Understanding Whale Communication : First Steps

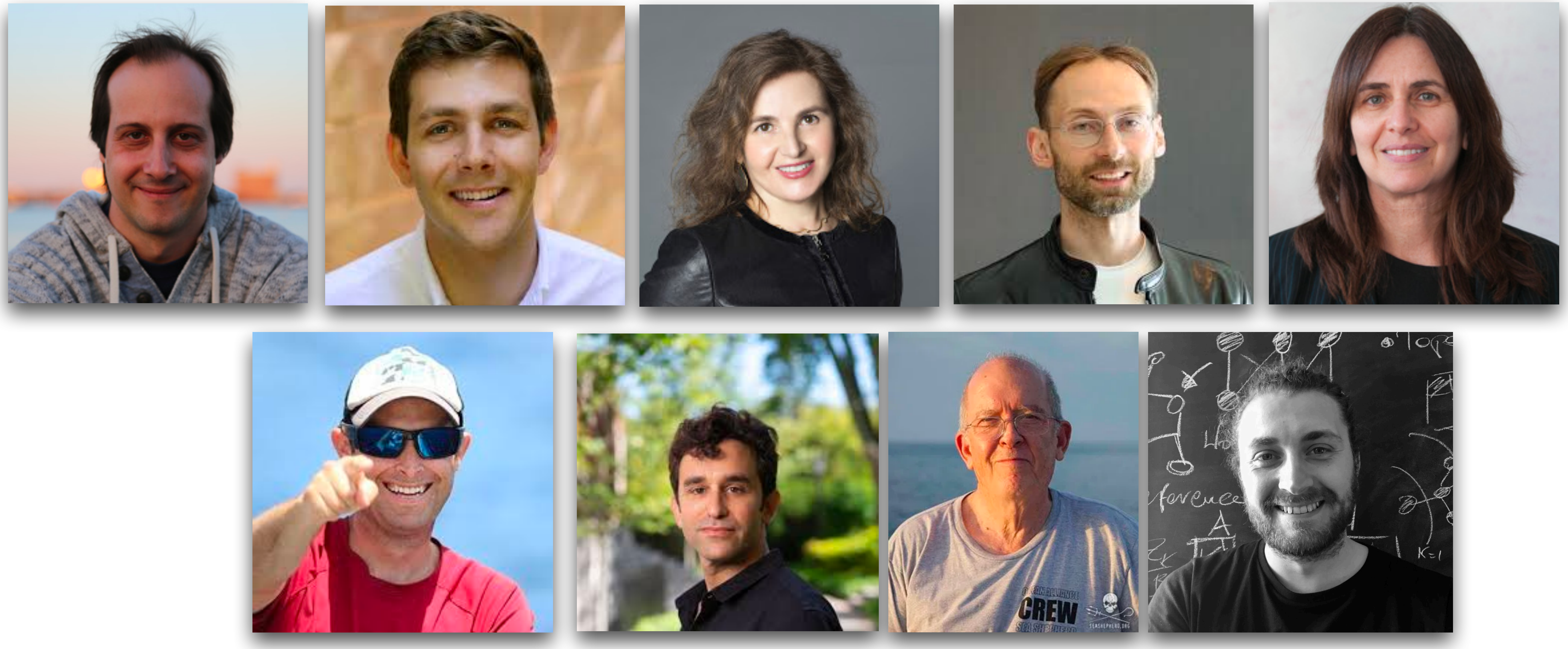
Pratyusha Sharma

CSAIL, Massachusetts Institute of Technology

Simons Institute, Berkeley
3rd August 2020

PROJECT CETI

Team



Flow

- Motivation
- What have we done so far?

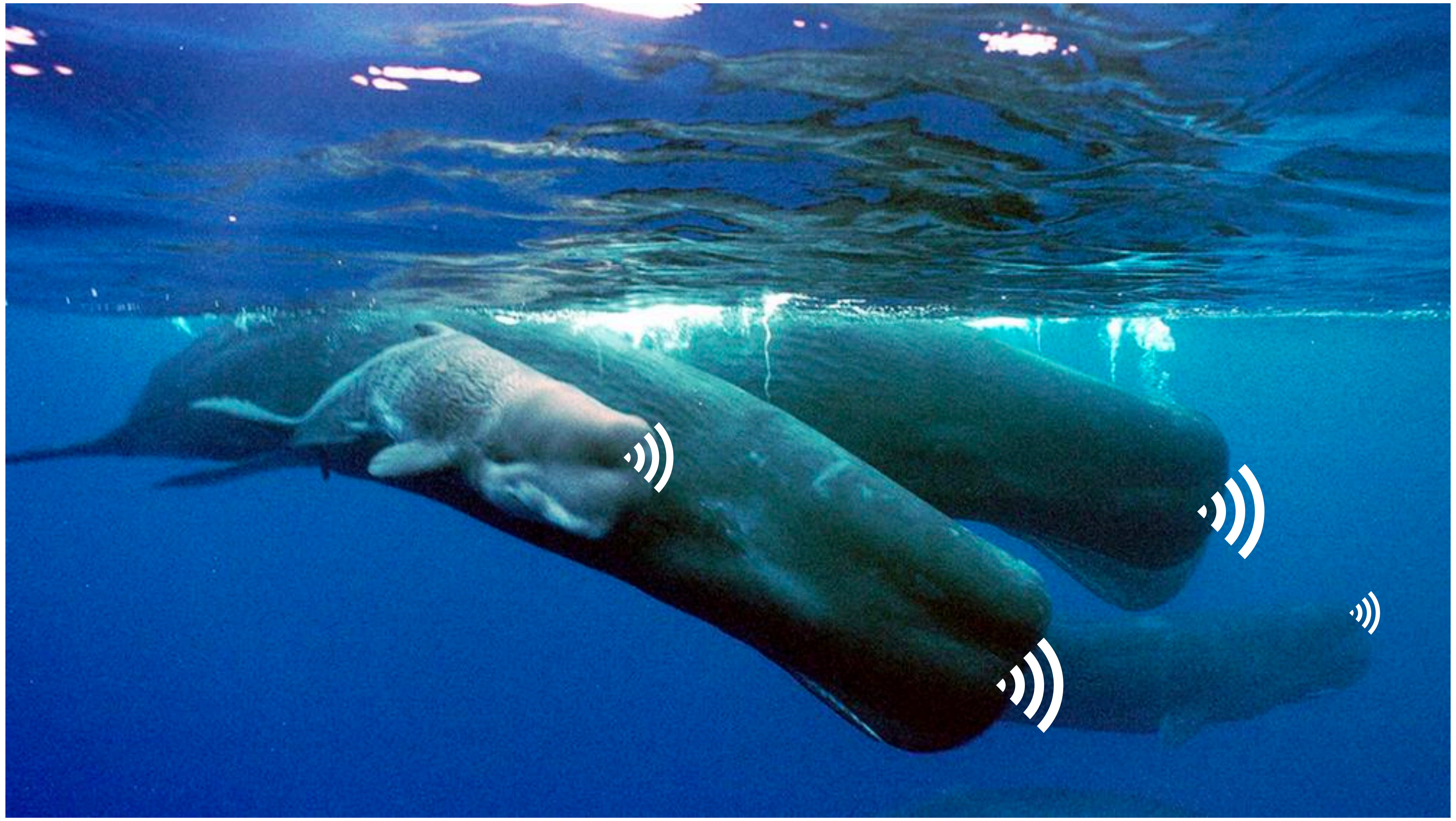
Why Whales?

- Largest brains in the world
- Sophisticated communication across large distances and cultures
- Underwater -> Sound is the major mode of sensing

Picture by: Amanda Cauden

Why study Communication?

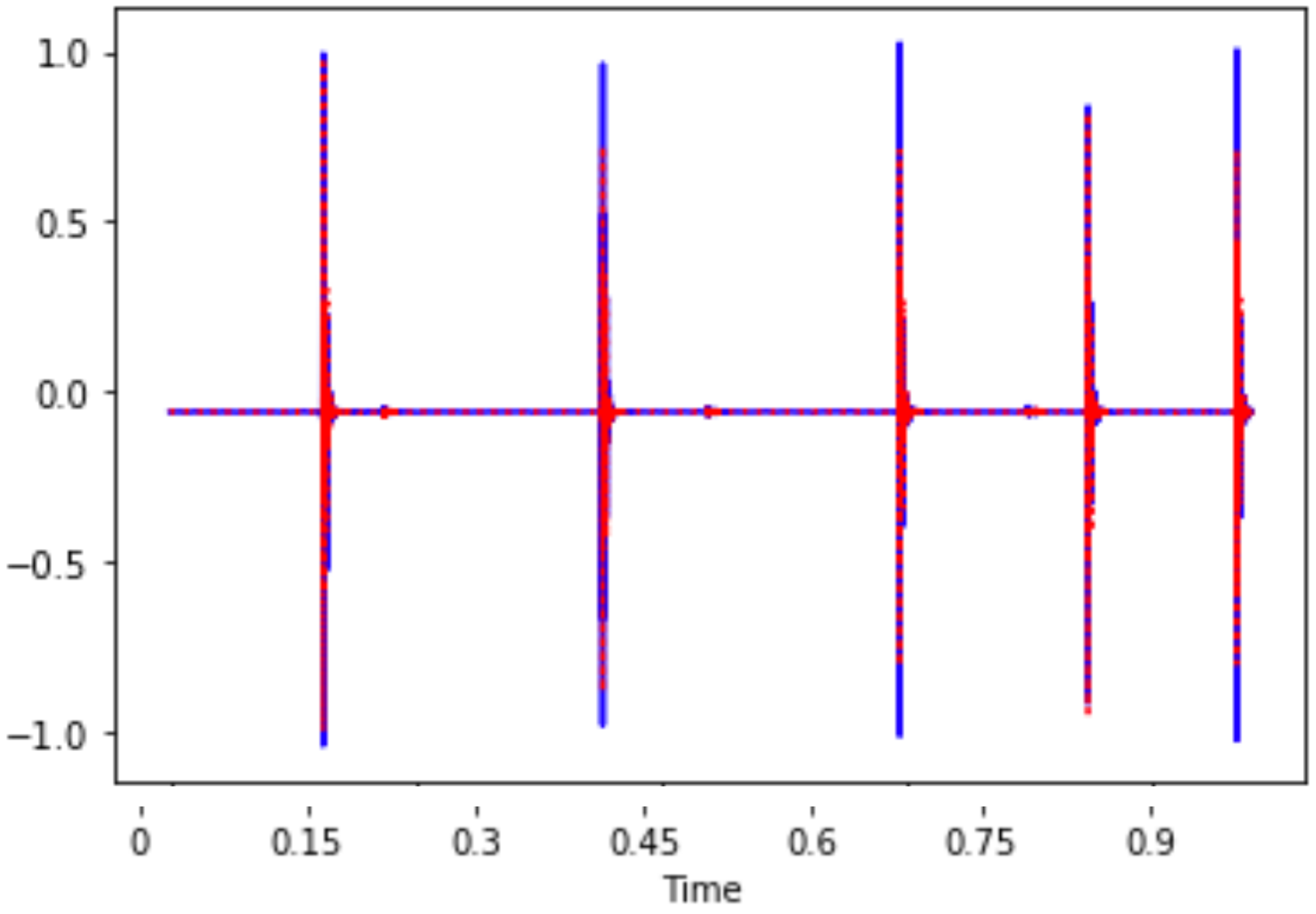
- A **major sign of intelligence**
- Presence of Language:
  - Discreteness
  - Grammar
  - Long range dependencies
  - Productivity
  - Displacement
- **No other animal except humans haven been proven to have a language so far**



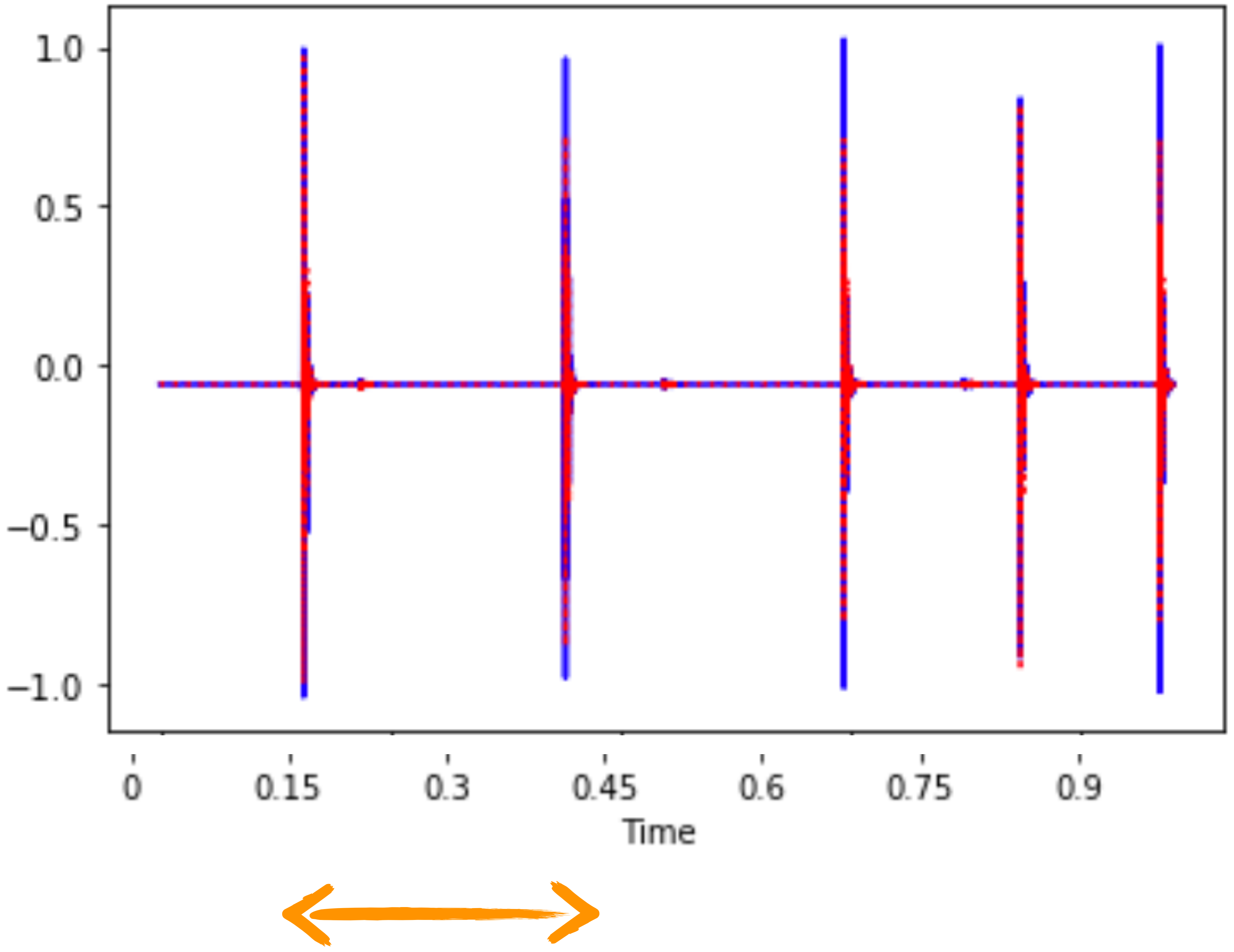
What do these sounds look like?

Short series of 3 to 20 or more clicks are produced by sperm whales, in stereotyped repetitive sequences or codas

🔊

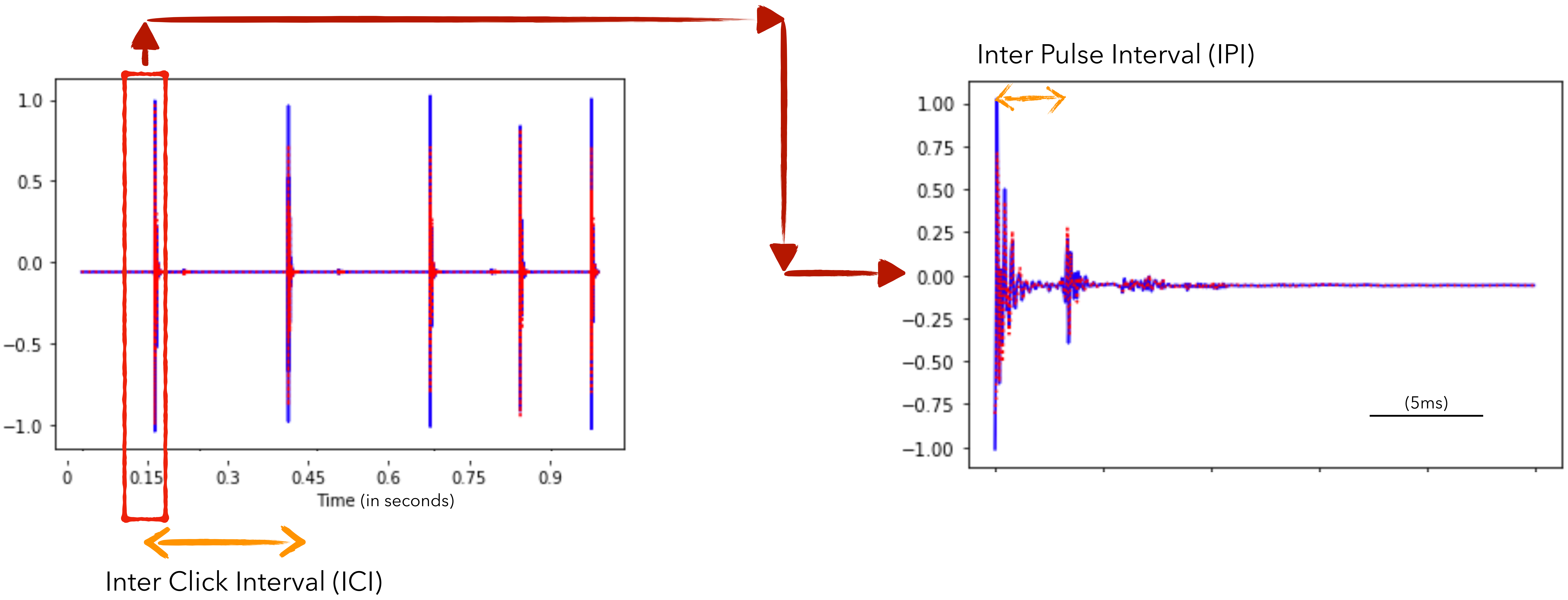


ICI and IPI



Inter Click Interval (ICI)

ICI and IPI



Inter Pulse Interval (IPI)

Inter Click Interval (ICI)

Data

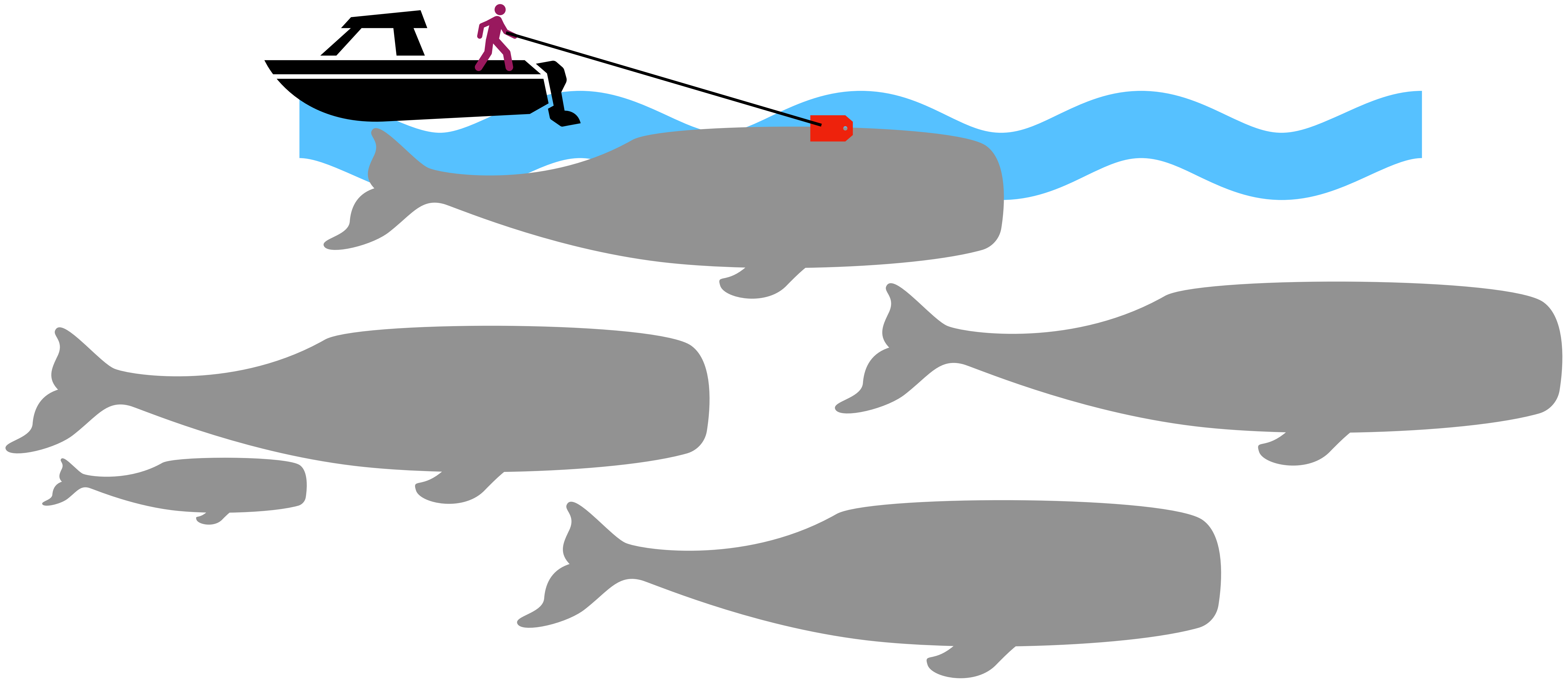
<u>Data</u>

3950 Codas ~ 22,386 clicks

- Stereo audio recordings

- Rich Annotations

- Gyroscope, Magnetometer, Accelerometer data



Data

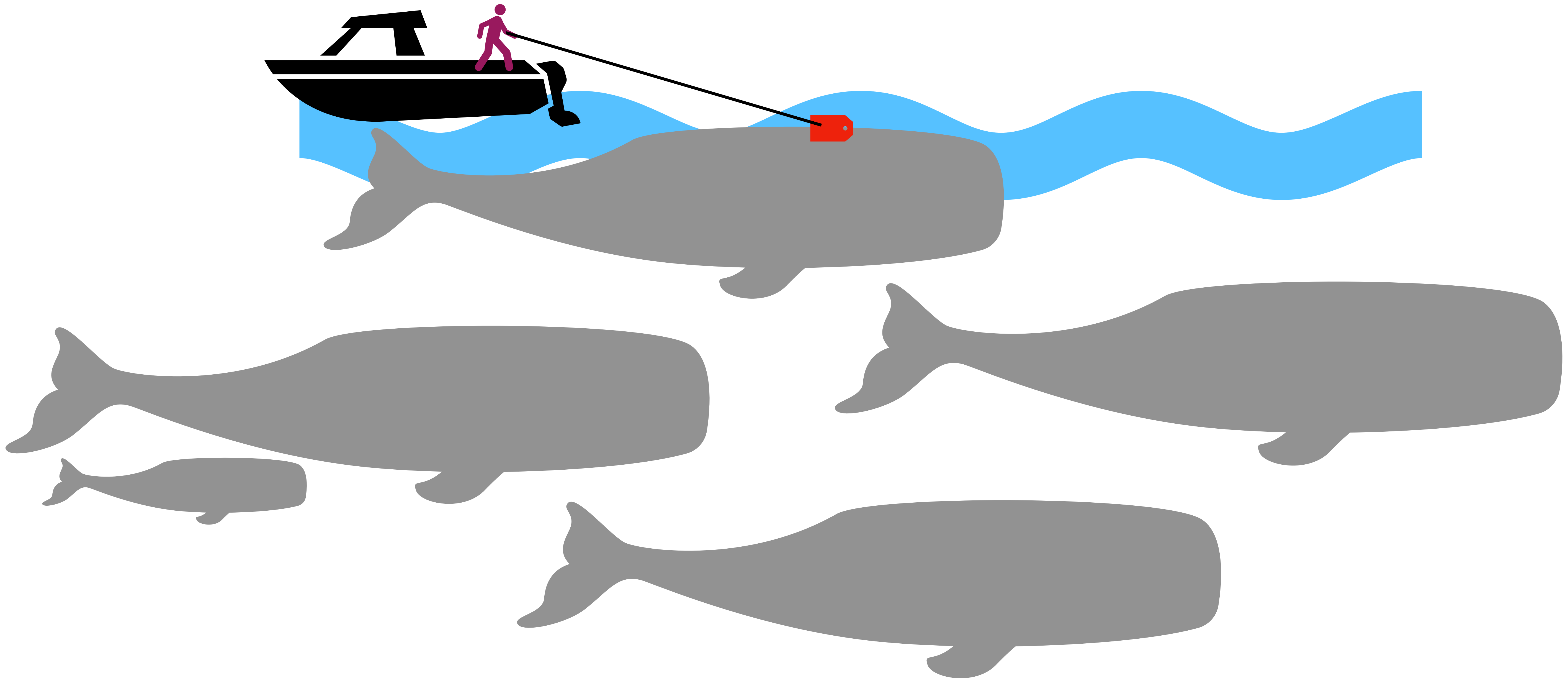
<u>Data</u>

3950 Codas ~ 22,386 clicks

- Stereo audio recordings

- Rich Annotations

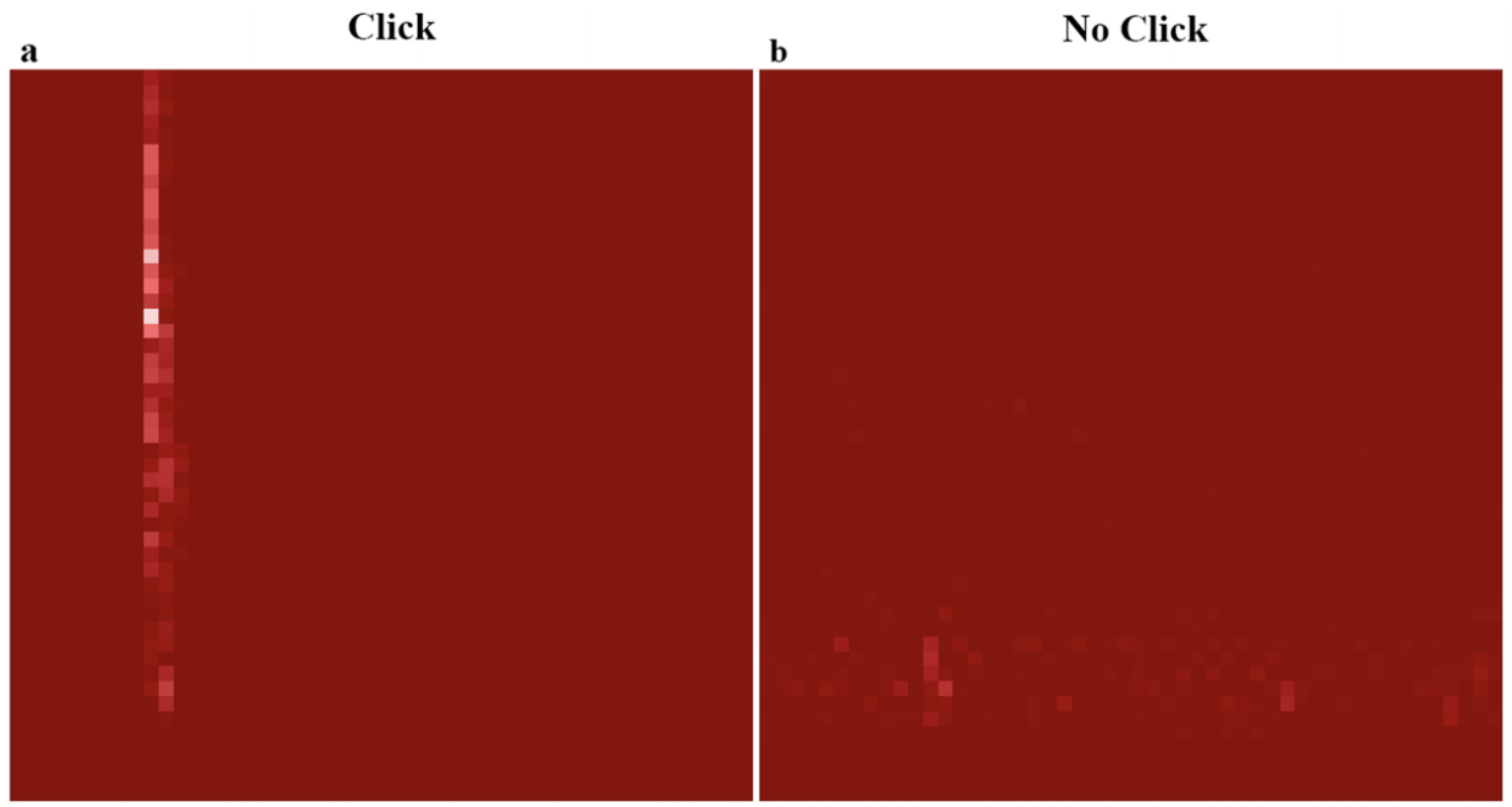
- Gyroscope, Magnetometer, Accelerometer data



Additional Audio data with no annotations

Previous Work

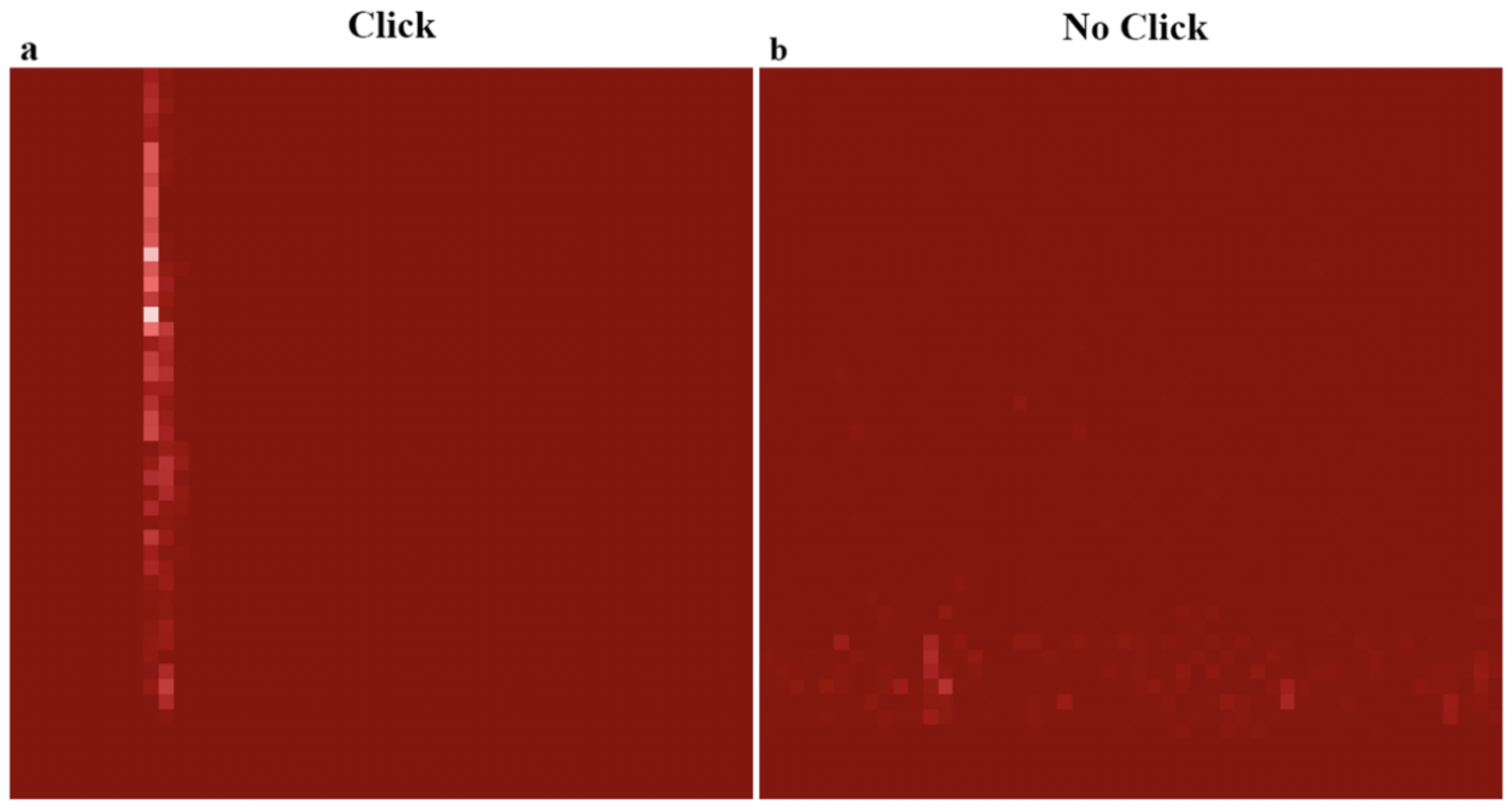
Click vs no-click



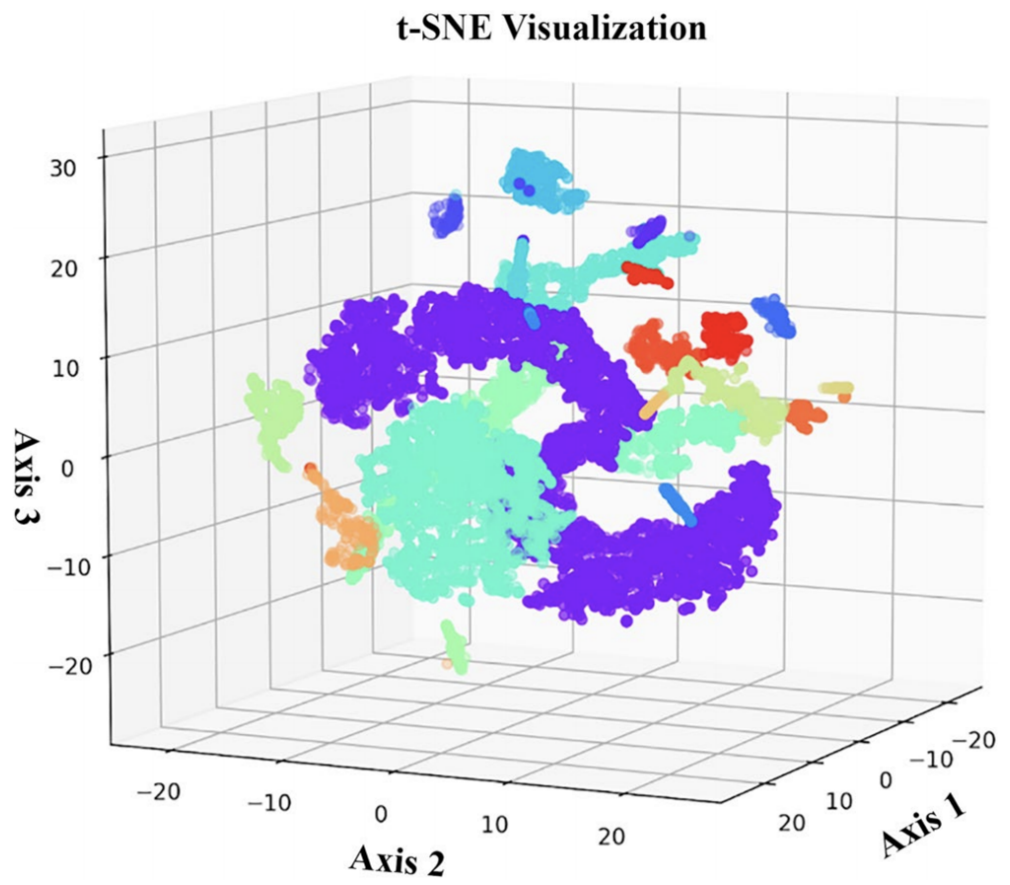
Paper: Deep Machine Learning Techniques for the Detection and Classification of Sperm Whale Bioacoustics - Peter C. Bermant, Michael M. Bronstein, Robert J. Wood, Shane Gero, David F. Gruber

