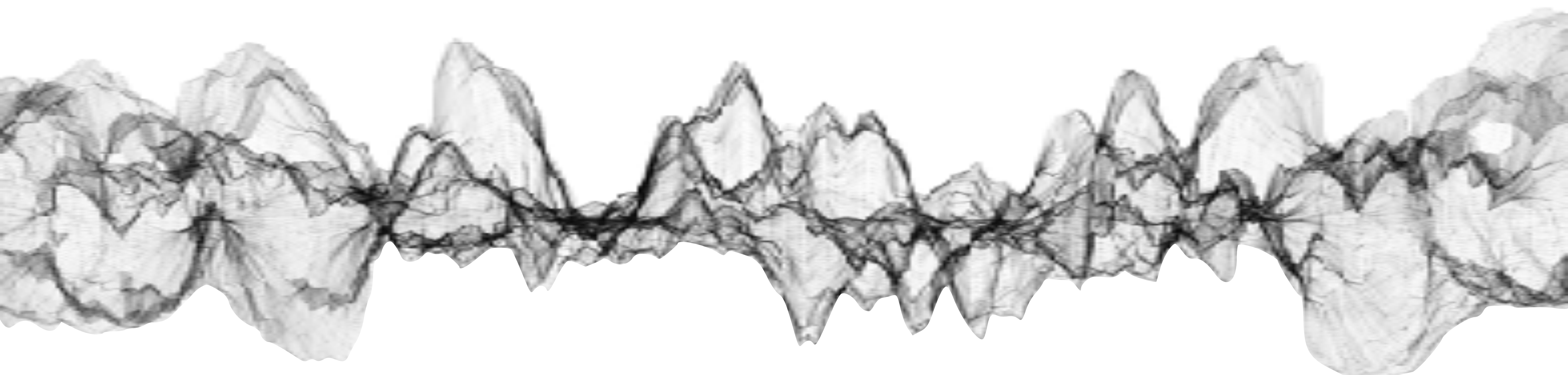
Stanford University

GLySN
Gwilliams Laboratory of Speech Neuroscience

Neural Algorithms of Human Language

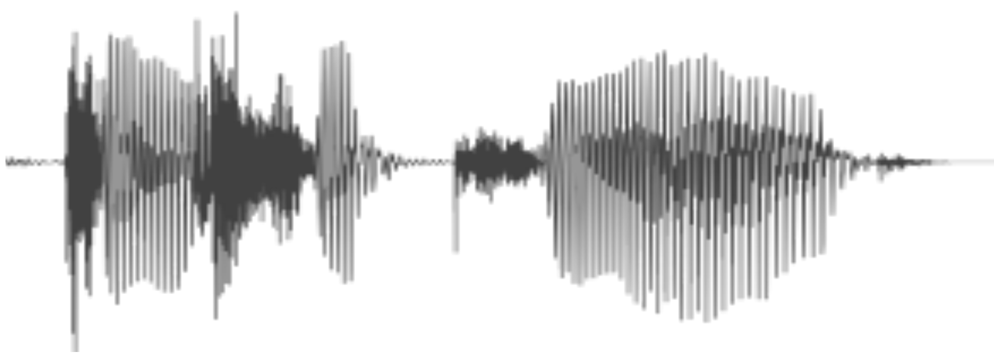
Laura Gwilliams

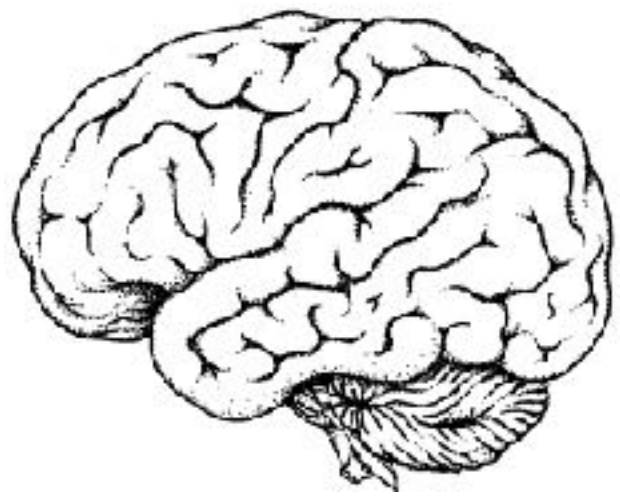
Gwilliams Lab of Speech Neuroscience

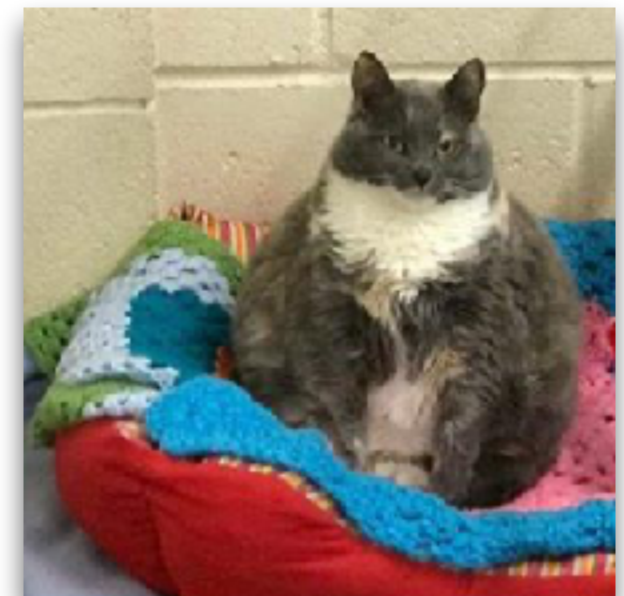


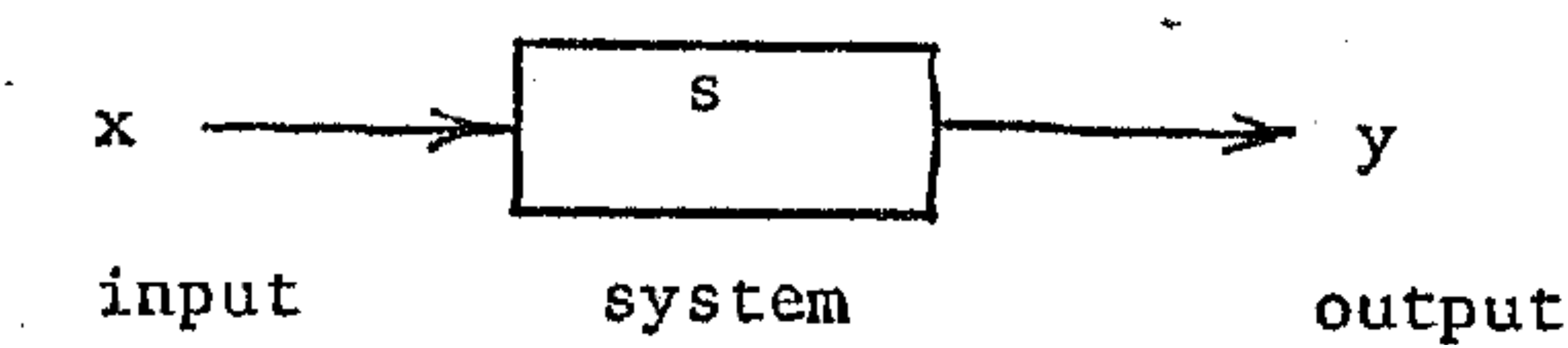
How does the brain translate between the sensory and the symbolic?

"the fat cat disappeared"





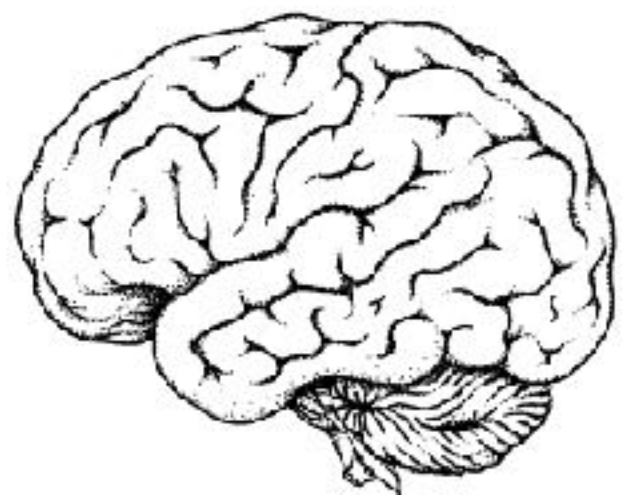


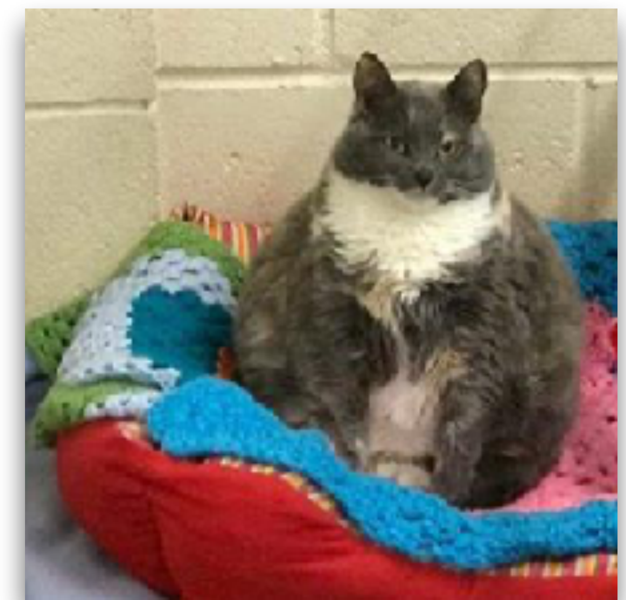


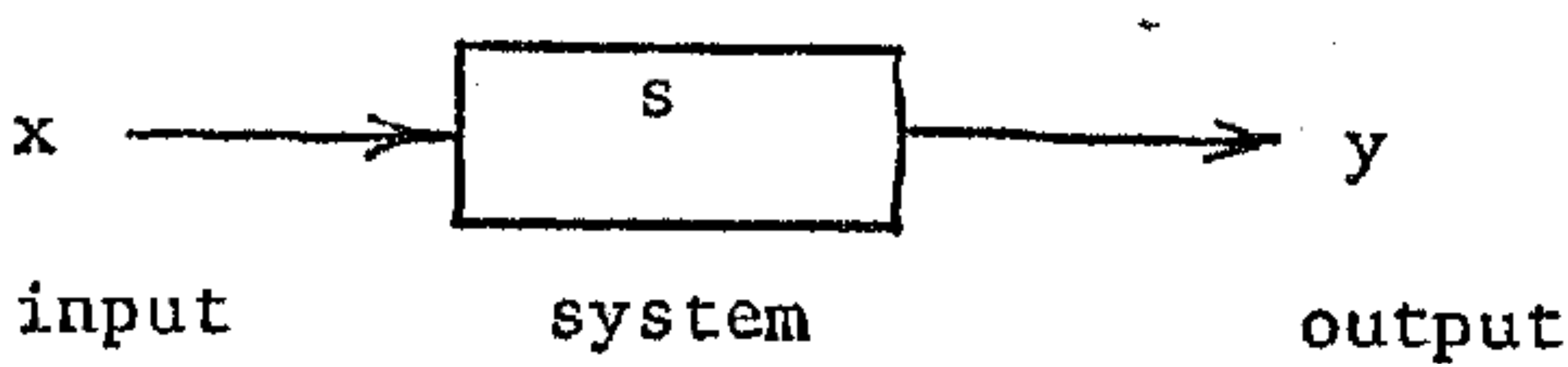
Laura Gwilliams | Stanford | @GwilliamsL

How does the brain translate between the sensory and the symbolic?

