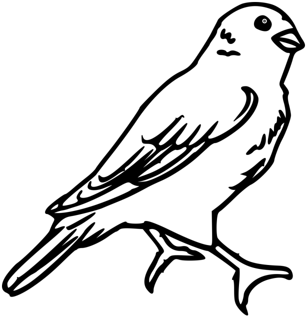
Reducing confounding factors in automatic acoustic recognition of individual birds

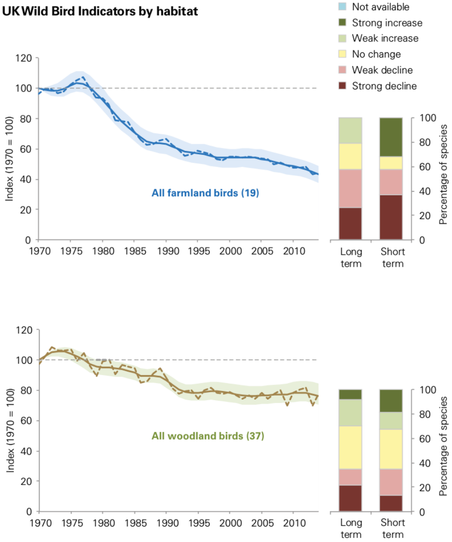
Dan Stowell

Machine Listening Lab
Centre for Digital Music



versus

Machine listening and bird sounds - why?



Machine listening and bird sounds - why?

- Changes in populations, in migration patterns
  —monitoring is important
- Intrusive vs. passive monitoring
  —behavioural impact of catching/ringing birds
- Many birds are most easily observed by sound
  - Manual (volunteer) monitoring common, but not scalable

In this talk...

Classification-based approaches to:

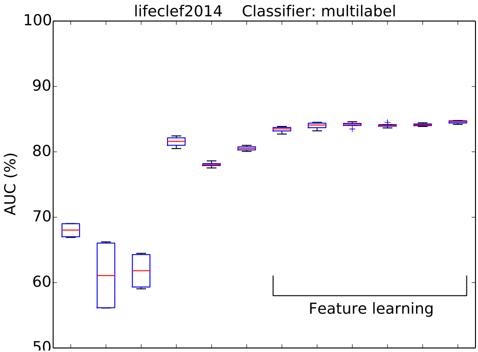
0. Bird species recognition
1. Bird sound detection (presence/absence)
2. Bird individual ID

(By the way: we do more than just classification!)

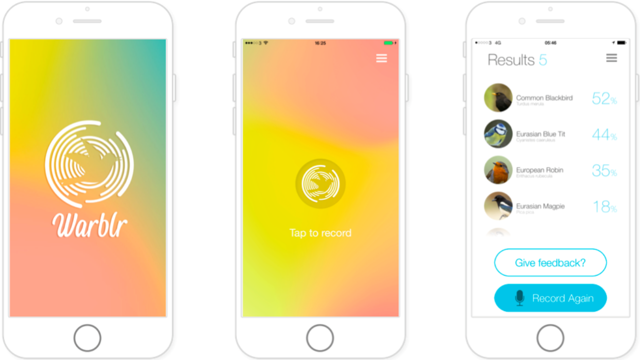
Species classification of bird sounds

In 2014: *feature-learning* approach to bird sound recognition

Dataset	Location	Total duration	Num items	Num classes	Labelling
lifeclef	Brazil	77.8 hours (12M frames)	9688	501	singlelabel



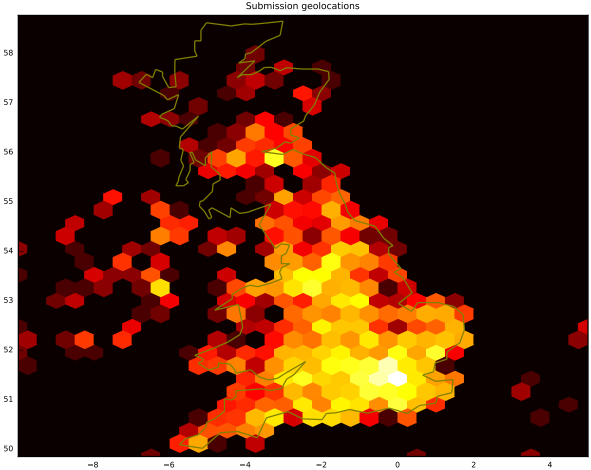
Bird species classification: Warblr



‘Warblr’ app – for Android and iOS

Bird species classification: Warblr

Over 45,000 recordings submitted to our database ($\approx$ 80/day)



Some of our users...