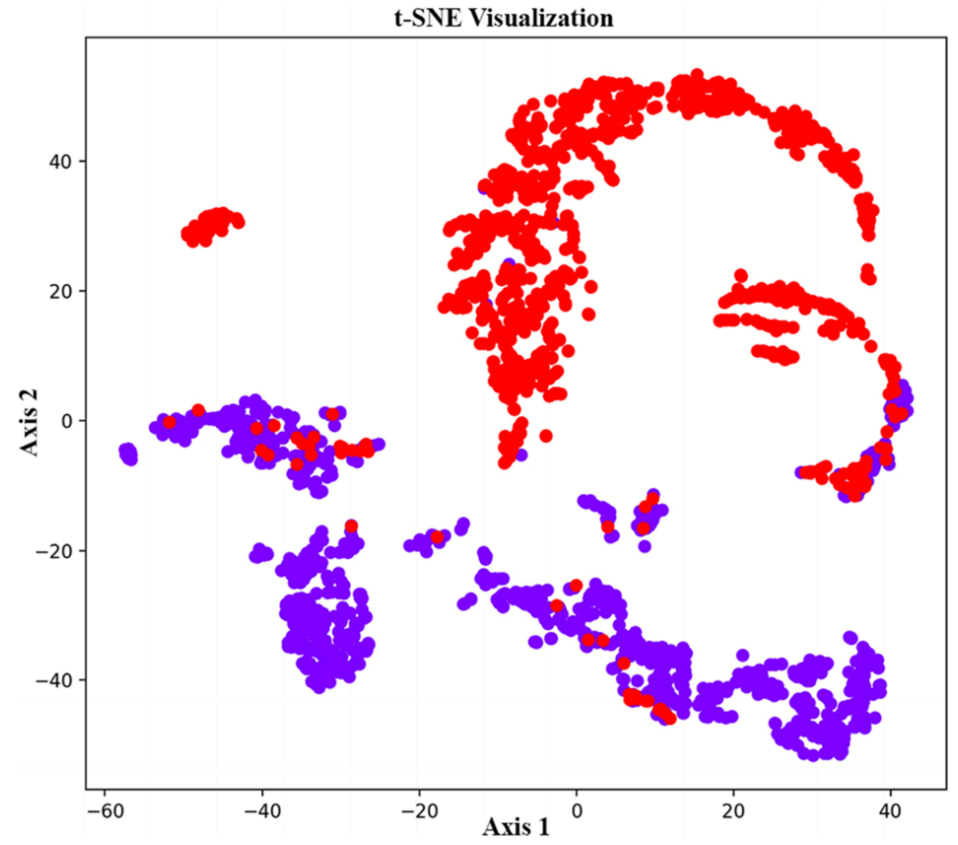
Previous Work

Click vs no-click



Clustering codas across clans and individuals

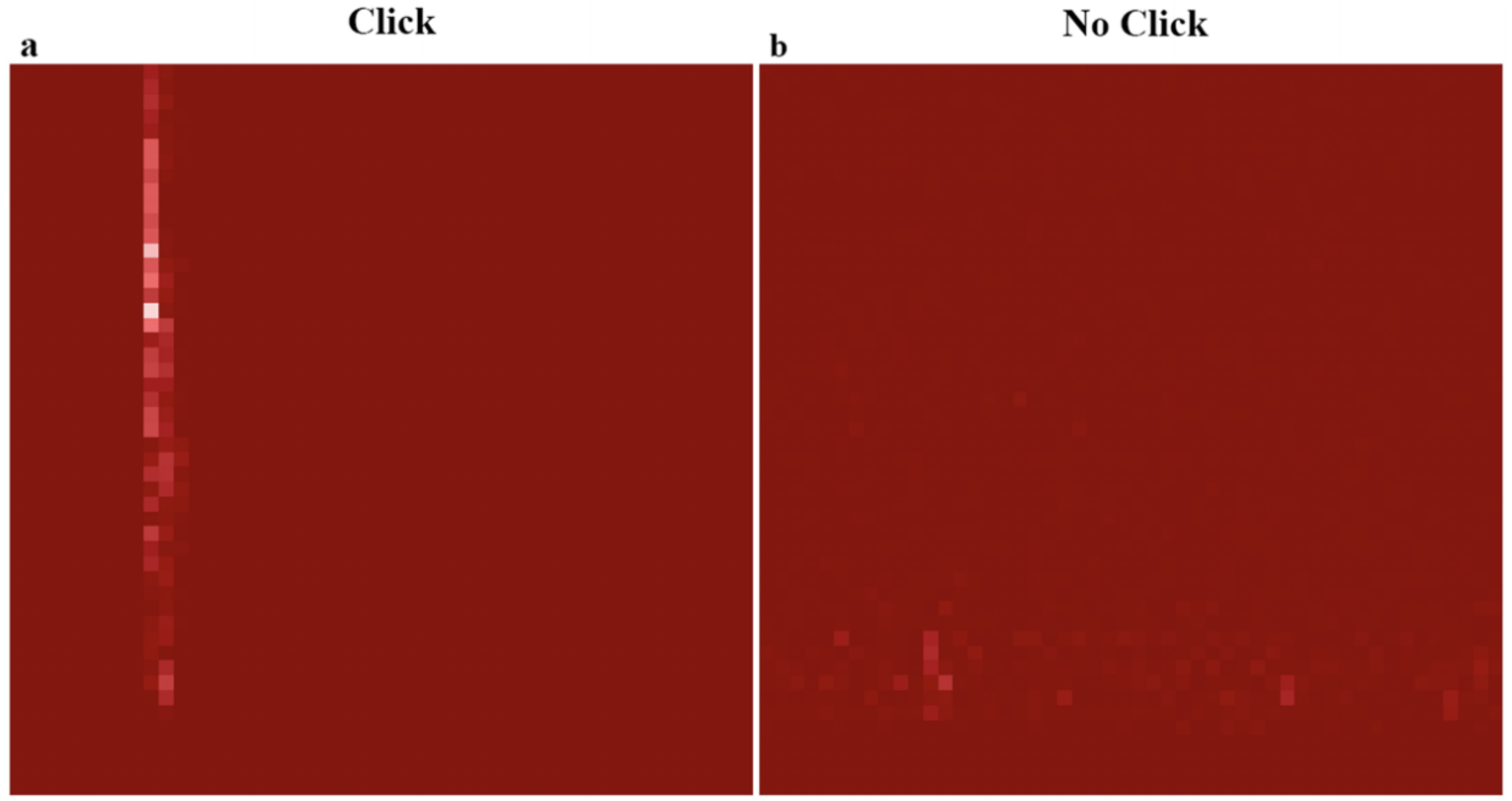




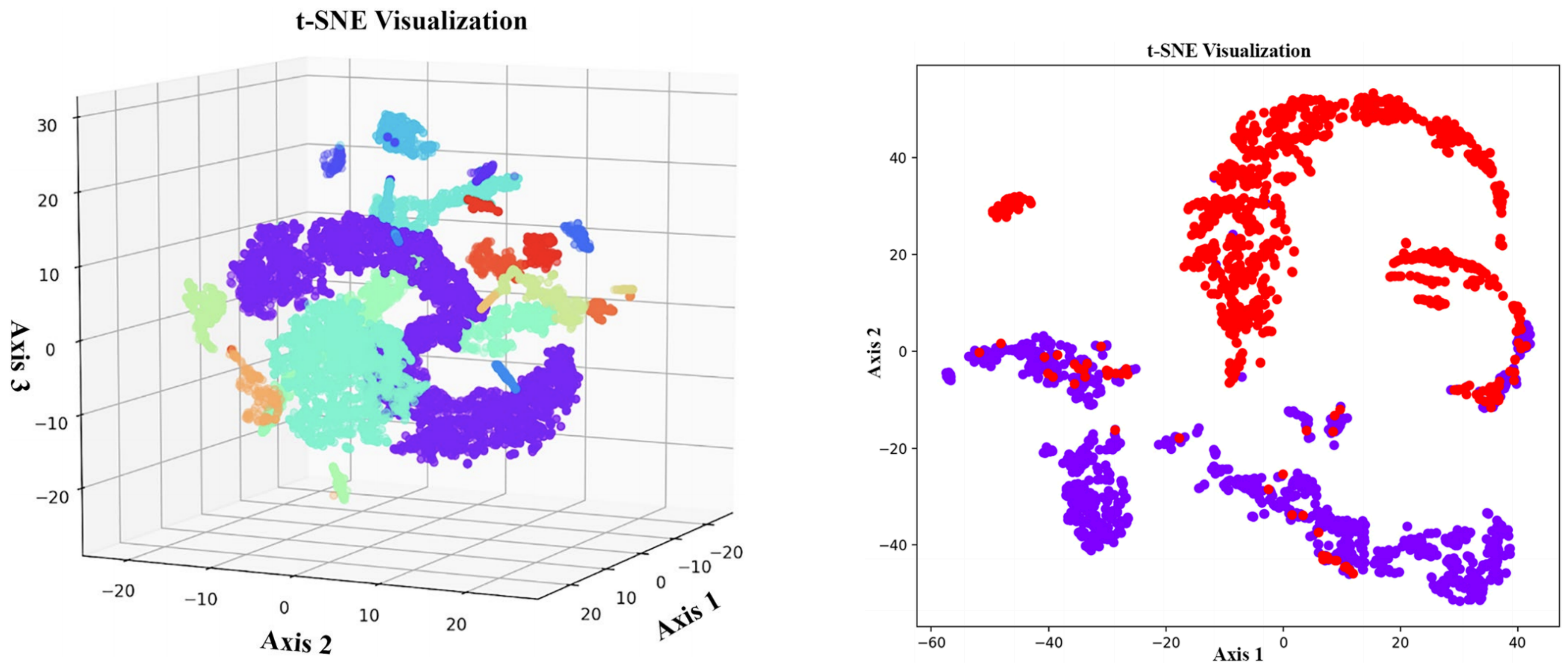
Paper: Deep Machine Learning Techniques for the Detection and Classification of Sperm Whale Bioacoustics - Peter C. Bermant, Michael M. Bronstein, Robert J. Wood, Shane Gero, David F. Gruber

Previous Work

Click vs no-click



Clustering codas across clans and individuals



Identifying coda type, vocal clan, and individual whale identity

- Coda type classification
- Vocal clan classification - 2 clans
- Individual whale identification - Across 2 whales

Paper: Deep Machine Learning Techniques for the Detection and Classification of Sperm Whale Bioacoustics - Peter C. Bermant, Michael M. Bronstein, Robert J. Wood, Shane Gero, David F. Gruber

Advantages and what is missing

Advantages

Assume clicks are binary signal - Clean first step

Can generate labels for the rest of the data almost as nicely as the human annotator

Advantages and what is missing

Advantages

Assume clicks are binary signal - Clean first step

Can generate labels for the rest of the data almost as nicely as the human annotator

What could be missing

Assume clicks are binary signal - Don’t use any other features in clicks (power, spectral features)

Make simplifying assumption about the coda types and variation

Generalization beyond heuristics

Advantages and what is missing

Advantages

Assume clicks are binary signal - Clean first step

Can generate labels for the rest of the data almost as nicely as the human annotator

What could be missing

Assume clicks are binary signal - Don’t use any other features in clicks (power, spectral features)

Make simplifying assumption about the coda types and variation

Generalization beyond heuristics

What further do we want to know?

What further do we want to know?

Question 1?

What are units of communication?

———————————————————————————————

Question 2?

Find the rules used to produce different combinations of these units?

———————————————————————————————

Question 3:

Do SWs have a communication with long range dependencies over the historic context of the sounds produced?

———————————————————————————————

Question 4:

Can we learn the meanings of their vocalizations?

What have we seen?

1. Data collection and annotation is expensive

2. How can we generalize beyond heuristics?

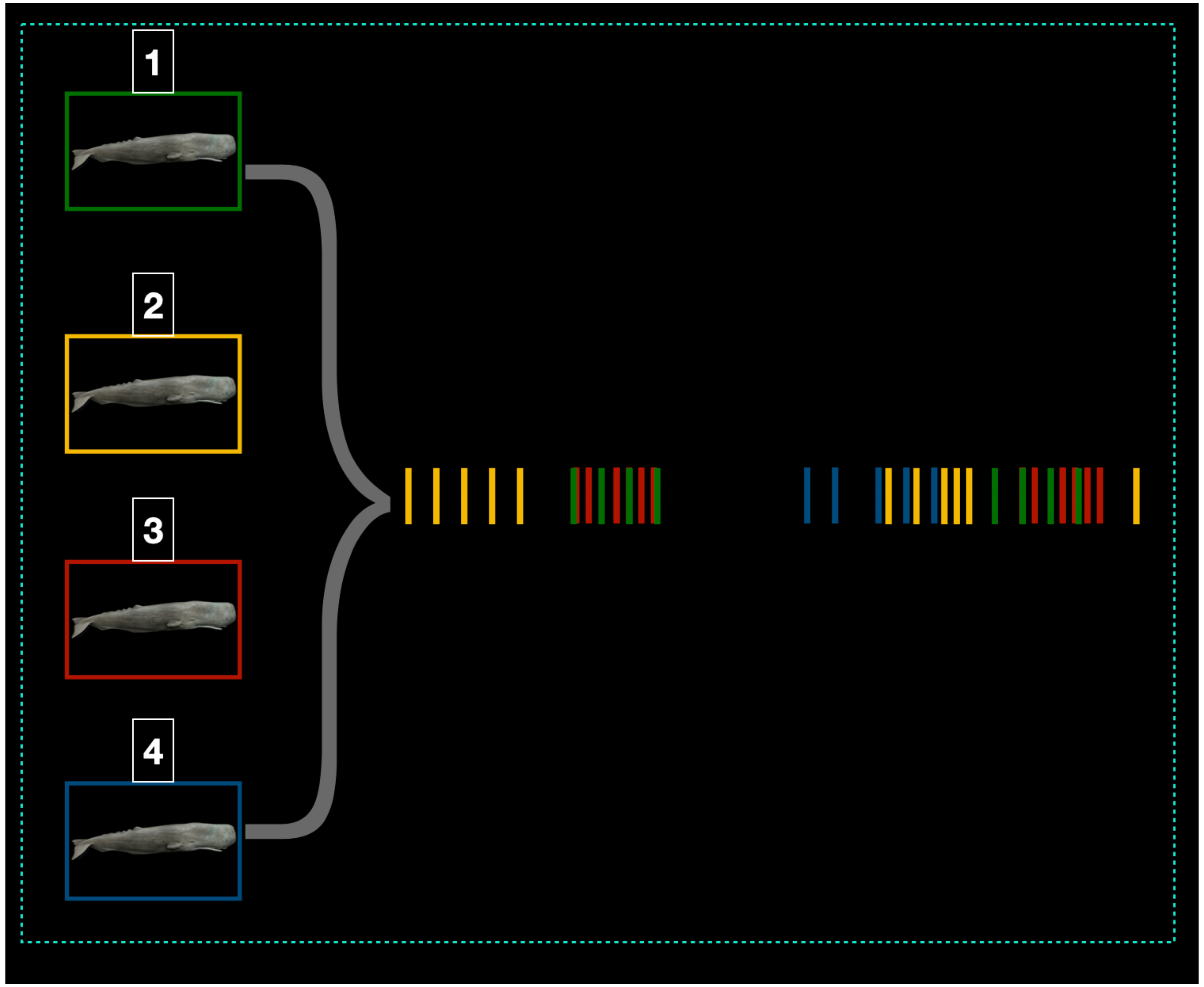
What do we want to do?

1. Automatic Annotation - Extract the portion of the audio files with the vocalizations and separate sources

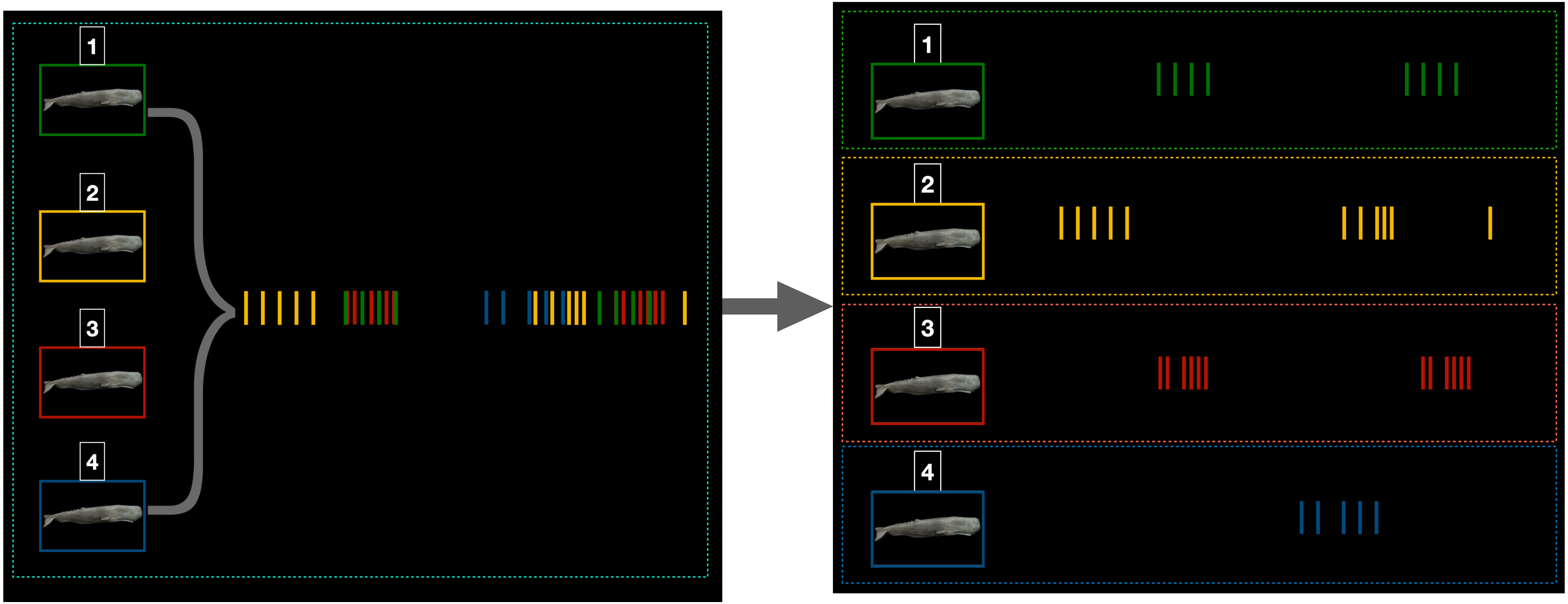
2. Identifying the underlying “Symbols” and “Rules” of the vocalizations that can help us communicate back with Sperm whales