"the fat cat disappeared"



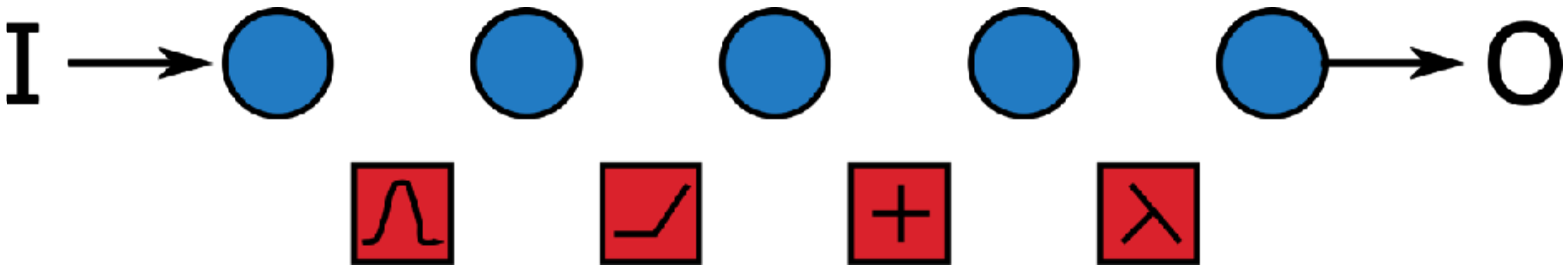




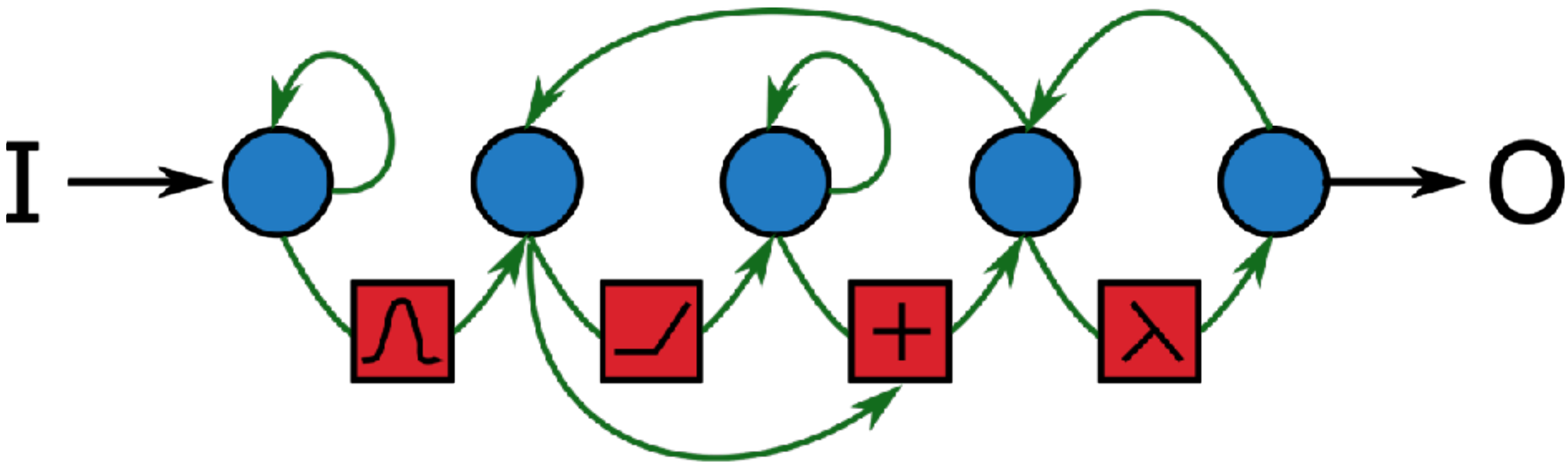
representations

representations

operations

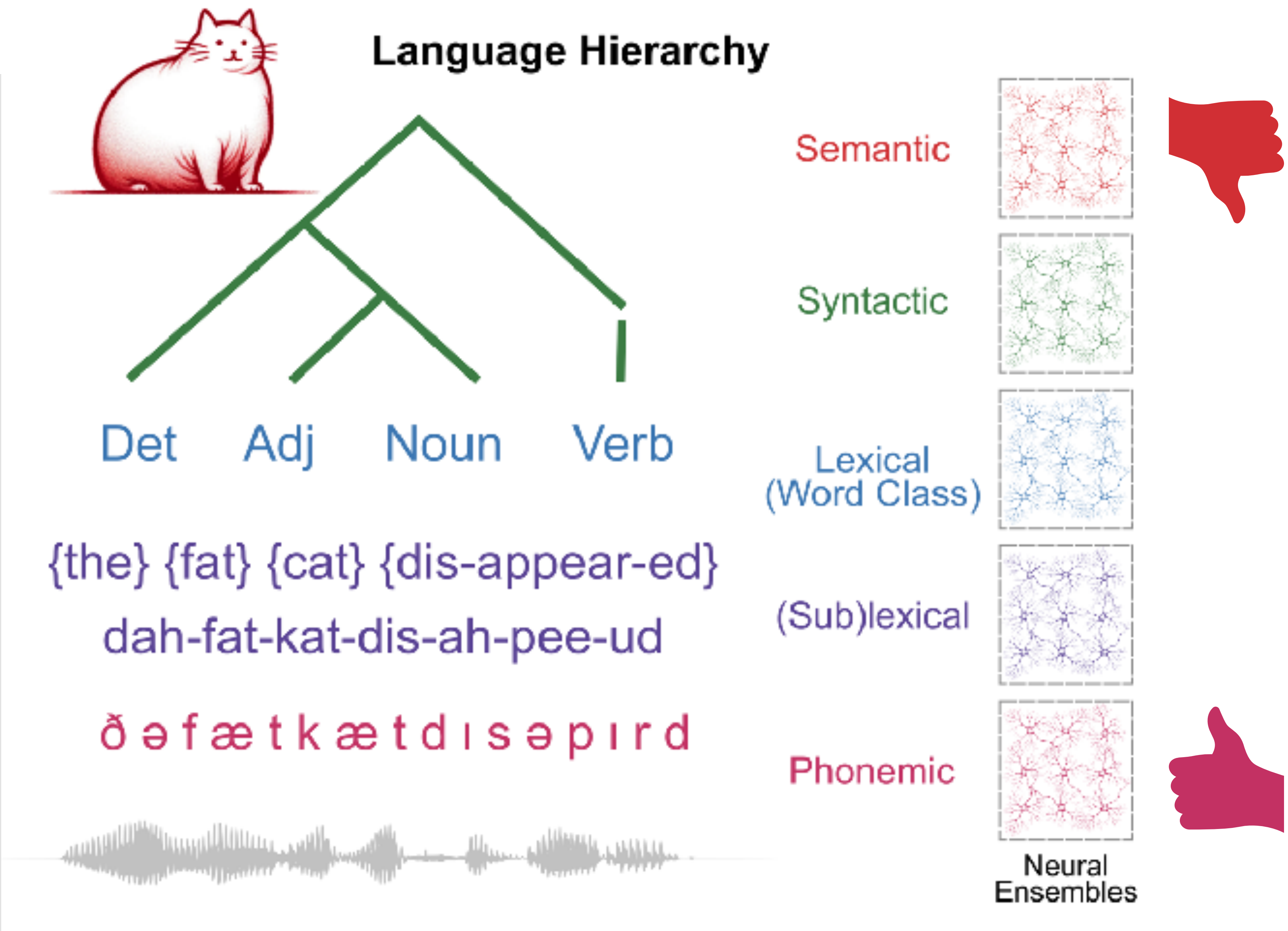


representations
operations
information flow



right NOW is the most exciting time for language neuroscience!

Understanding Symbolic Processing



- No animal models(?)
- How to represent numerically?
- Response variability

- Animal models
- Measurable input signal
- Response consistency
- Focal coding

Understanding Symbolic Processing

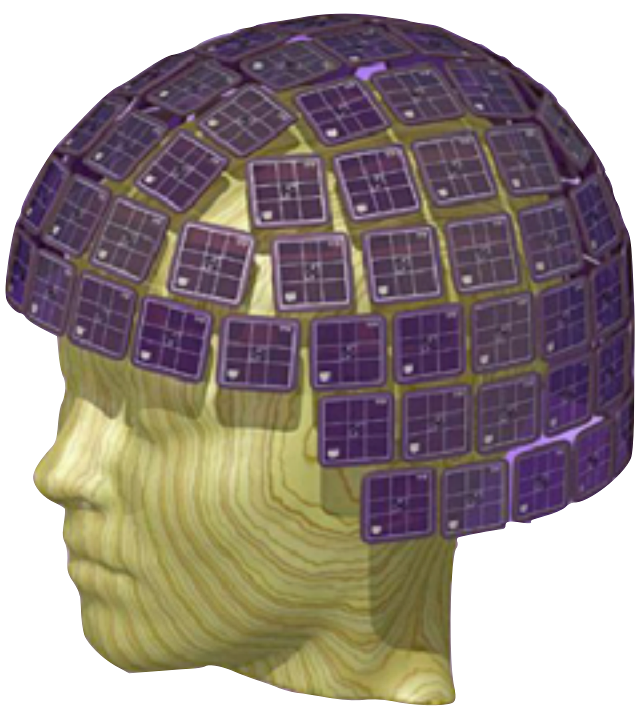
- No animal models(?)

