Automated acoustic observatories

Scaling for the future

Ed Baker

University of York

What's the big goal?

Historically there are many.

- How many species are there?
- What are the distributions of these species?
- What do these species do?

- Can we accurately understand (model) an ecosystem?
- Can we do this for the entire biosphere?

Can we do everything?

Can we do everything?

No.

How many species are there?

From an old sound recording to a new species in the genus *Horatosphaga* (Orthoptera: Tettigonioidea: Phaneropterinae: Acrometopini)

K.-G. HELLER[1,3] & ED BAKER[2]

[1]*Grillenstieg 18, 39120 Magdeburg, Germany. E-mail: heller.volleth@t-online.de*

[2]*Department of Electronic Engineering, University of York, York, YO10 5DD, United Kingdom & Department of Life Sciences, Natural History Museum, London, SW7 5BD, United Kingdom*

[3]*Corresponding author*

What are the distributions of species?

Remarkably and immediately preceding this discovery, a large colony of **Tree Cricket** *Oecanthus pellucens* was also discovered at Dungeness in 2015. With at least 50 singing males and a similar number of females as well as nymphs this population may have a good chance of persisting (Walker, *loc. cit.*). This species had previously been recorded as singletons in Cambridge in 1996 and at Sittingbourne in Kent in 2005 (Beckmann & Sutton, 2015), and in 2010 the first breeding colony was reported from Jersey (David, 2013). In 2016, the Dungeness colony was again observed in August by David Walker, who estimated that over 100 singing males were present at the site (Sutton, 2016; Beckmann & Sutton, 2016).

Sutton, Beckmann & Nelson (2017)

What do these species do?

Abstract

1. Sexual signals may be acquired or lost over evolutionary time, and are tempered in their exaggeration by natural selection.
2. In the Pacific field cricket, *Teleogryllus oceanicus*, a mutation ("flatwing") causing loss of the sexual signal, the song, spread in <20 generations in two of three Hawaiian islands where the crickets have been introduced. Flatwing (as well as some normal-wing) males behave as satellites, moving towards and settling near calling males to intercept phonotactic females.

Zuk, Bailey, Gray & Rotenberry (2018)

The case for automation

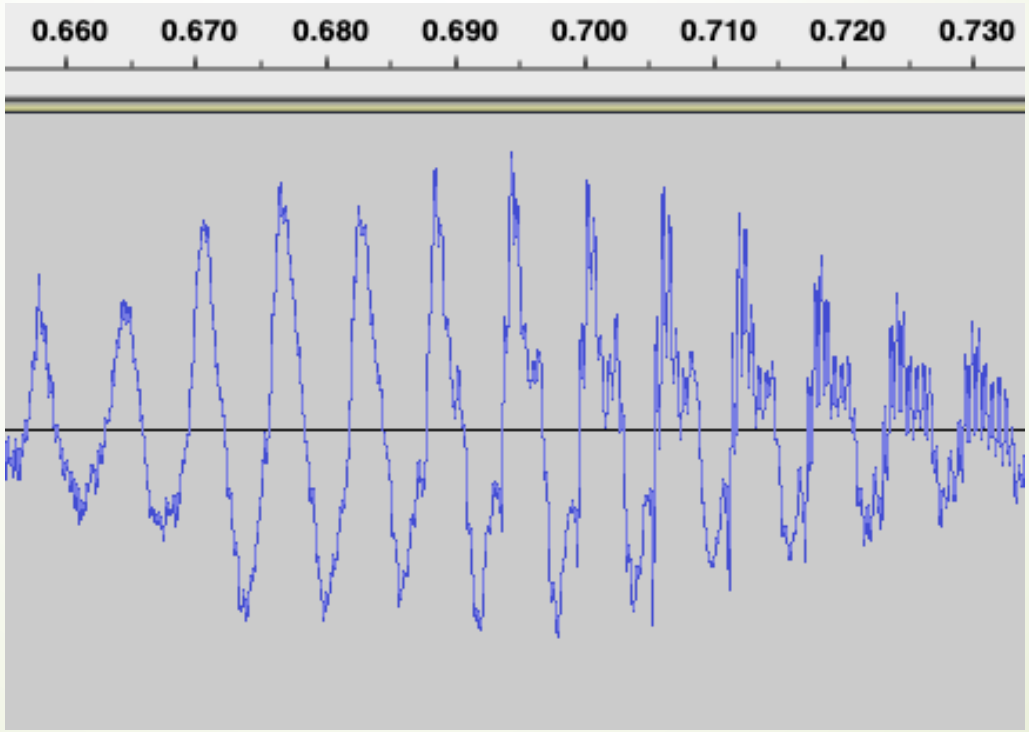
- A desire to do more with the current resources
- Our collective abilities are insufficient to fulfil the major goals for every species

The case for sound

- At it's most basic it is a 2-dimensional problem (amplitude varying with time)