1. Automatic Annotation

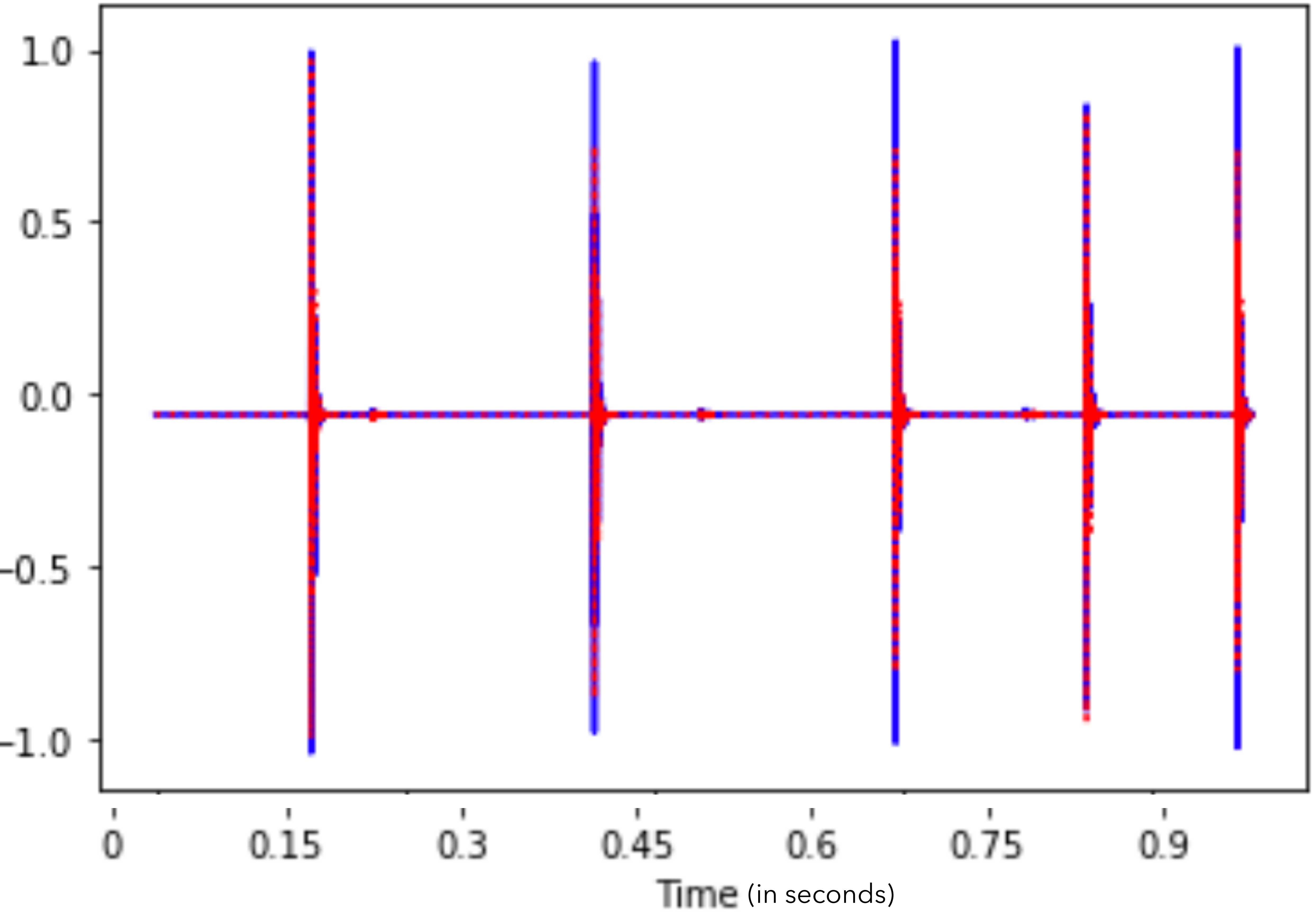
Automatic Annotation

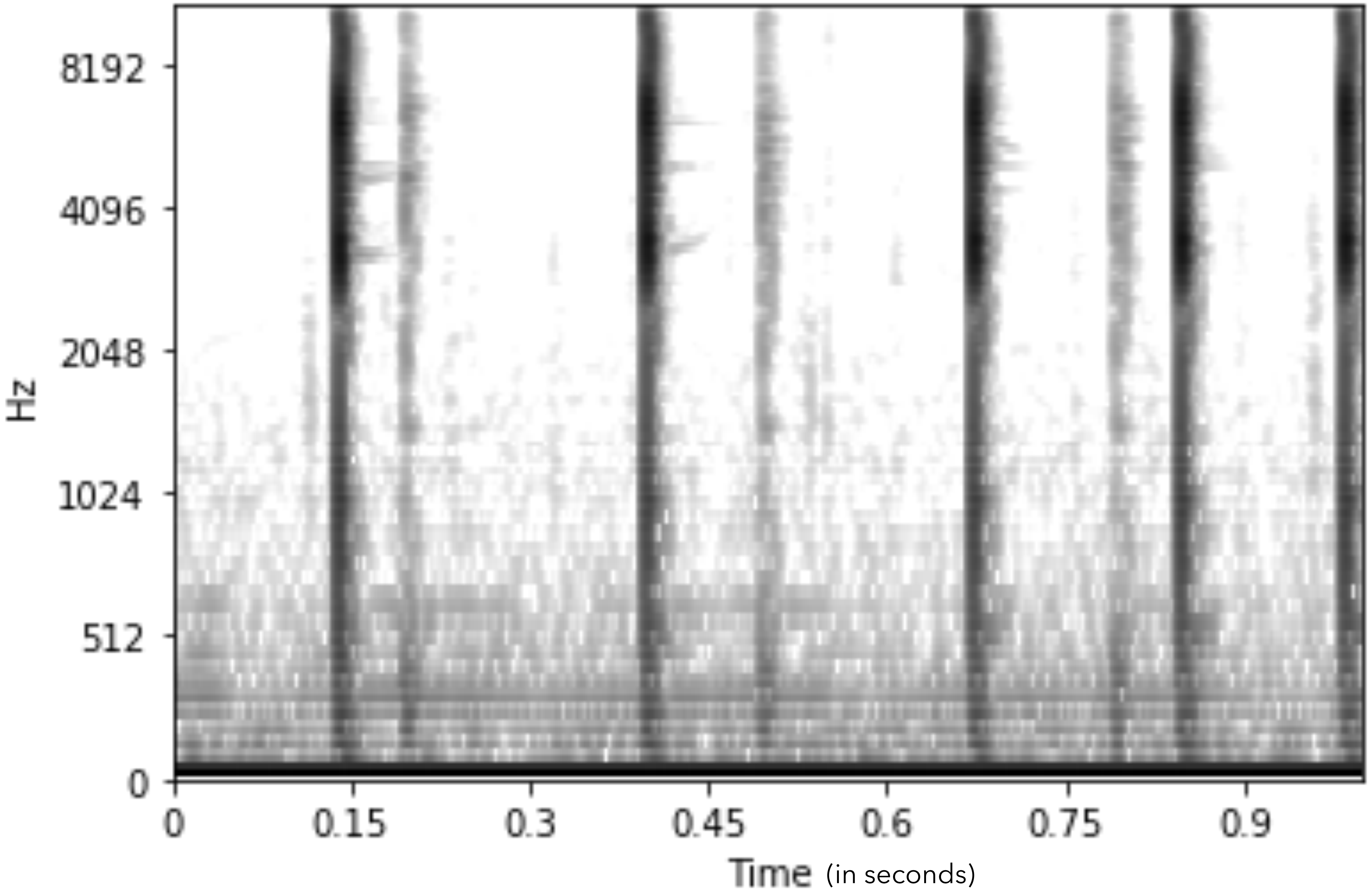


Automatic Annotation

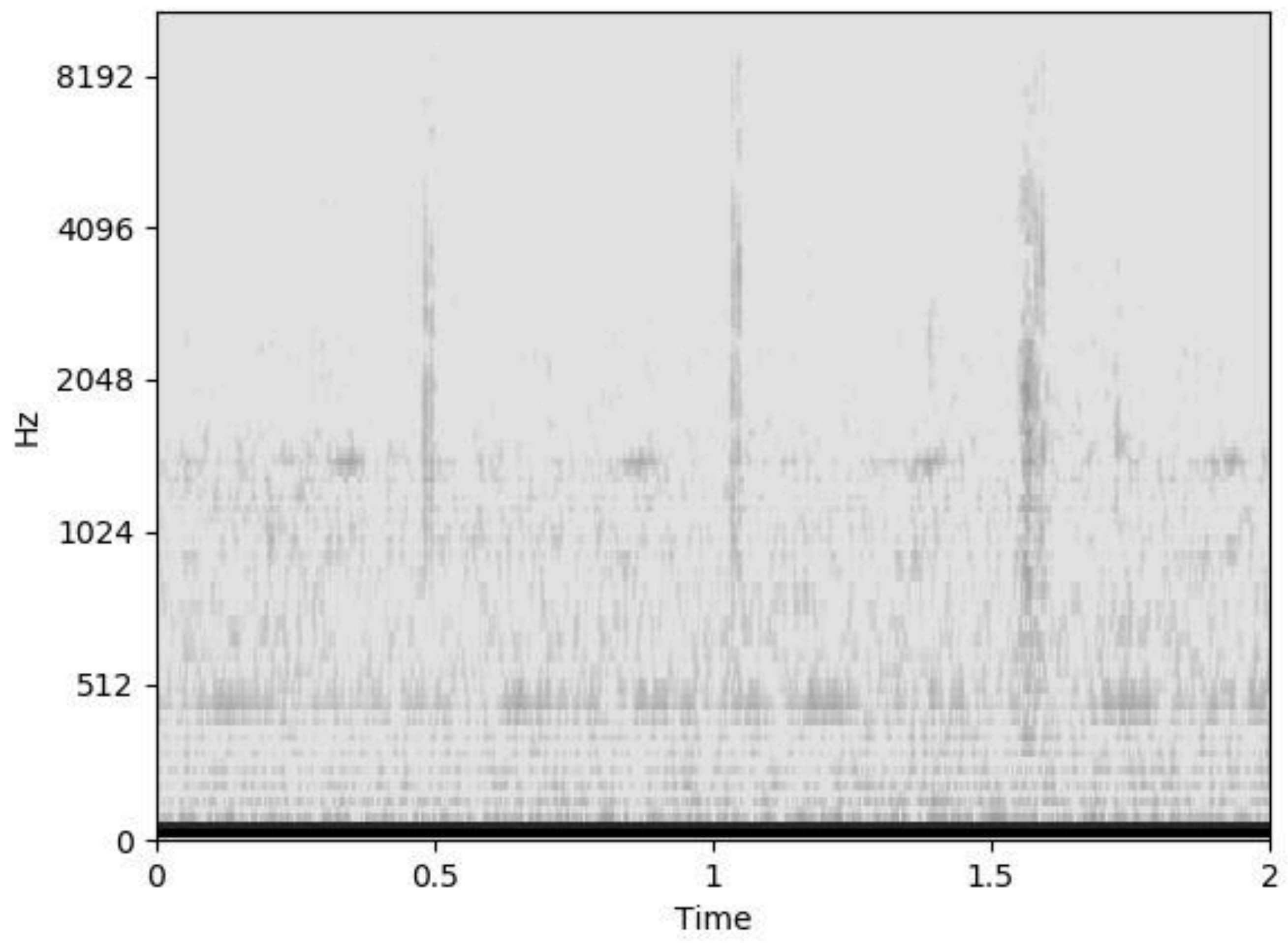


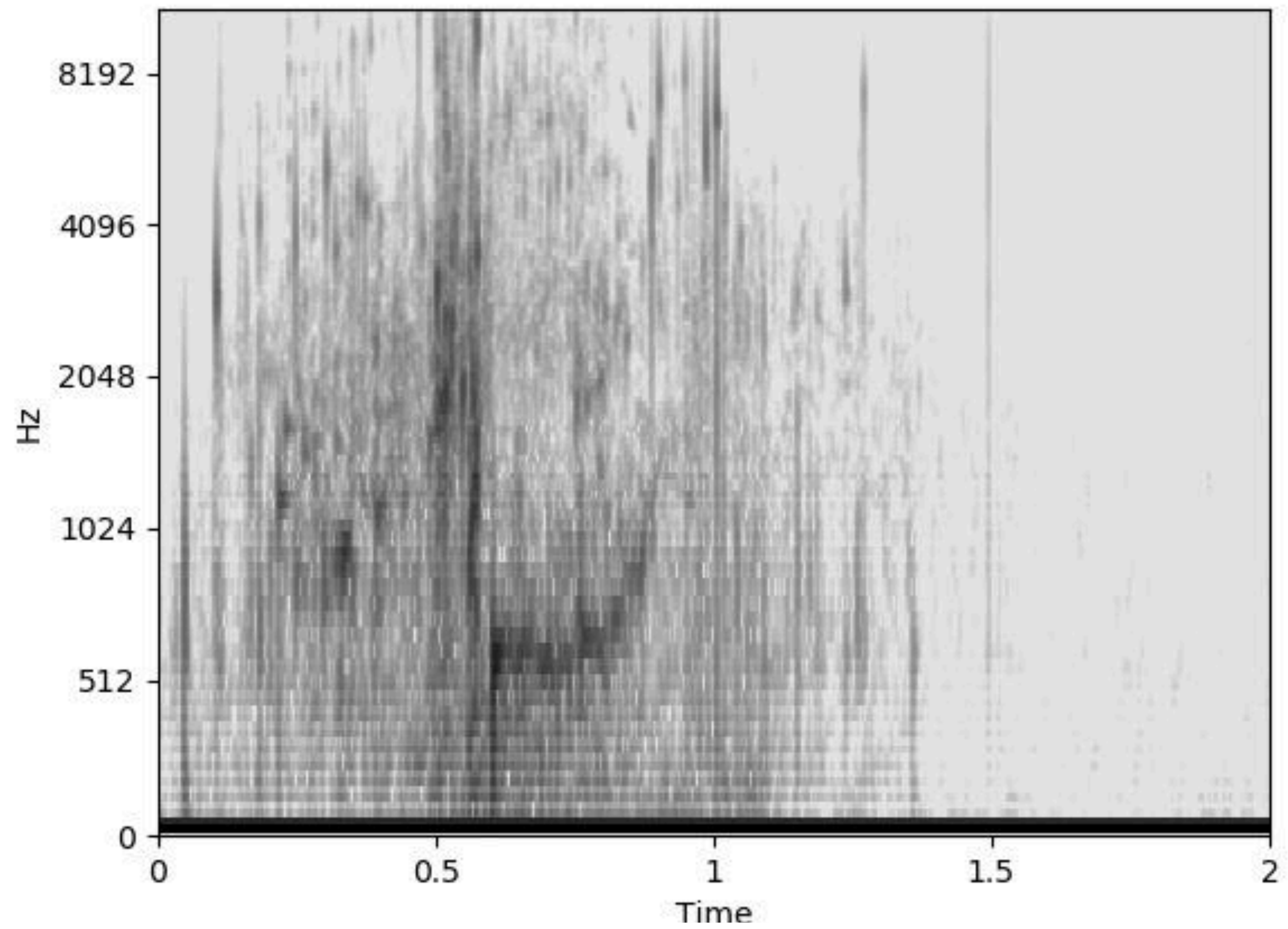
Click Detection

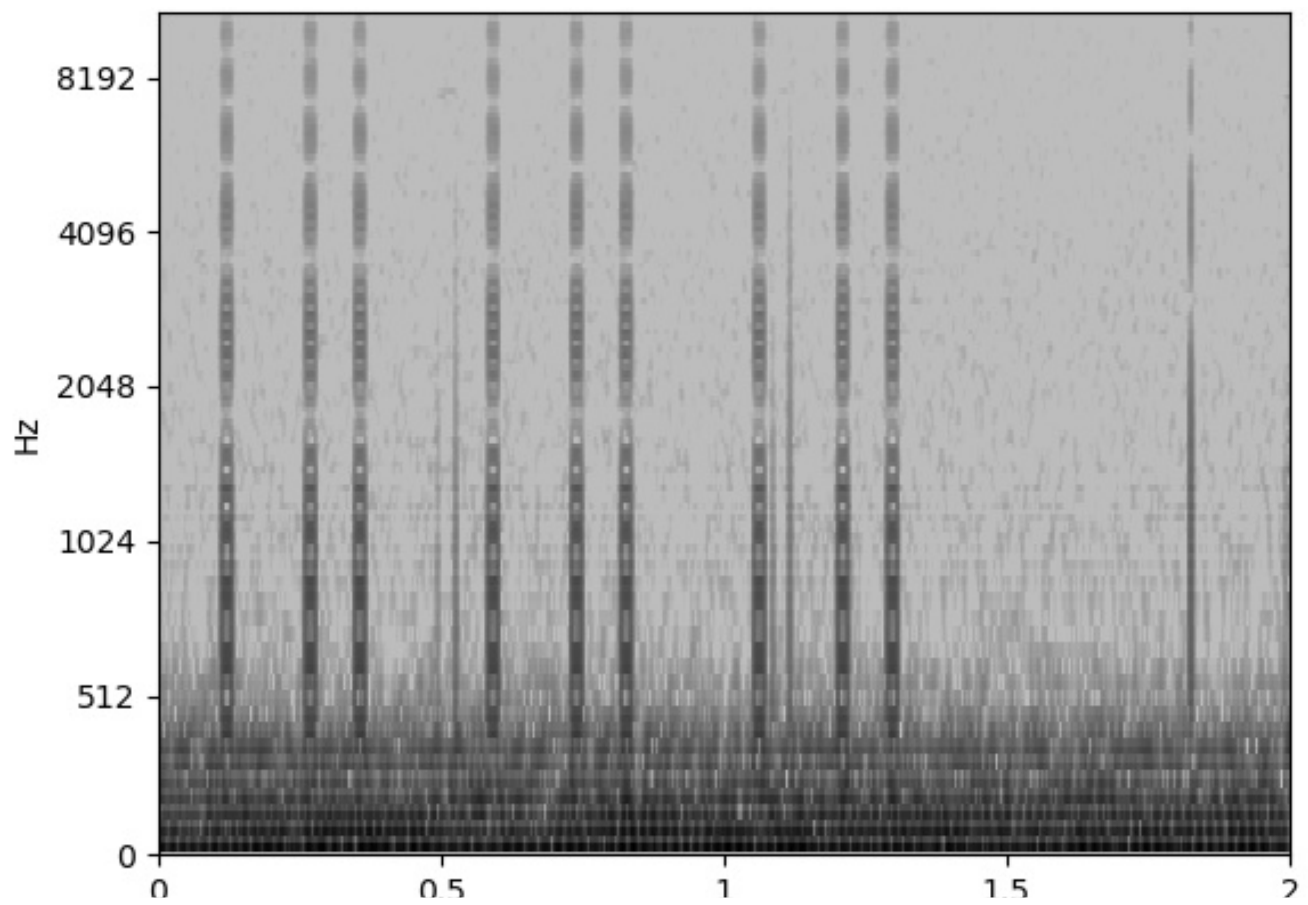


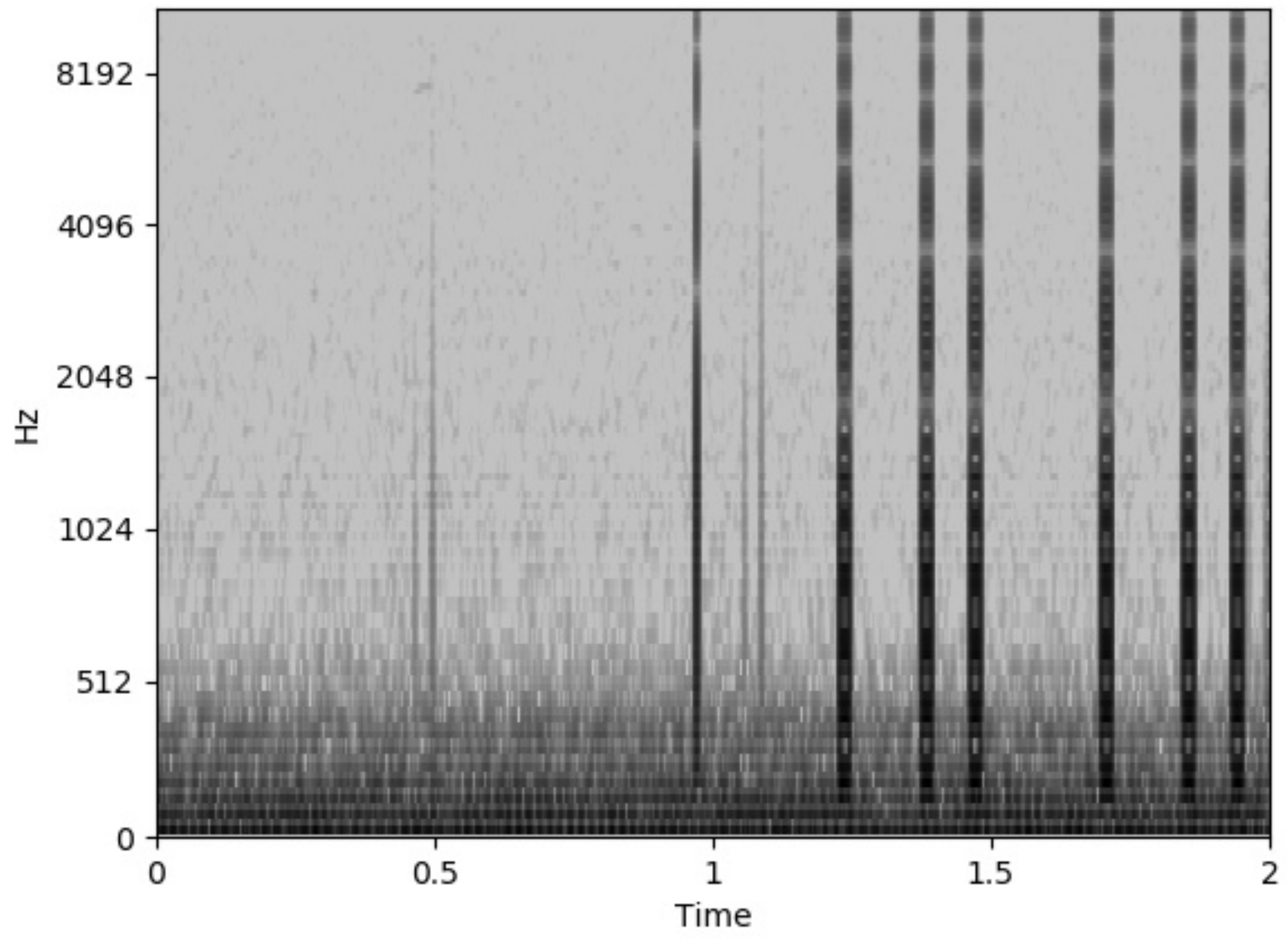


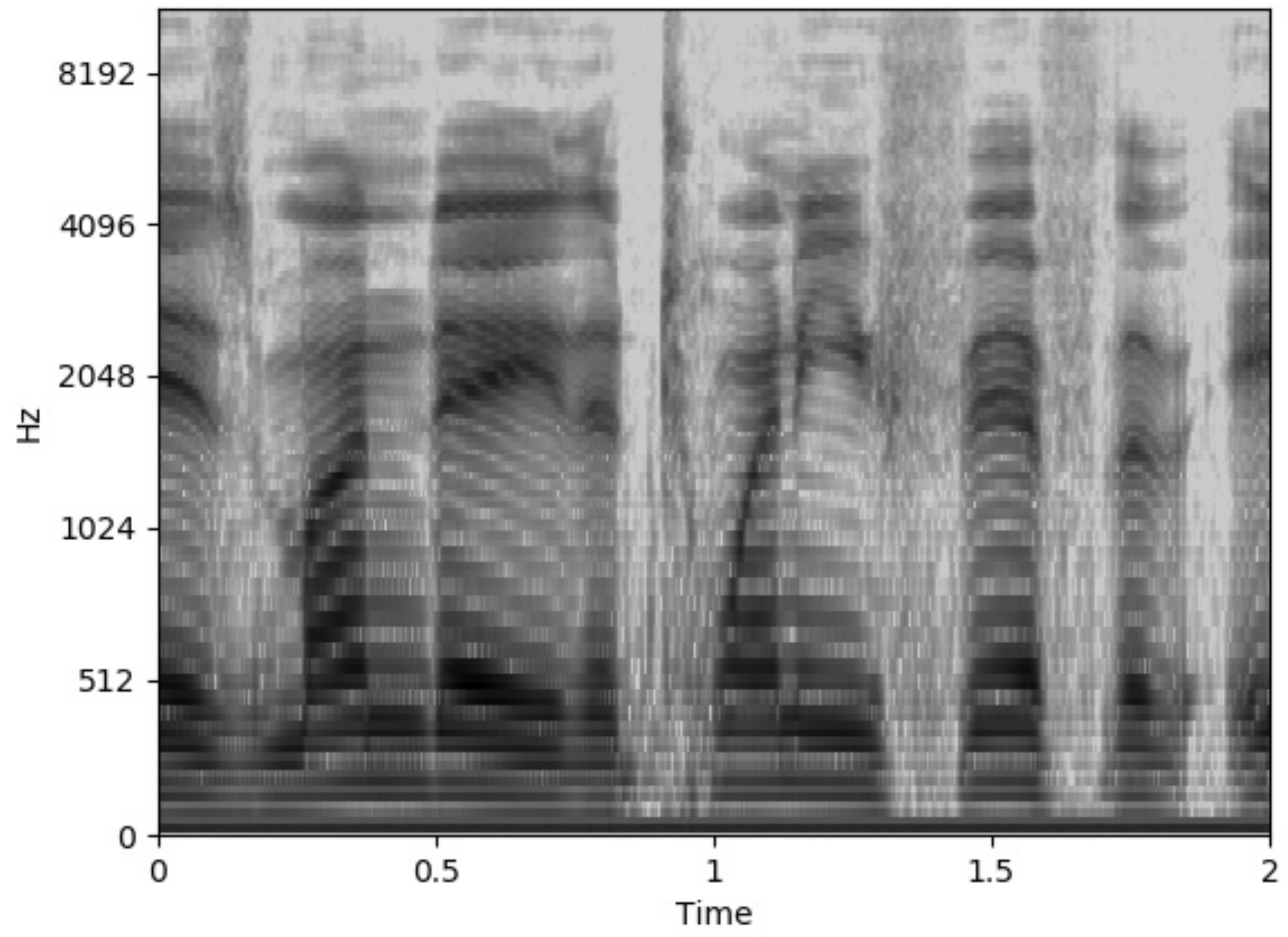
Some images of noise

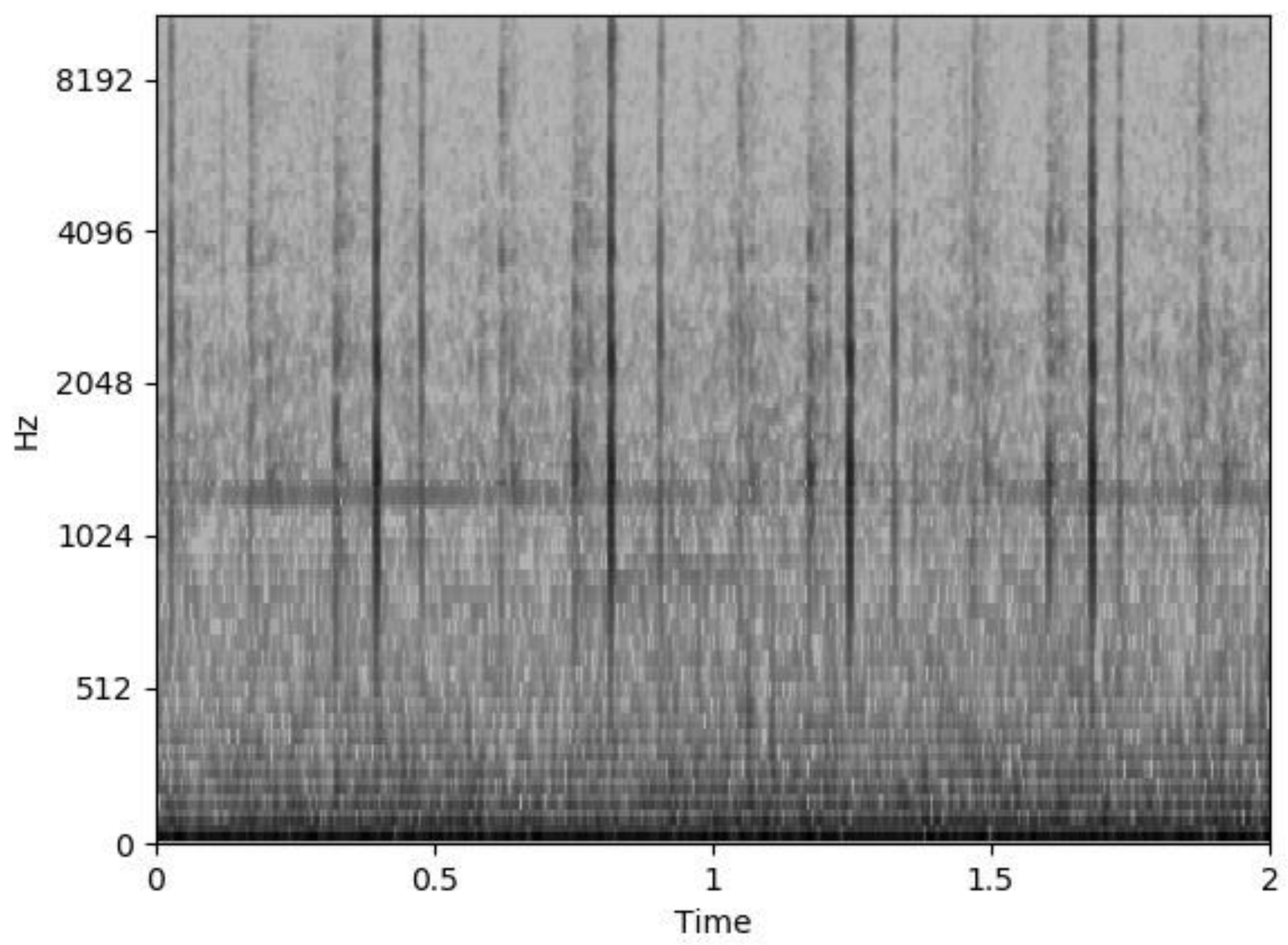




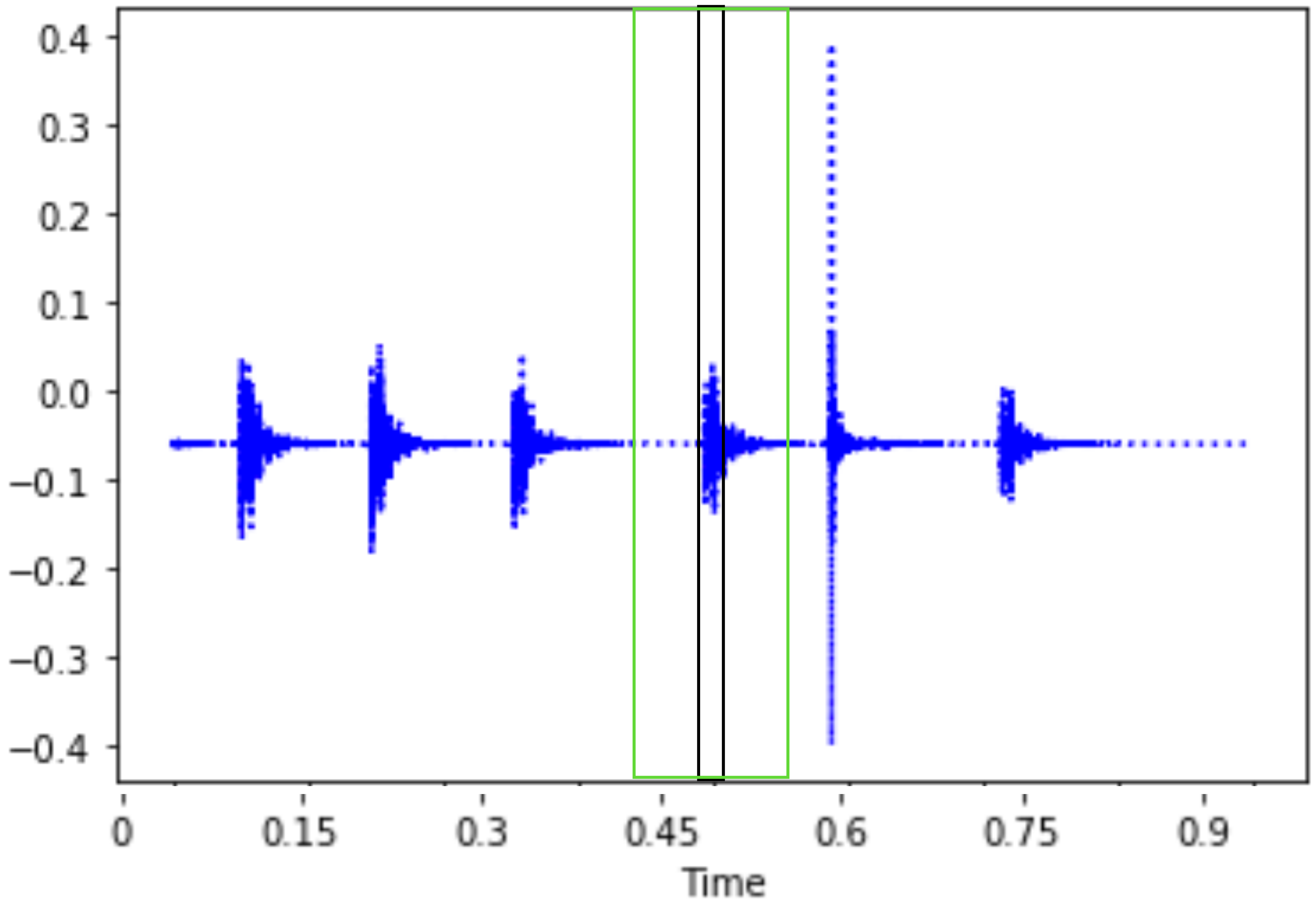






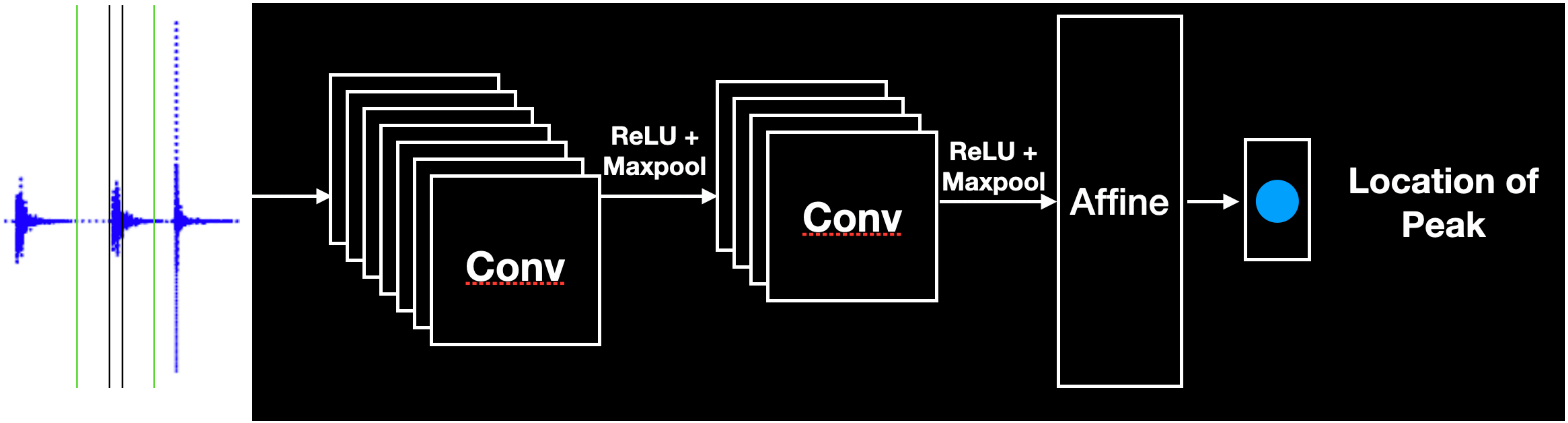


Click Detection

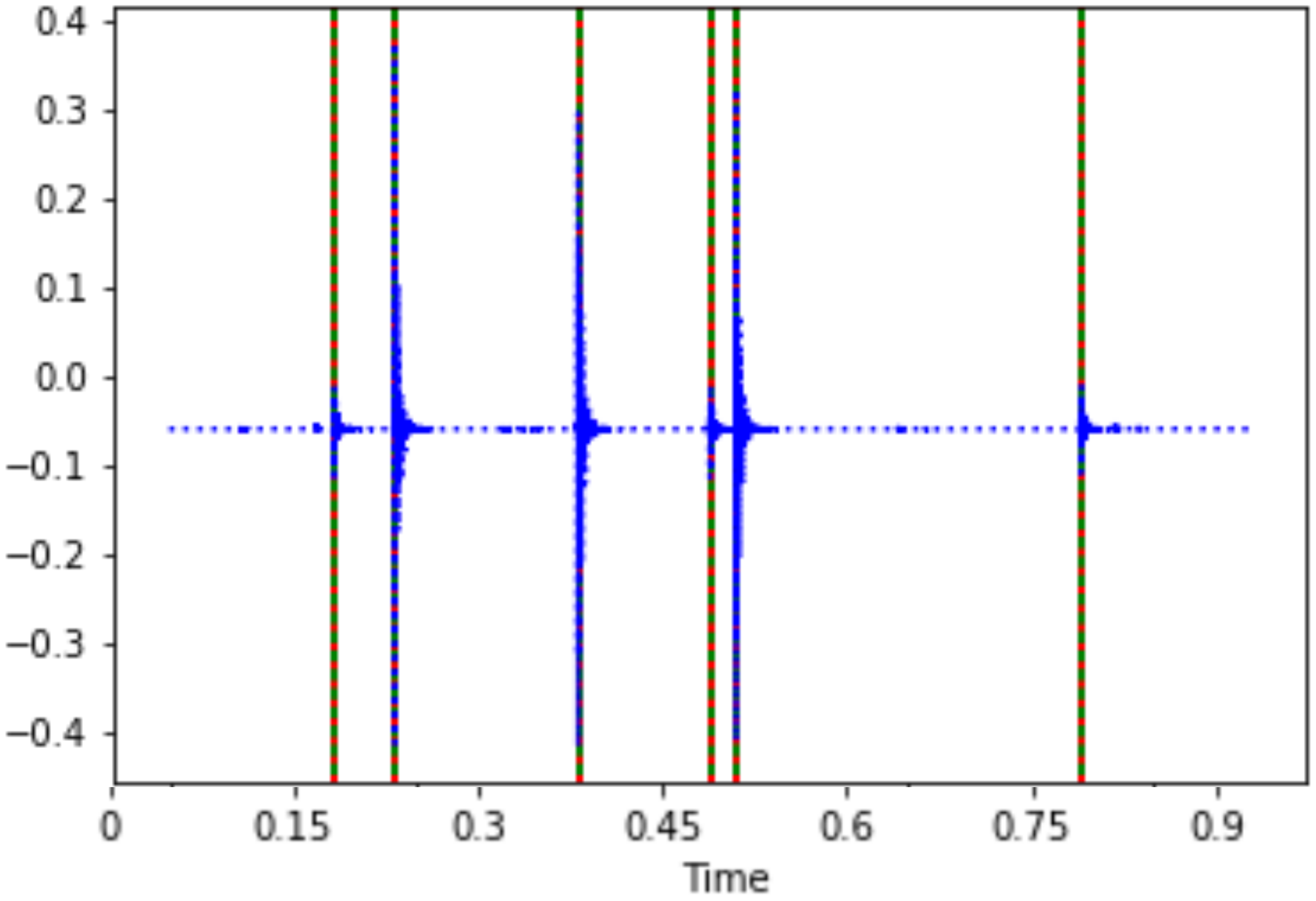


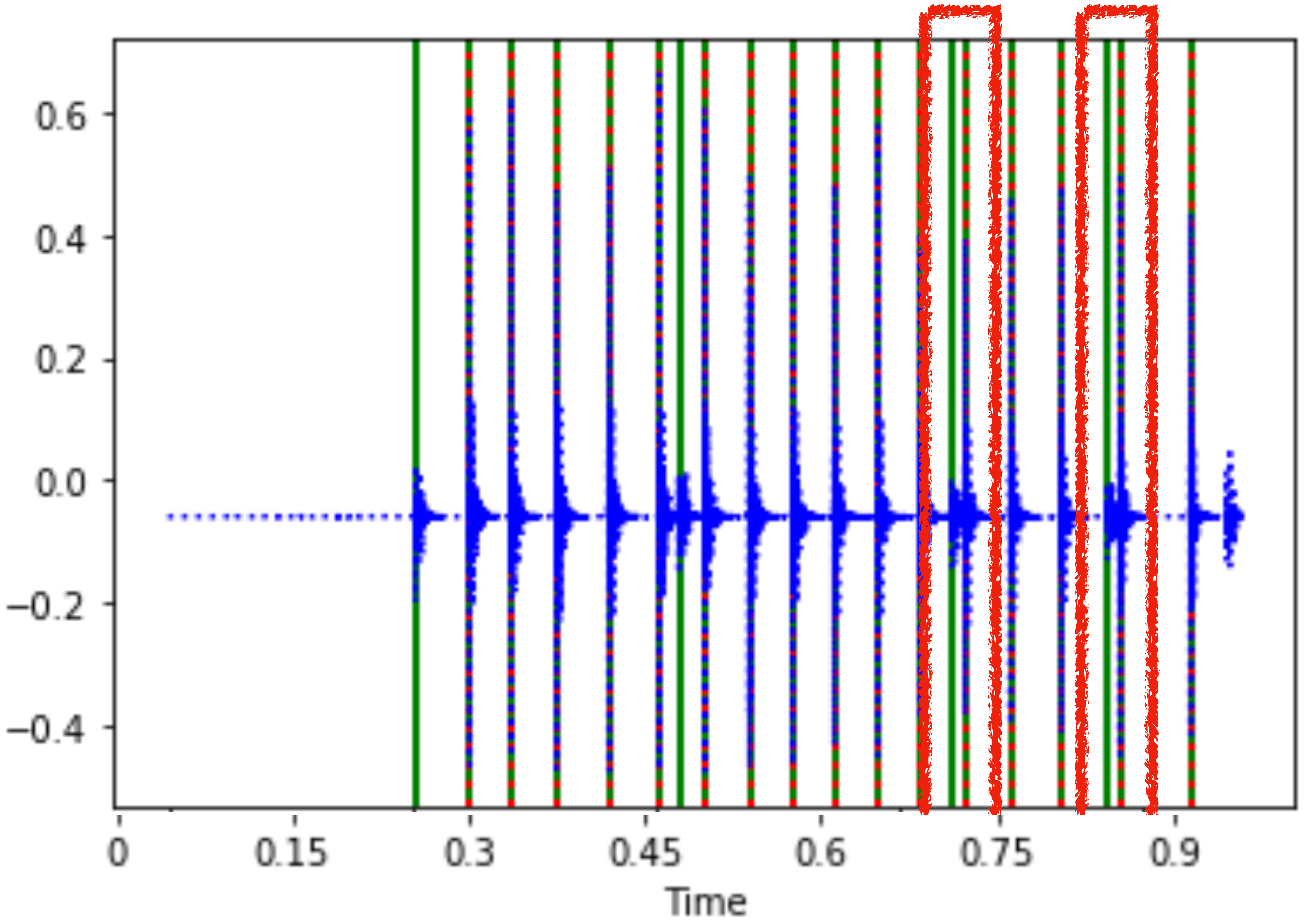
Where is the peak?

Model


Conv
ReLU +
Maxpool
Conv
ReLU +
Maxpool
Affine
Location of
Peak

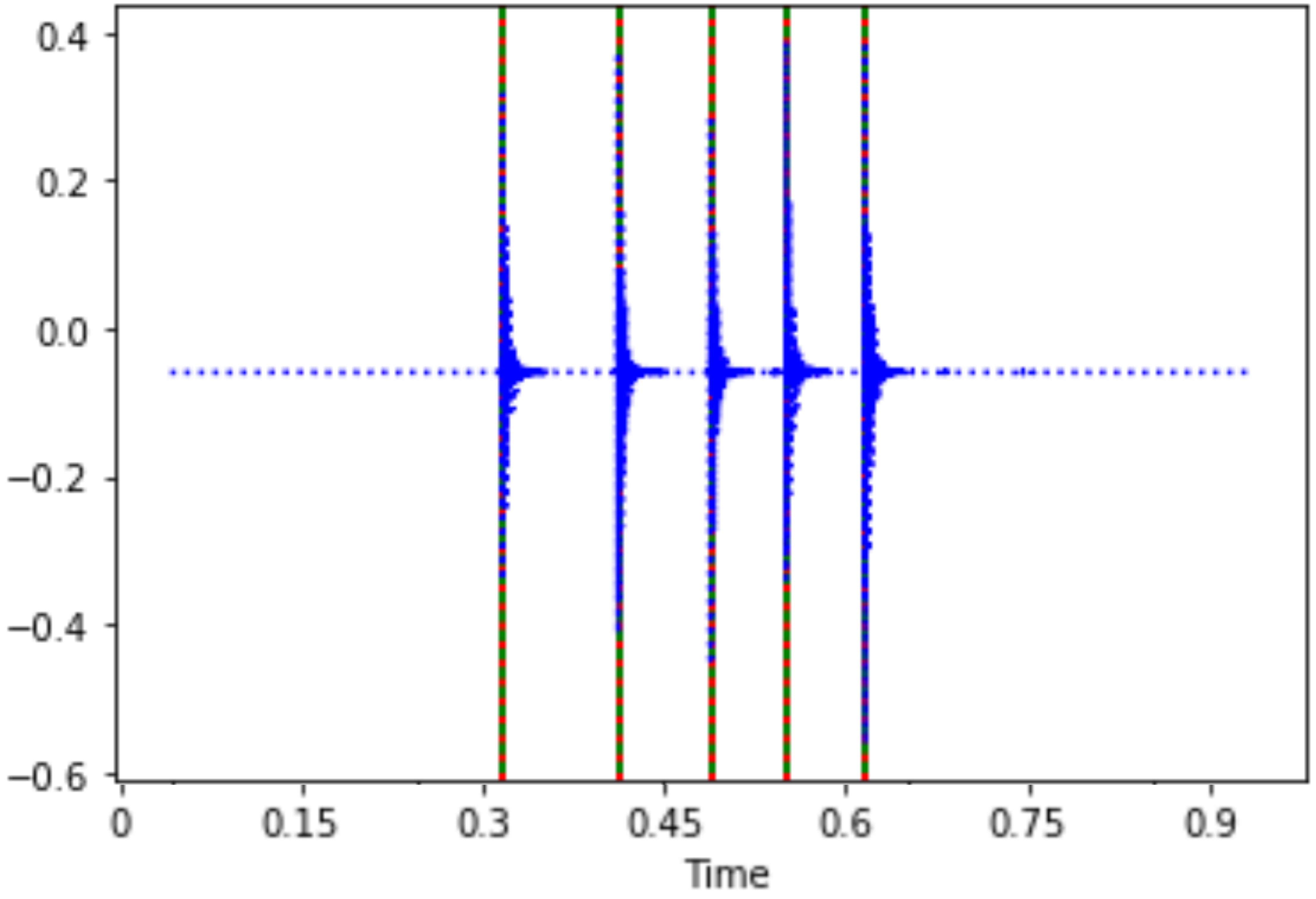
Click Detection

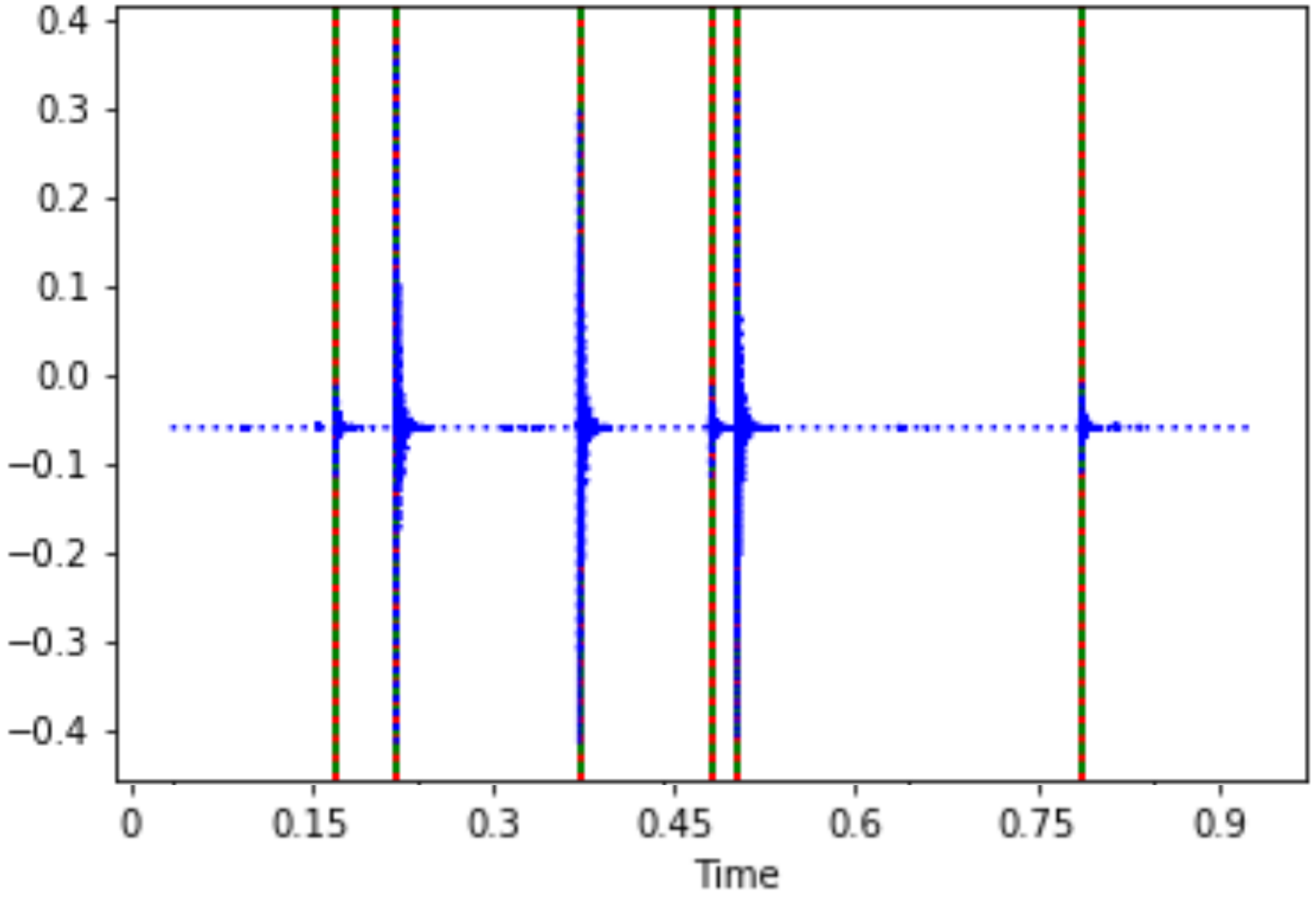




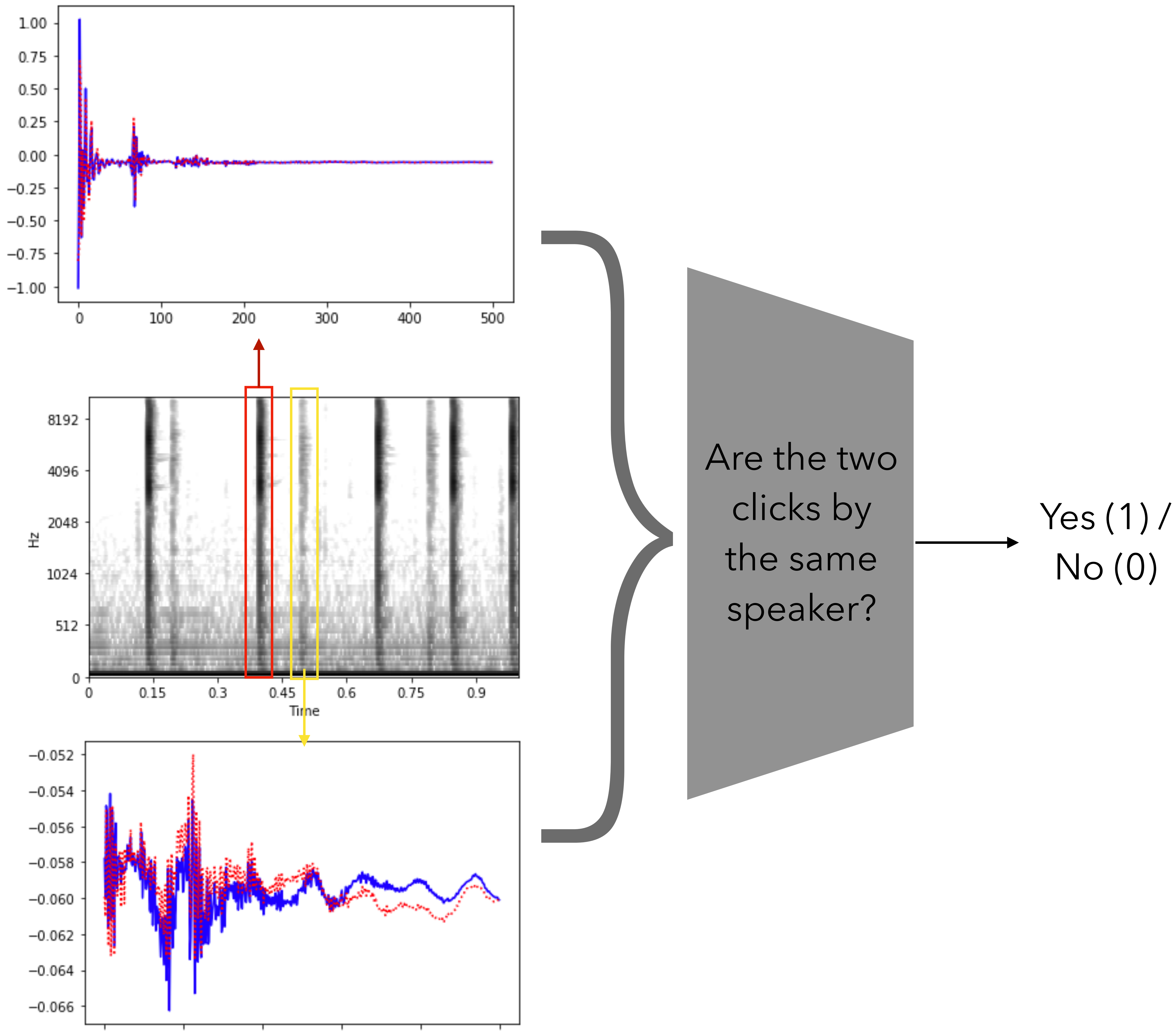
- Can detect the onset of both soft/ loud clicks >96% accuracy
- Can also recover previously unannotated/unidentified signal!

Click Detection (some more)





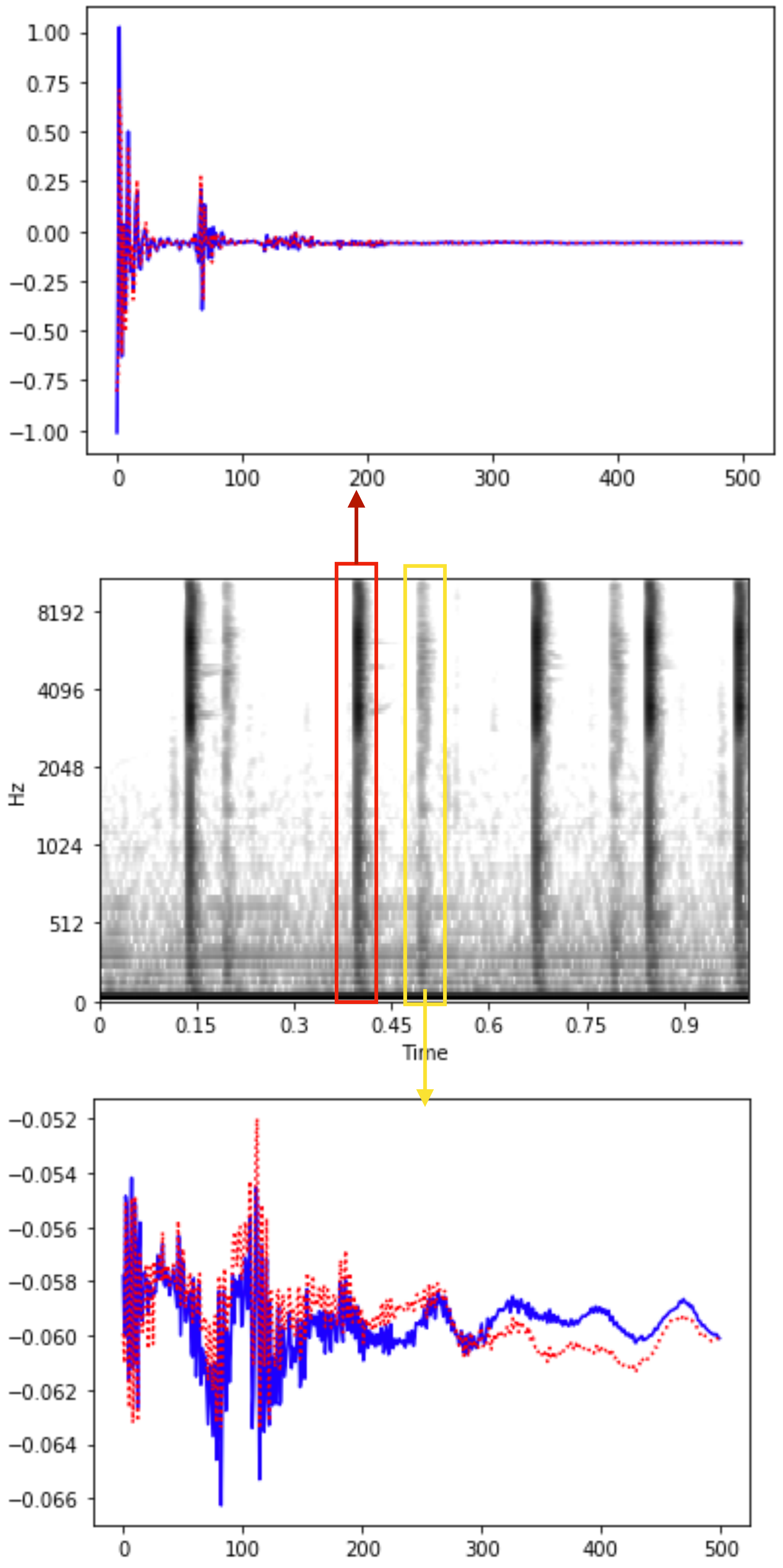
Click Separation



Are the two clicks by the same speaker?