LLMs as cognitive models

- How to represent numerically?

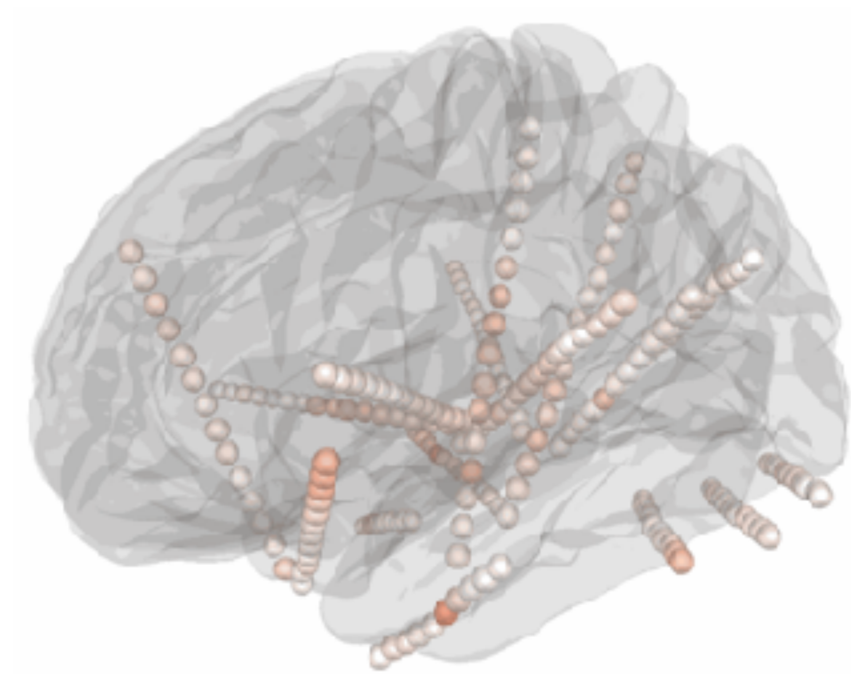
LLMs as feature generators

Human Neuroscience Technology



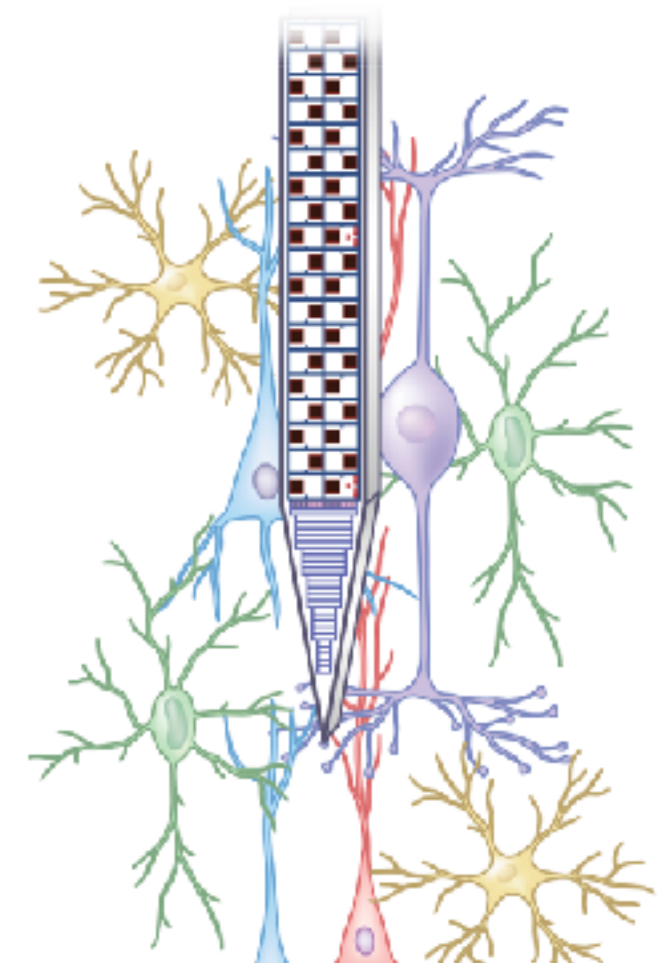
MEG / EEG

~100,000 neurons



iEEG

~10,000 neurons



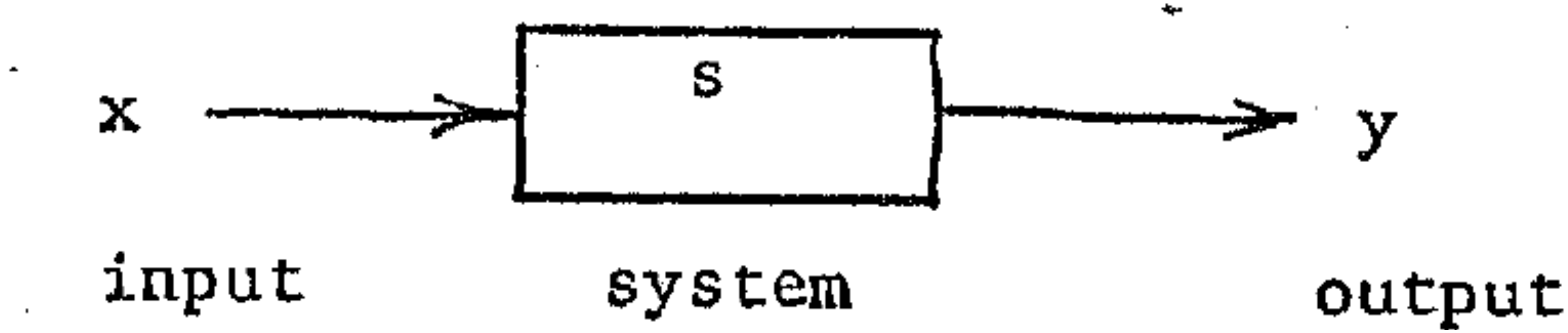
Neuropixels

~1 neurons

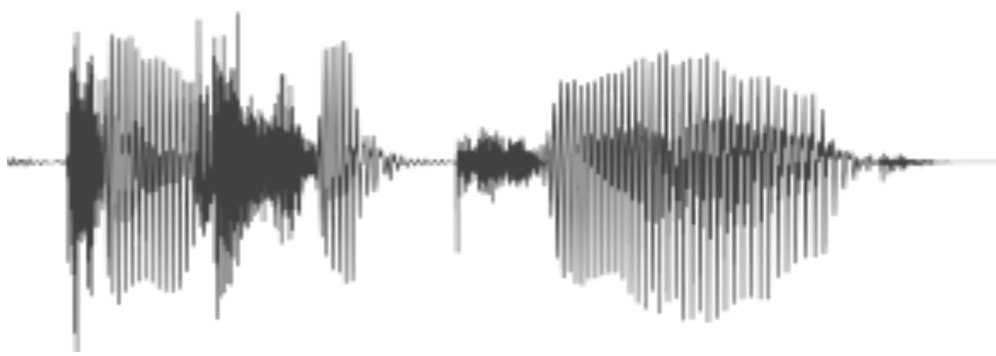
Menu

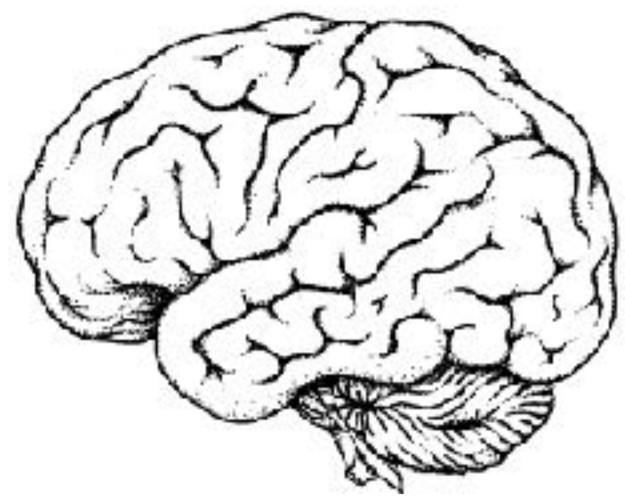
- Neural **representation**: From sensory to symbolic
- Neural **computation**: Sequencing sensory inputs
- Neural **implementation** at the level of single neurons in humans

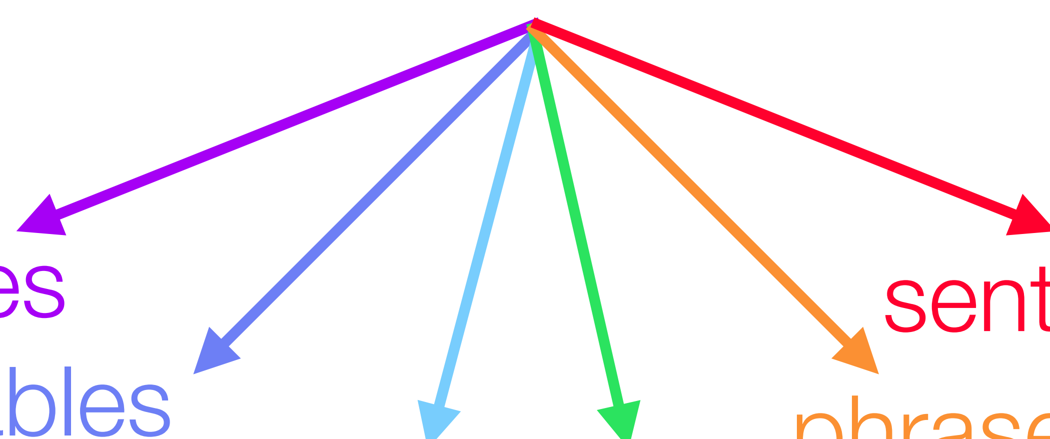
sensory to symbolic



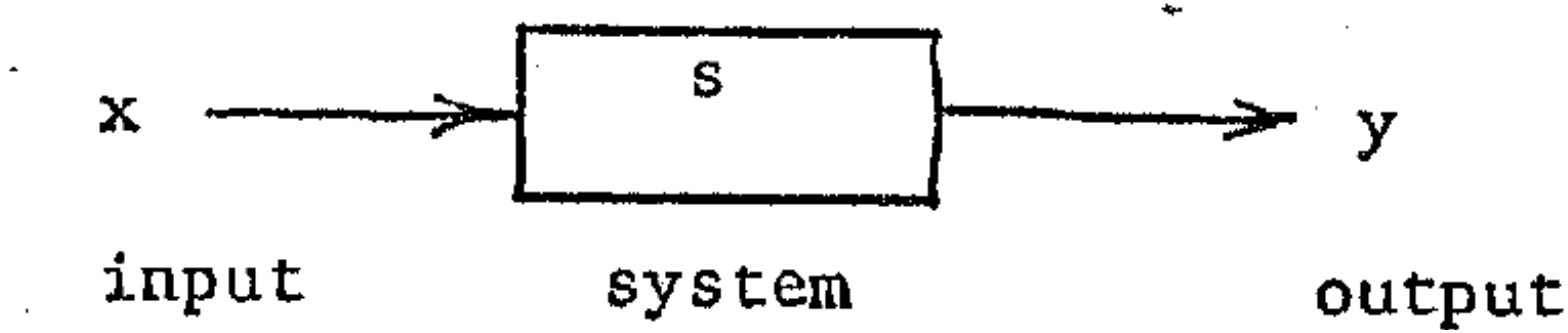
"the fat cat disappeared"