The case for sound

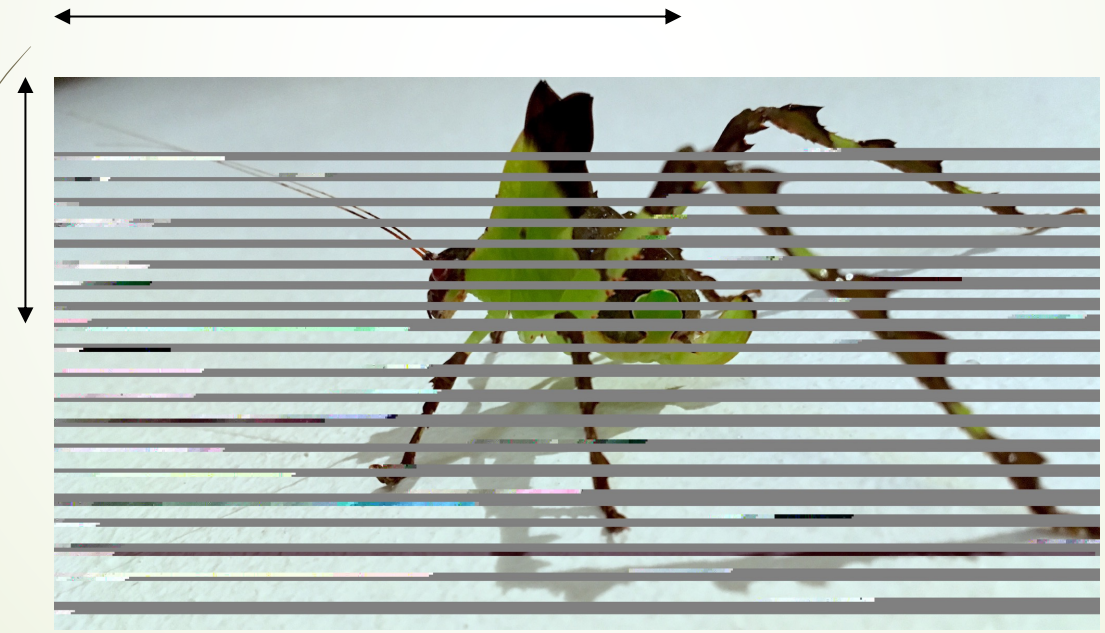
- At it's most basic it is a 2-dimensional problem (amplitude varying with time)