Yes (1) / No (0)

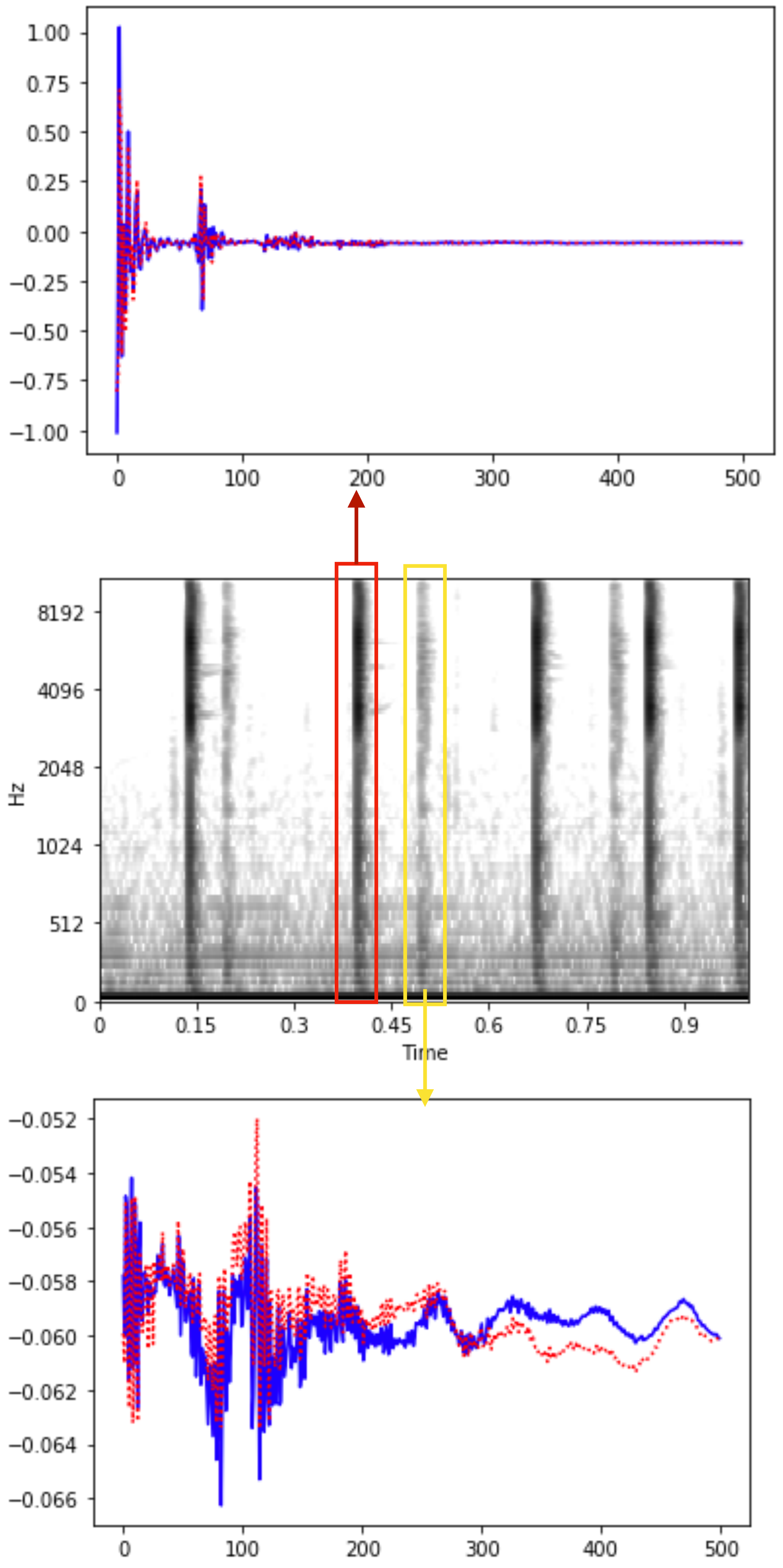
Click Separation



- IPI info + angle of arrival info

Accuracy: 59%

Click Separation



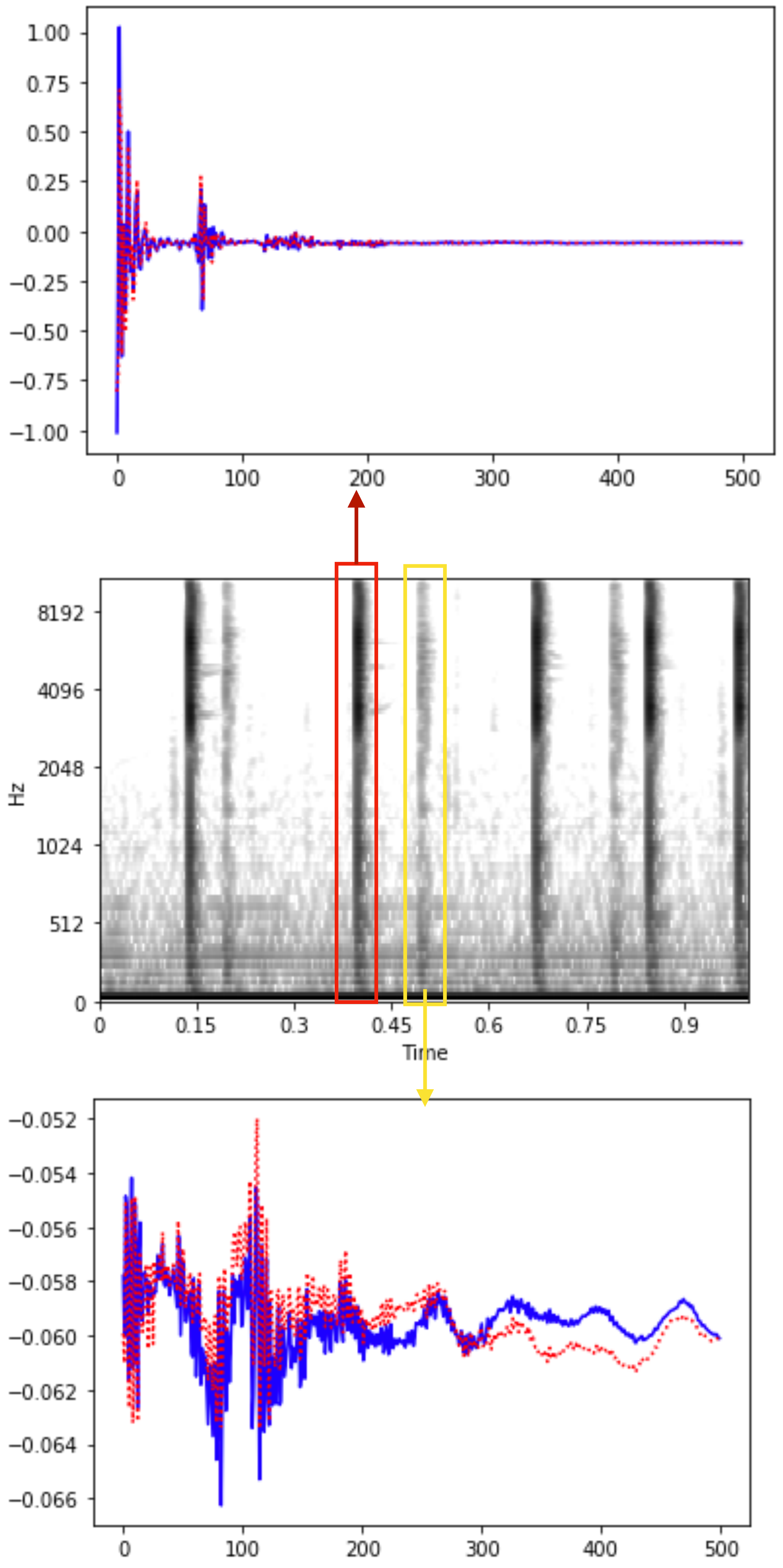
- IPI info + angle of arrival info

Accuracy: 59%

- Raw wav

Accuracy: 69.8%

Click Separation



- IPI info + angle of arrival info

Accuracy: 59%

- Raw wav

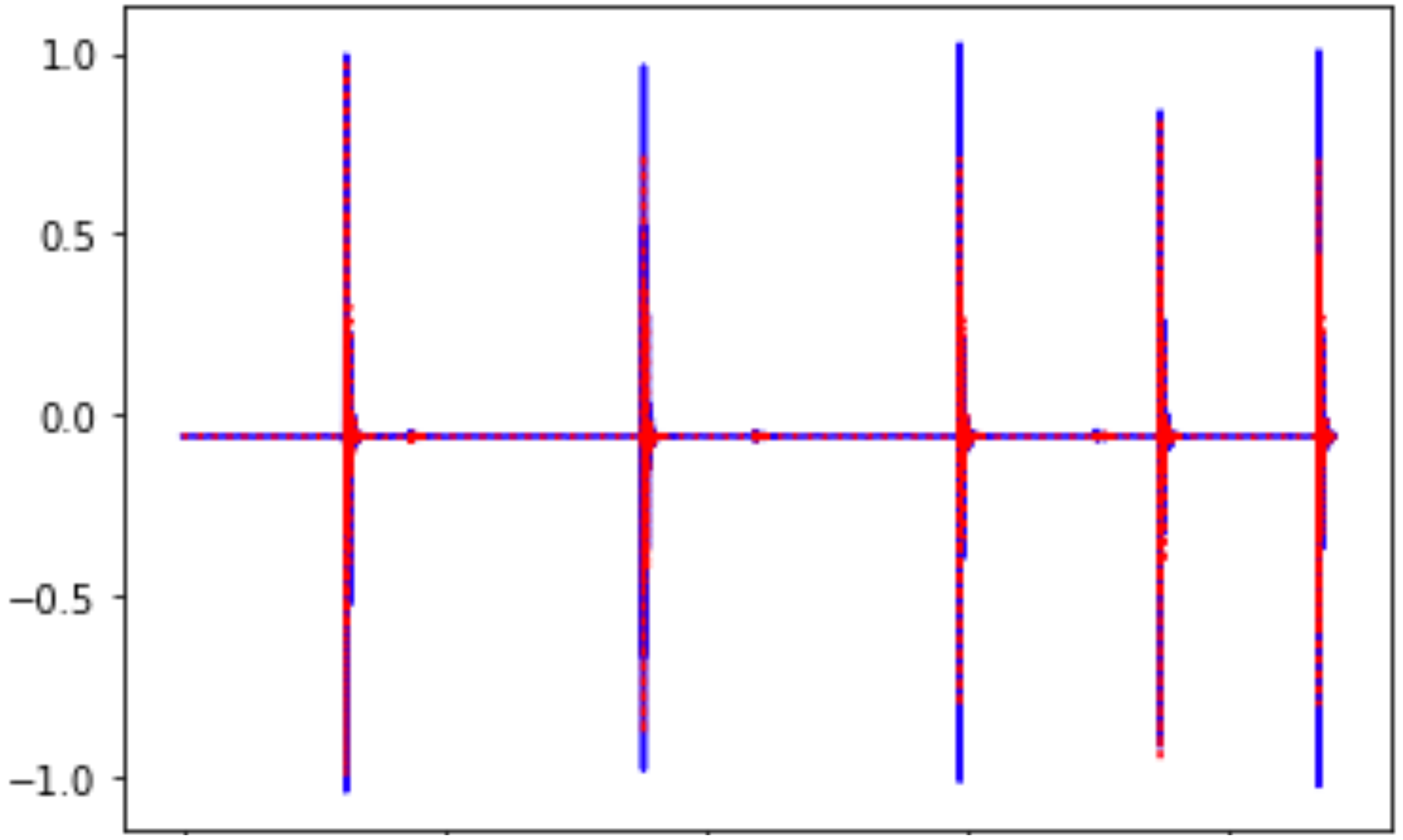
Accuracy: 69.8%

- Input: Just raw wav of one click: Output: Does the click belong to the whale wearing the mic?

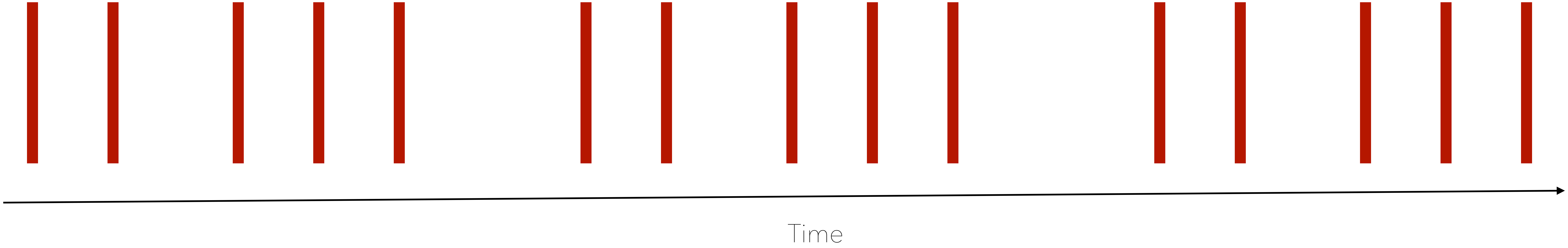
Accuracy: 88%

2. Identifying the underlying “Symbols” and “Rules”

Could looking at the sound differently help us build better hypotheses?