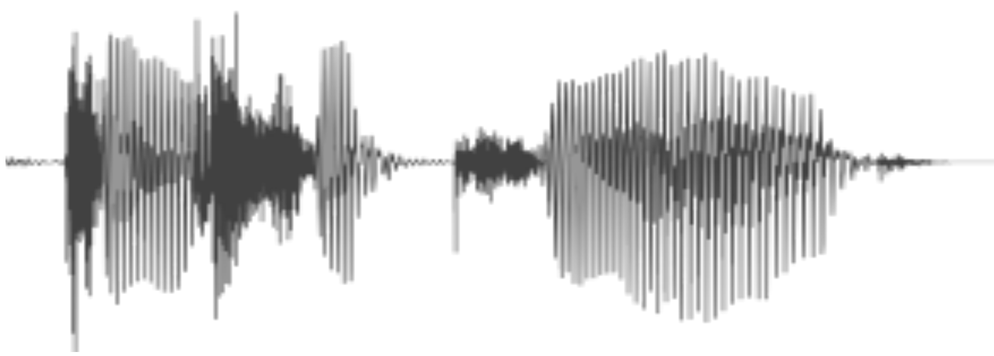


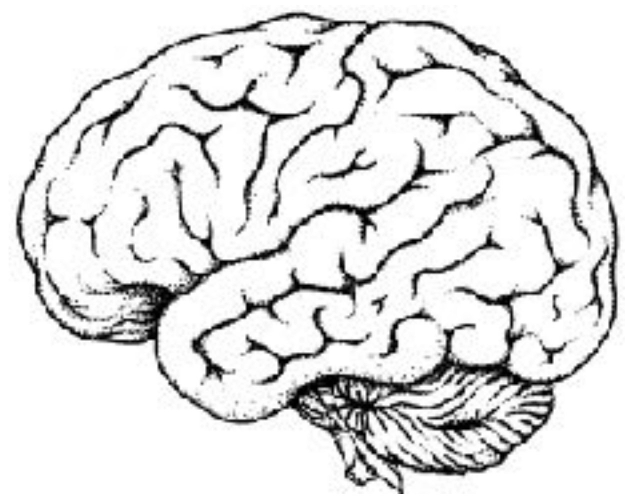


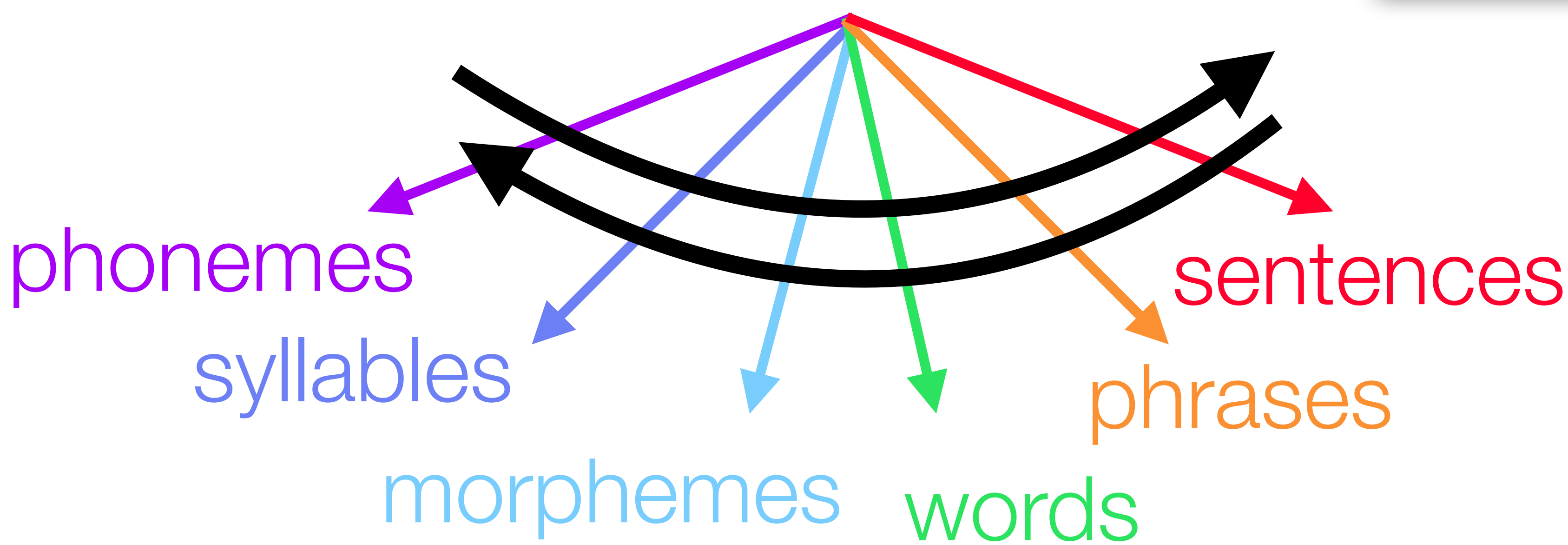
phonemes
syllables
morphemes
words
phrases
sentences



“the fat cat disappeared”







Setup

scientific **data**

OPEN

DATA DESCRIPTOR

Check for updates

Introducing MEG-MASC a high-quality magneto-encephalography dataset for evaluating natural speech processing

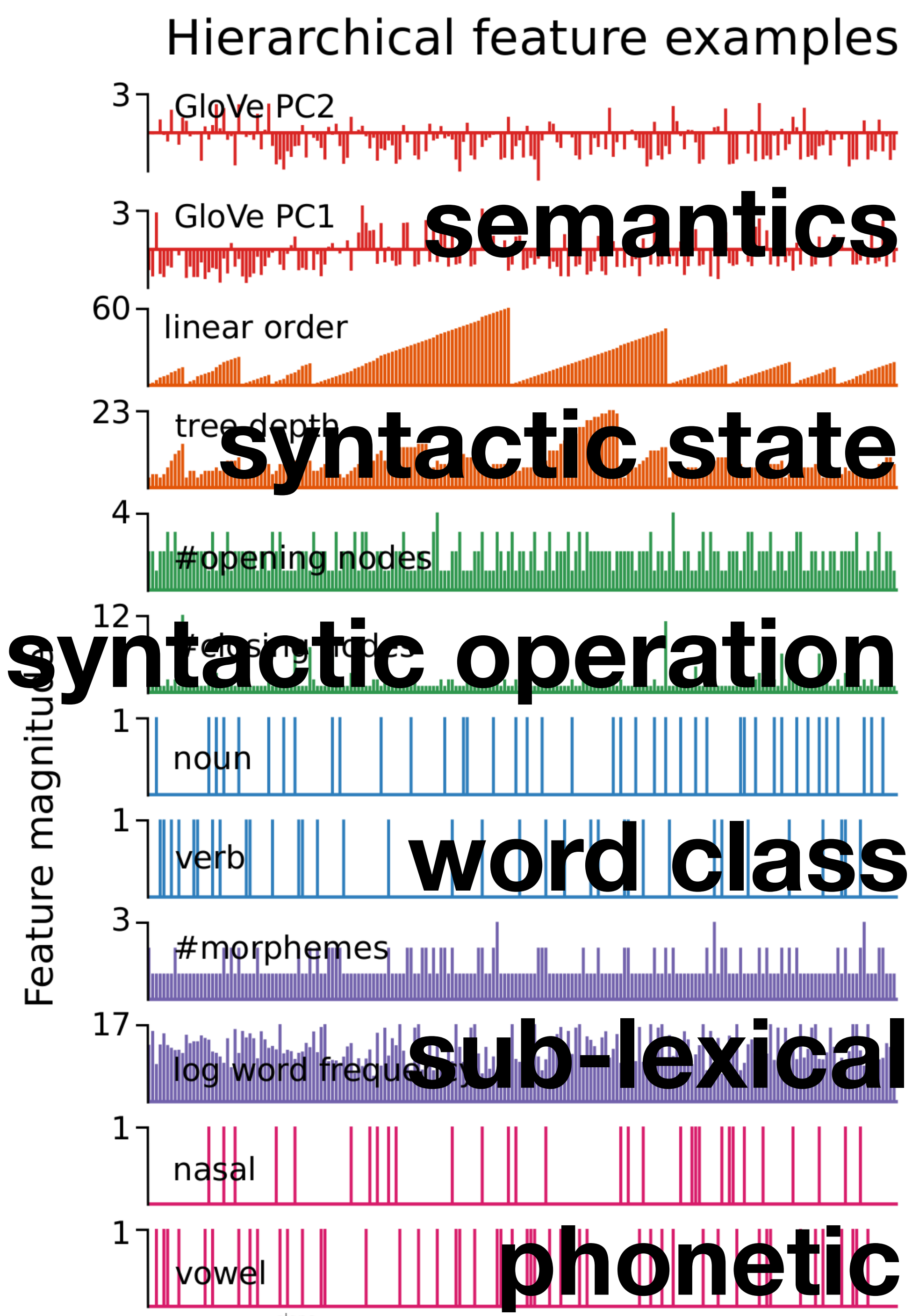
Laura Gwilliams[1,2,3], Graham Flick[2,3,4,5], Alec Marantz[2,3,4], Liina Pylkkänen[2,3,4], David Poeppel[2,6] & Jean-Rémi King[2,7]

