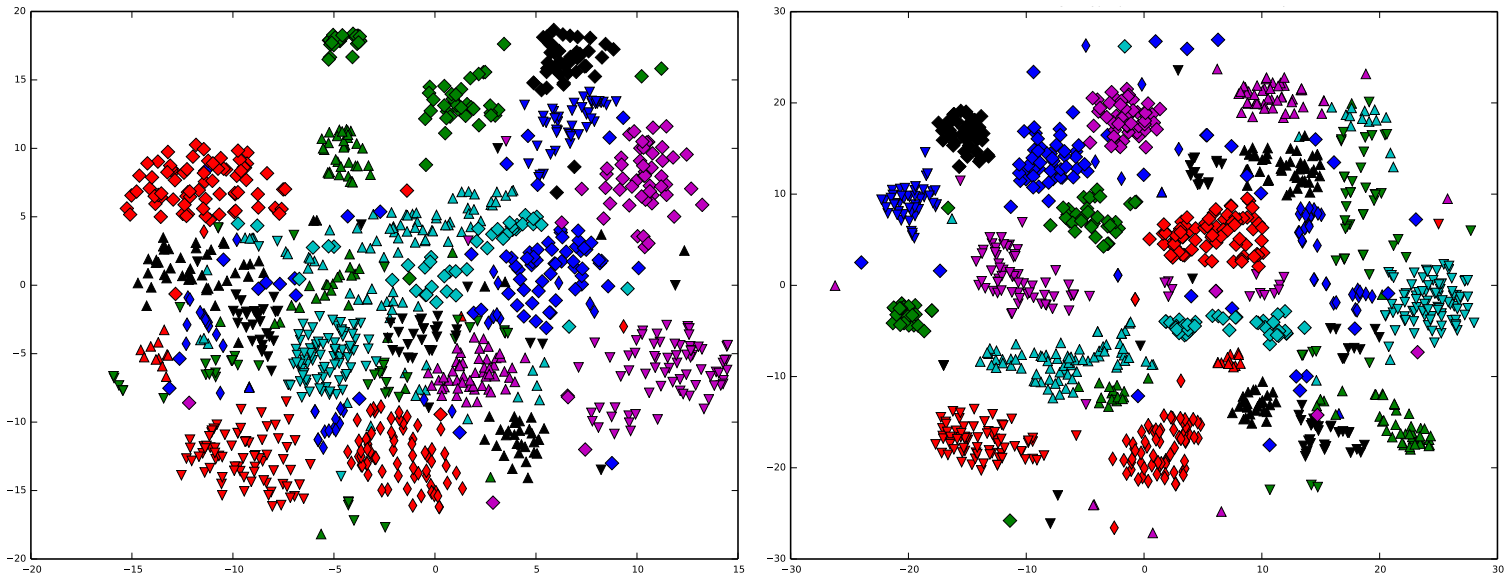
Auto-ID: jackdaws

LPC preprocessing before/after:



Auto-ID: jackdaws

LPC preprocessing before/after:



t-SNE distance plots
(each point is one call; marker style indicates individual)