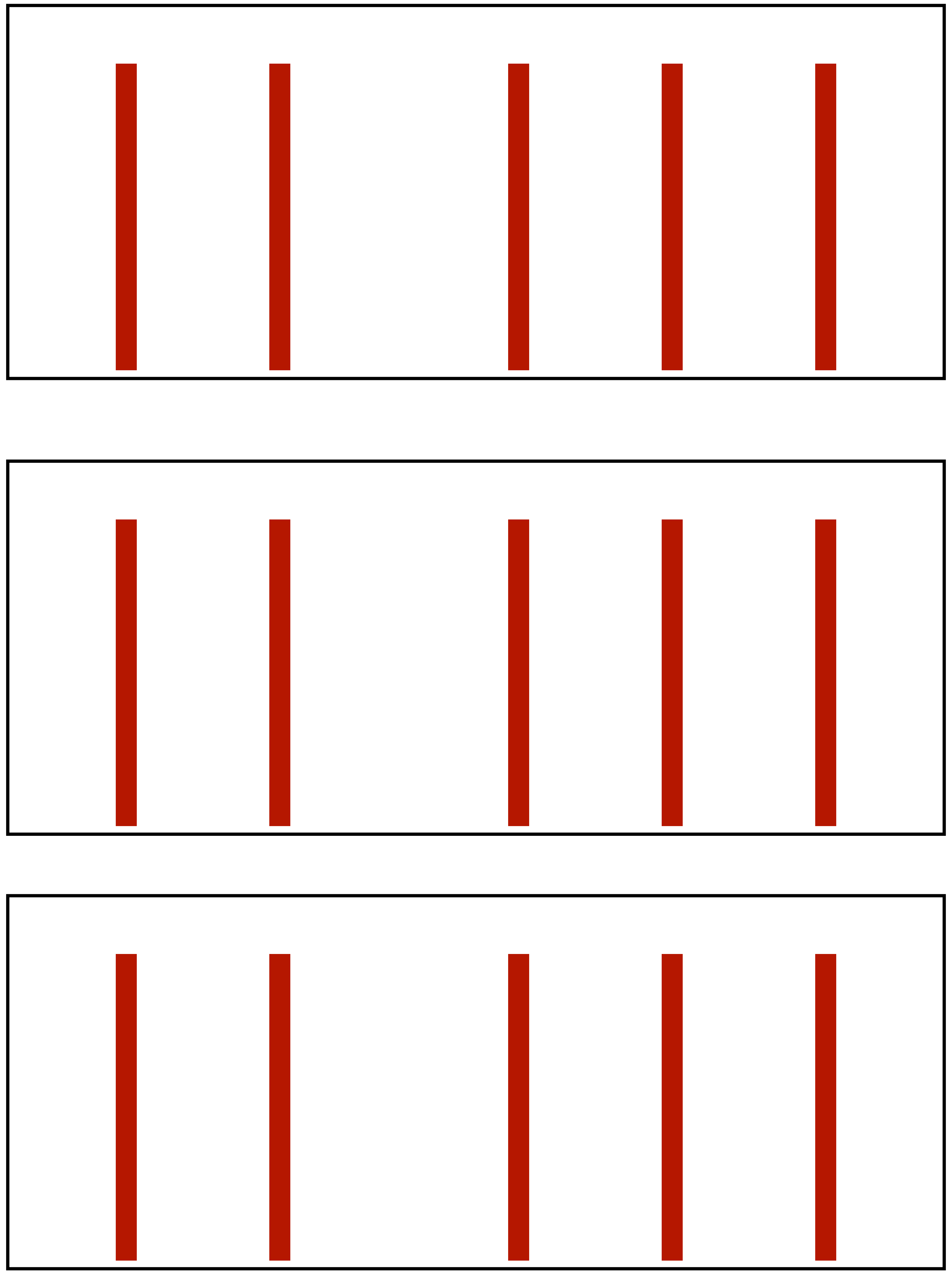


Understanding codas

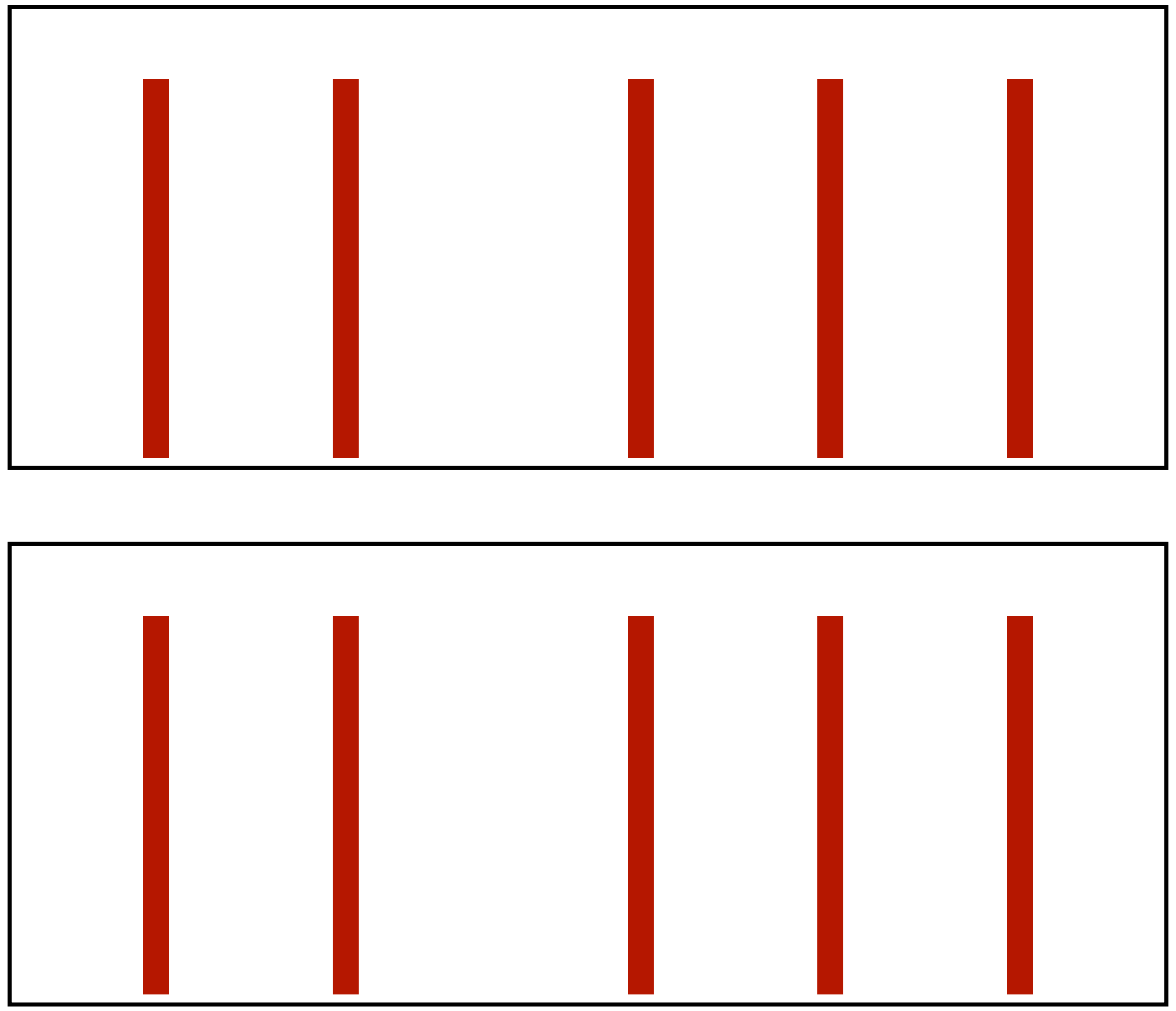


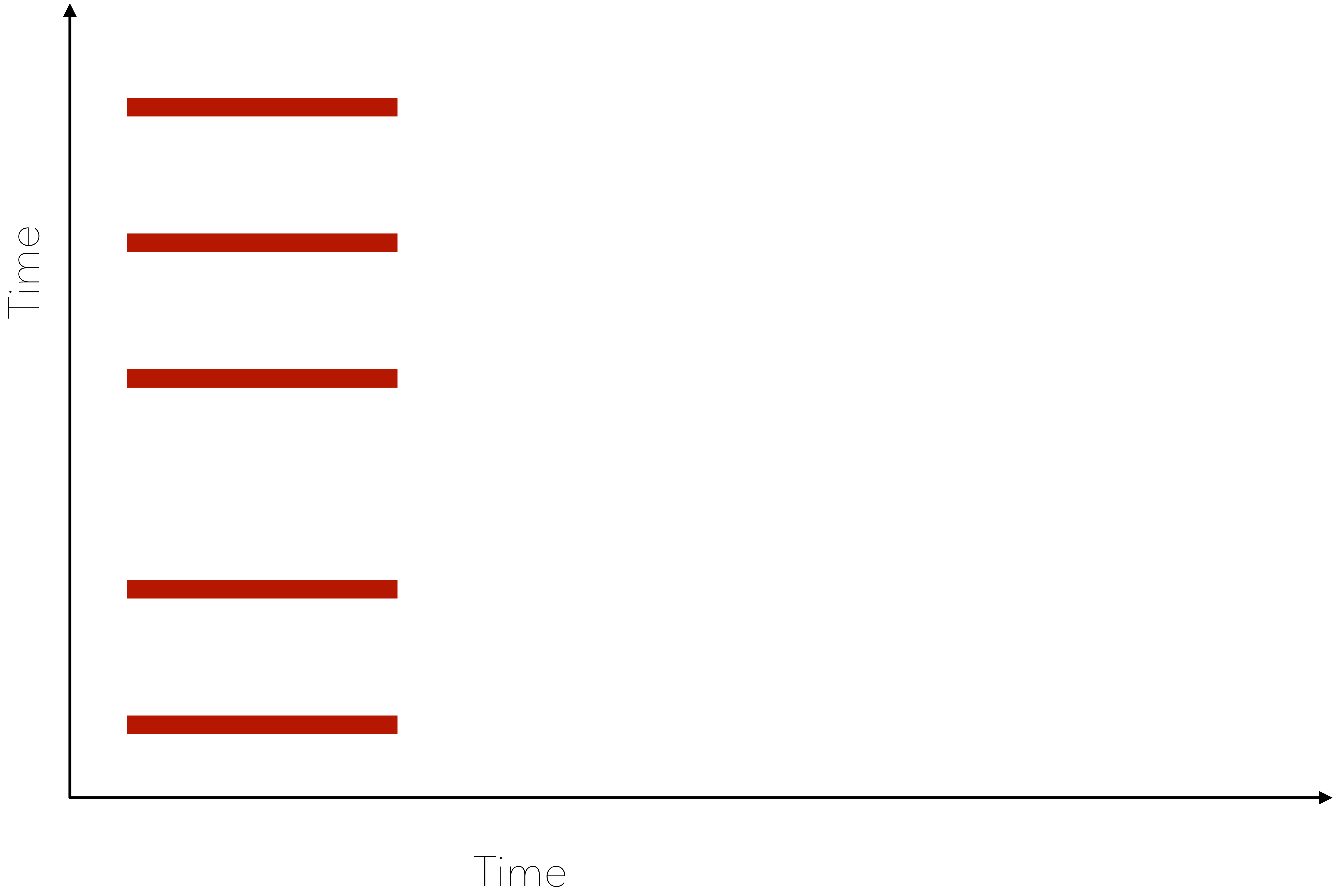
Understanding codas

Understanding codas

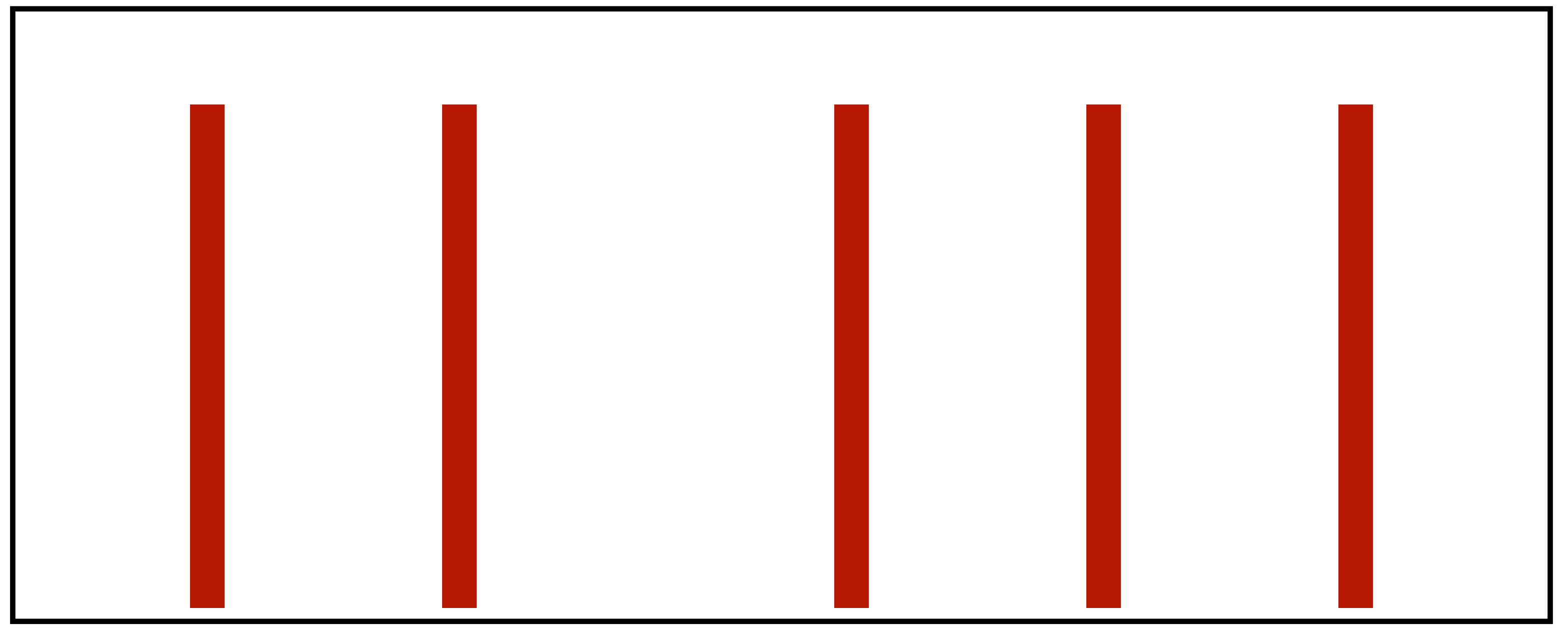


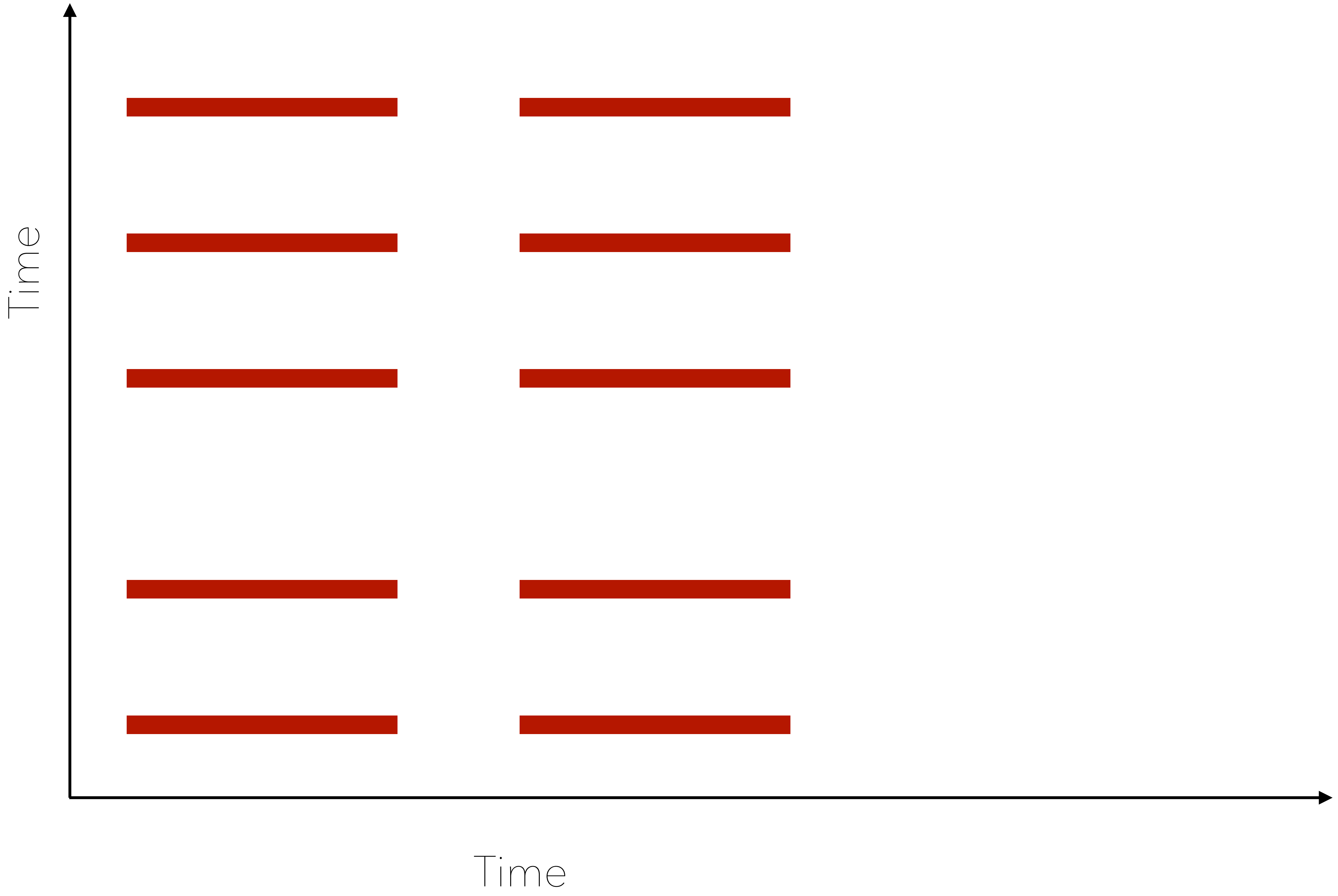
Understanding codas



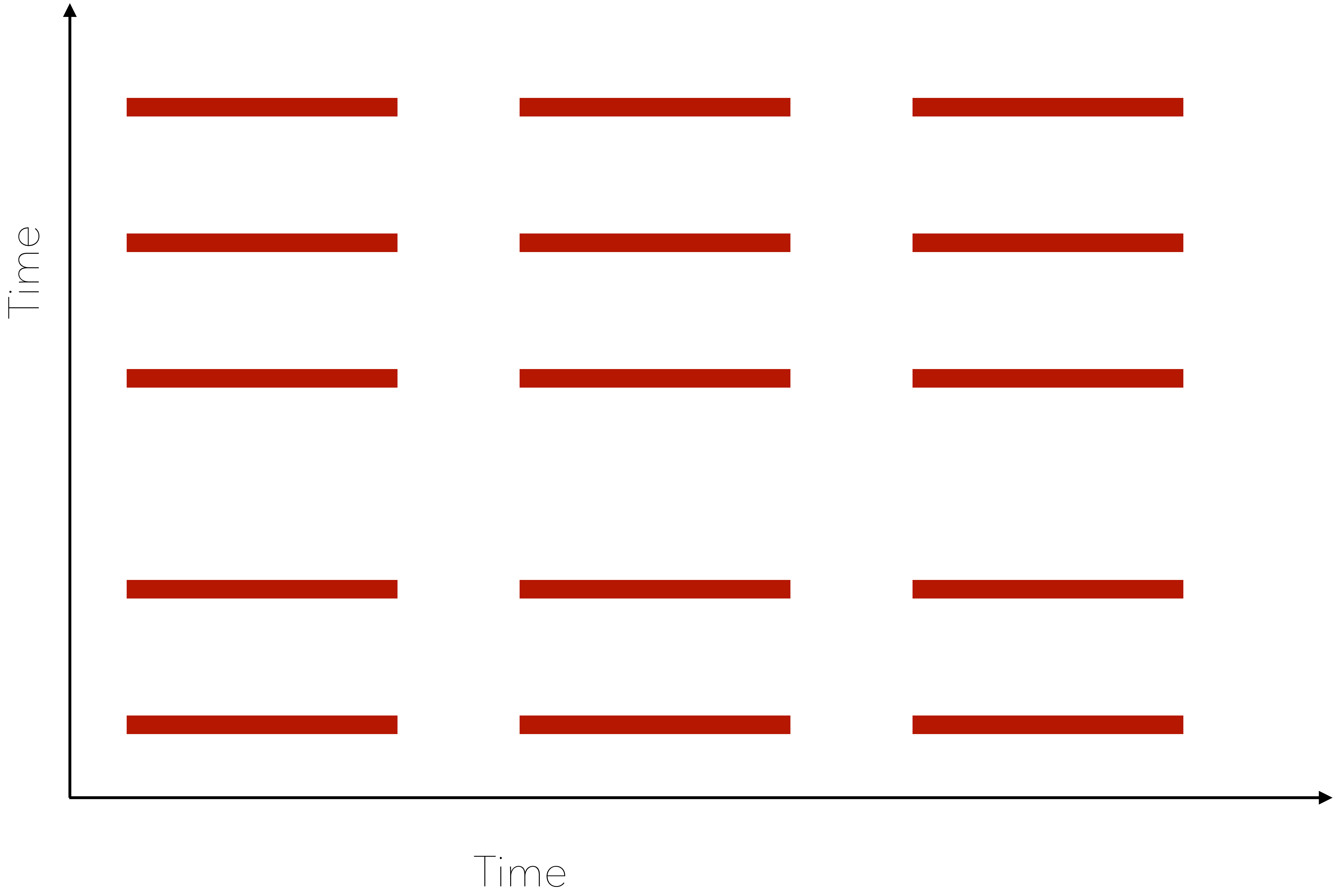


Understanding codas

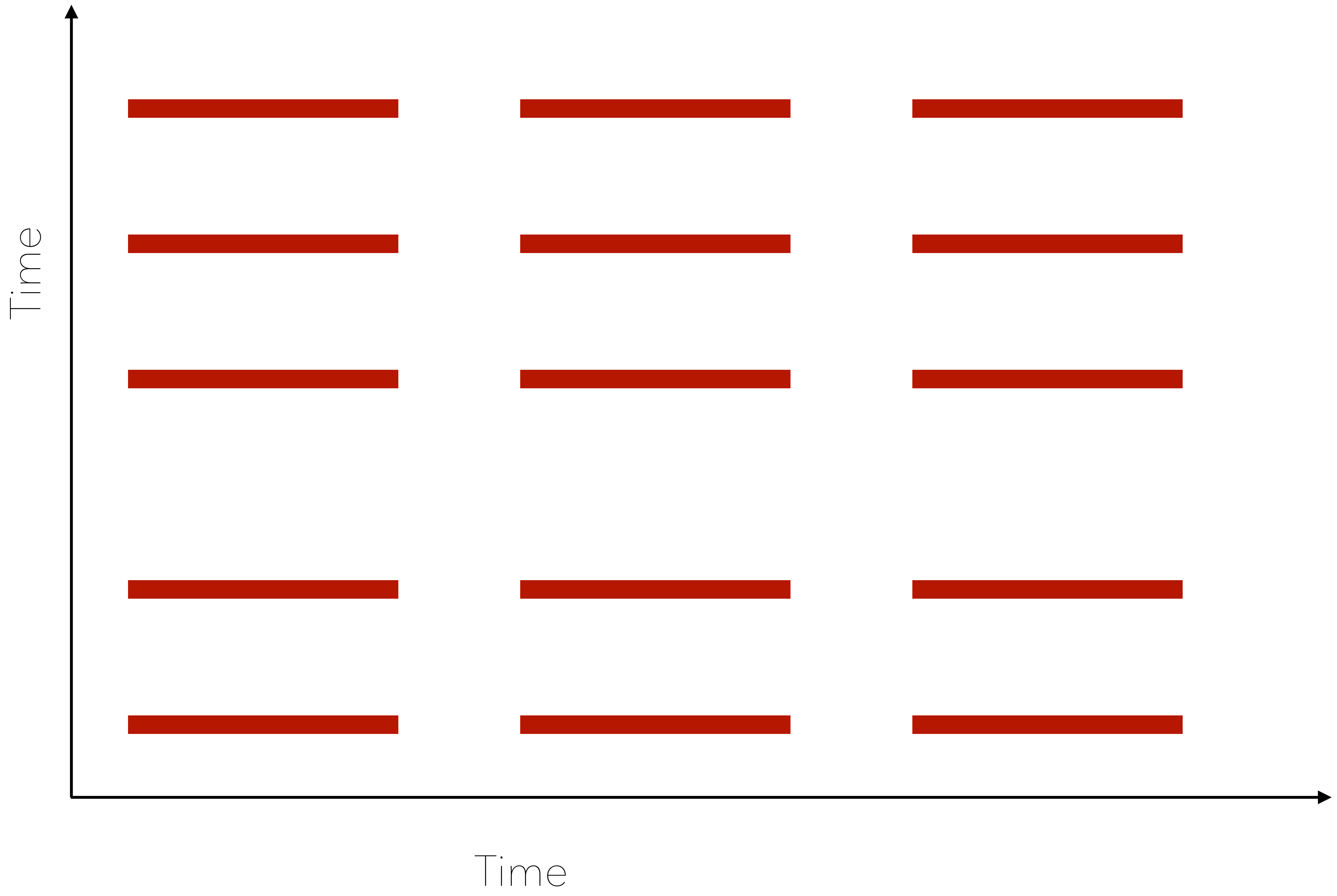




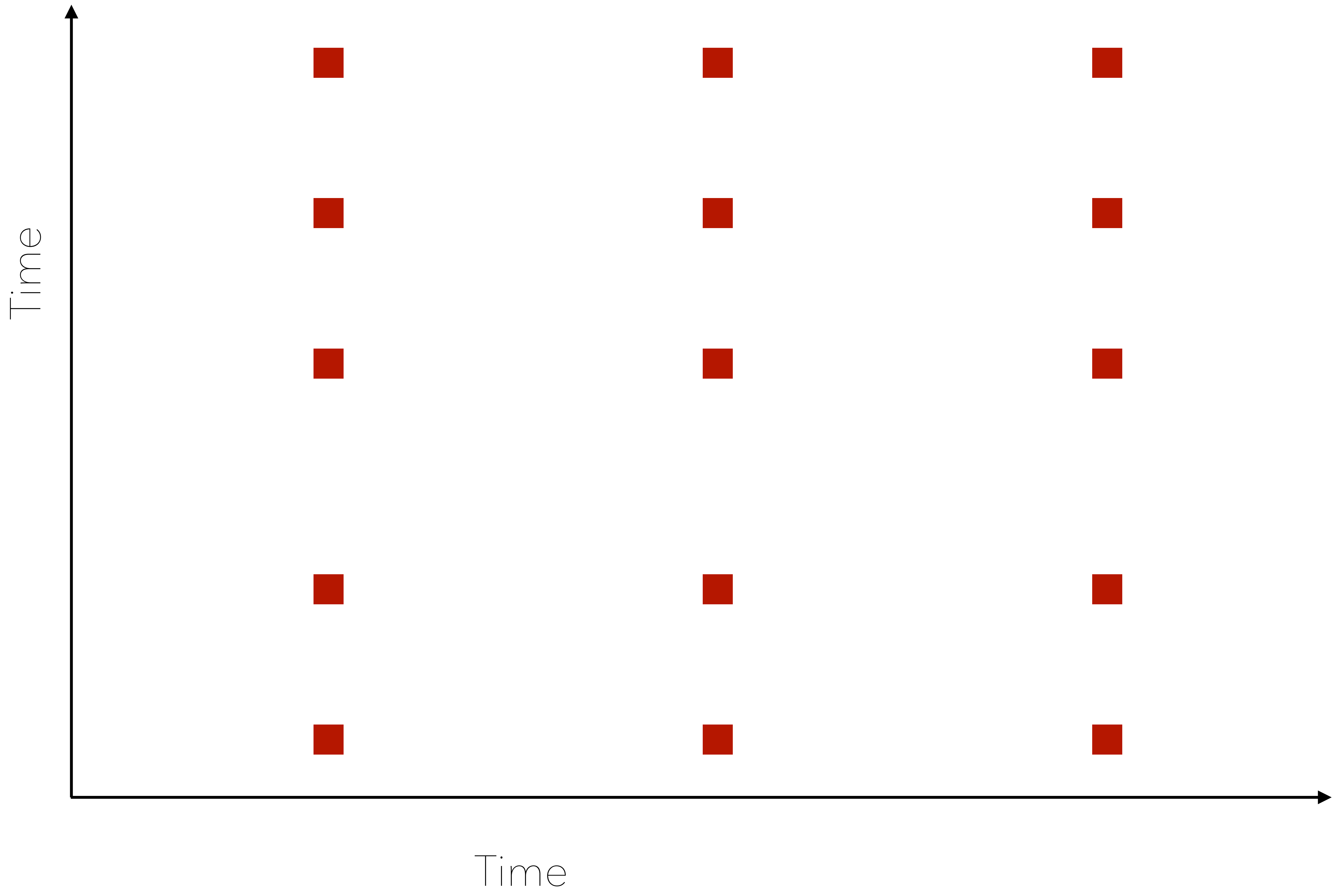
Understanding codas



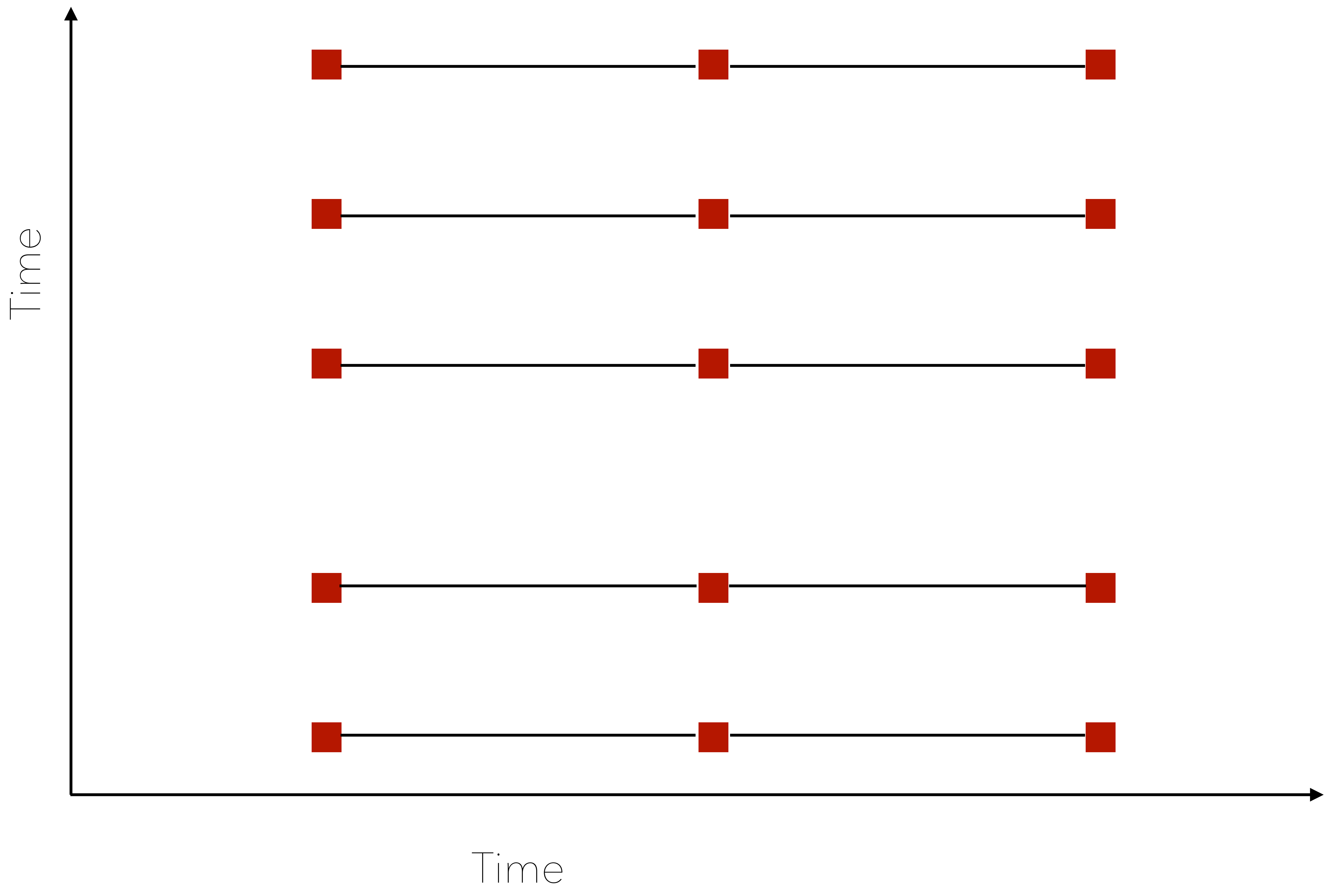
Understanding codas



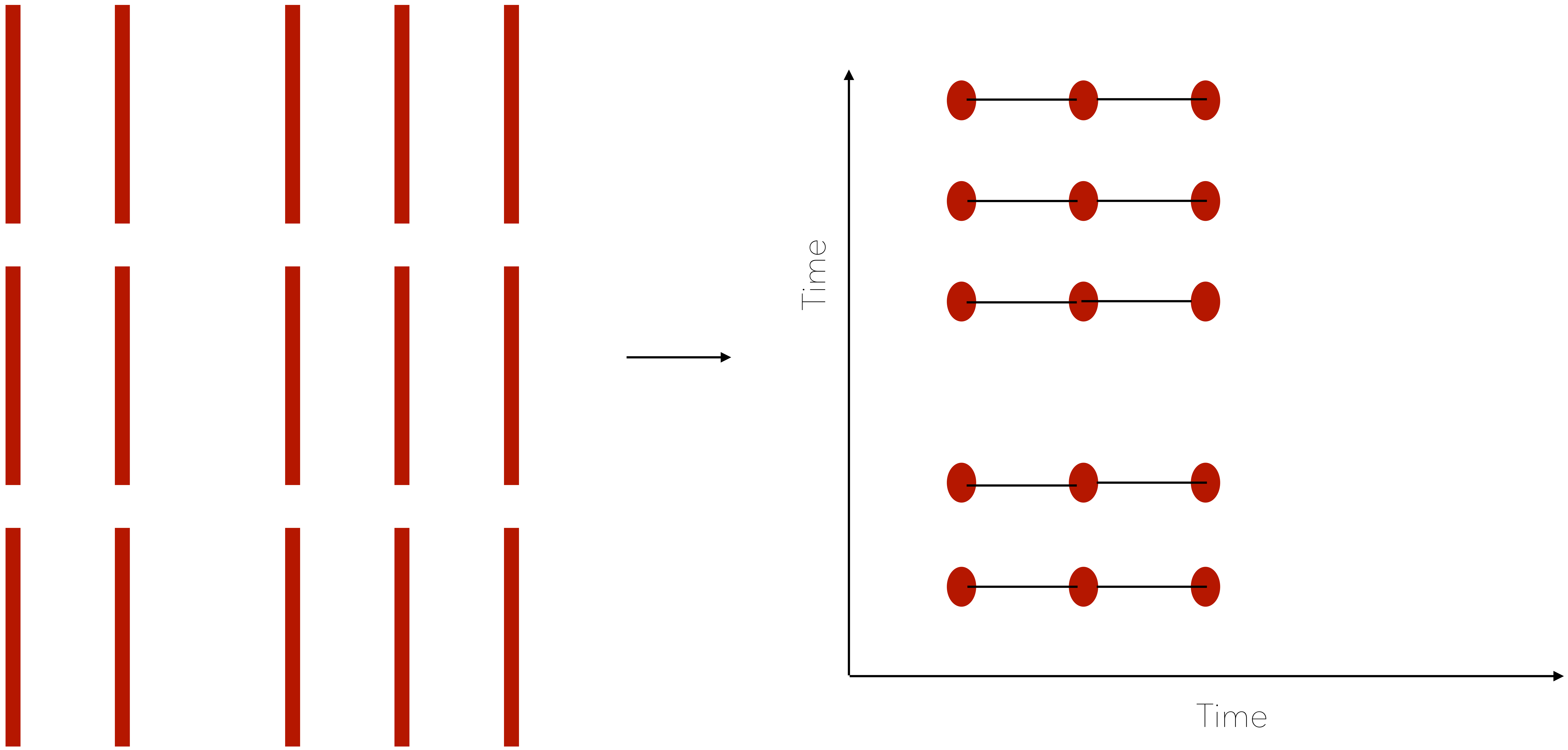
Understanding codas

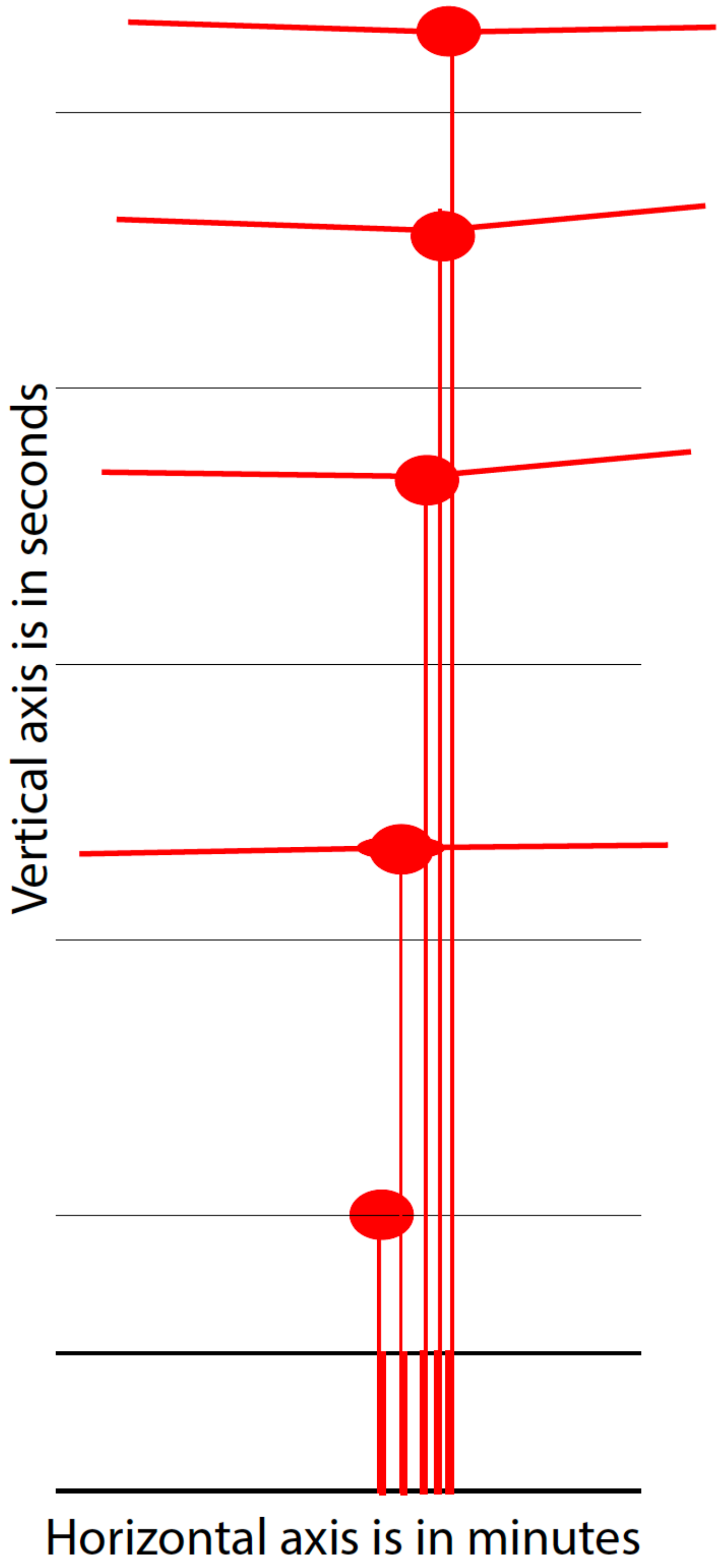


Understanding codas

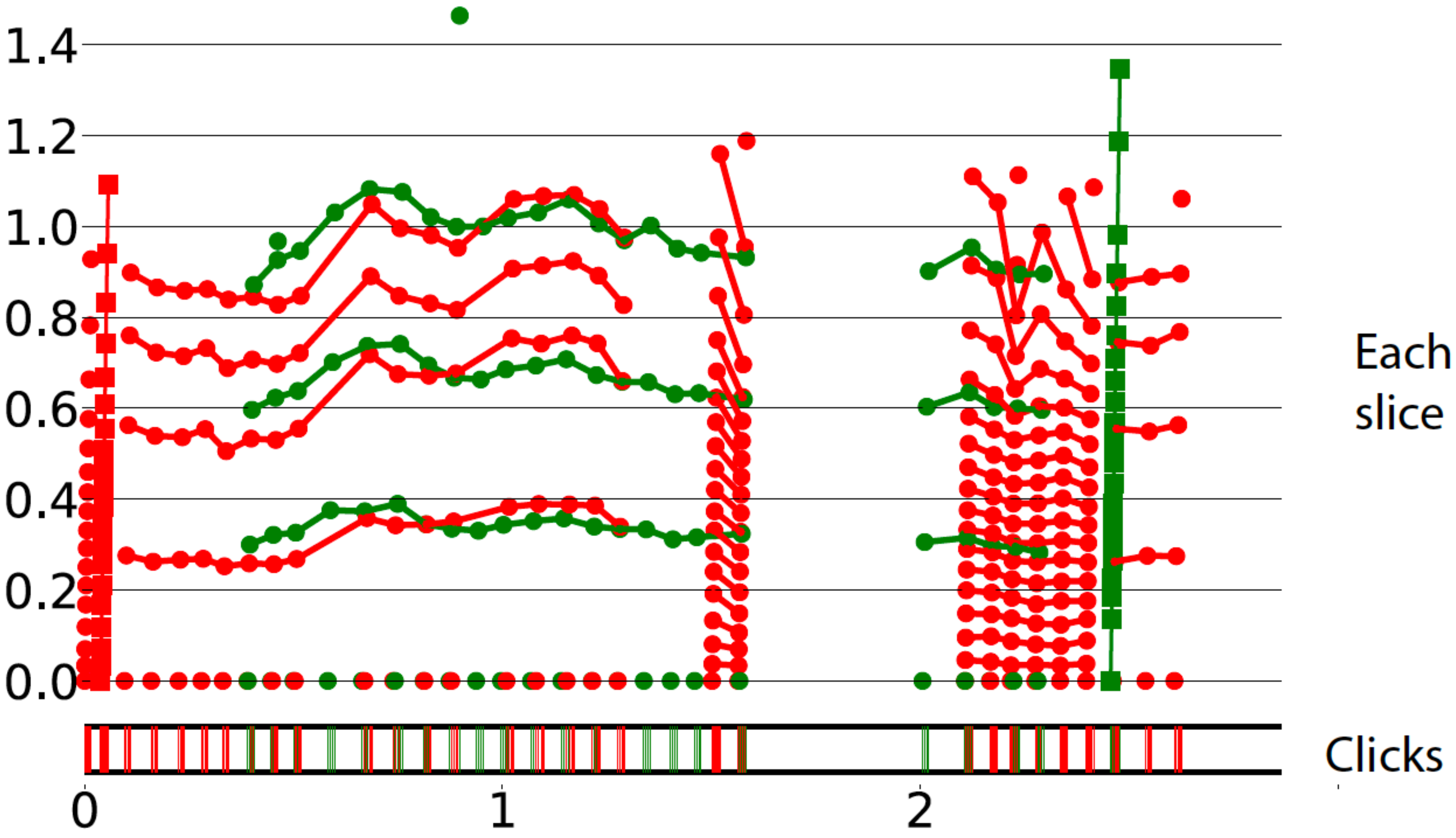


Understanding codas





Vertical axis is in seconds



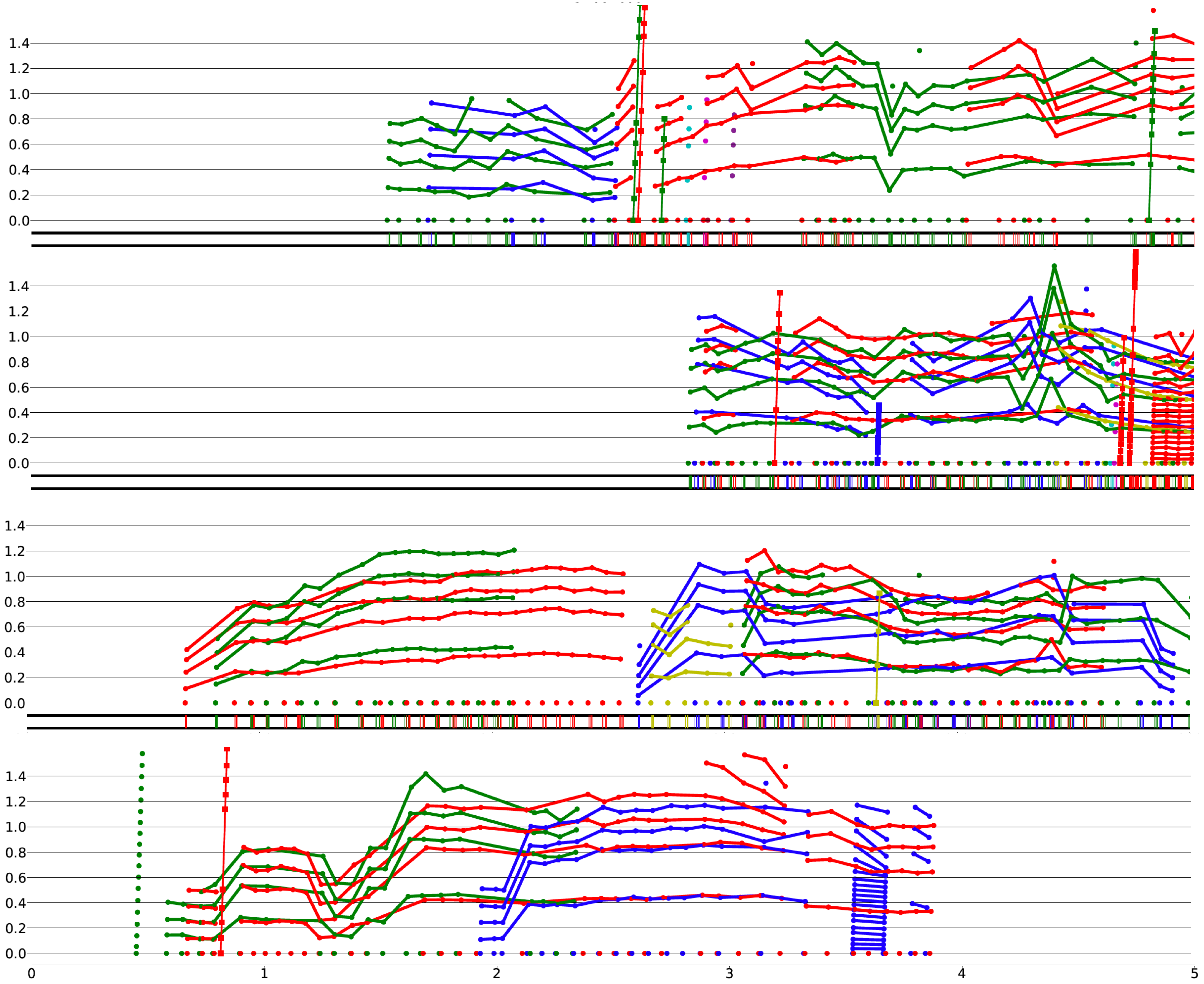
Each vertical slice is one CODA

Horizontal axis is in minutes

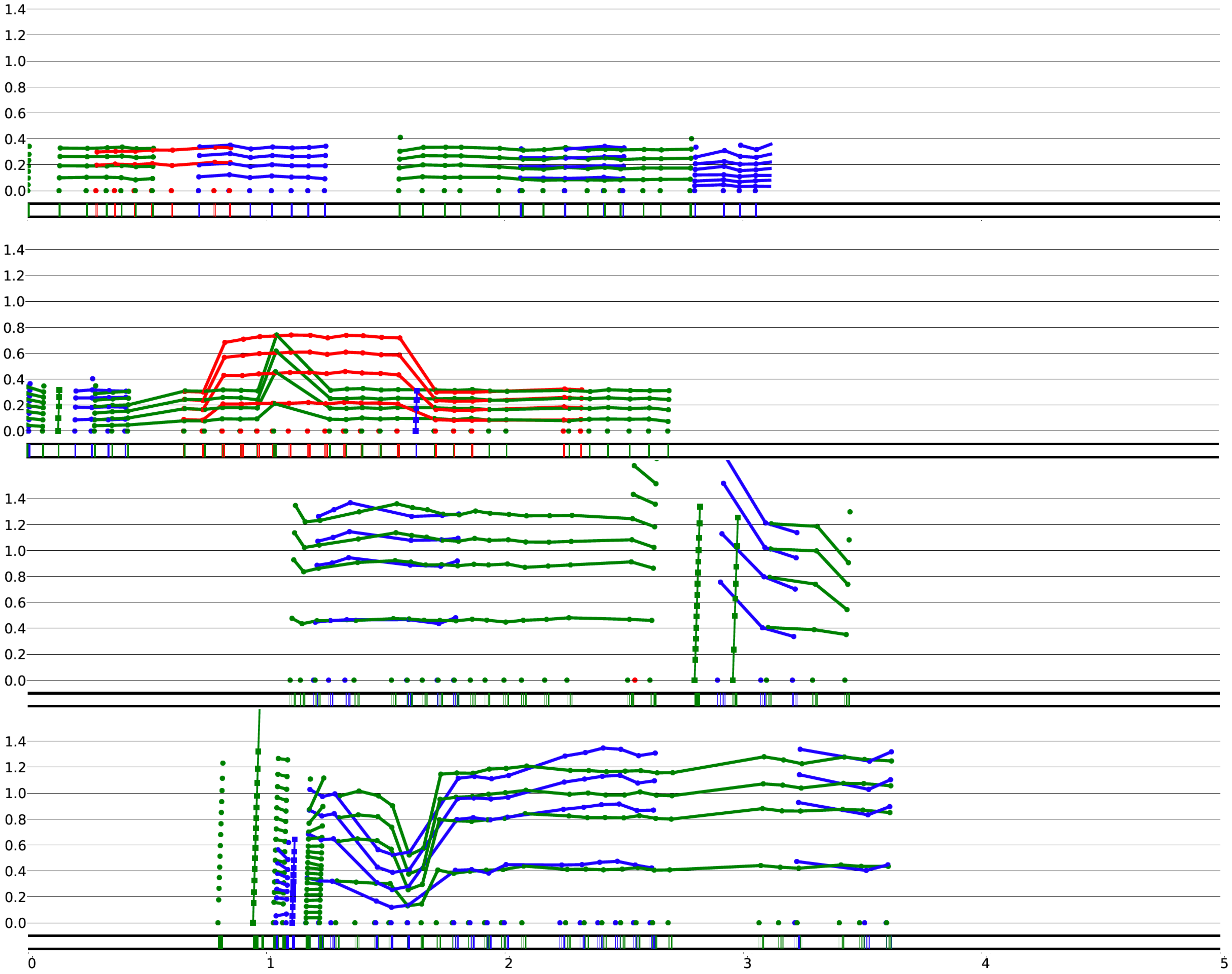
!! Let’s say the little amount of variation in the lengths was noise. Then why does the red whale’s pattern follow the green whale’s pattern?

The book of whales

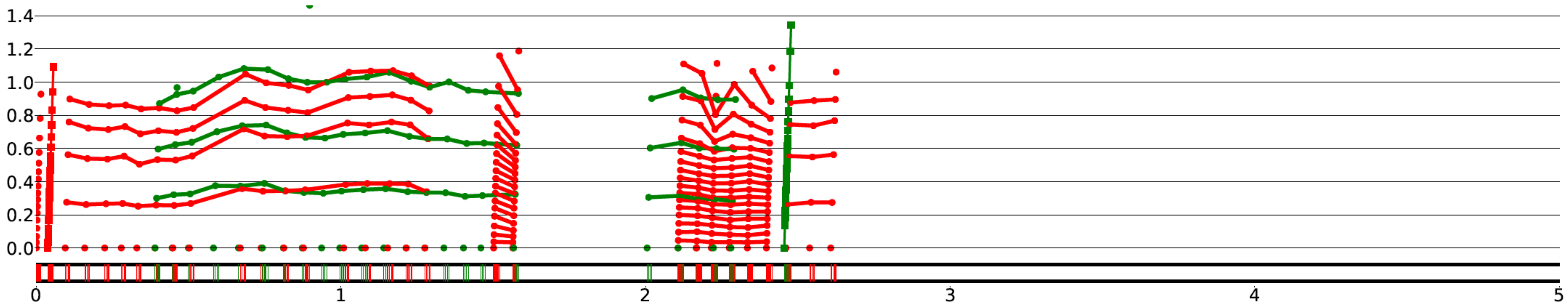
A book of whales talking about... well, we do not know yet



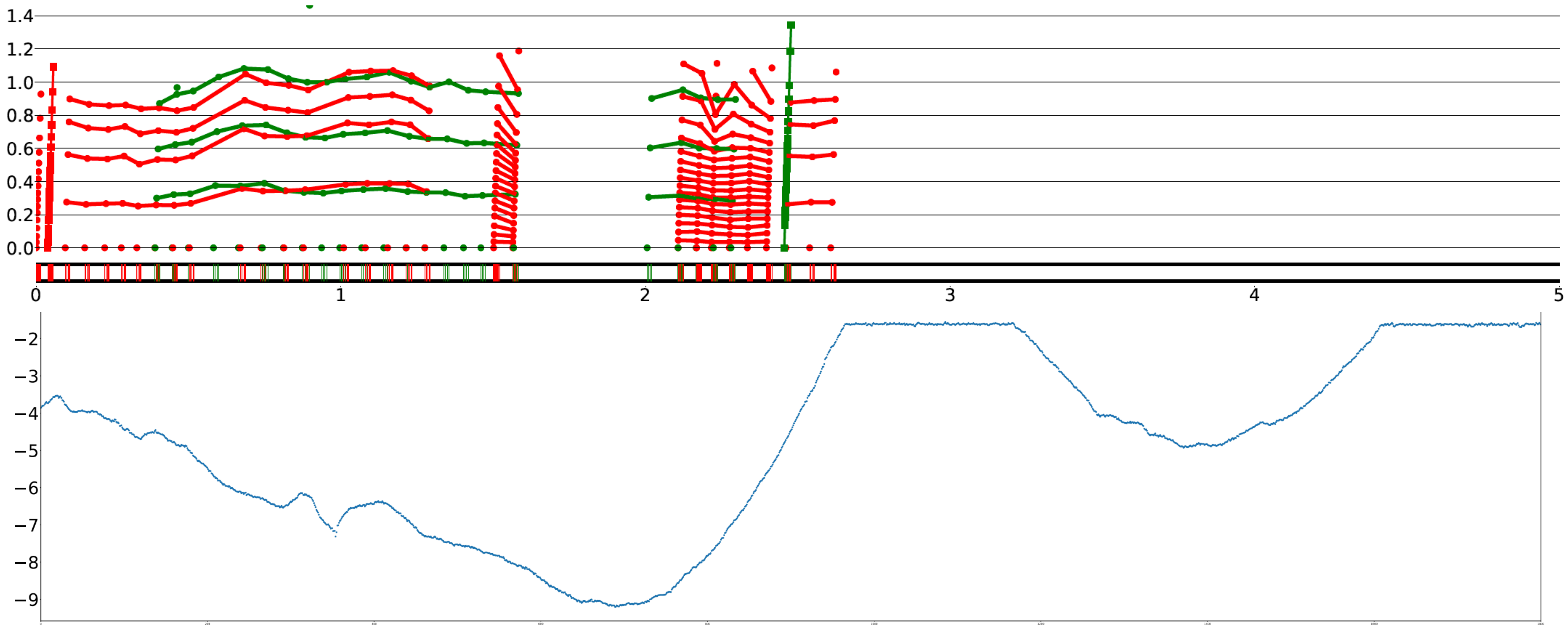
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
1
2
3
4
5


1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5

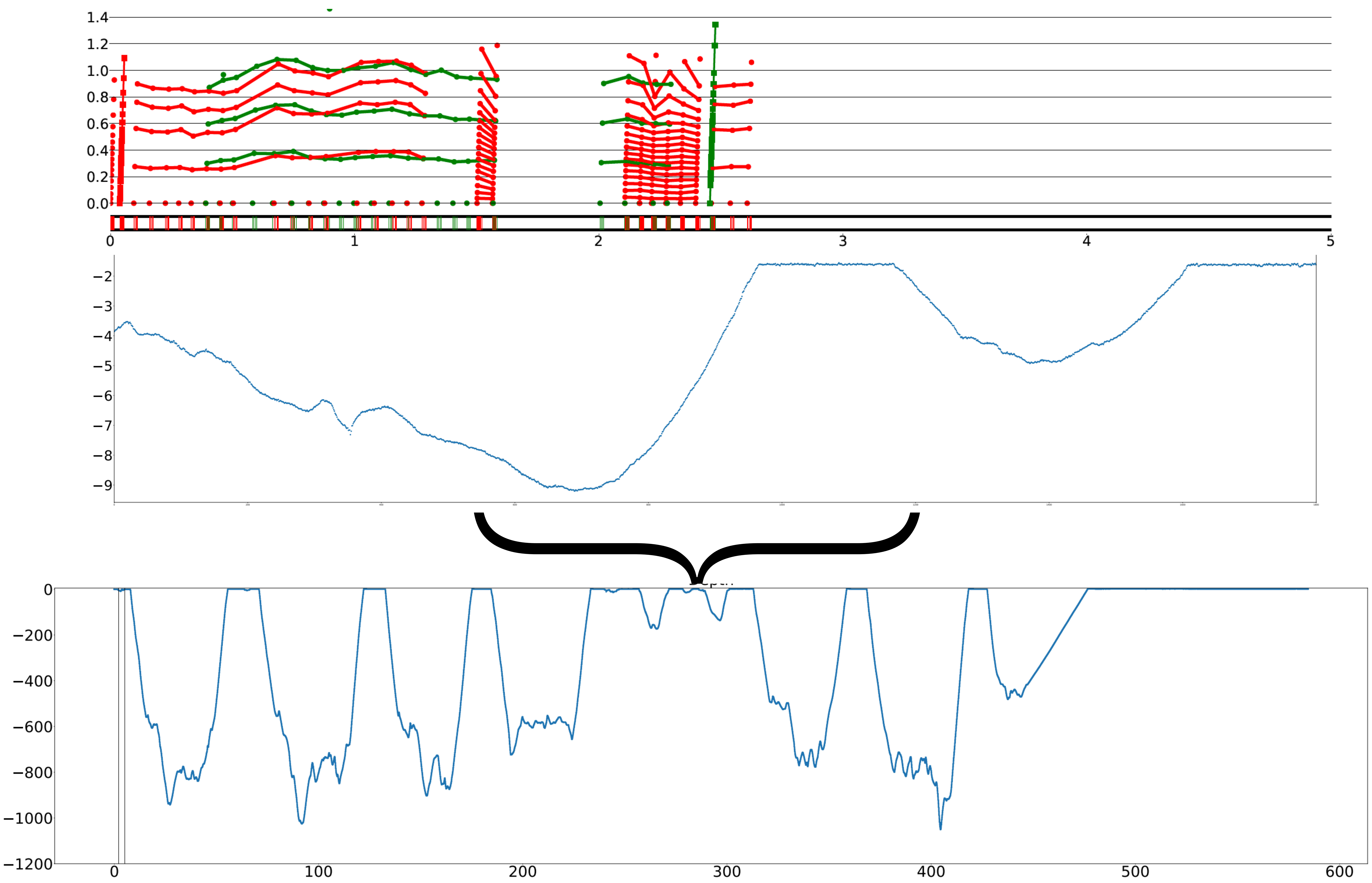
Meta-data



Meta-data

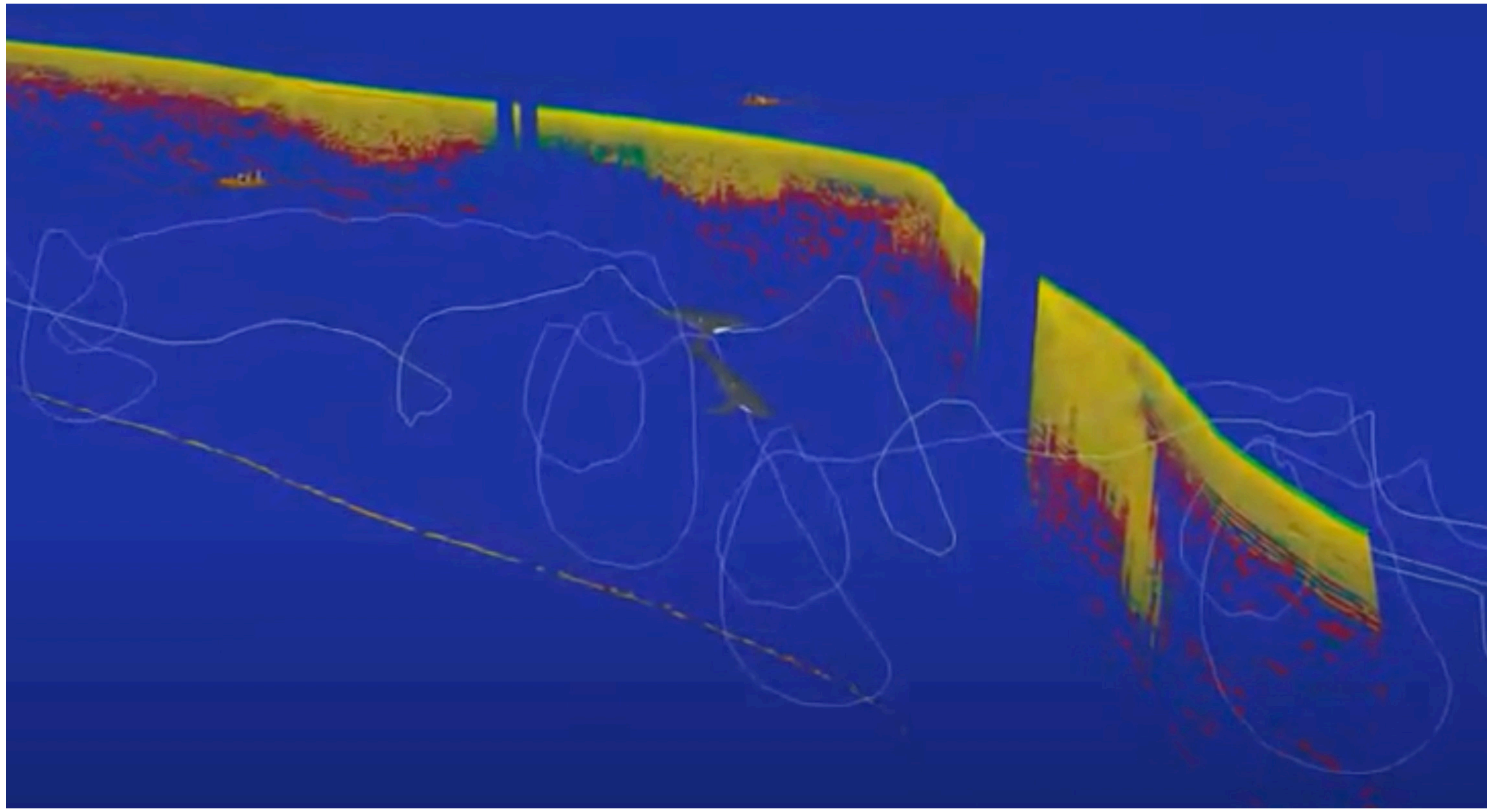


Meta-data

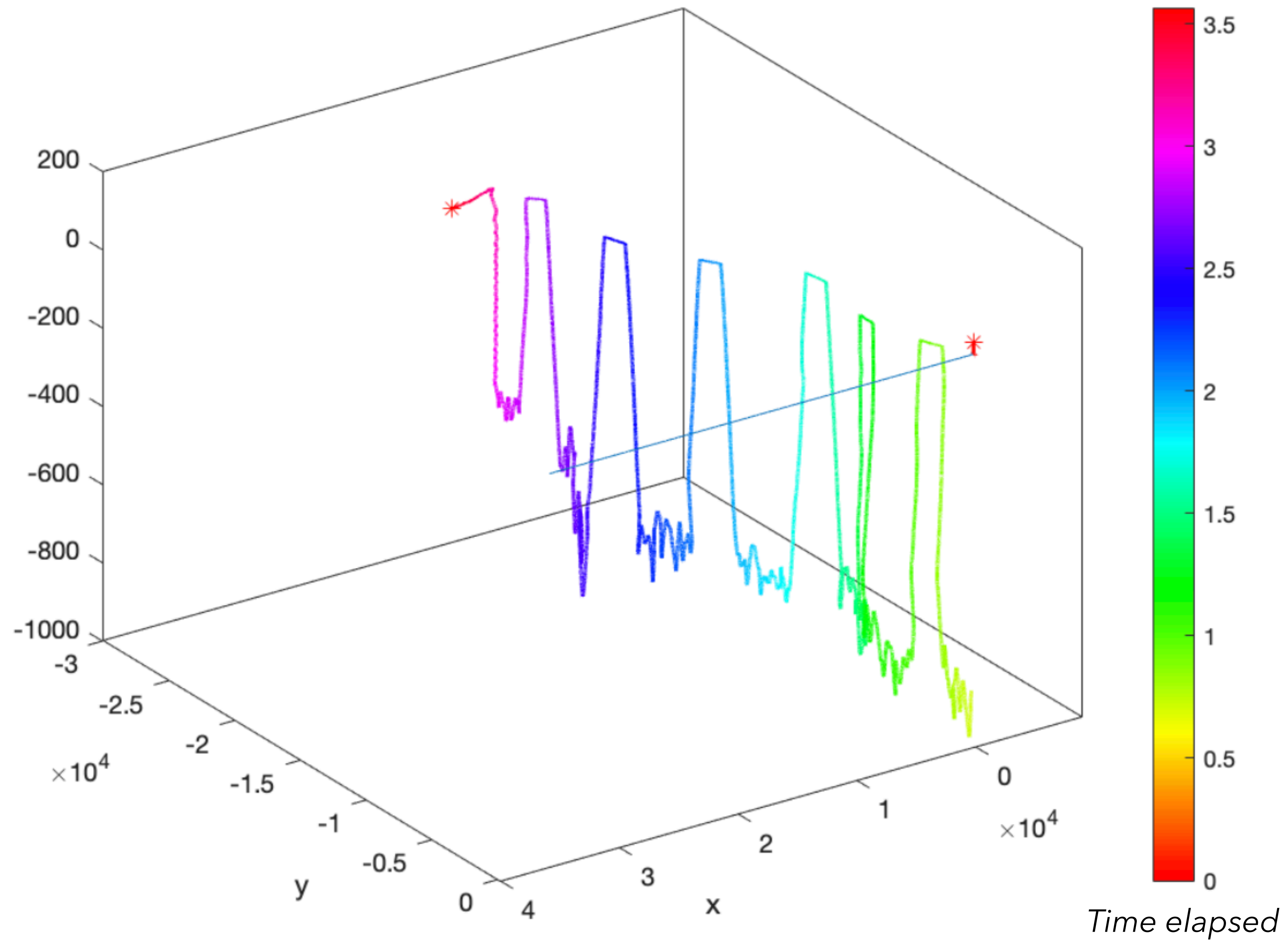


1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
-2
-3
-4
-5
-6
-7
-8
-9
0
−200
−400
−600
−800
−1000
−1200
0
100
200
300
400
500
600

Meta Data



Source : GeoZui4D, Data and Visualization Research Lab, University of New Hampshire