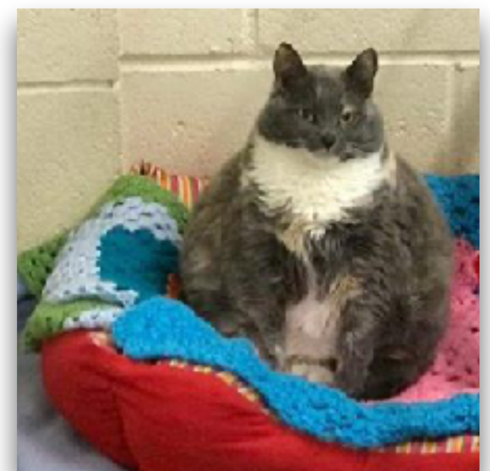
- ~16,000 words per participant

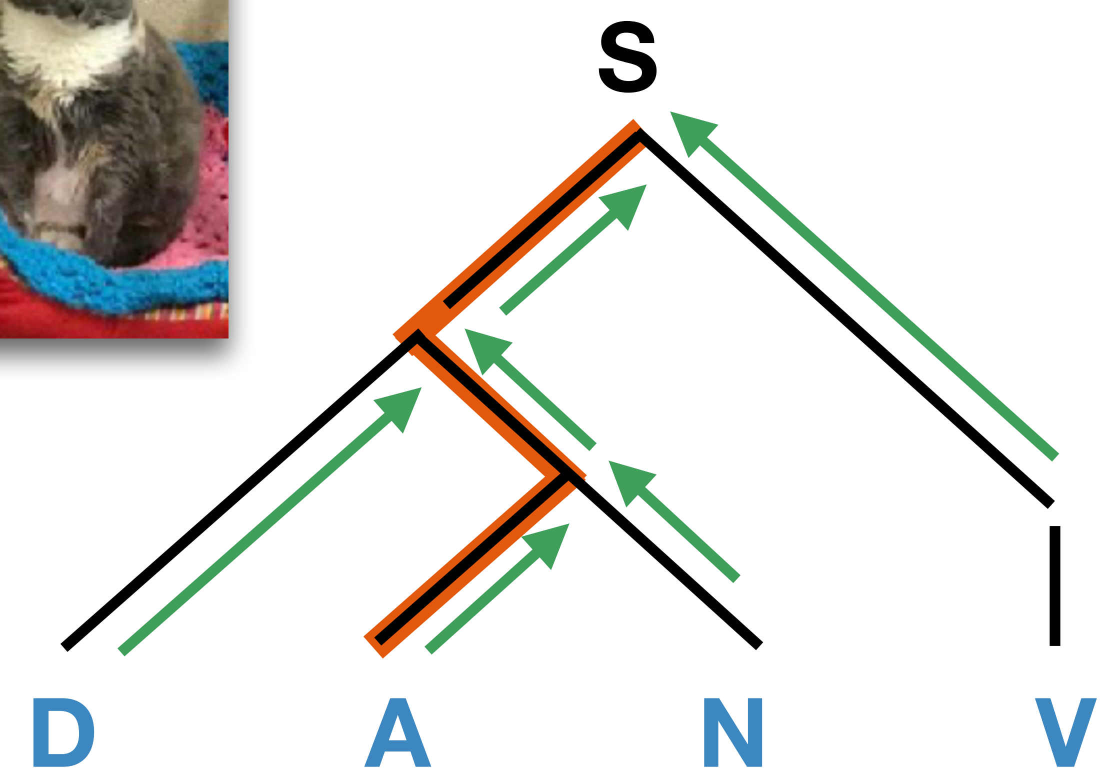
Gwilliams et al. (2024) *Submitted*

Setup

in what **order** are representations generated?







{the} {fat} {cat} {dis-appear-ed}

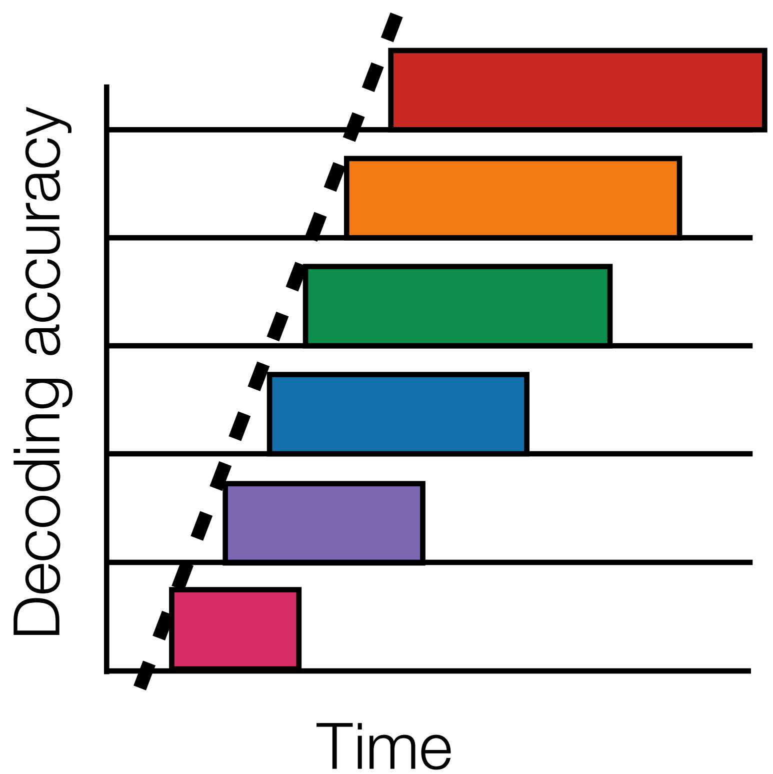
dah-fat-kat-dis-ah-pee-ud

- \+ - + - - + + -
- \- + - - + - - +
- \+ + + + + + + + +
- \- - - - + - - + +

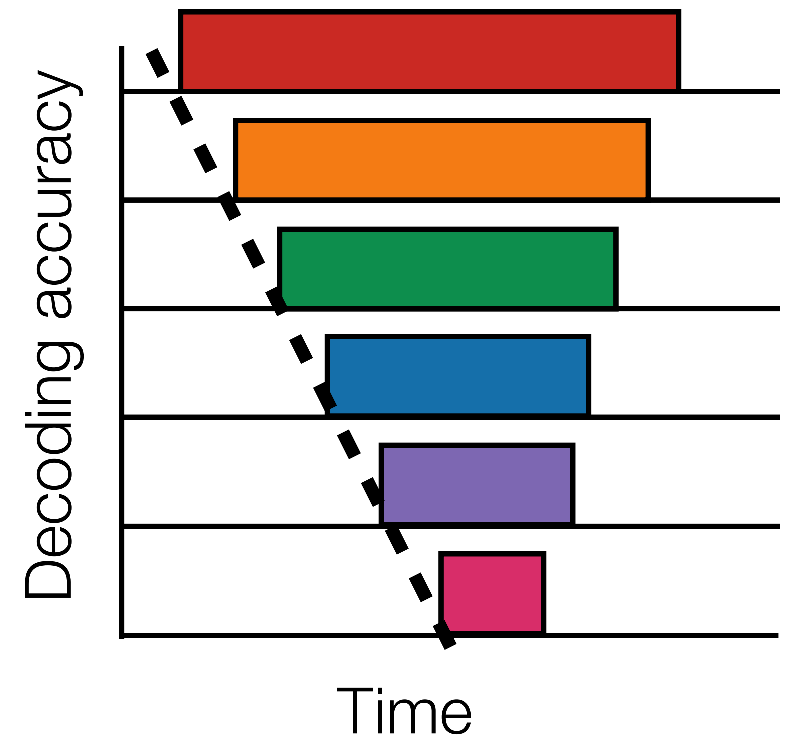
Gwilliams et al. (2024) *Submitted*

Hypotheses

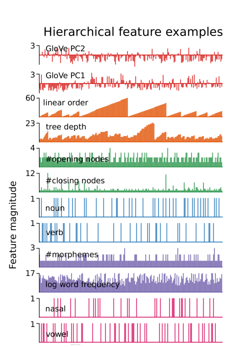
in what **order** are representations generated?



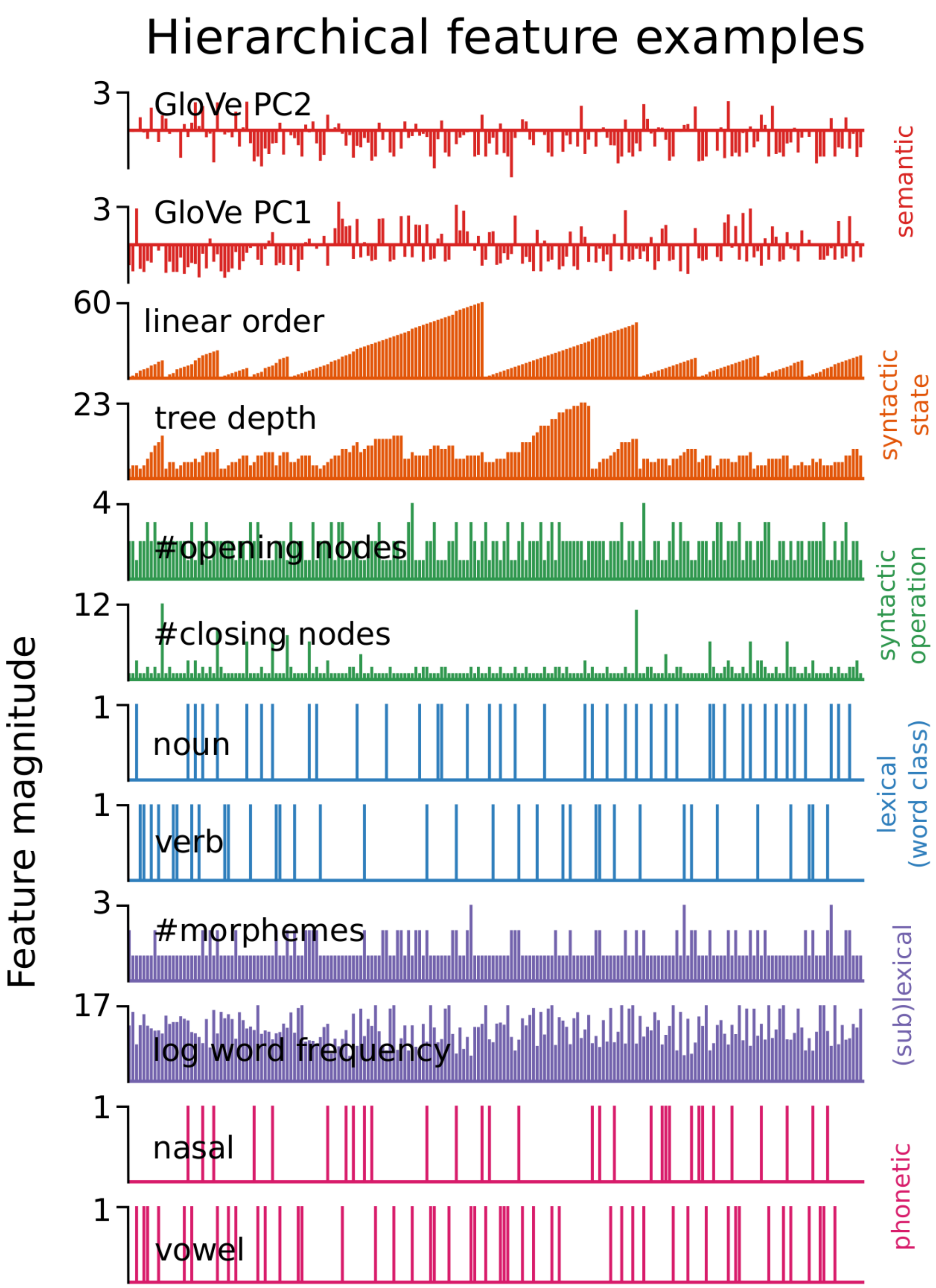
(e.g. Gaskell & Marslen-Wilson, 1997)