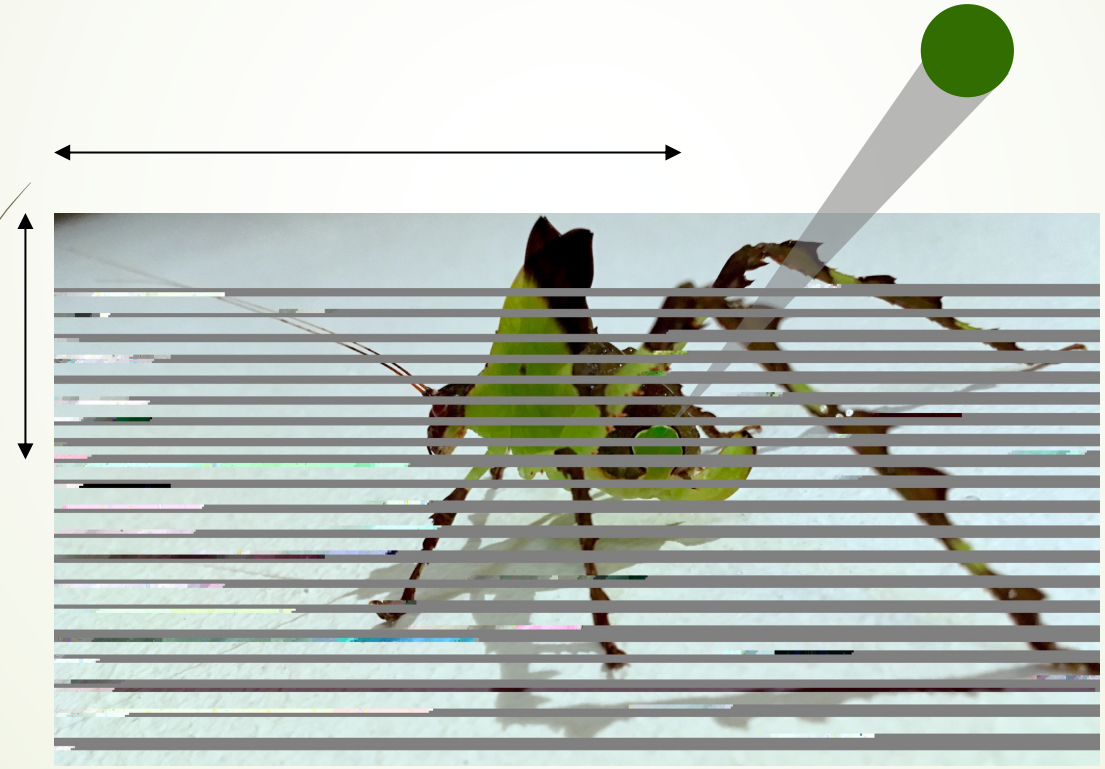
The case for sound

The case for sound

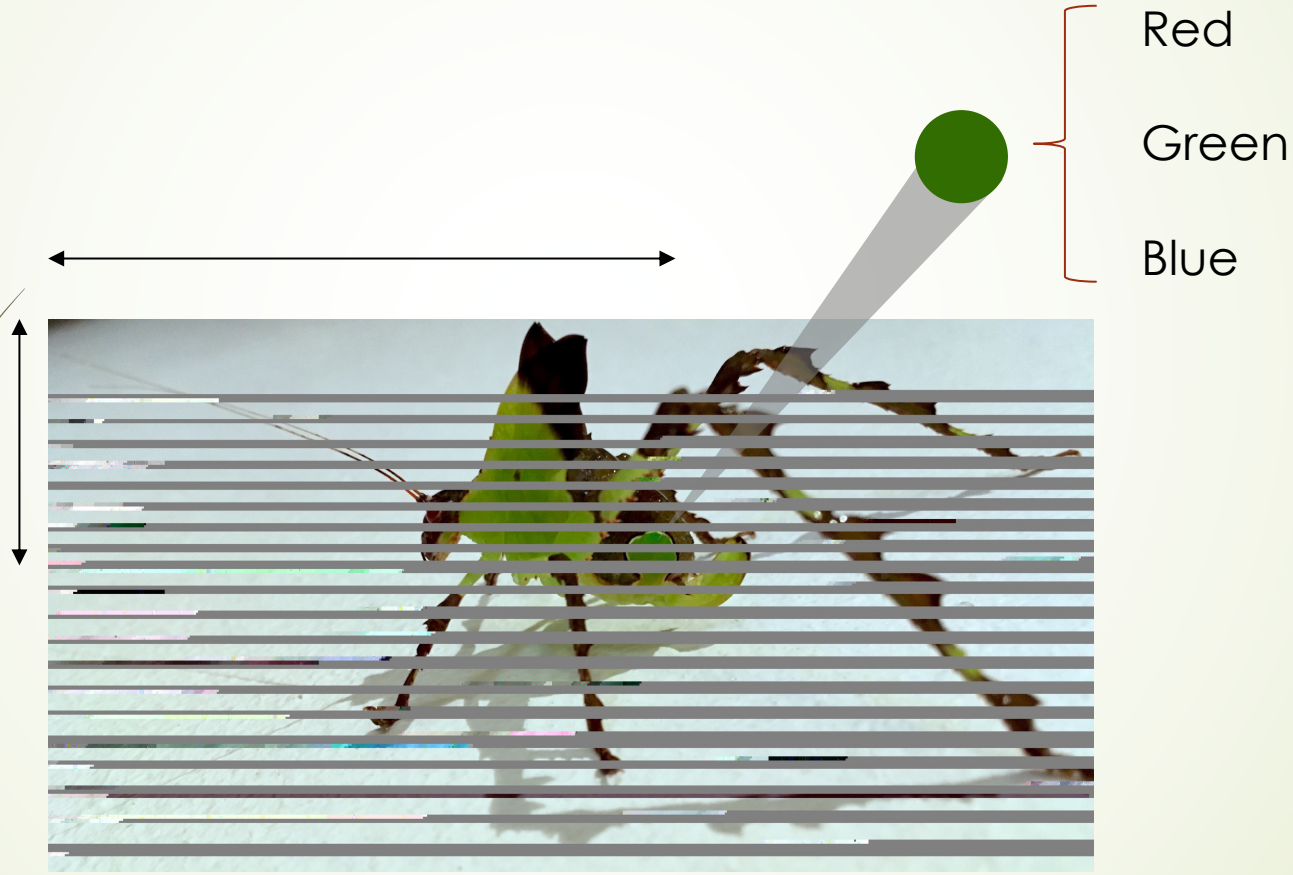
The case for sound



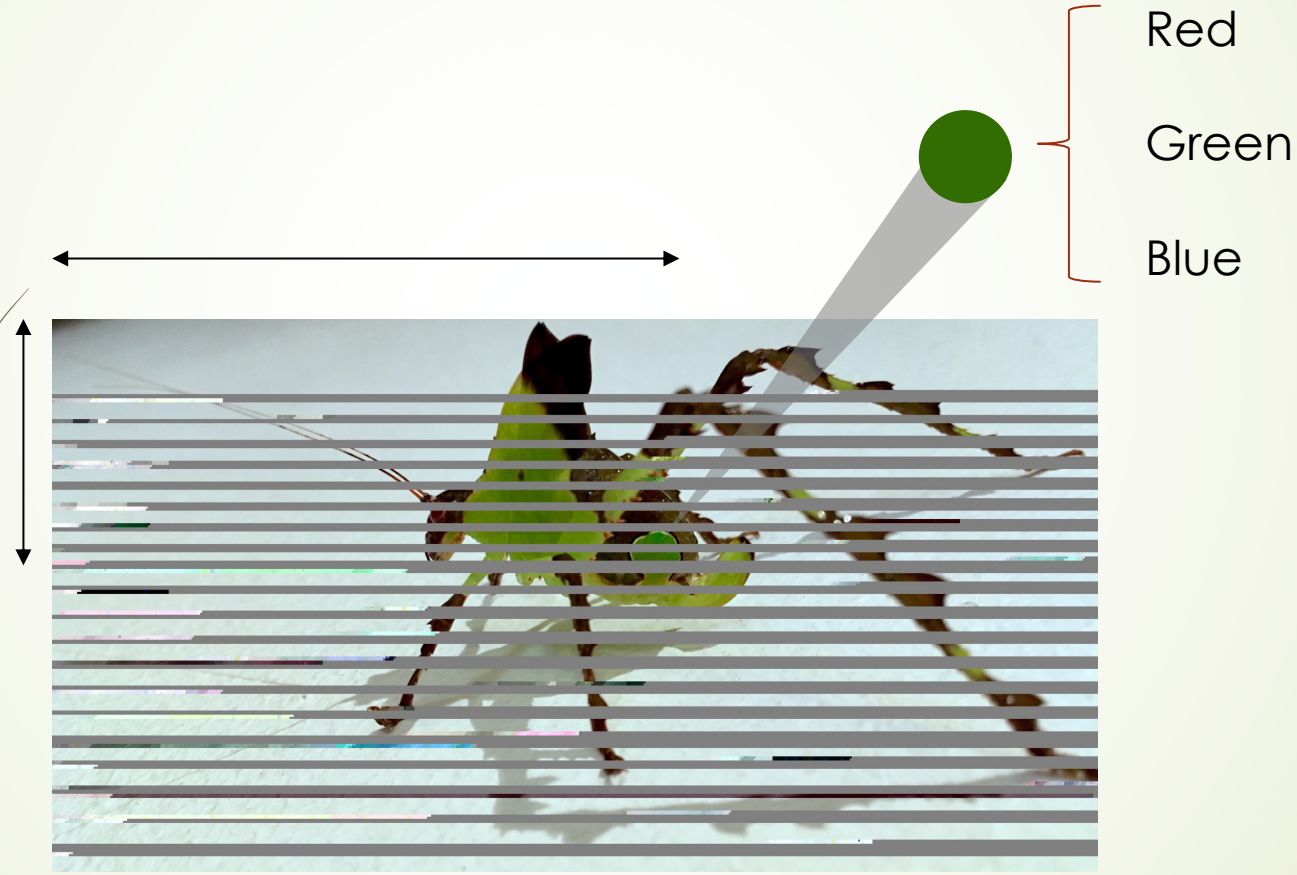
The case for sound



The case for sound



The case for sound



5-dimensional

Why is this important?

Why is this important?

- Computational complexity ≈ x^n

Why is this important?

- Computational complexity ≈ x^n
- Fewer dimensions allows for significantly faster analysis

Why is this important?