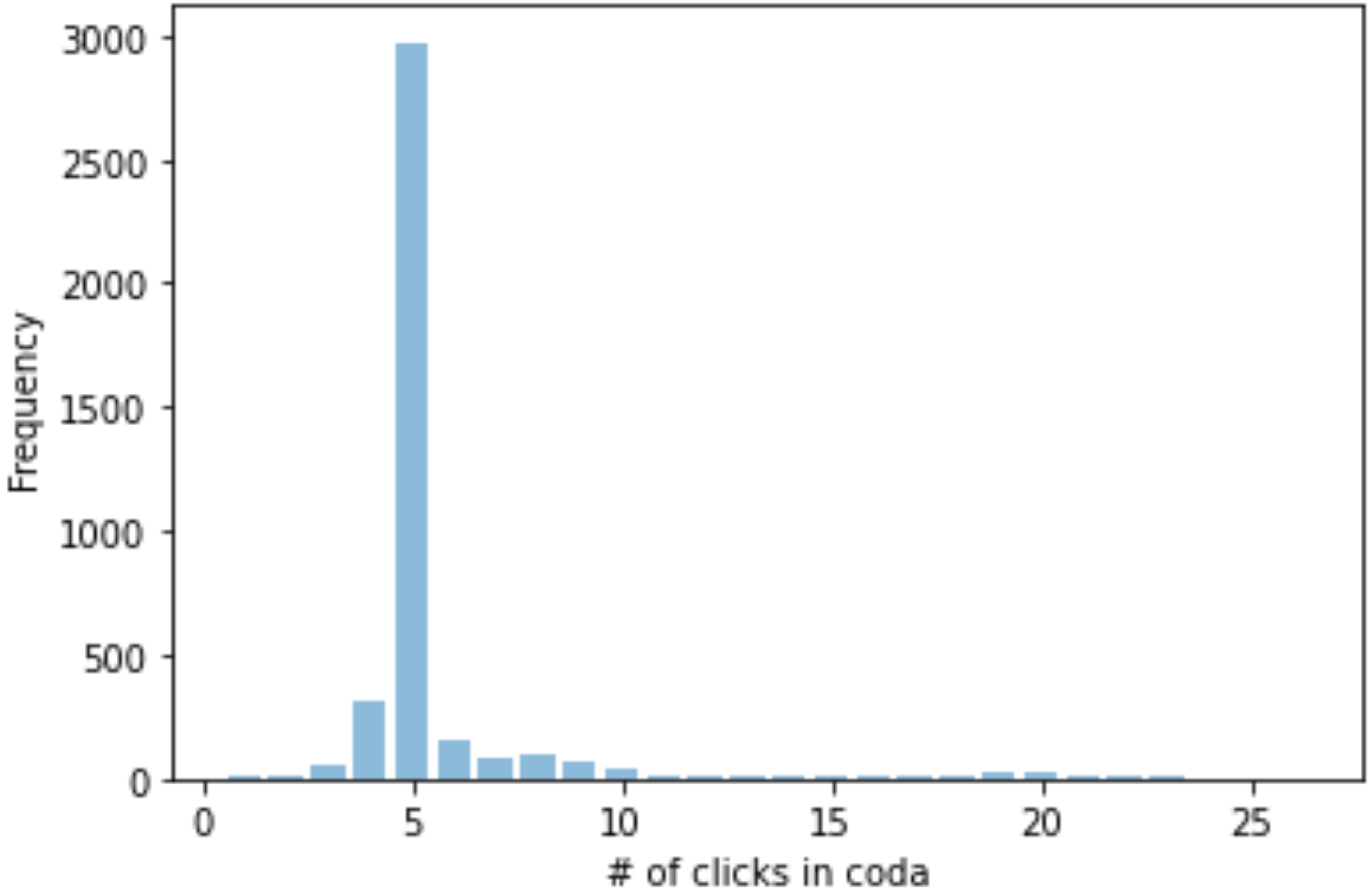


Trajectory of the whale

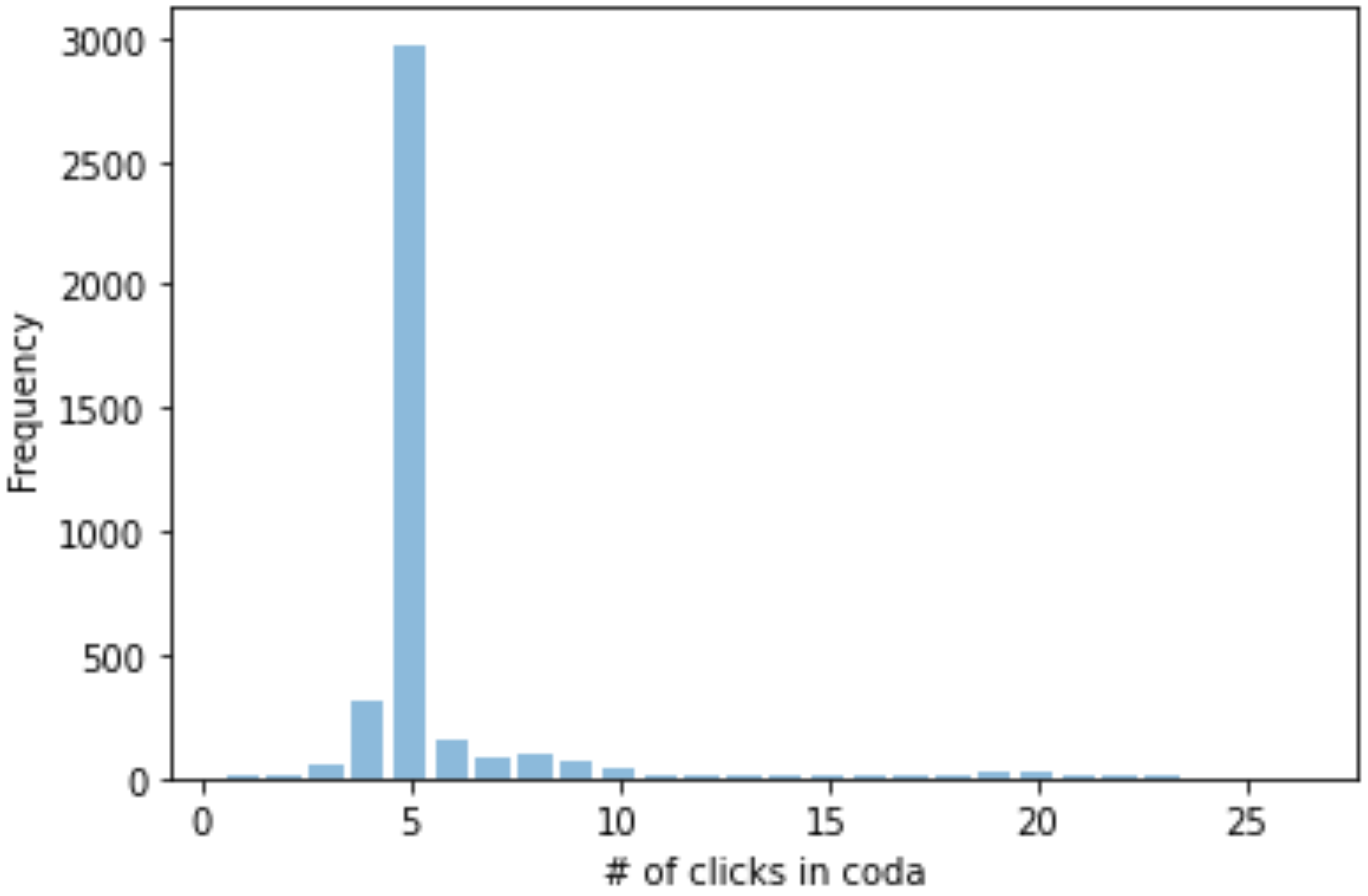
We want to be able to contextualize the vocalizations with the behavior.

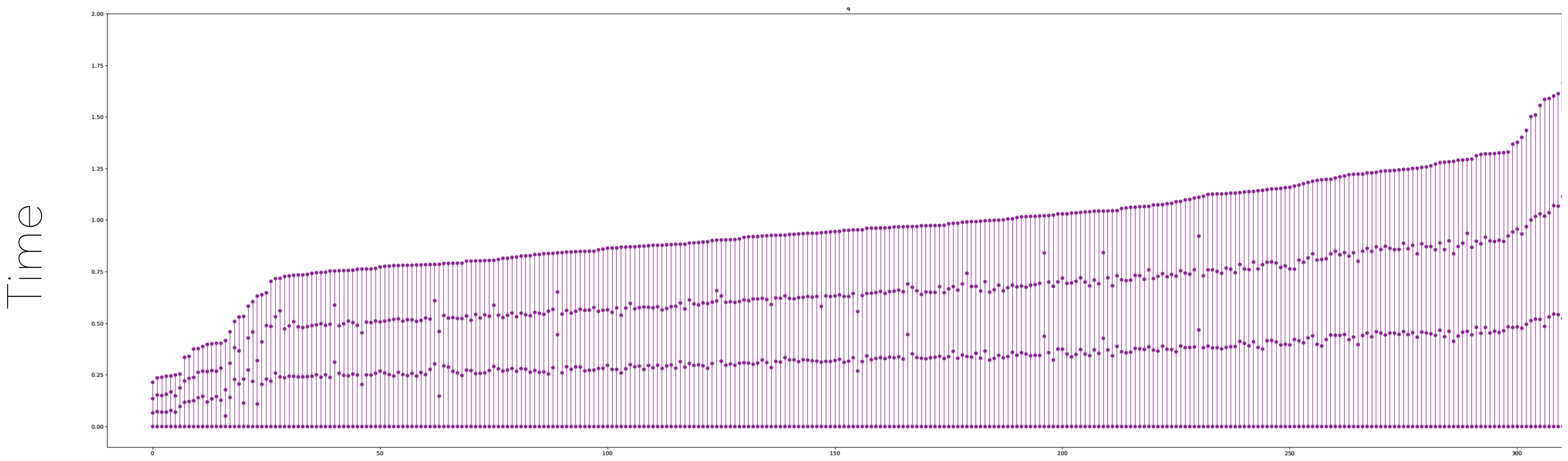
What are the differences between different codas?

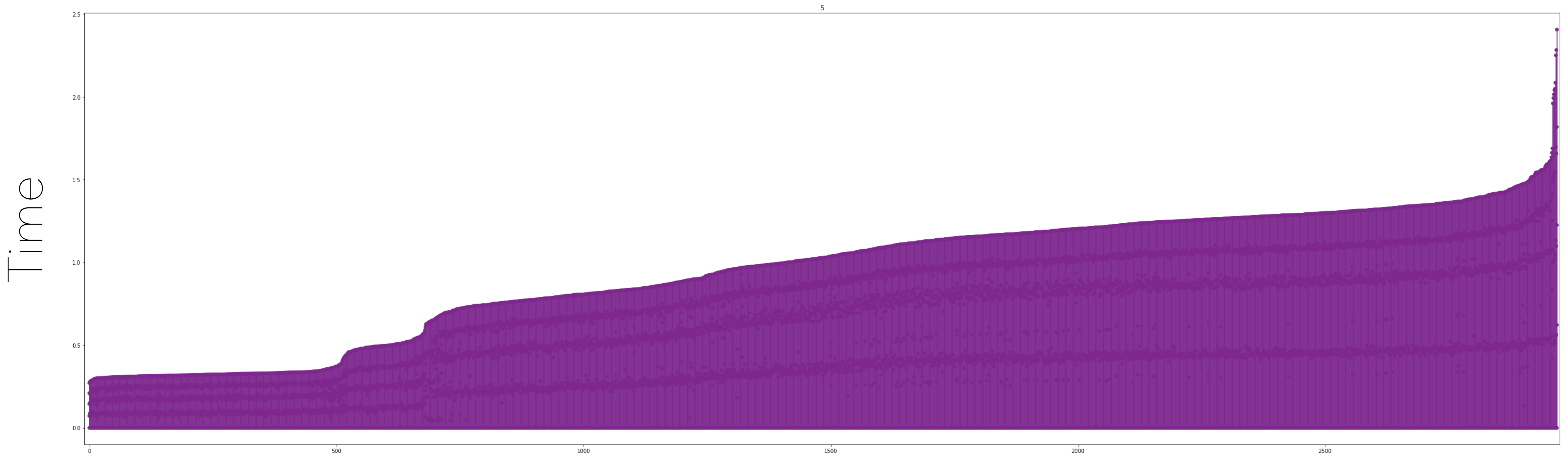
Variation in the Codas

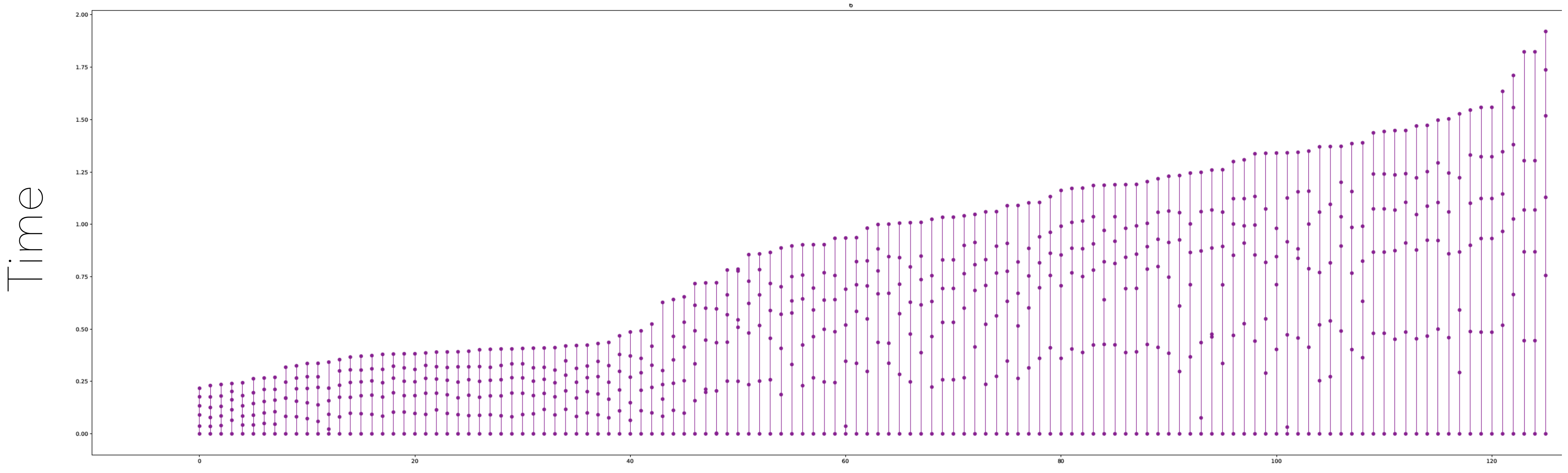


Variation in the Codas







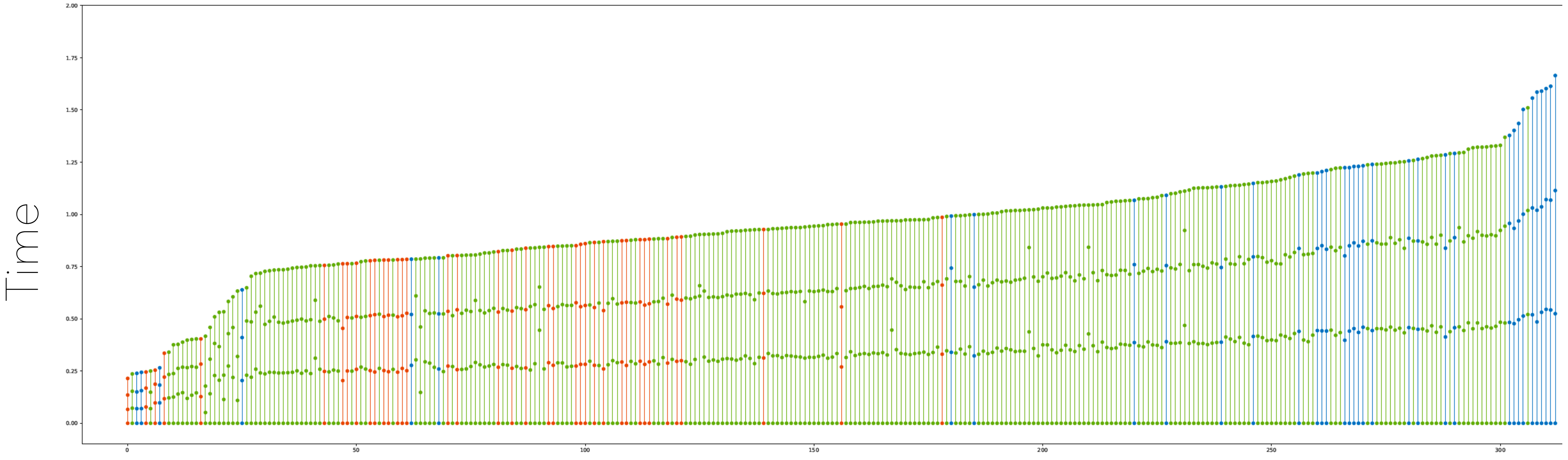


Variation in the Codas

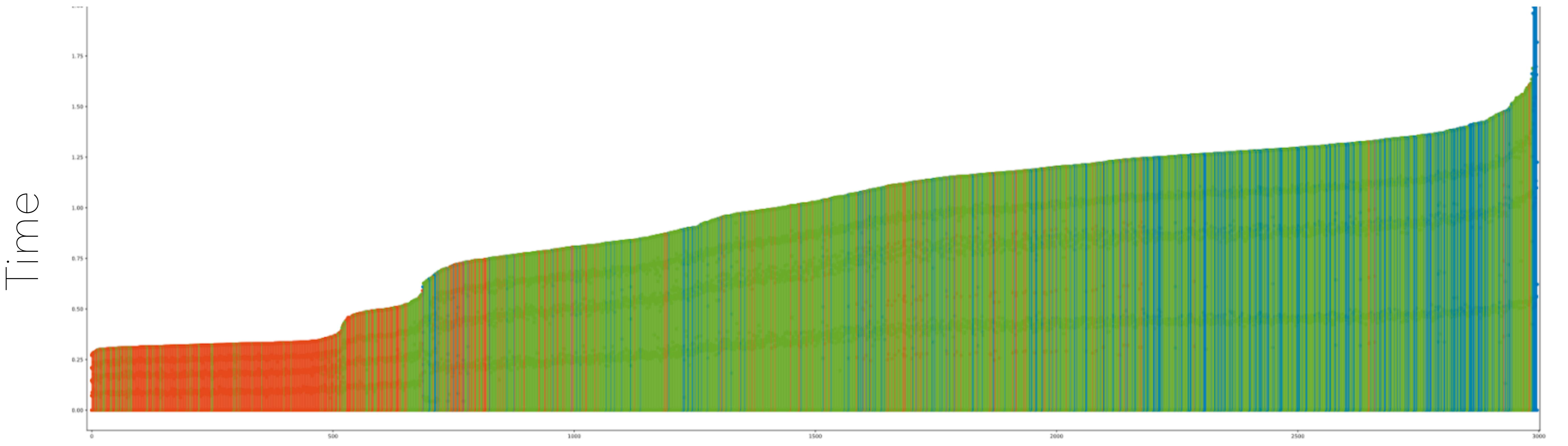
🟥 -> Diving ascent

🟩 -> codas on the surface

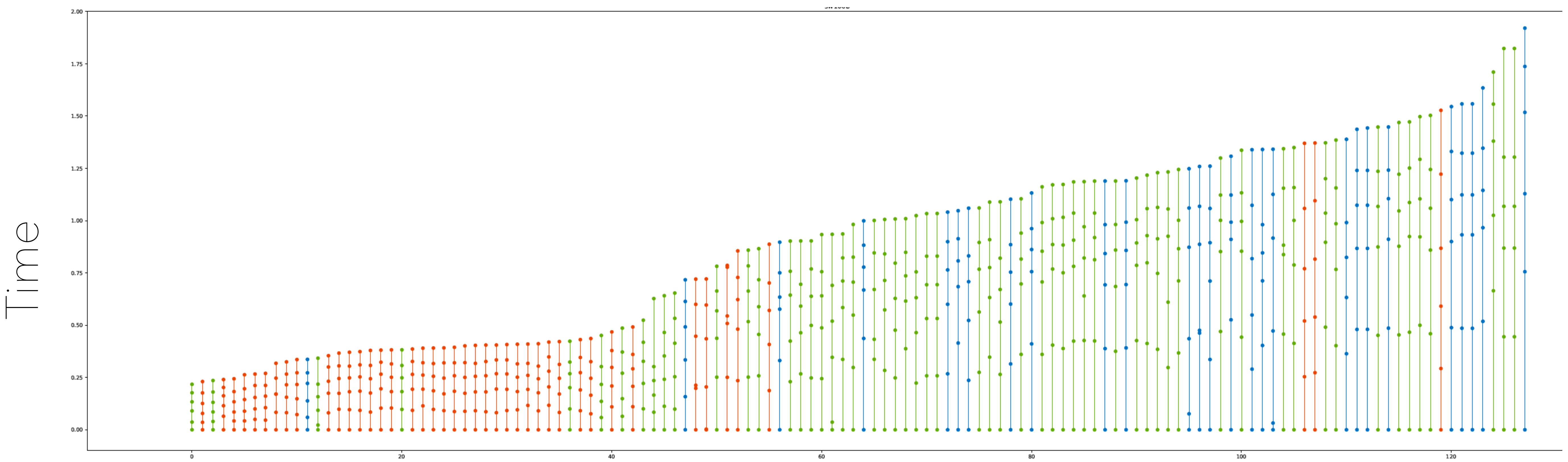
🟦 -> Codas on diving descent



4 click codas



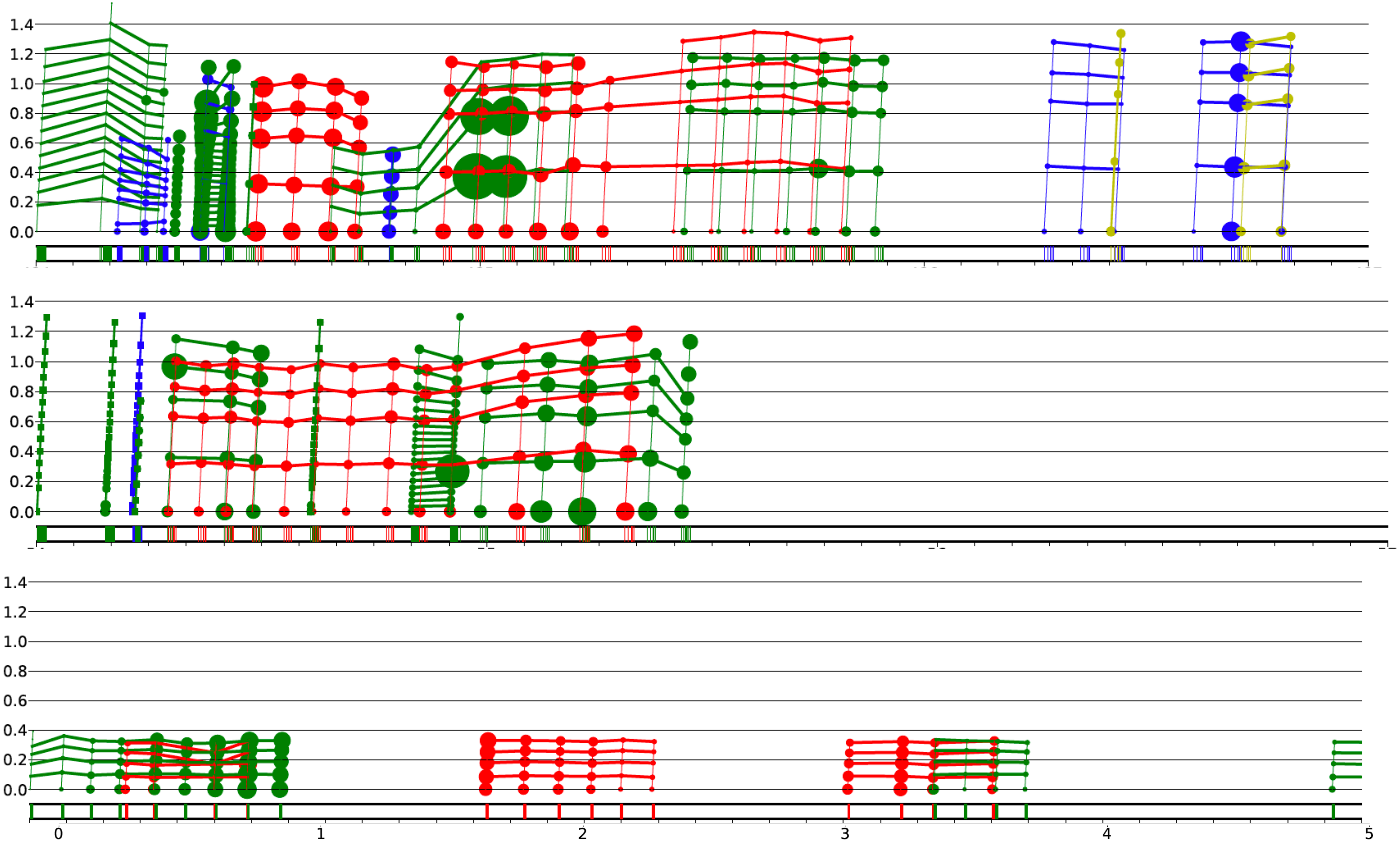
5 click codas



6 click codas

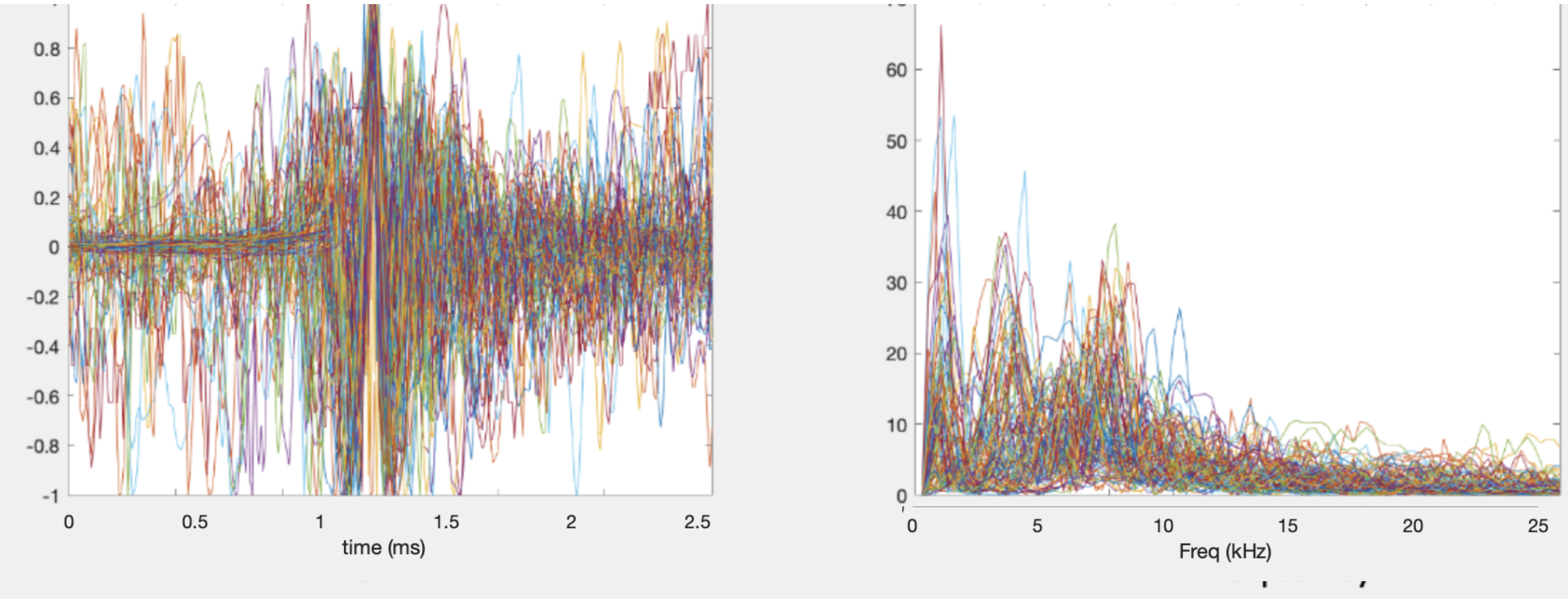
Is when the click starts all that there is to a click?

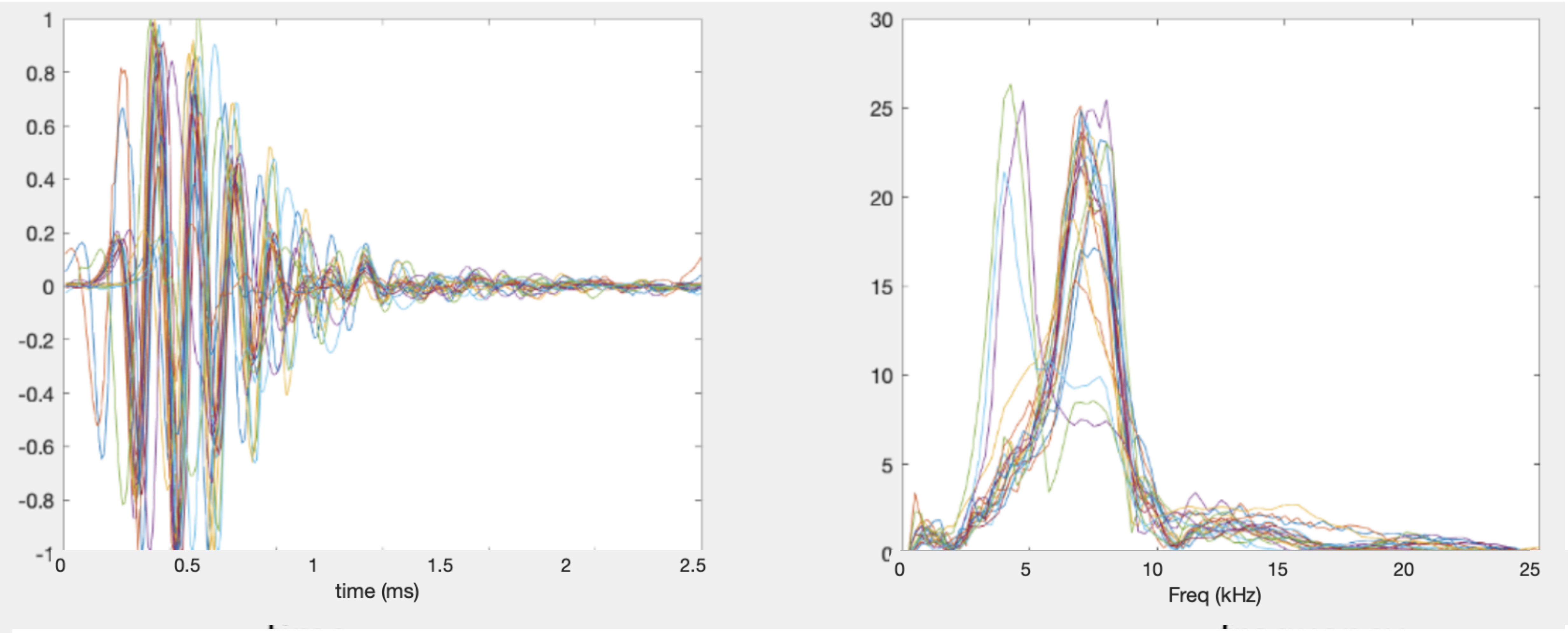
Power

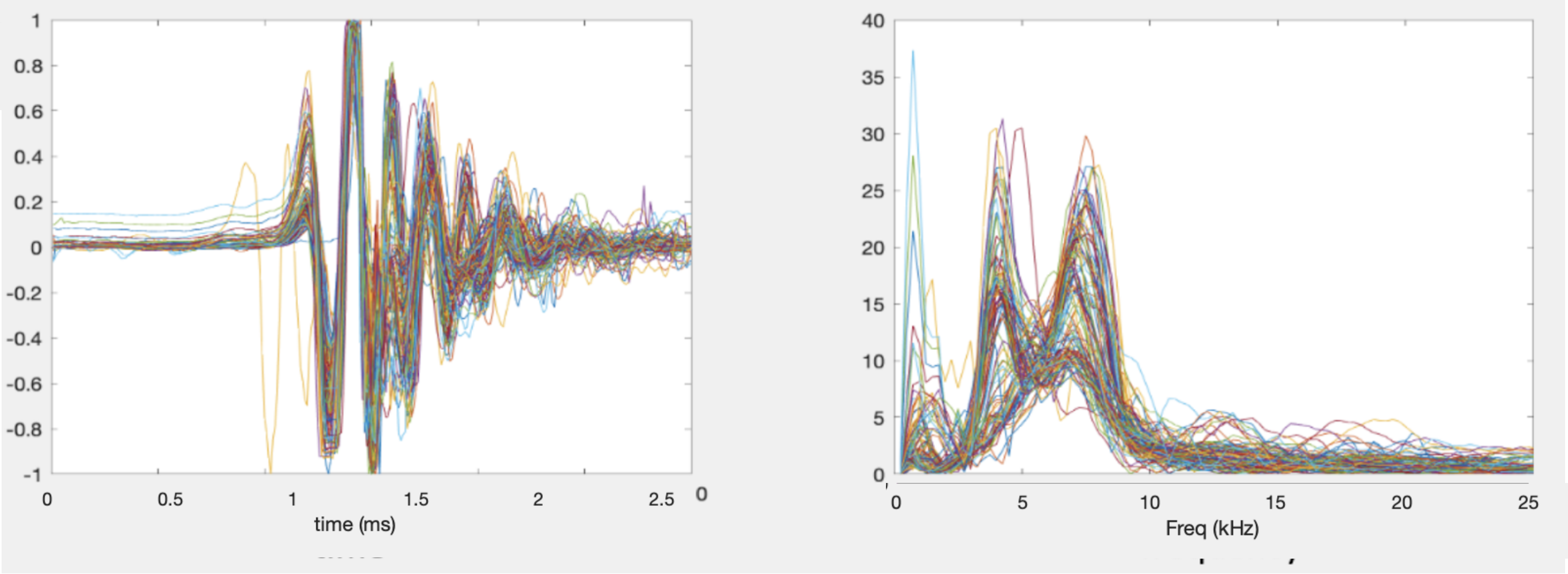


Radius of the circle ∝ power of the click

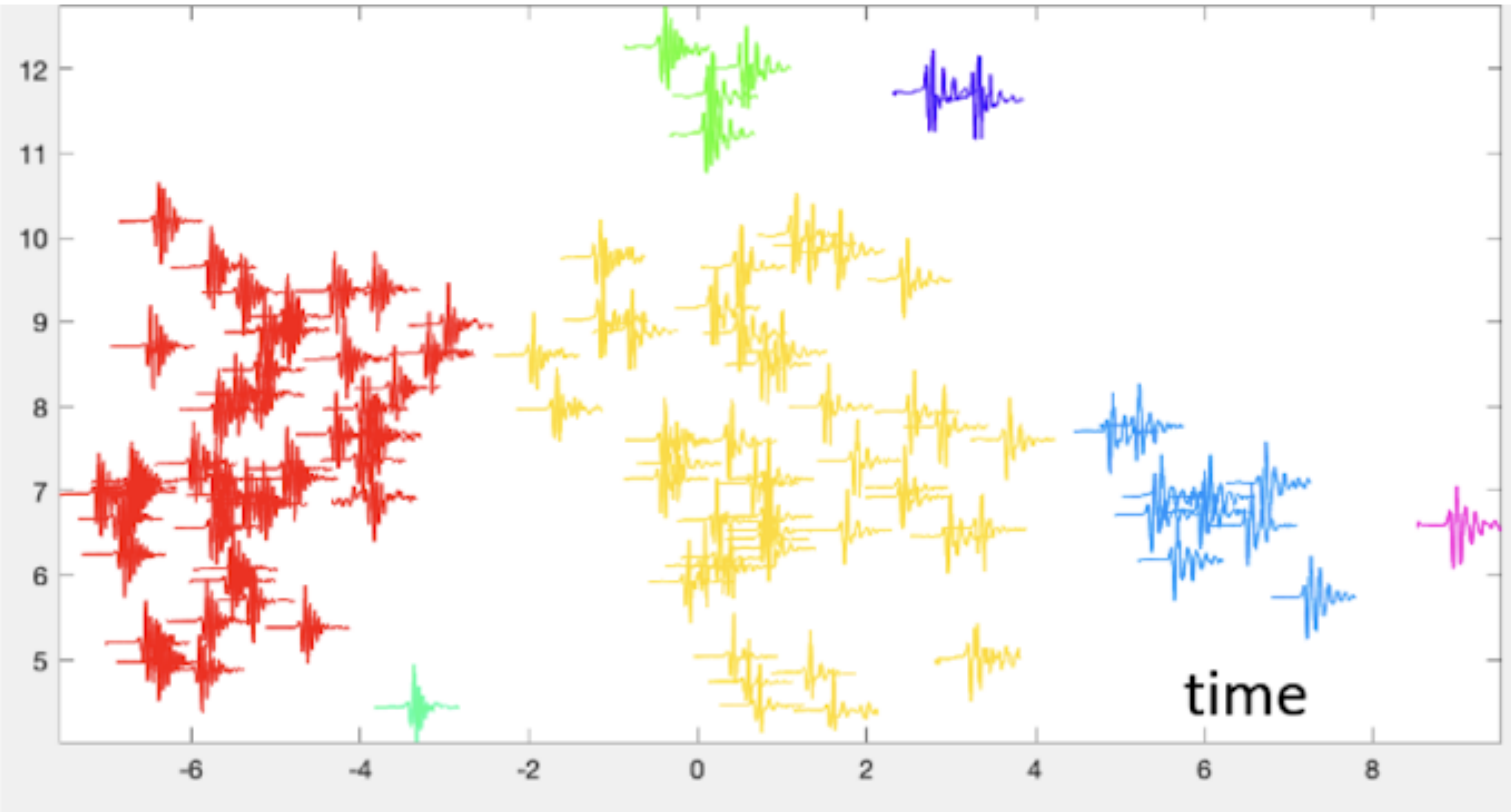
Other voice cues?