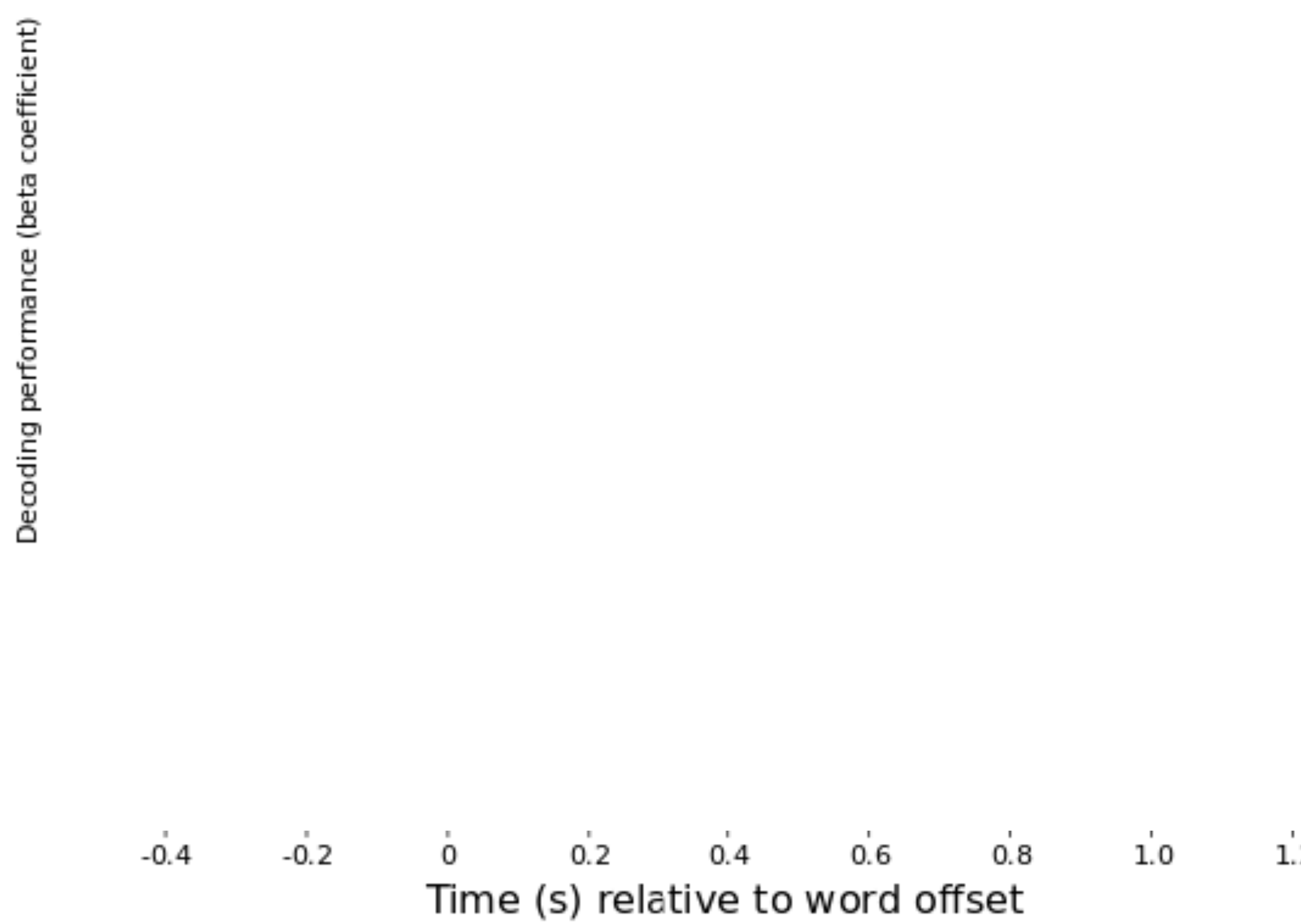


(e.g. Ahissar & Hochstein, 2004)



Hierarchical processing



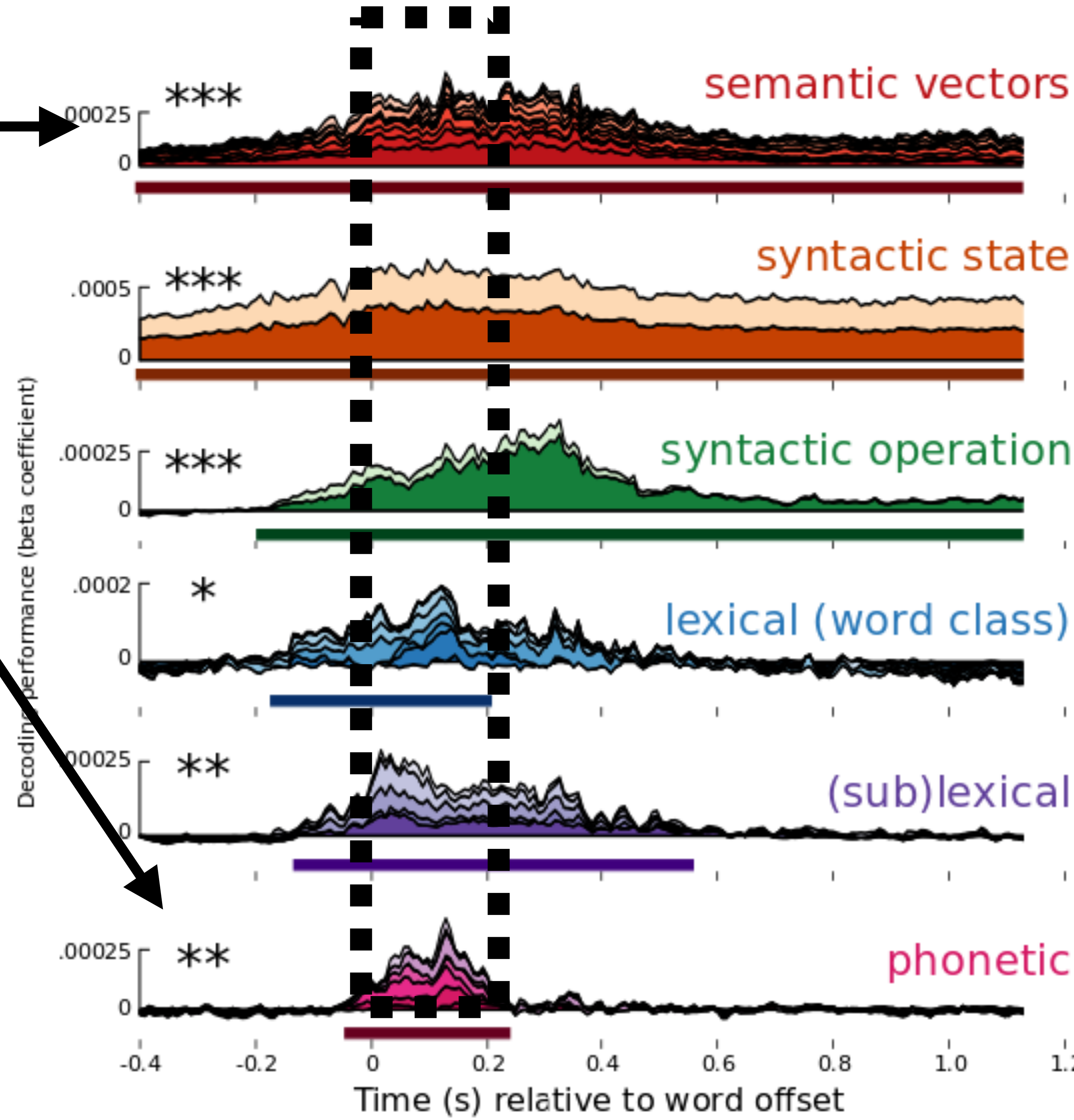


Gwilliams et al. (2024) *Submitted*

Feature Dynamics

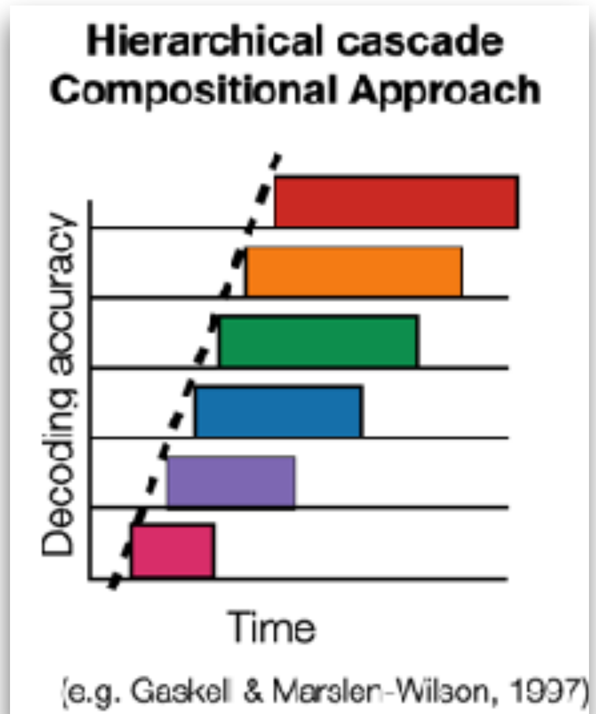
high-level representations are activated first, and maintained the longest