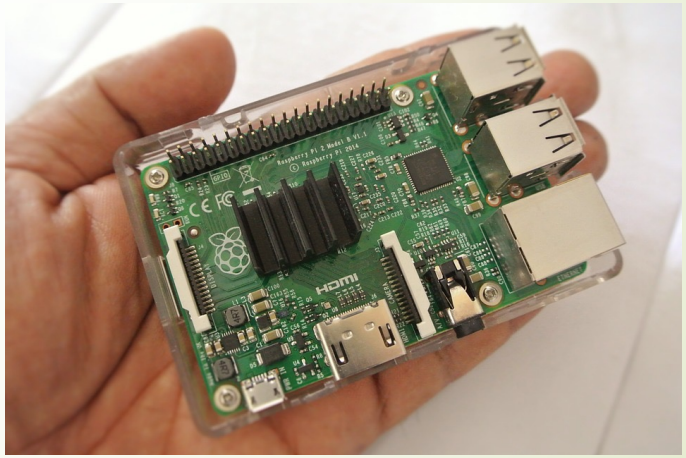
- Computational complexity ≈ x^n
- Fewer dimensions allows for significantly faster analysis

1990s

Why is this important?

- Computational complexity ≈ x^n
- Fewer dimensions allows for significantly faster analysis

1990s

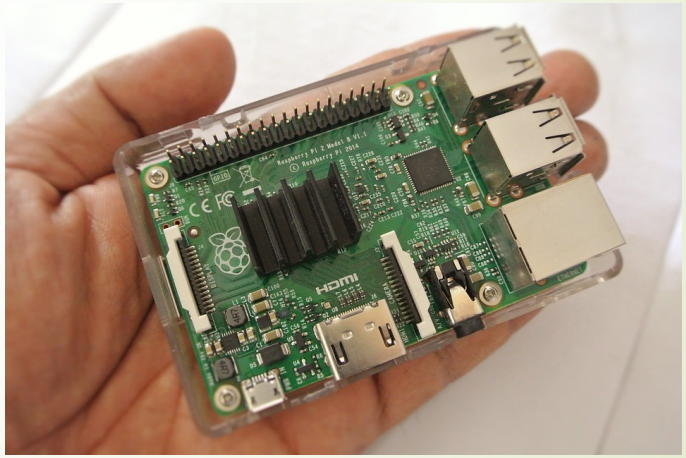


2010s

Why is this important?

- Computational complexity ≈ x^n
- Fewer dimensions allows for significantly faster analysis

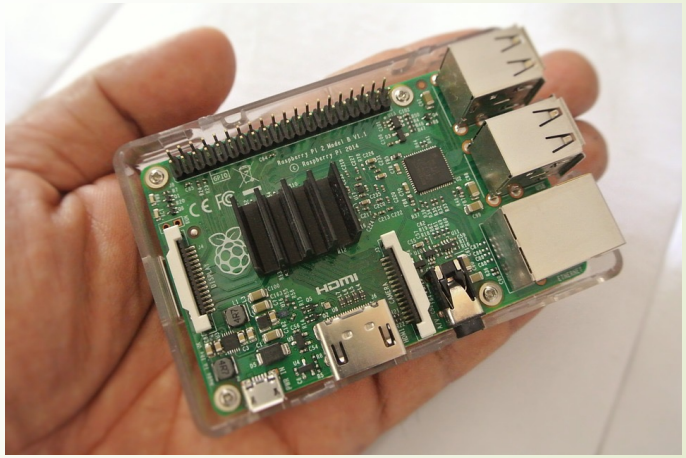
£2,000



£25

Why is this important?