Some of our users...

Part 1: Bird Audio Detection challenge

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

But existing methods are not robust, not general-purpose enough, and need lots of manual tweaking/post-processing

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

Bird Audio Detection challenge

- Training/testing sets differ in:
  - location
  - recording eqpt
  - species
  - class balance
  - background sounds
  - time of day
  - time of year
  - weather
  - ...
- How is a classifier meant to work in such mismatched conditions???

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 averaging
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
Elias	88.0 %	88.0 %	CNN	no	Model averaging
kdrosos	86.1 %	85.8 %			

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

So why do we evaluate using matched conditions?

- To study the classifier's behaviour
- *Sometimes* a practical application is in matched conditions
- Pragmatic reasons: only one dataset available; free choice of bootstrap/n-fold crossvalidation
- ...because our algorithms aren't good enough at avoiding confounds?

Machine learning workflow

train → validate → test

Machine learning workflow

train → validate → test → reallytest

Part 2: Identifying individual bird ID

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

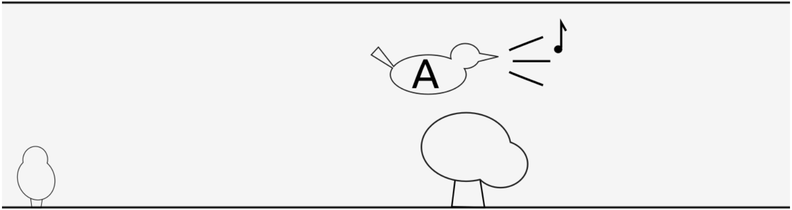
Many birds do have individual signature

centre for digital music

Queen Mary University of London

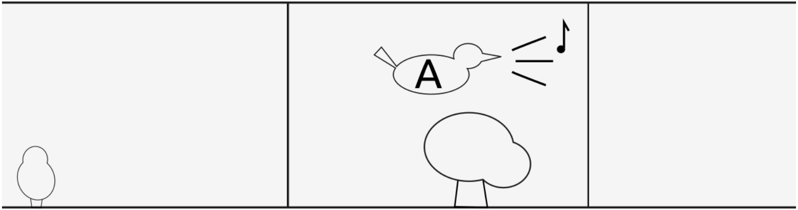
Identifying individual bird ID

Data collection:



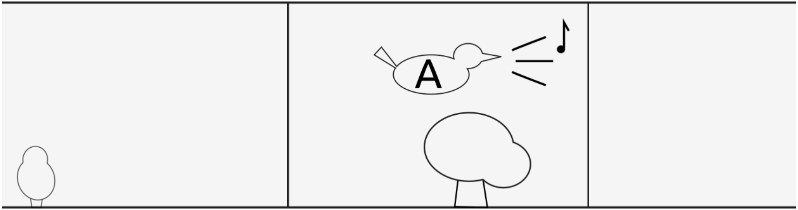
Identifying individual bird ID

Data collection:



Identifying individual bird ID

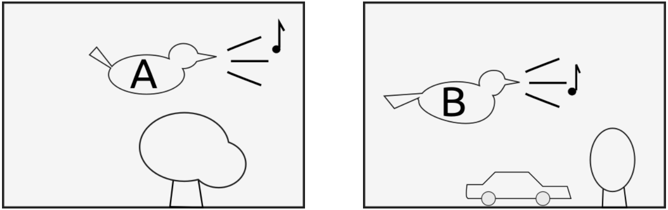
Data collection:



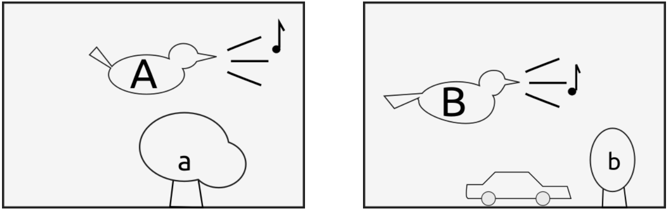
Bird ID: categorical label.