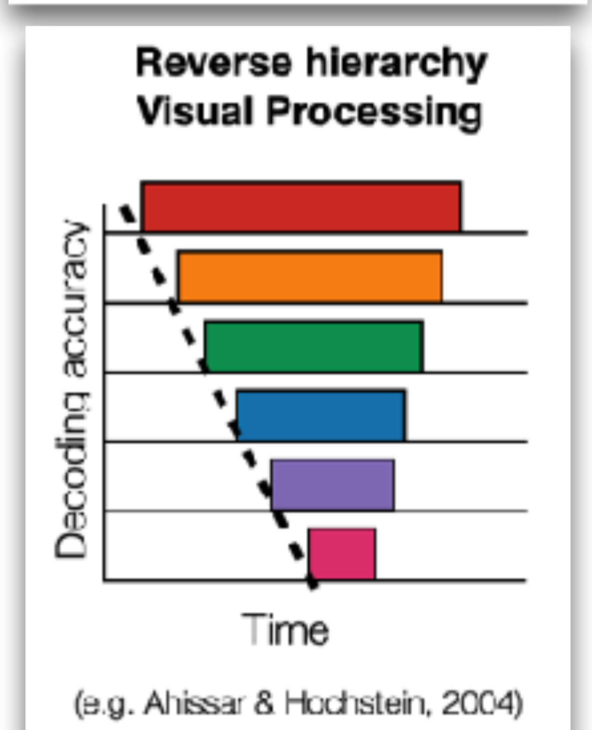
local details of the input are encoded later, and for much less time