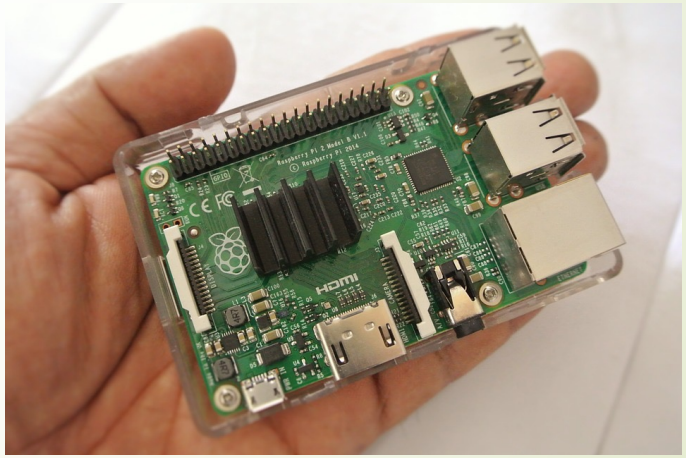
- Computational complexity ≈ x^n
- Fewer dimensions allows for significantly faster analysis



8x cheaper

Why is this important?

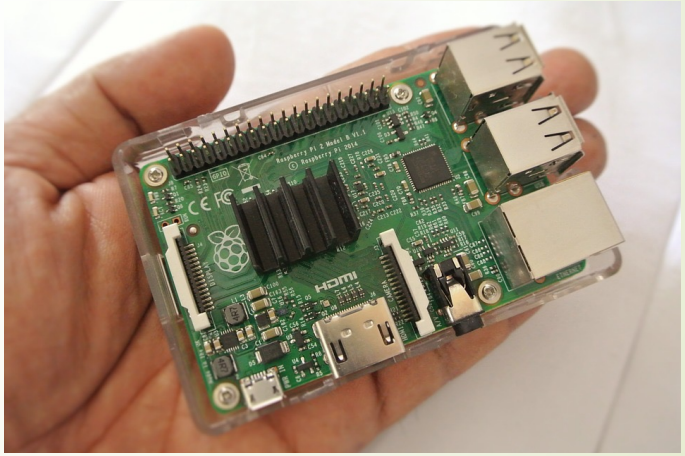
- Computational complexity ≈ x^n
- Fewer dimensions allows for significantly faster analysis



720x more calculations per second

Why is this important?

- Computational complexity ≈ x^n
- Fewer dimensions allows for significantly faster analysis



> 1,500x more RAM

Correcting for inflation..

Increase in computational power per £: 57,600x

Using about 1/4000 of the energy

Correcting for inflation..

Increase in computational power per £: 57,600x

Using about 1/4000 of the energy

Disposable super computer than can be battery/solar powered.

The case for sound

The case for sound

Limited resource, pressure to partition

The case for sound

Limited resource, pressure to partition

- Frequency of calls

The case for sound

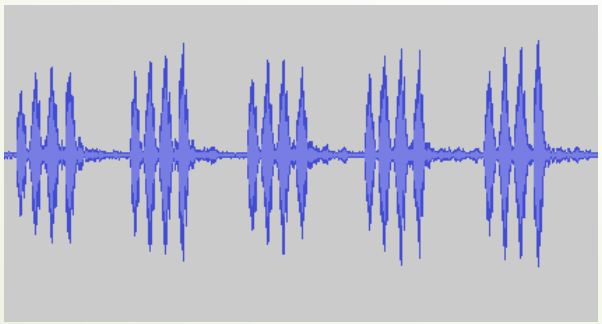
Limited resource, pressure to partition

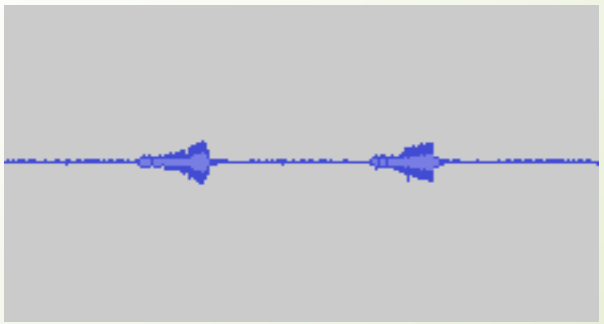
- Frequency of calls
- Time of calls (time of year; time of day)

The case for sound

Limited resource, pressure to partition

- Frequency of calls
- Time of calls (time of year; time of day)
- Structure of calls





The case for Orthoptera

- Tradition as model organisms (ALRC)

The case for Orthoptera

- Tradition as model organisms (ALRC)
- Song is stereotypical/neurologically controlled

The case for Orthoptera

- Tradition as model organisms (ALRC)
- Song is stereotypical/neurologically controlled INATE not LEARNED

The case for Orthoptera

- Tradition as model organisms (ALRC)
- Song is stereotypical/neurologically controlled INATE not LEARNED
- Major cause of variation is temperature

The case for Orthoptera

- Tradition as model organisms (ALRC)
- Song is stereotypical/neurologically controlled INATE not LEARNED
- Major cause of variation is temperature PREDICTABLE

The case for Orthoptera

- Tradition as model organisms (ALRC)
- Song is stereotypical/neurologically controlled INATE not LEARNED
- Major cause of variation is temperature PREDICTABLE
- Calling song attracts mates

What should we automate?