Is this the "same" task as species classification?

Identifying individual bird ID

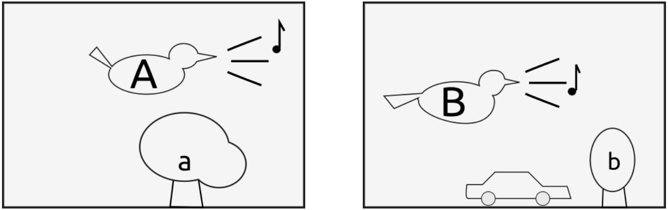


Identifying individual bird ID



Making use of silence (1)

Training set:

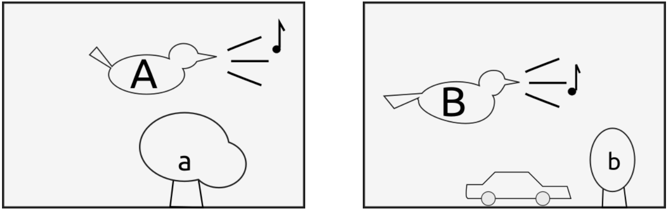


centre for digital music

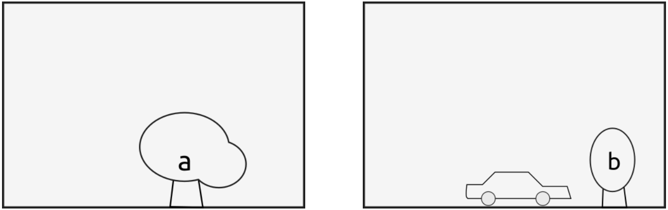
Queen Mary
University of London

Making use of silence (1)

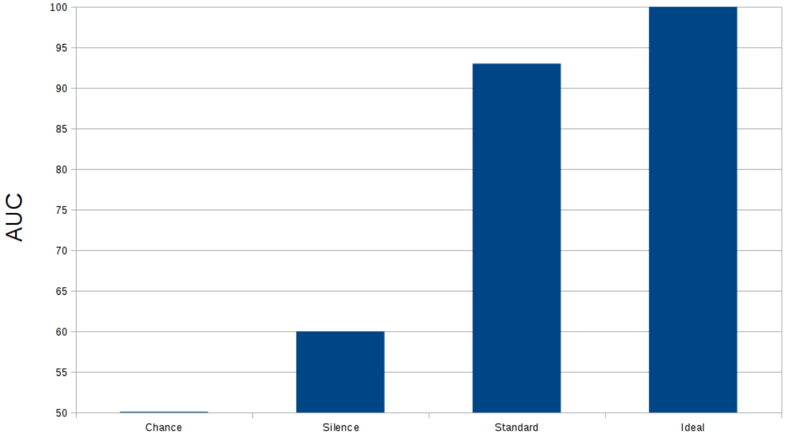
Training set:



Testing set:

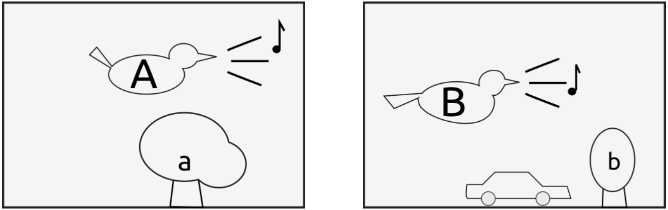


Making use of silence (1)

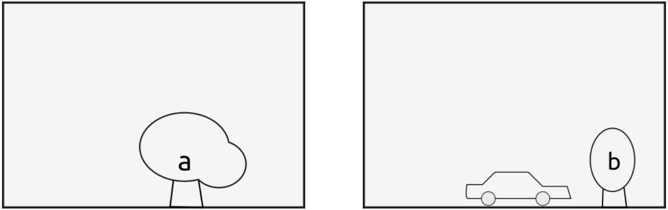


Making use of silence (1)