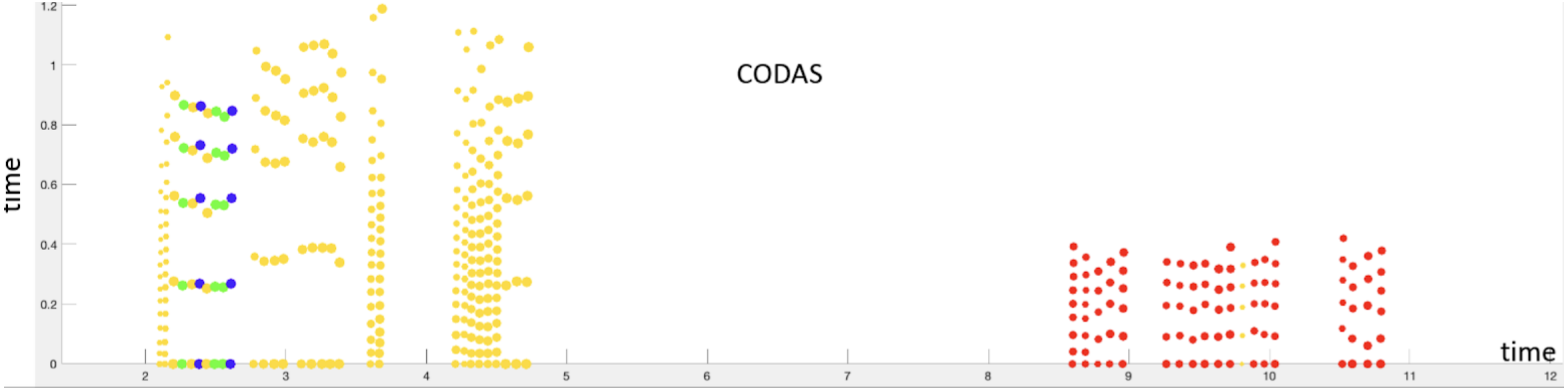




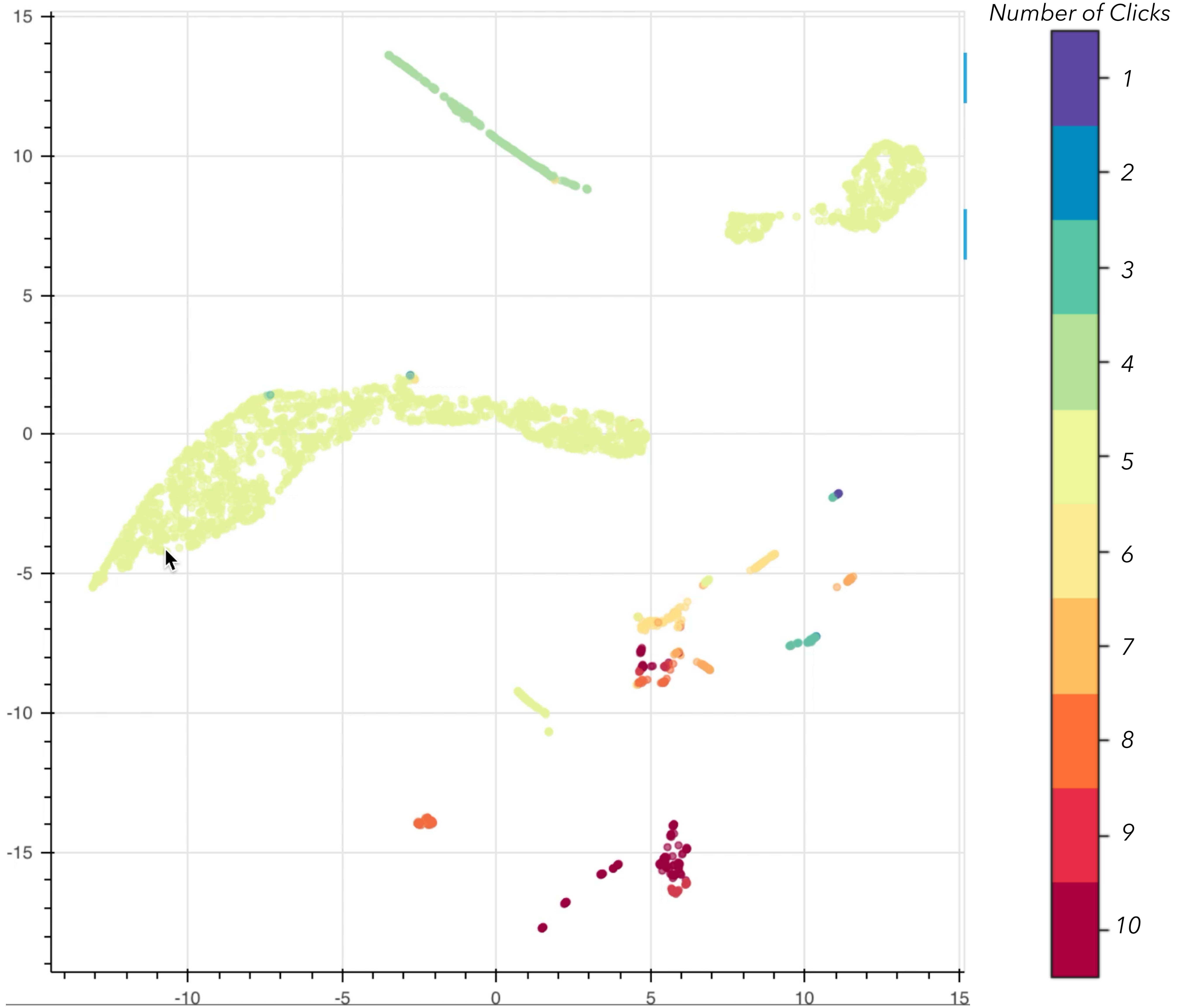


Other voice cues?





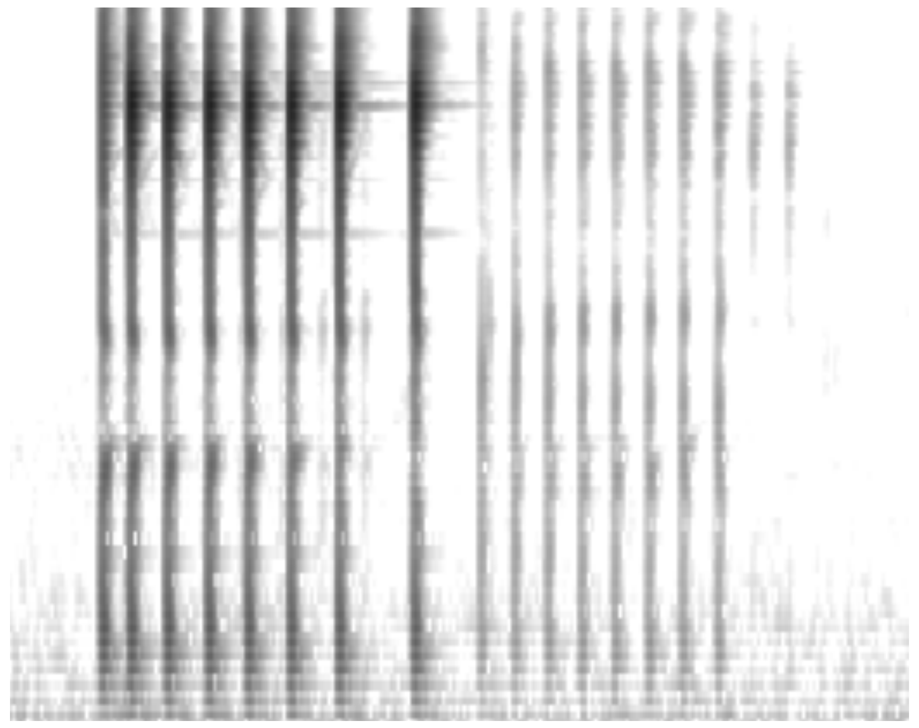
Structure of Codas



What is the smallest # discrete units that may explain the data distribution the best?

Representation

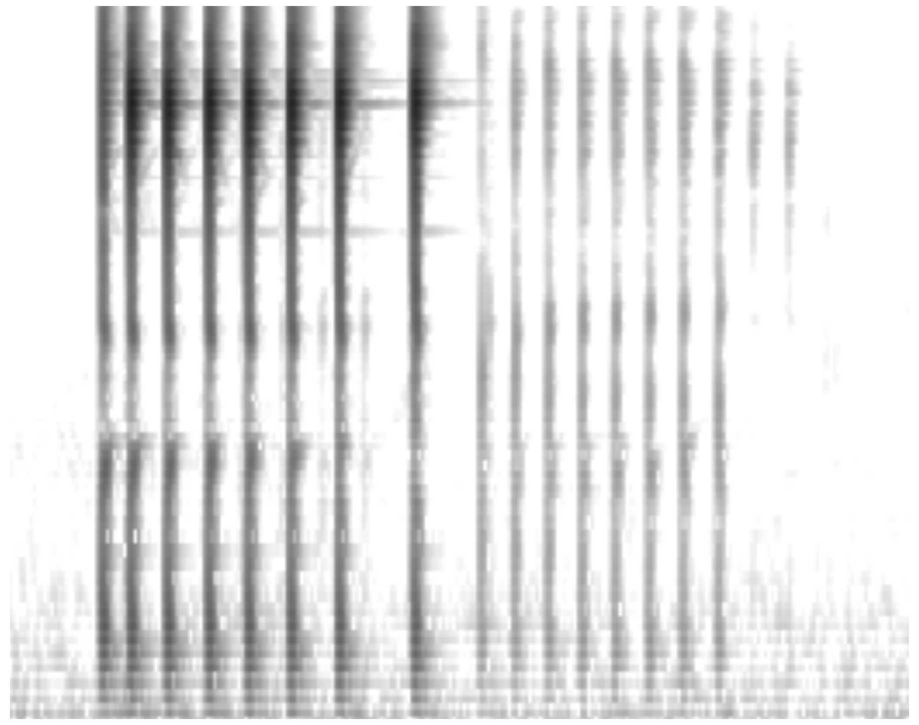
How should we represent the vocalizations?



Attributes: [which whale?,what coda? At what depth?, How loud?, Which direction was it facing?, at what time?]

Representation

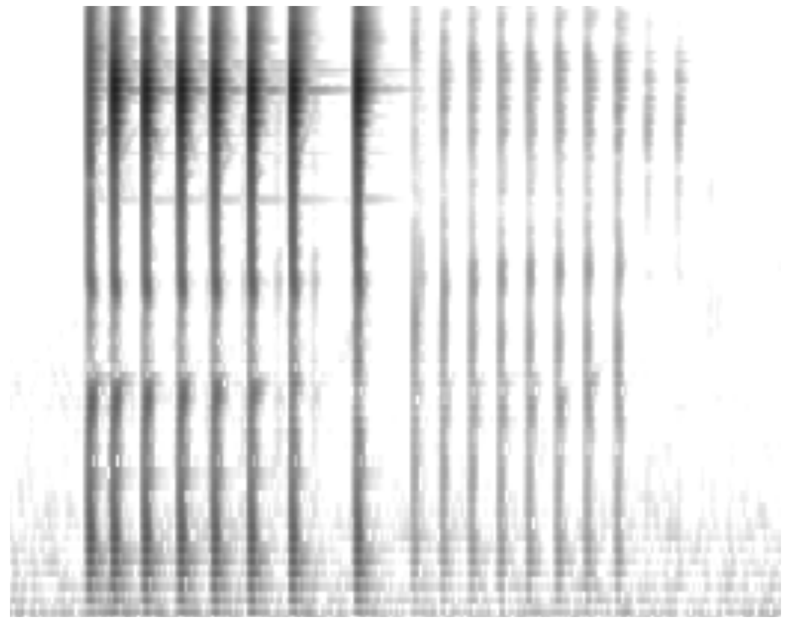
How should we represent the vocalizations?



Attributes: [which whale?,what coda? At what depth?, How loud?, Which direction was it facing?, at what time?]

Continuous like Music?

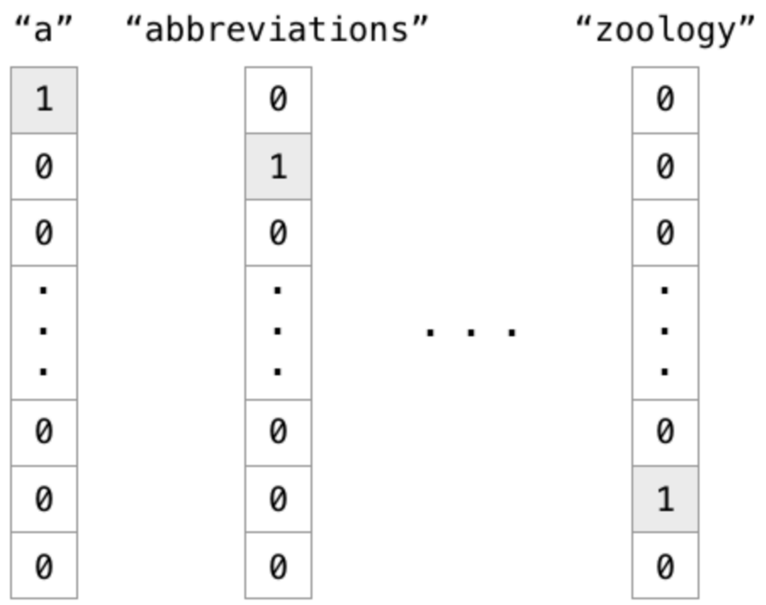
Option1



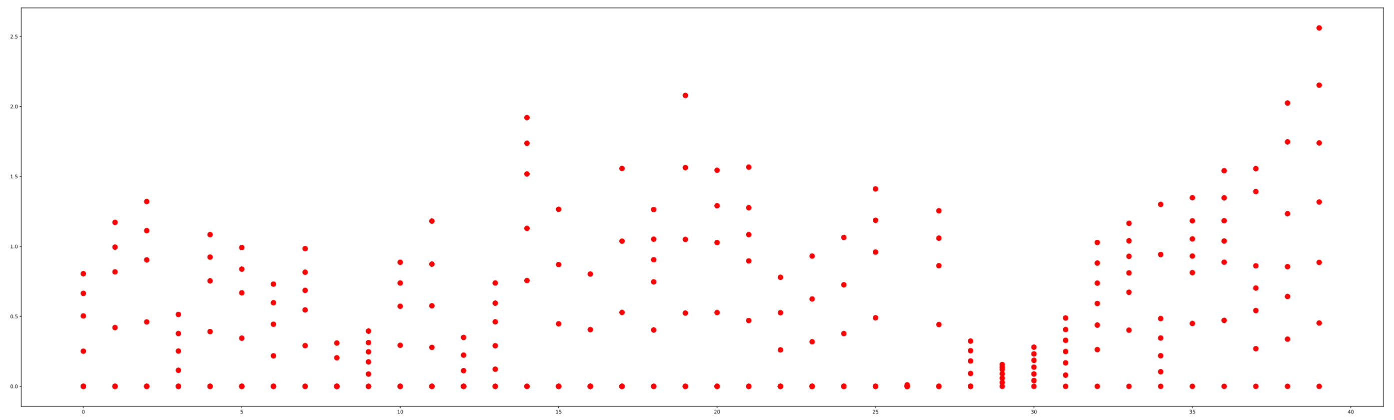
Coda: [8D] :
[ICI1,ICI2,ICI3…,IC7,1/0]

Discrete like Language?

Option2

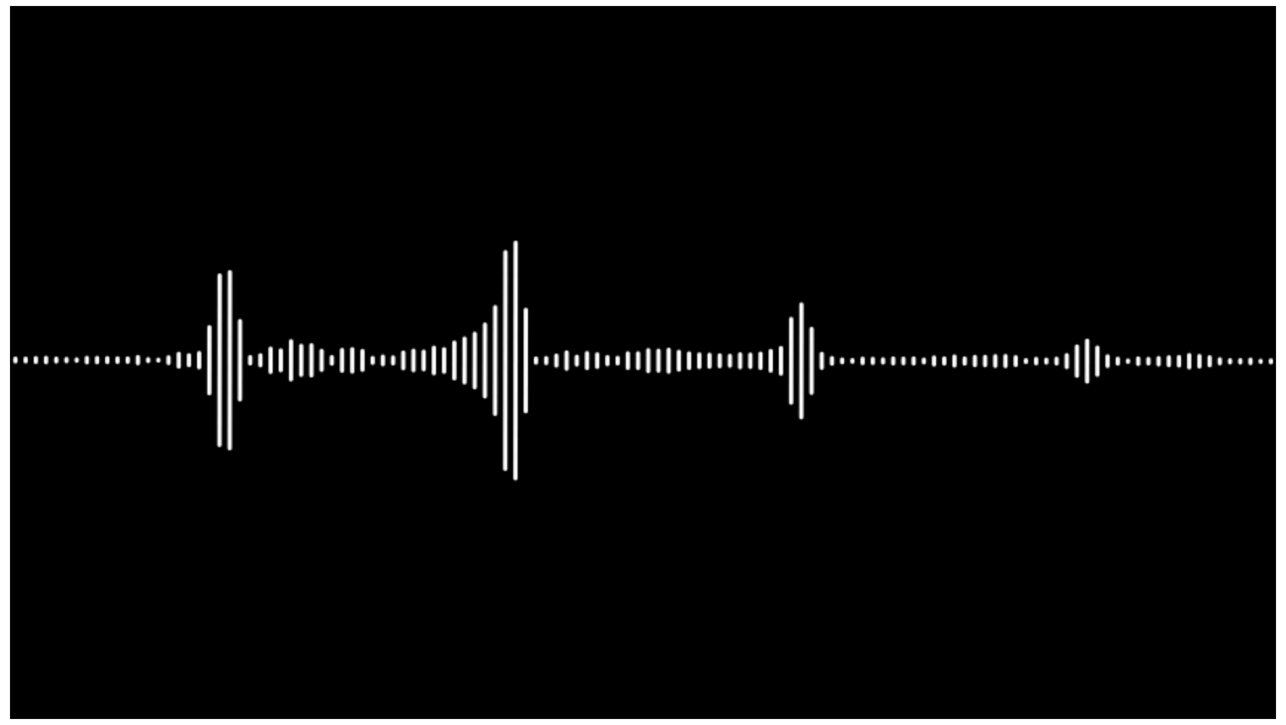


"a"	"abbreviations"		"zoology"
1	0		0
0	1		0
0	0		0
.	.		.
.	.	. . .	.
.	.		.
0	0		0
0	0		1
0	0		0



Cluster: 1 , 2 , 3, 4 …,*n*

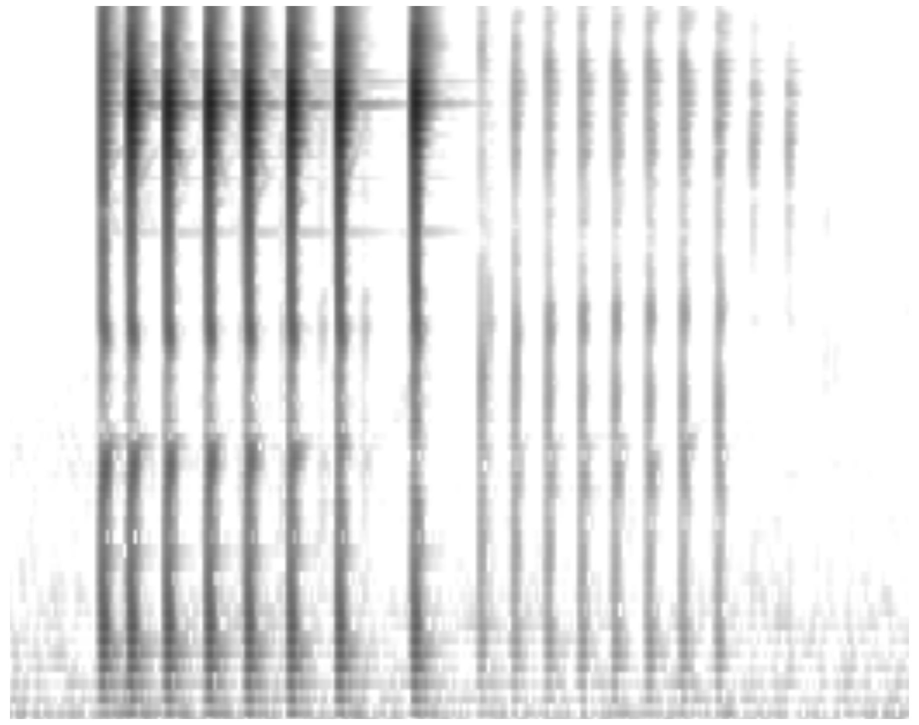
Representation



0001000100001000001000

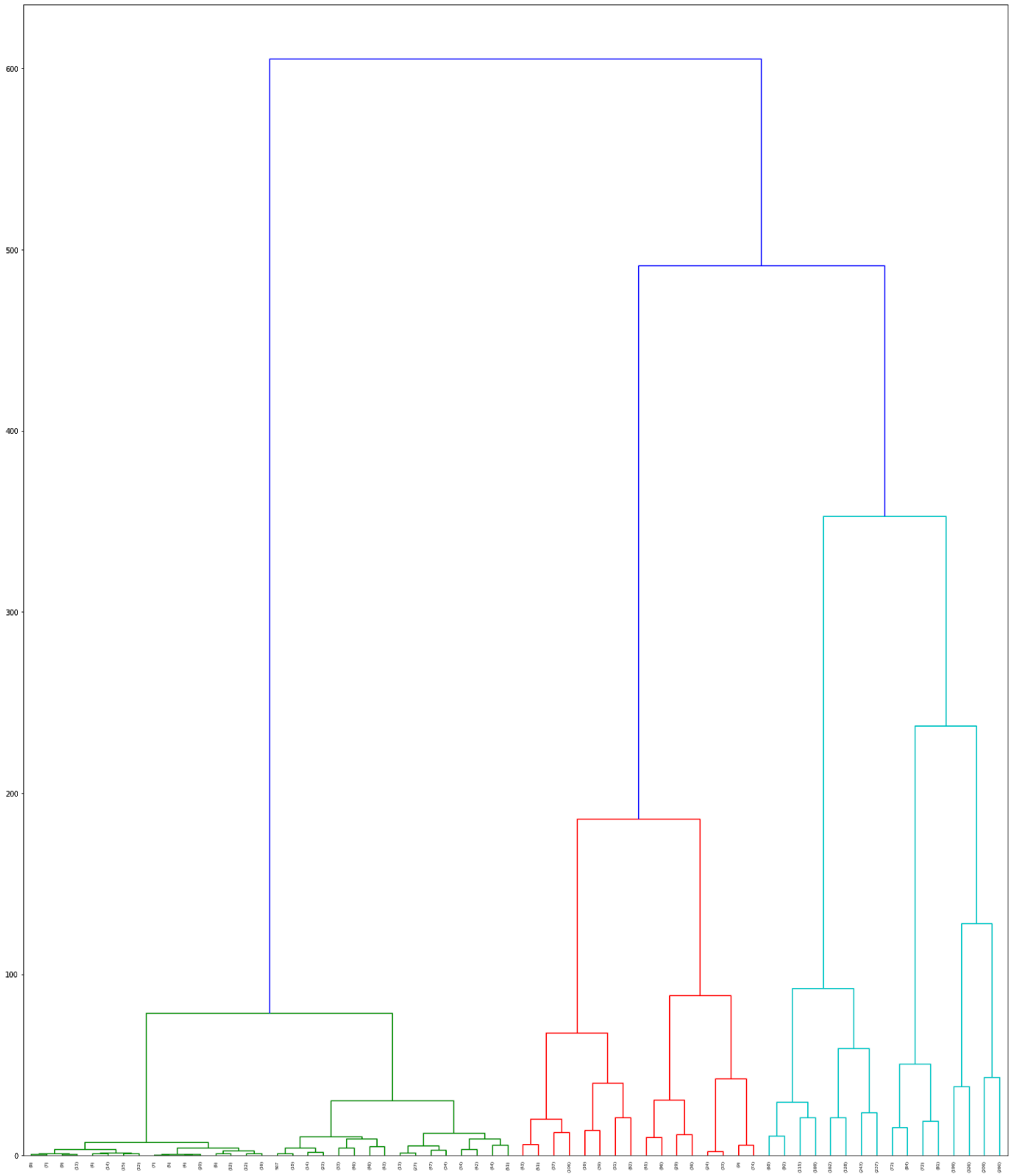
Lower dimensionality better!

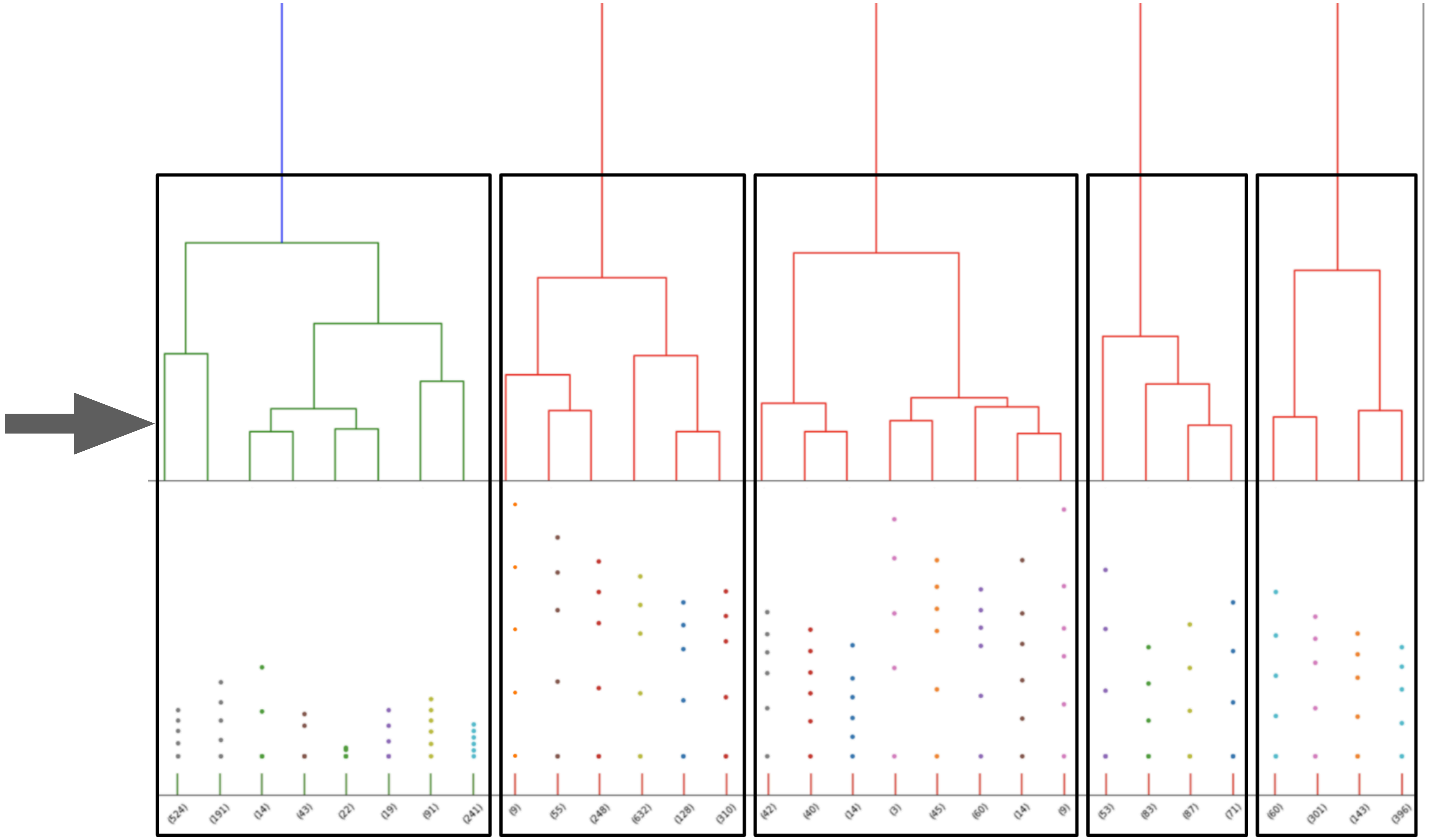
More information better!



Attributes: [which whale?,what coda? At what depth?, How loud?, Which direction was it facing?, at what time?]

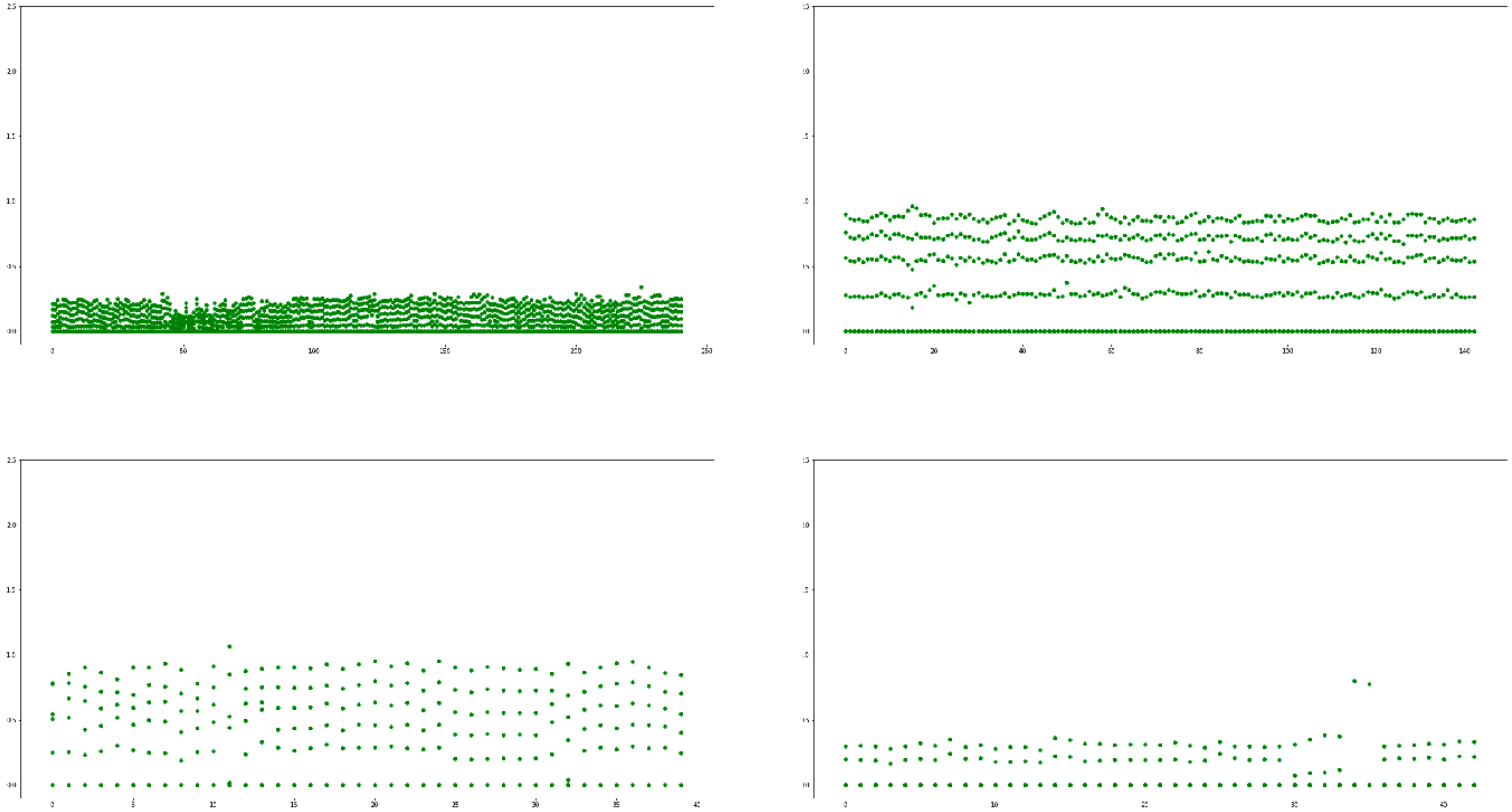
Clustering



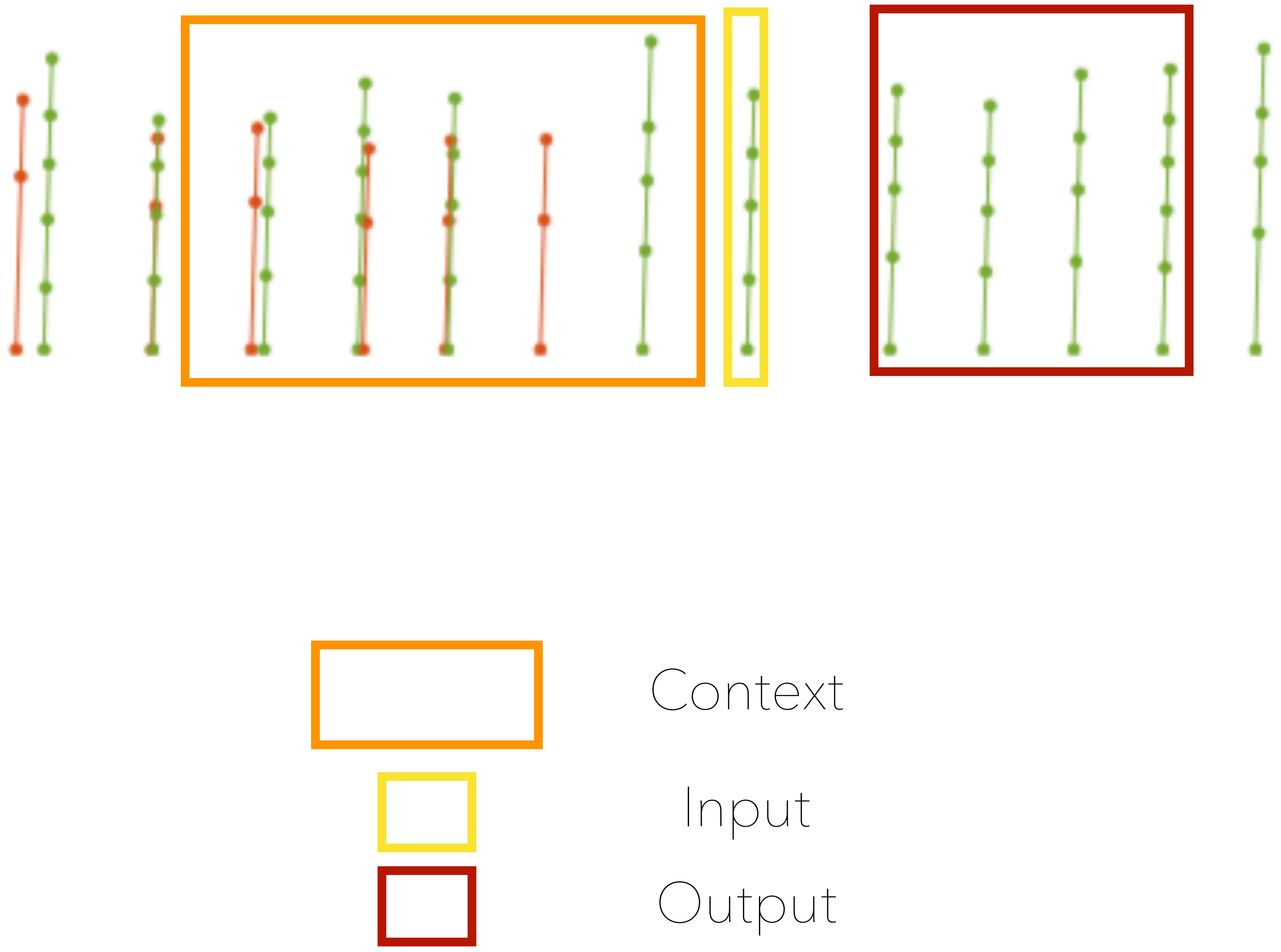


Hierarchical Agglomerative Clusters of CODAs

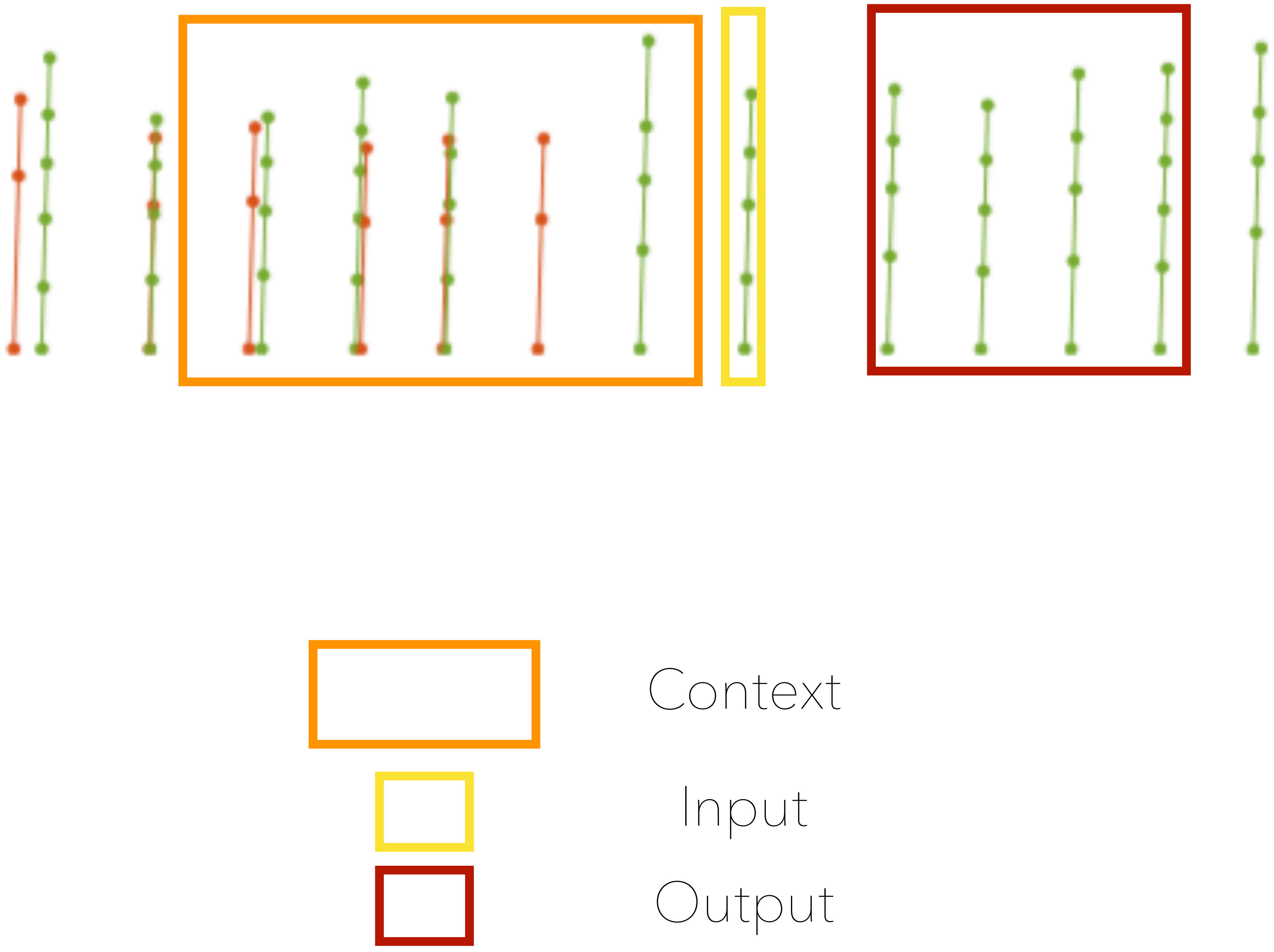
Variability within clusters

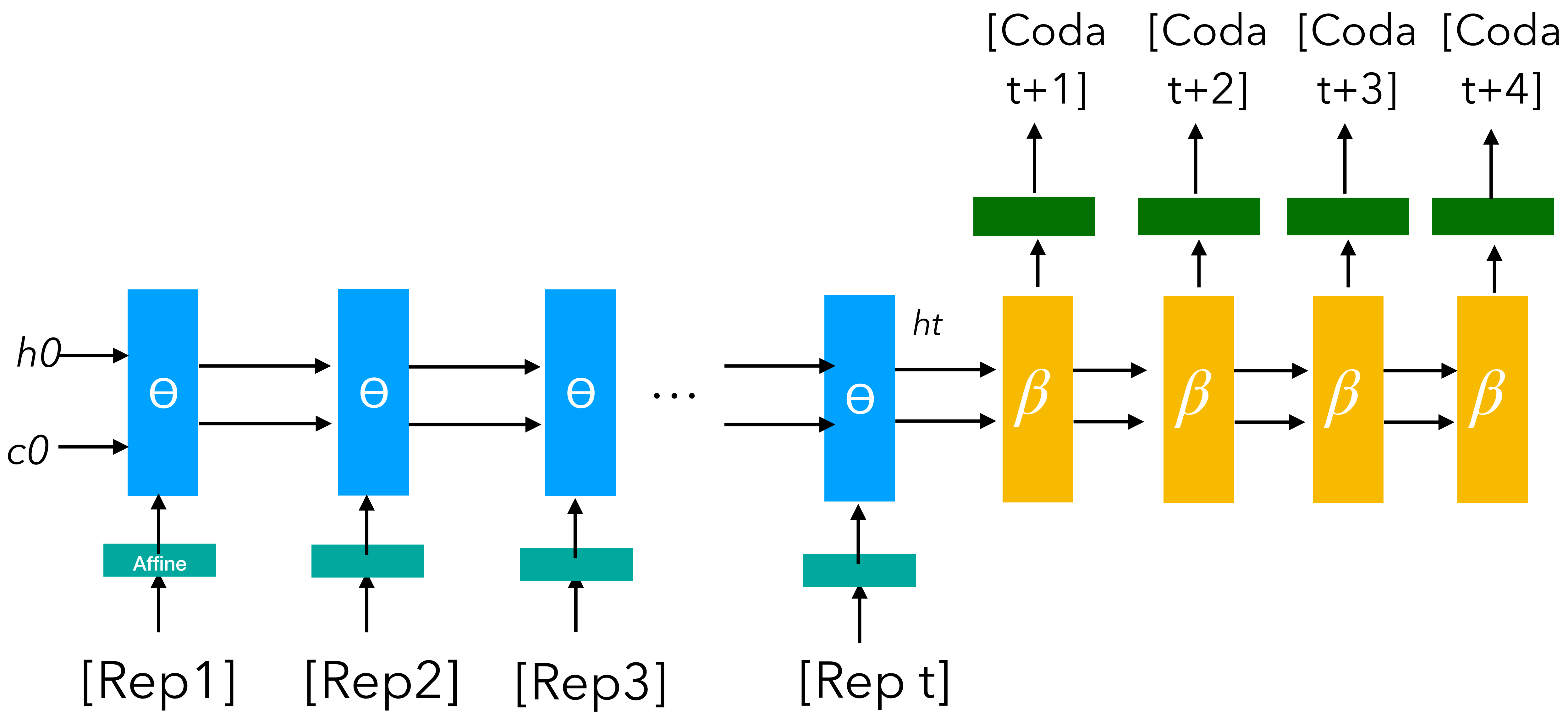


Can we build a model that can predict the vocalizations?