Start with what people can't do

- Can work in human-audible and ultrasound simultaneously

What should we automate?

Start with what people can't do

- Can work in human-audible and ultrasound simultaneously
- Can do 24/7 for a week without sleeping

Major concerns

- Unaccountable

Major concerns

- Unaccountable
  How did the computer reach that conclusion?

- Error rate
  How do we know the computer is right? (False positive / false negative)

Major concerns

- Unaccountable
  How did the computer reach that conclusion?
  If we choose how it makes a decision, we can make it explain how I got there

- Error rate
  How do we know the computer is right? (False positive / false negative)

  We can test this and get reliable numbers

What do people do?

- Unaccountable
  How did the computer reach that conclusion?
  If we choose how it makes a decision, we can make it explain how I got there
  Cite the published key / collection used to make ID
- Error rate
  How do we know the computer is right? (False positive / false negative)

  We can test this and get reliable numbers
  Relies on every study depositing voucher specimens

There are an awful lot of publications that don't link their results to voucher specimens, keys, sound recordings, or anything else

Automated systems can do this on their own

Major concerns

- Unaccountable
  How did the computer reach that conclusion?
  If we choose how it makes a decision, we can make it explain how I got there
  Cite the published key / collection used to make ID
  Cite the algorithm and it's version used to make the ID
- Error rate
  How do we know the computer is right? (False positive / false negative)

  We can test this and get reliable numbers
  Relies on every study depositing voucher specimens
  Automatically submit sample sound files to a repository

Current state

Current state

- Accuracy of species ID for a UK site is 70-100%

Current state

- Accuracy of species ID for a UK site is 70-100%
  - Several estimates of expert ID are inside this range

Current state

- Accuracy of species ID for a UK site is 70-100%
  - Several estimates of expert ID are inside this range
- Limited to presence or absence

Current state

- Accuracy of species ID for a UK site is 70-100%
  - Several estimates of expert ID are inside this range
- Limited to presence or absence
  - Hard to estimate abundance

Current state

- Accuracy of species ID for a UK site is 70-100%
  - Several estimates of expert ID are inside this range
- Limited to presence or absence
  - Hard to estimate abundance
- Systems are naive

Current state

- Accuracy of species ID for a UK site is 70-100%
  - Several estimates of expert ID are inside this range
- Limited to presence or absence
  - Hard to estimate abundance
- Systems are naïve
  - Not aware of the distribution/behavior/etc of the species they monitor

Current state

- Accuracy of species ID for a UK site is 70-100%
  - Several estimates of expert ID are inside this range
- Limited to presence or absence
  - Hard to estimate abundance
- Systems are naïve
  - Not aware of the distribution/behavior/etc of the species they monitor
  - Only aware of a handful of species

Where do we go from here?

Where do we go from here?

Make systems less naive

Where do we go from here?

Make systems less naïve

- Account for species distribution and acoustic behaviour

Where do we go from here?

Make systems less naïve

- Account for species distribution and acoustic behaviour

Figure out how to get a measure of abundance

Where do we go from here?

Make systems less naïve

- Account for species distribution and acoustic behaviour

Figure out how to get a measure of abundance

- Some work happening

Where do we go from here?

Make systems less naïve

- Account for species distribution and acoustic behaviour

Figure out how to get a measure of abundance

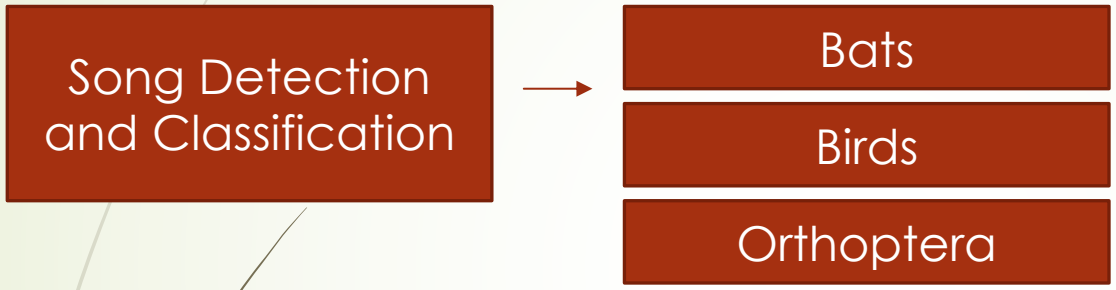
- Some work happening

How do we deal with songs of unknown species?