Training set:

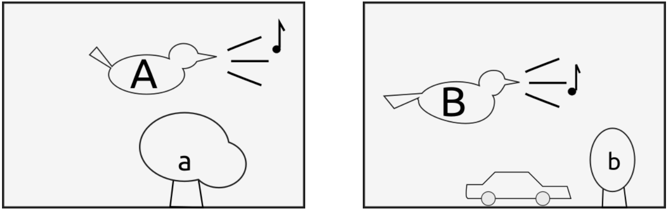


Testing set:

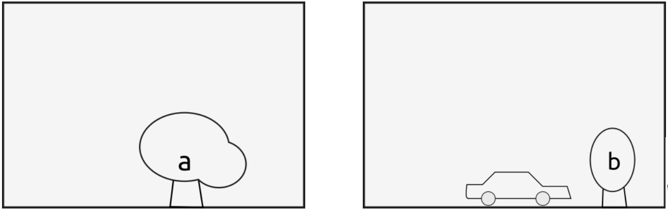


Making use of silence (1)

Training set:



Testing set:



Analogy: the 'album effect' in music artist ID

Training set:

Express Yourself

Bad

Analogy: the ‘album effect’ in music artist ID

Training set:

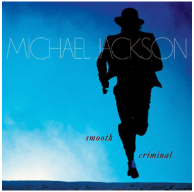
Express Yourself

Bad

Testing set:

Like a Prayer

Smooth Criminal



centre for digital music

Queen Mary University of London

Analogy: the 'album effect' in music artist ID

Training set:

Express Yourself

Bad

Testing set:

Like a Prayer

Smooth Criminal

Analogy: the 'album effect' in music artist ID

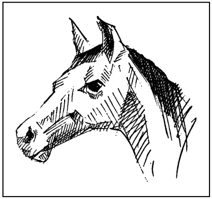
Training set:

Express Yourself

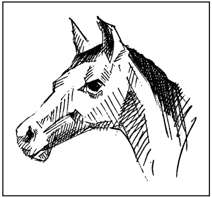
Bad

Testing set:

Like a Prayer



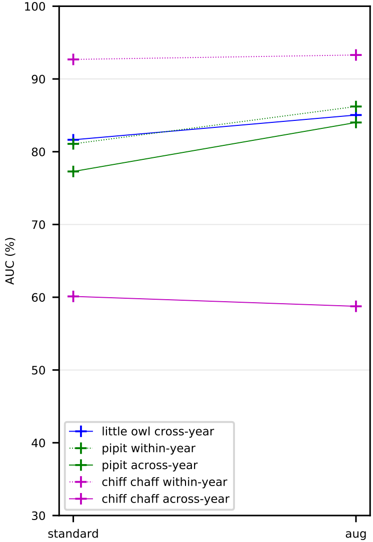
Smooth Criminal



centre for digital music

Queen Mary
University of London

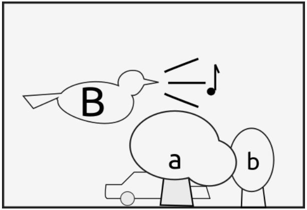
Territorial birds: the territory is the 'album'



Making use of silence (2)

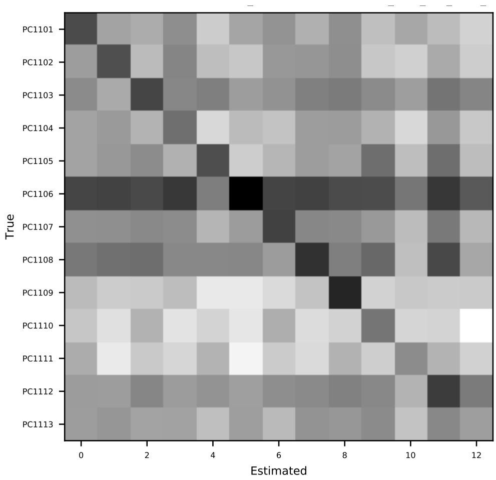
Data augmentation of the TESTING set (adversarial)