Can we build a model that can predict the vocalizations?





Coda:[Rep]: [Whale ID, start time,ICI1,…,IC7,1/0, Power1,… Power7,Depth]

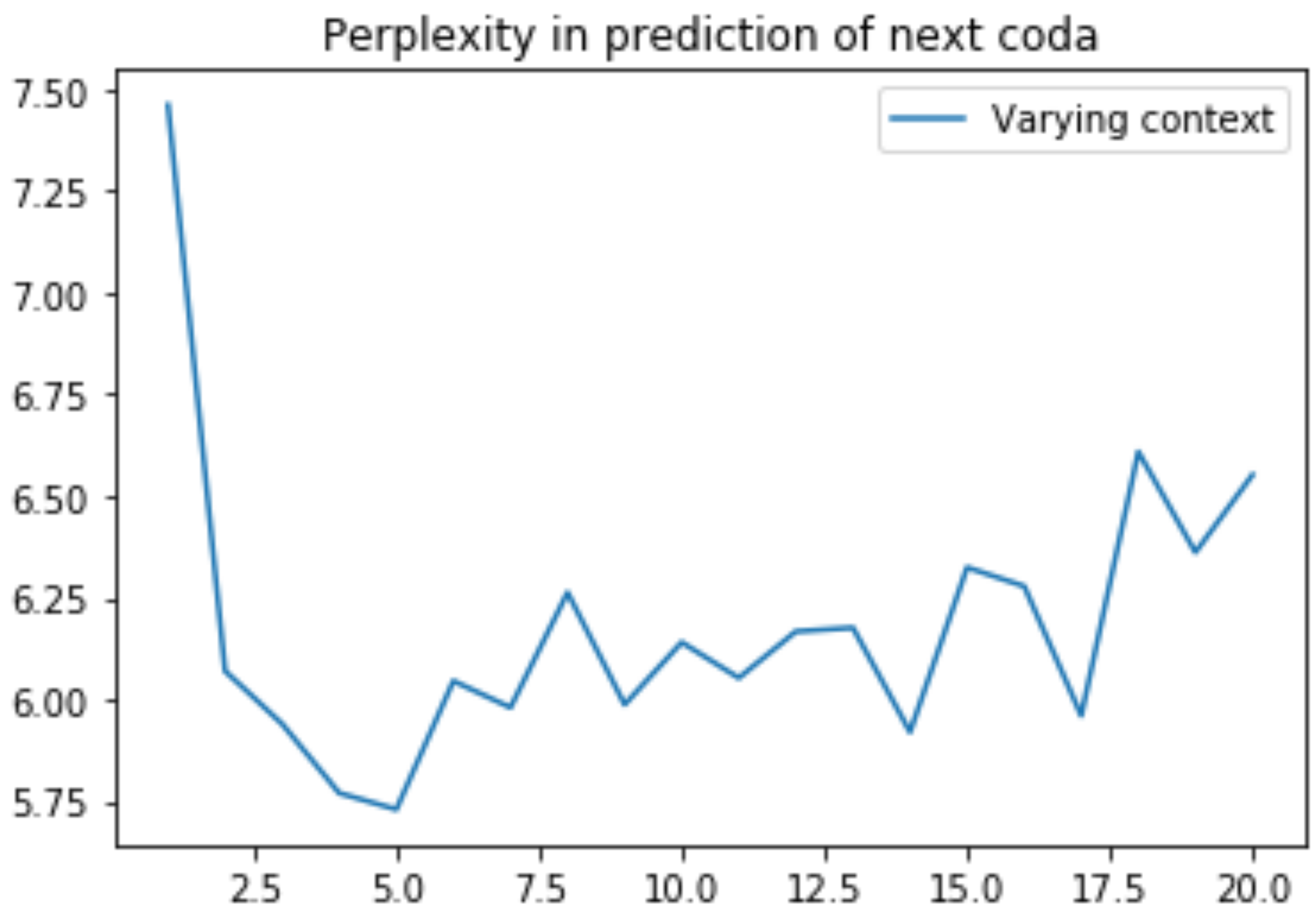
Evaluation

$$PP(X) = 2^{-\frac{1}{n}\log P(x_1, x_2, \ldots, x_n; \theta)}$$

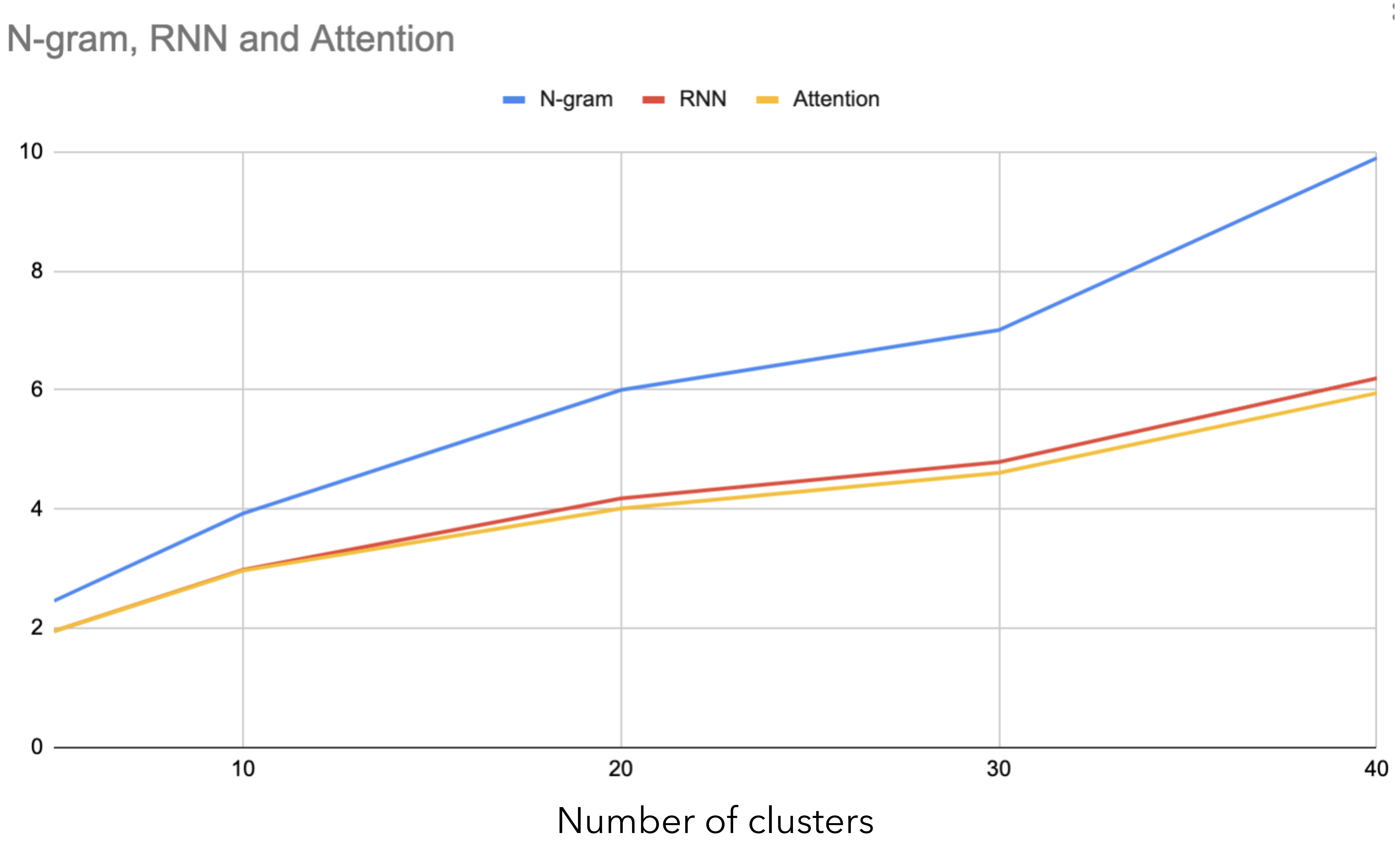
Perplexity: Inverse probability of the test set normalized by the number of words

Minimizing perplexity => Maximizing probability

Results

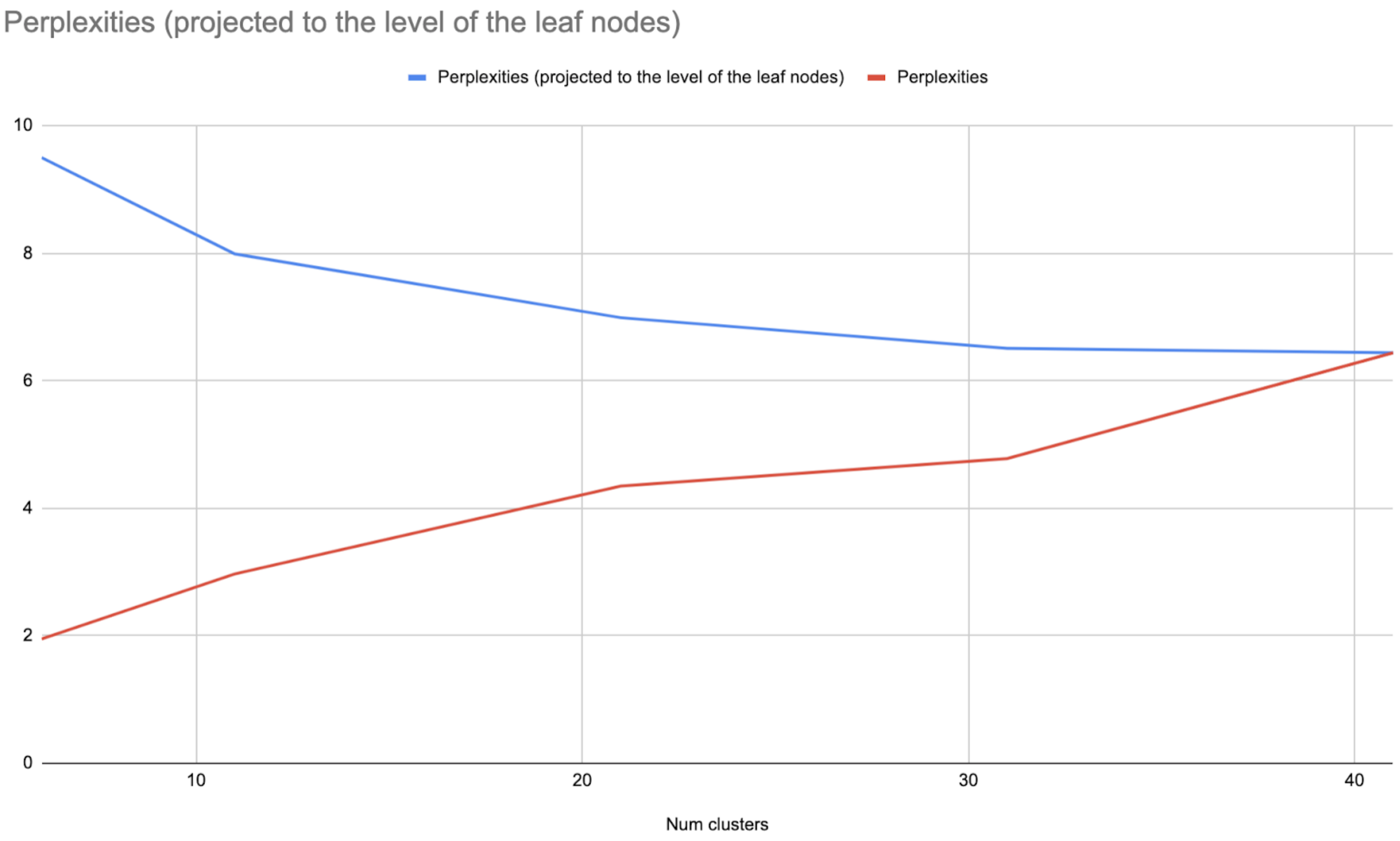


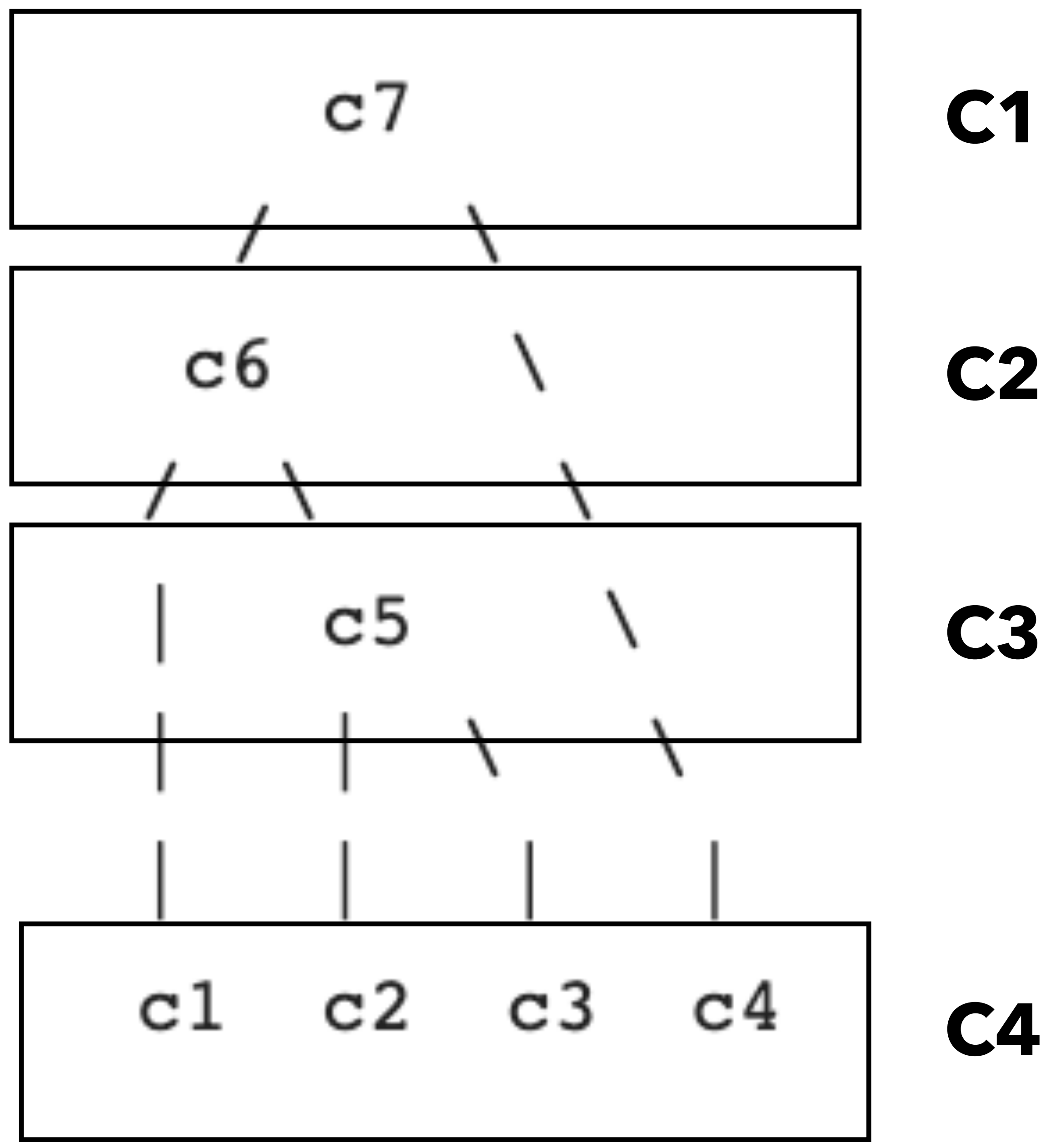
Affect of context size



Affect of model complexity

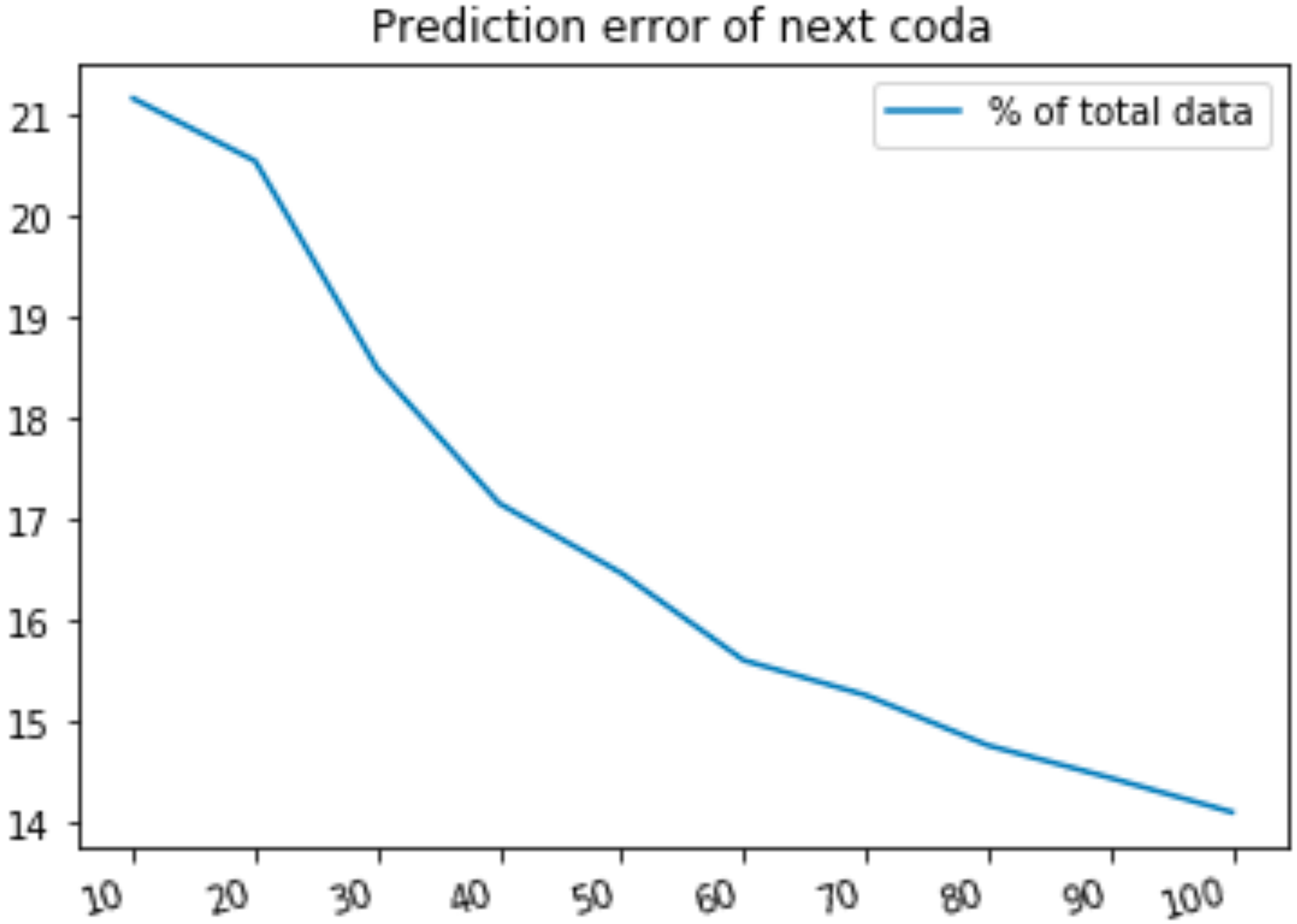
How do we decide how many clusters we should have?


Perplexities (projected to the level of the leaf nodes)



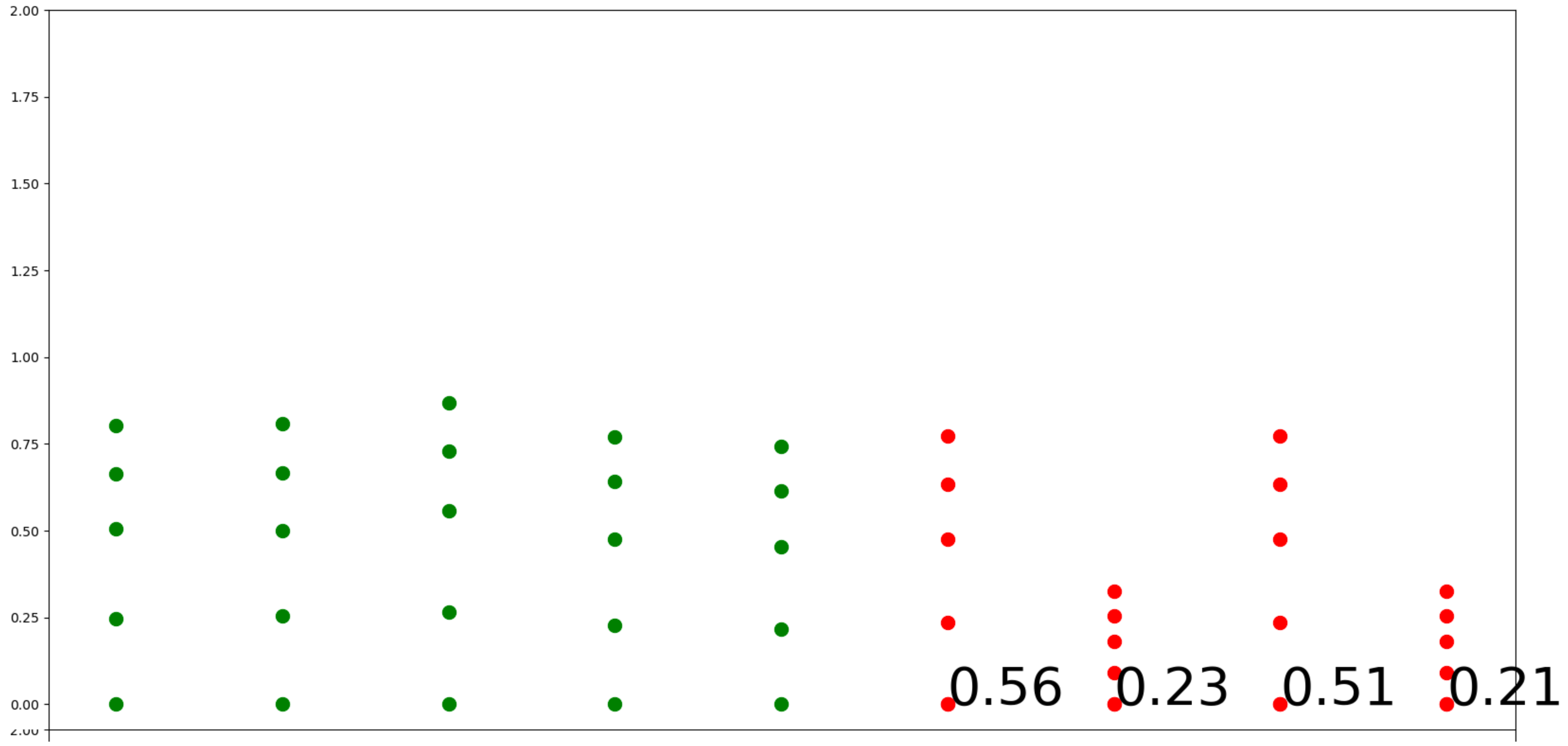
$$p(c3 \mid context) = p(c3 \mid C2) x p_m(c6 \mid context, \theta)$$

Affect of size of the training data

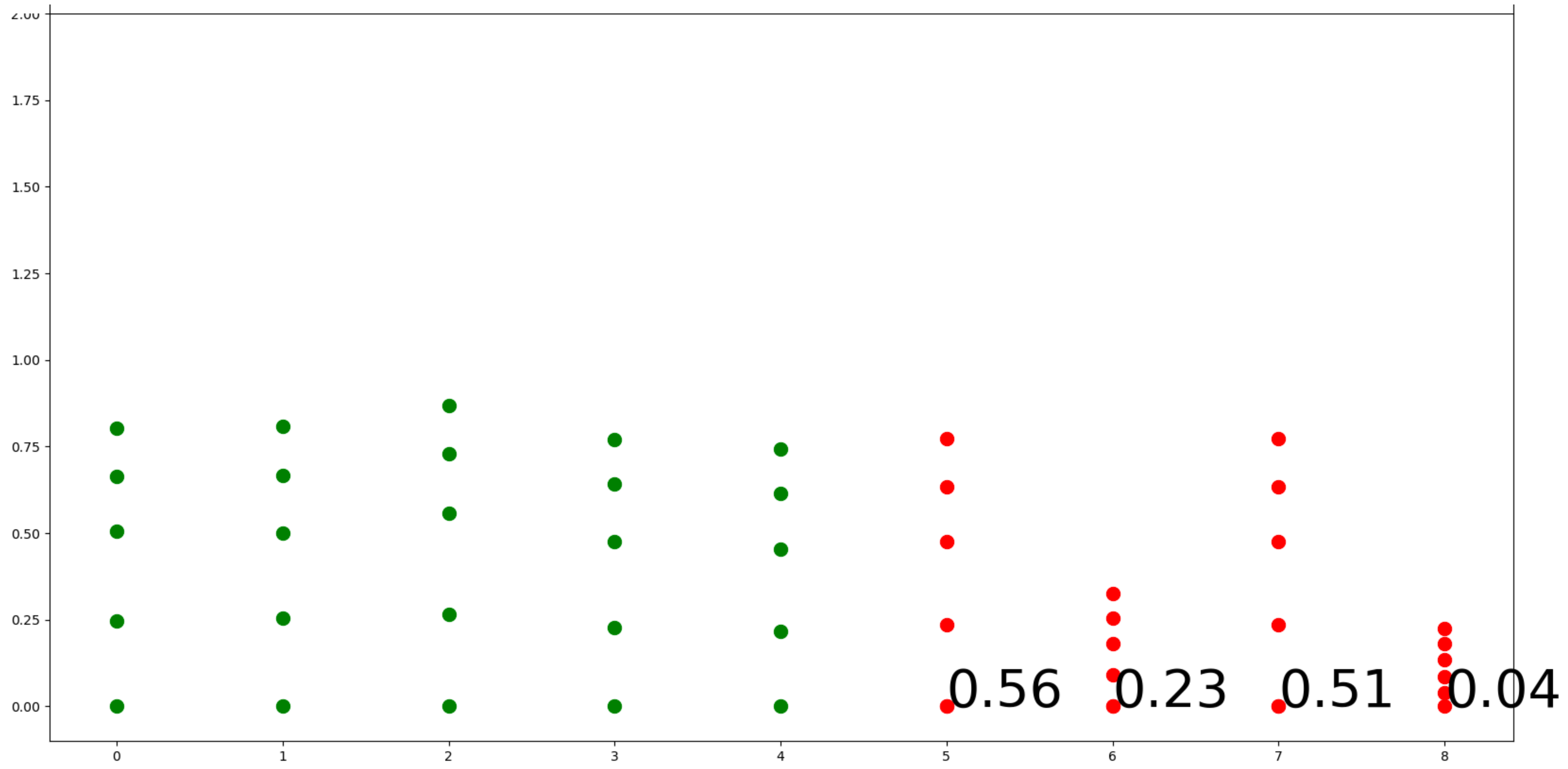


Predictions by the model

Ground Truth

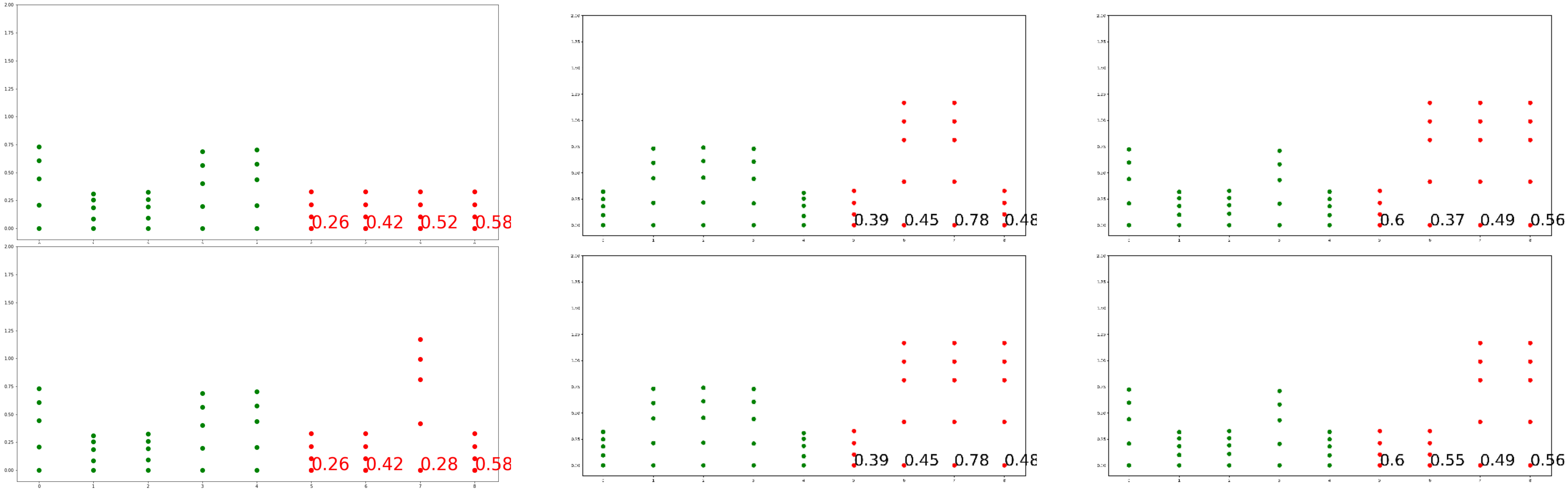


Prediction

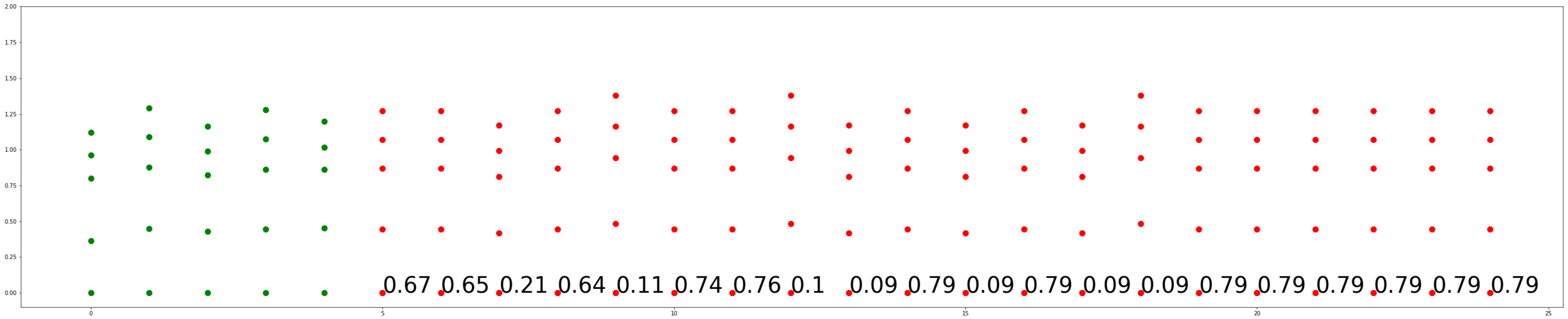


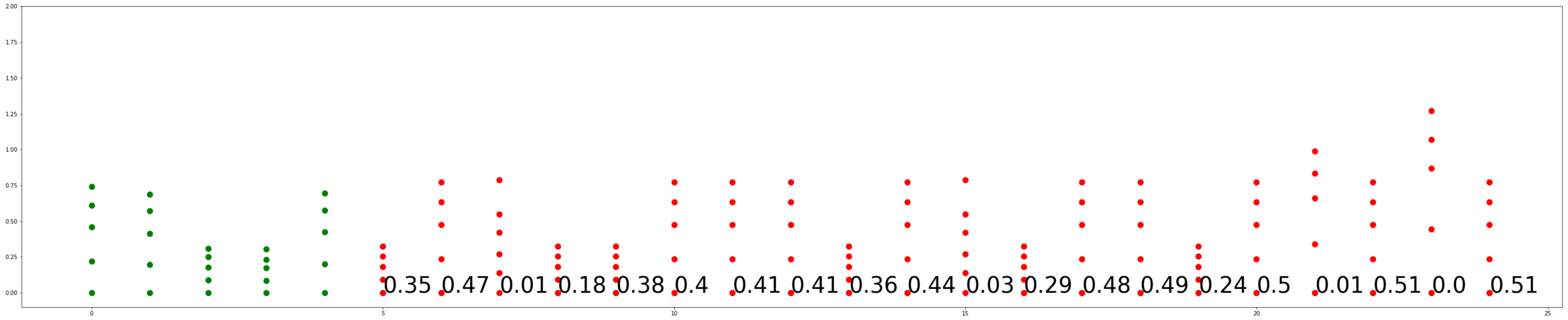
Ground Truth

Prediction



Some longer generated whale conversations





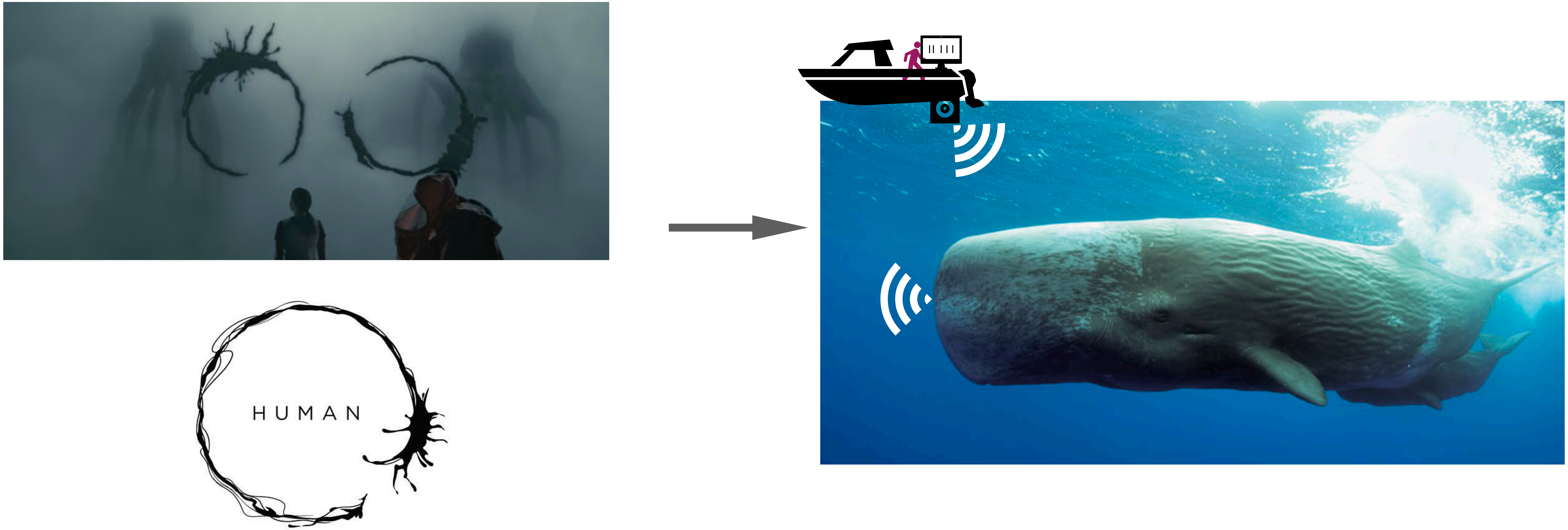
What we have found out so far

- Our visualizations have helped us find patterns of variation within the vocalizations - Imitation of rhythm and interruption (which were earlier treated us mere repetitions of roughly the same coda by an individual)

- With increase in amount of history as context for prediction the ability of models to predict the next coda improves (Evidence of non Markovian behavior in the vocalizations!)

- We can generate good / highly probable responses to sounds by sperm whales which could help us conduct interventional studies.

Next Steps?



Thank you!