Automated Acoustic Observatories

Song Detection and Classification

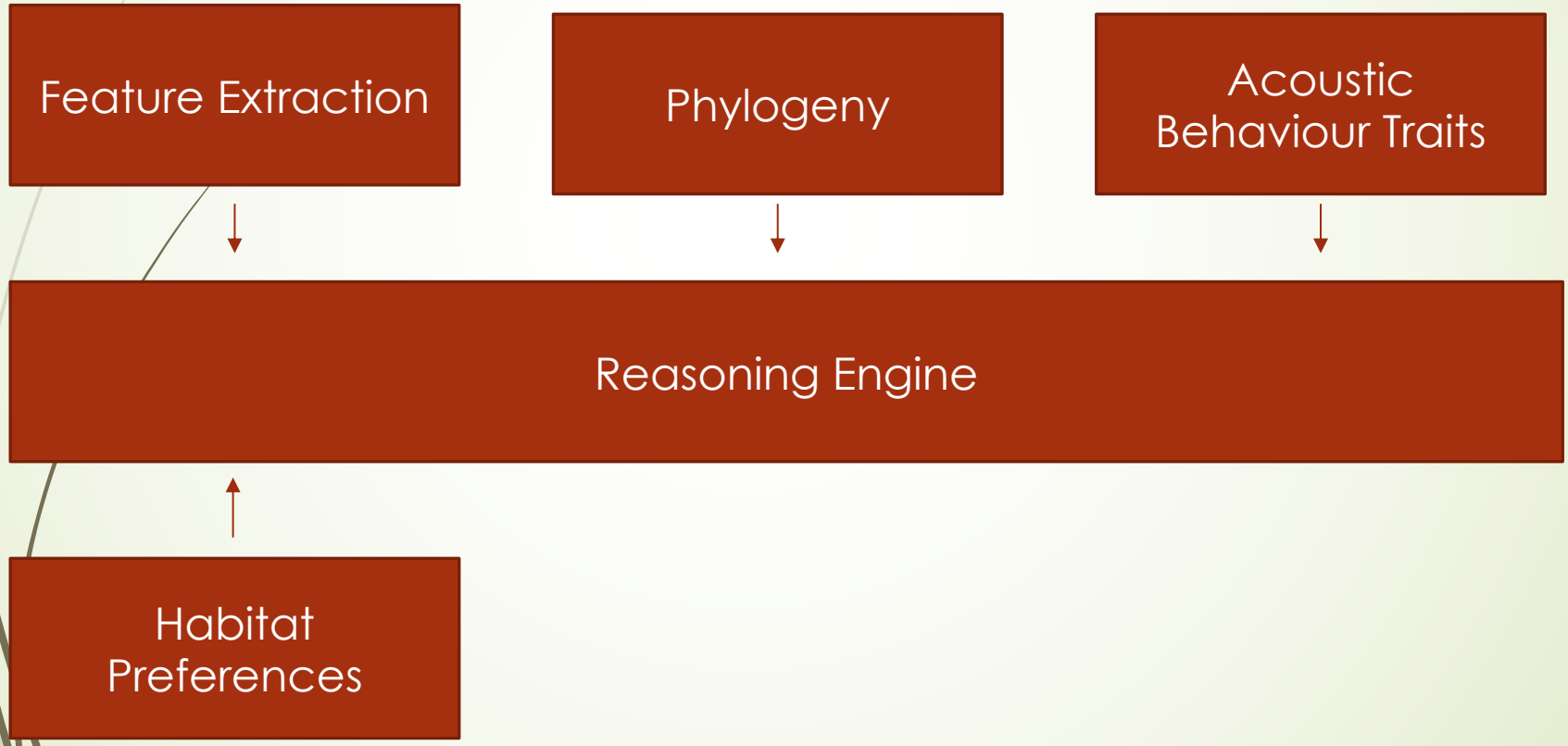
Automated Acoustic Observatories



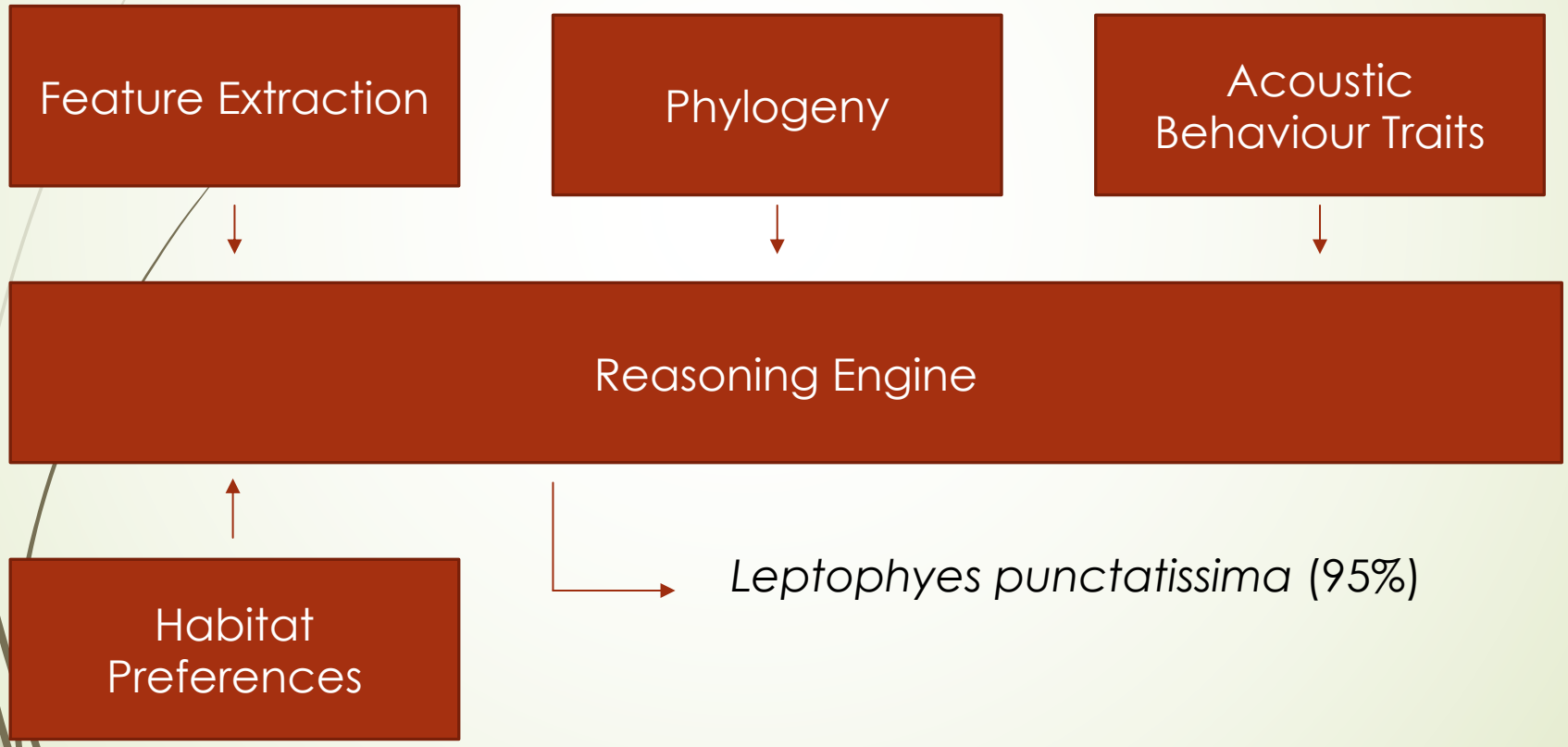
Automated Acoustic Observatories