Measure the 'distractability' of the classifier
when mismatched silence is added



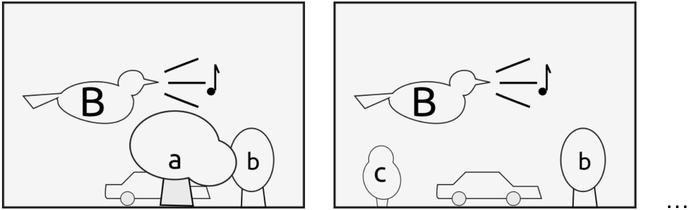
Measure RMSE in classifier decisions

Making use of silence (2)



Making use of silence (3)

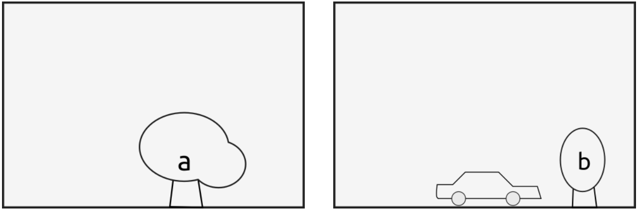
Data augmentation of the TRAINING set



Each item gets new versions with added silence from each class

Making use of silence (4)

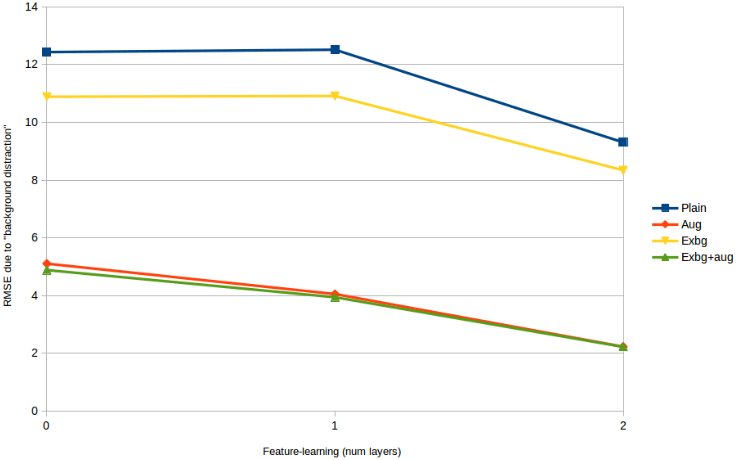
Finally we can add a new wastebasket class



…

NB not using the a/b labels here

Results



Plus silence-test result: 50% AUC

Conclusions

Outdoor bird sound recognition is tricky:

- The sounds (classes) are highly variable
- Many potential confounding factors for black-box ML

1. Bird Audio Detection Challenge:
   - Good detection, even in strongly mismatched conditions
   - Adaptation methods useful—though, not always needed?
2. Recognising individual bird ID:
   - Strong recognition possible (depending on species)
   - Silence is surprisingly useful for sound recognition!

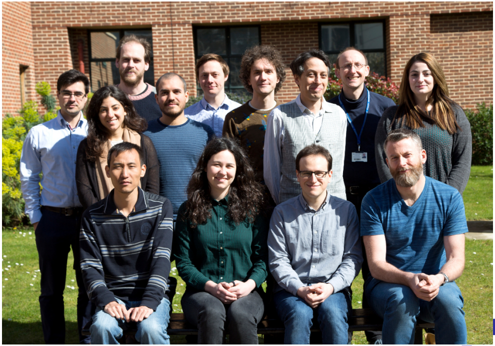
Generally: make more use of mismatched-condition testing

Thank you

Collaborators:

1. Bird Audio Detection Challenge:
   Mike Wood (U of Salford), Yannis Stylianou (U of Crete), Hervé Glotin (U of Toulon), IEEE Signal Processing Society
2. Recognising individual bird ID:
   Pavel Linhart (Adam Mickiewicz U / Praha U)

Machine Listening Lab:



for digital music

Queen Mary University of London