Automated Acoustic Observatories

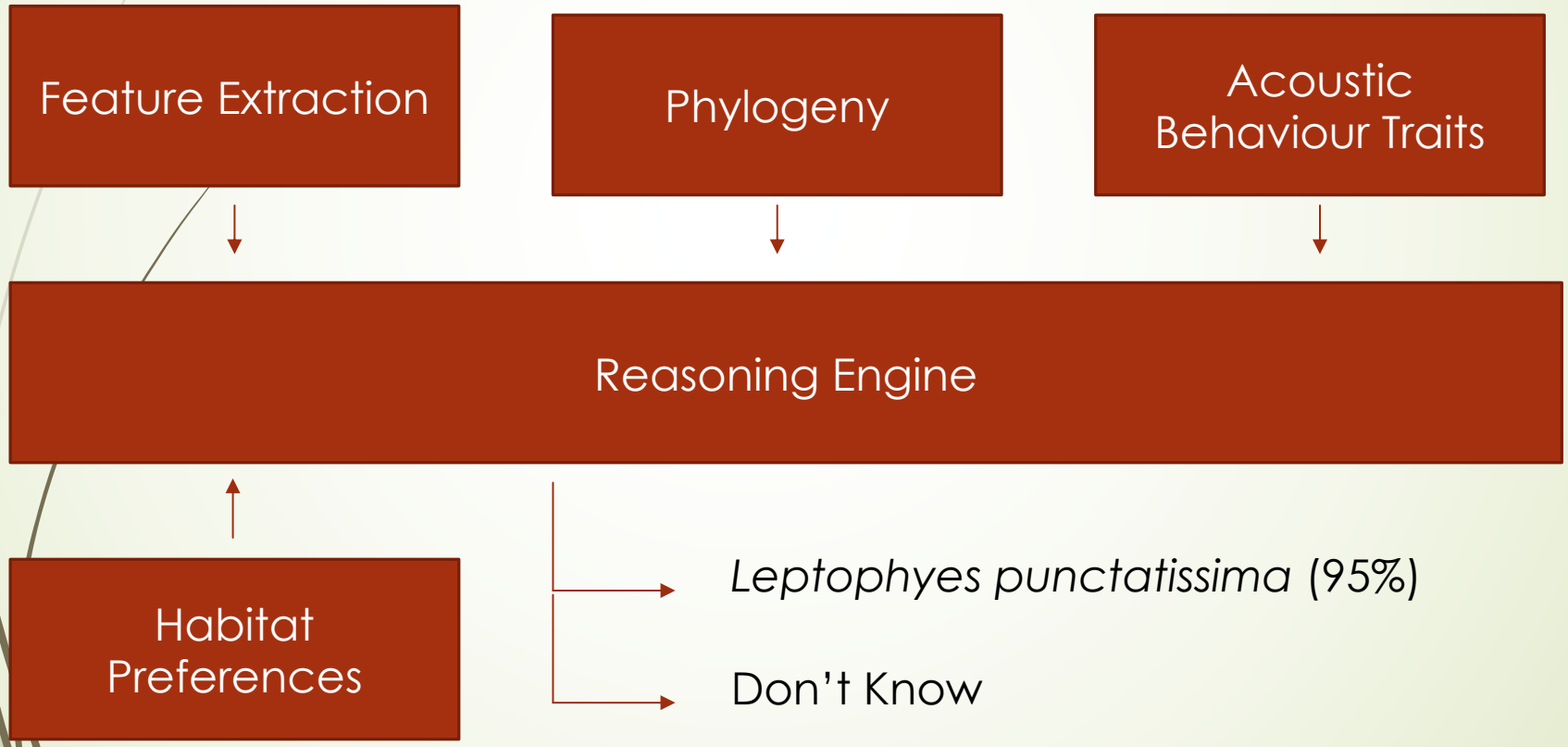
Automated Acoustic Observatories



Automated Acoustic Observatories



Automated Acoustic Observatories



Automated Acoustic Observatories