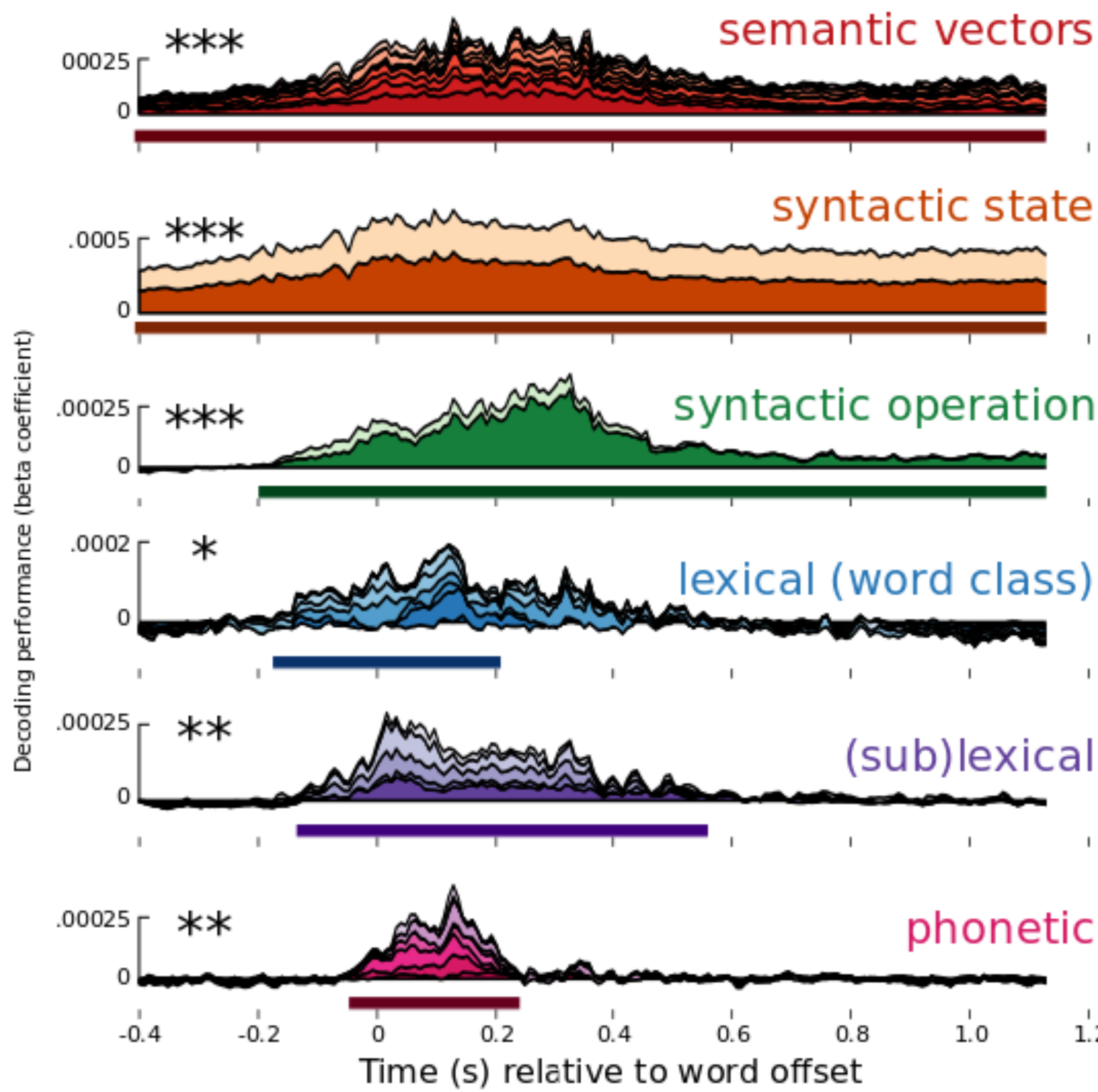


Gwilliams et al. (2024) *Submitted*

Interpretation







Gwilliams et al. (2024) *Submitted*

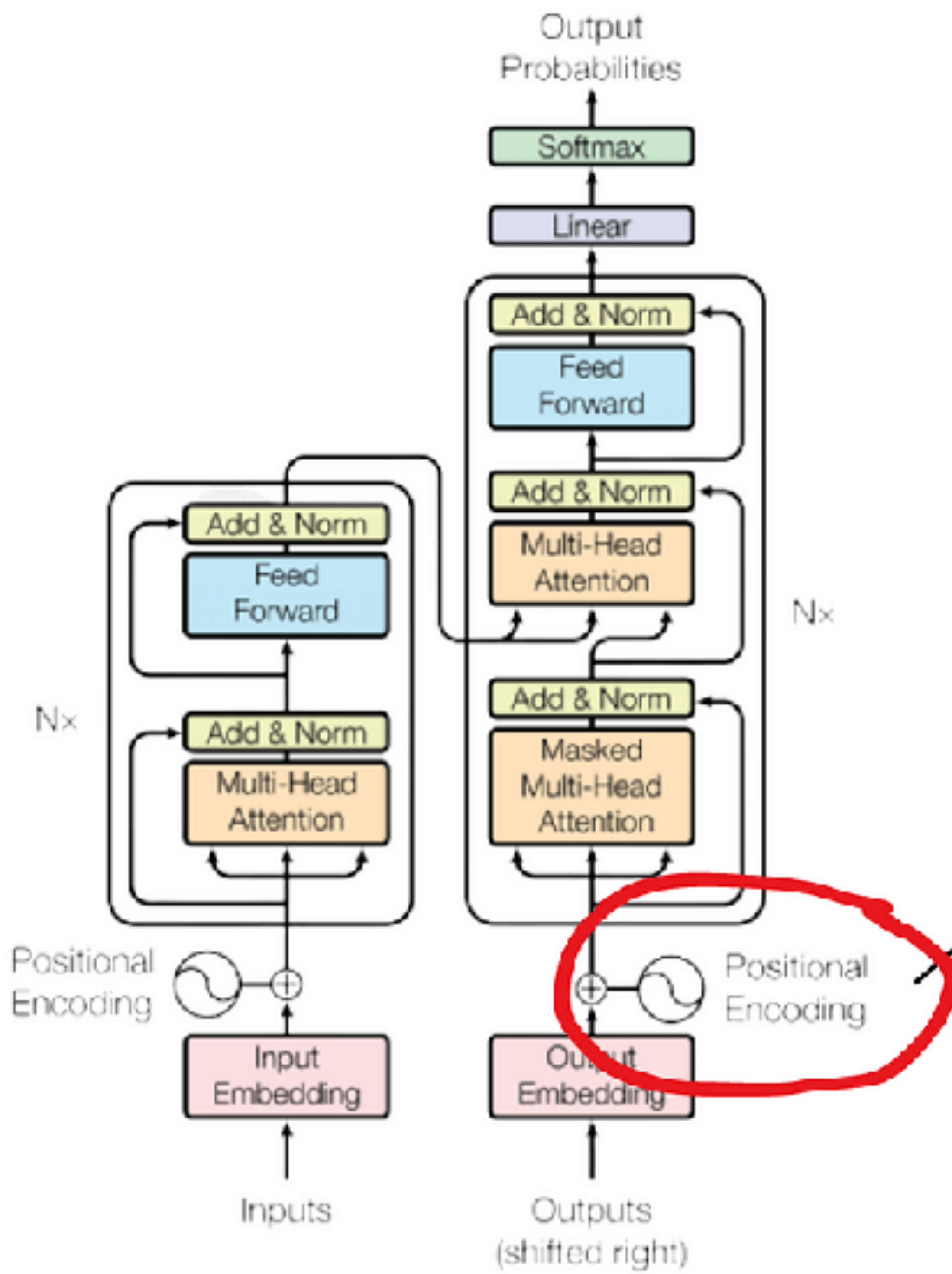
Spatial Localization Across Development

Jill Kries
Postdoc

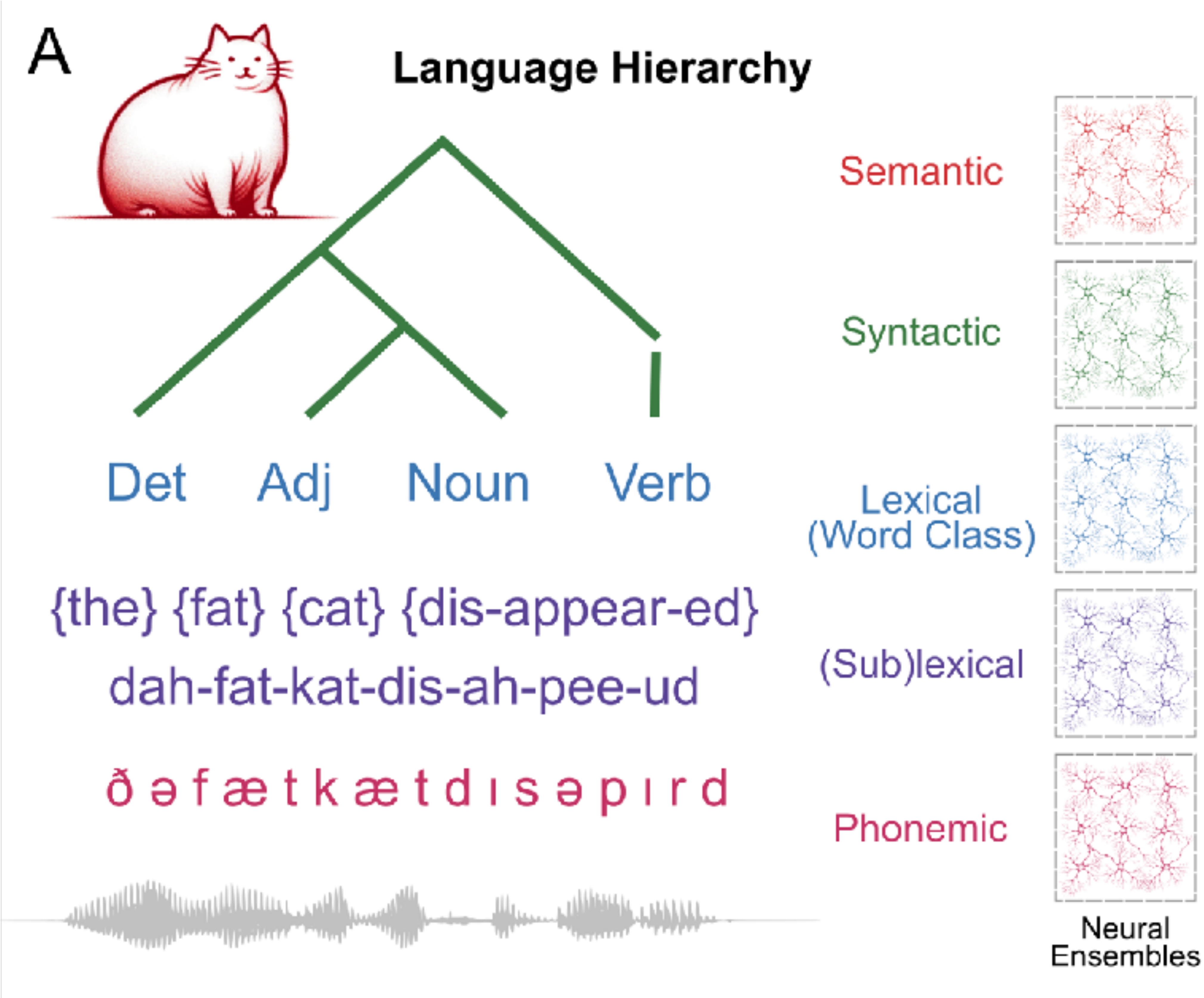
44 children
From ages 4-17

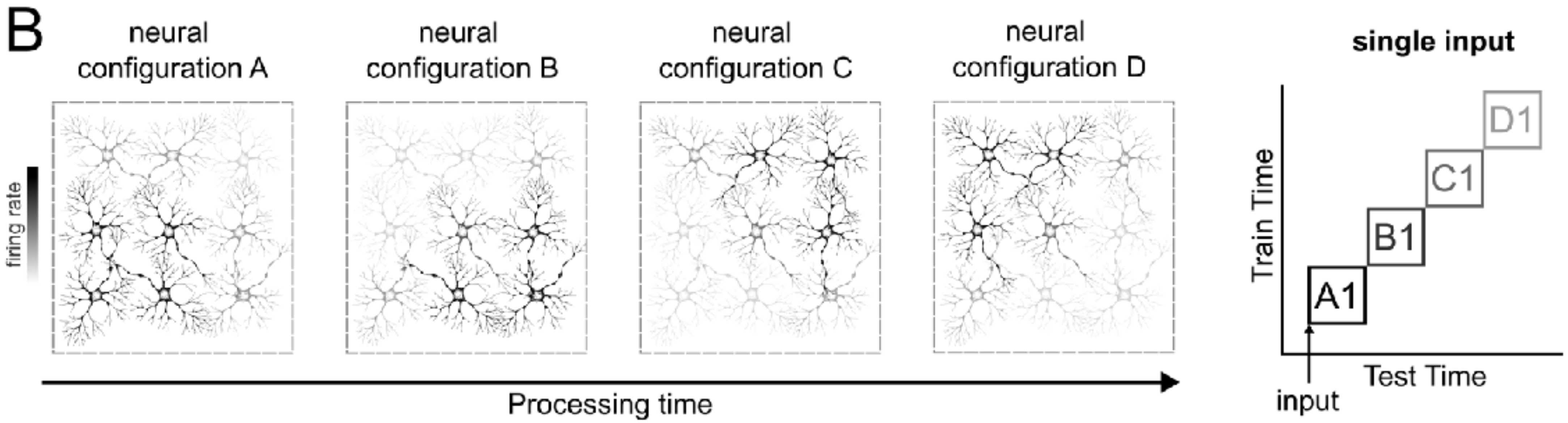
Kries, Arya, Hickok & Gwilliams (in prep)

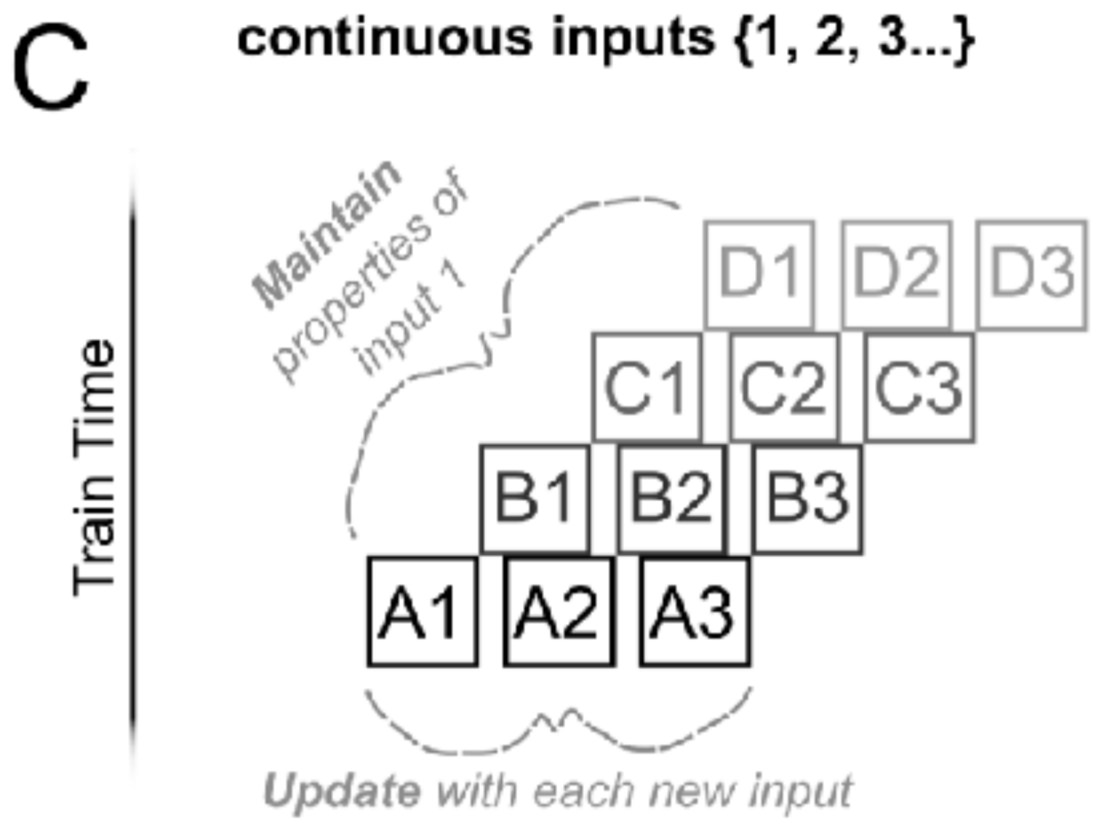
sequencing continuous speech input


Output
Probabilities
Softmax
Linear
Add & Norm
Feed
Forward
Add & Norm
Multi-Head
Attention
Add & Norm
Masked
Multi-Head
Attention
Nx
Add & Norm
Feed
Forward
Add & Norm
Multi-Head
Attention
Nx
Positional
Encoding
Positional
Encoding
Input
Embedding
Output
Embedding
Inputs
Outputs
(shifted right)

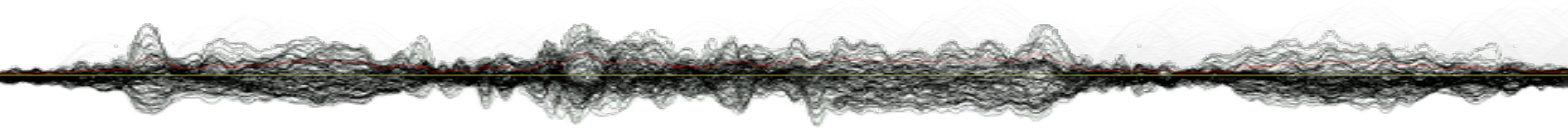
RoPE


A
Language Hierarchy
Det
Adj
Noun
Verb
{the} {fat} {cat} {dis-appear-ed}
dah-fat-kat-dis-ah-pee-ud
ð ə f æ t k æ t d ɪ s ə p ɪ r d
Semantic
Syntactic
Lexical
(Word Class)
(Sub)lexical
Phonemic
Neural
Ensembles


B
neural configuration A
neural configuration B
neural configuration C
neural configuration D
firing rate
Processing time
single input
Train Time
Test Time
input
A1
B1
C1
D1


C
continuous inputs {1, 2, 3...}
Train Time
Maintain properties of input 1
A1 A2 A3
B1 B2 B3
C1 C2 C3
D1 D2 D3
Update with each new input

Setup



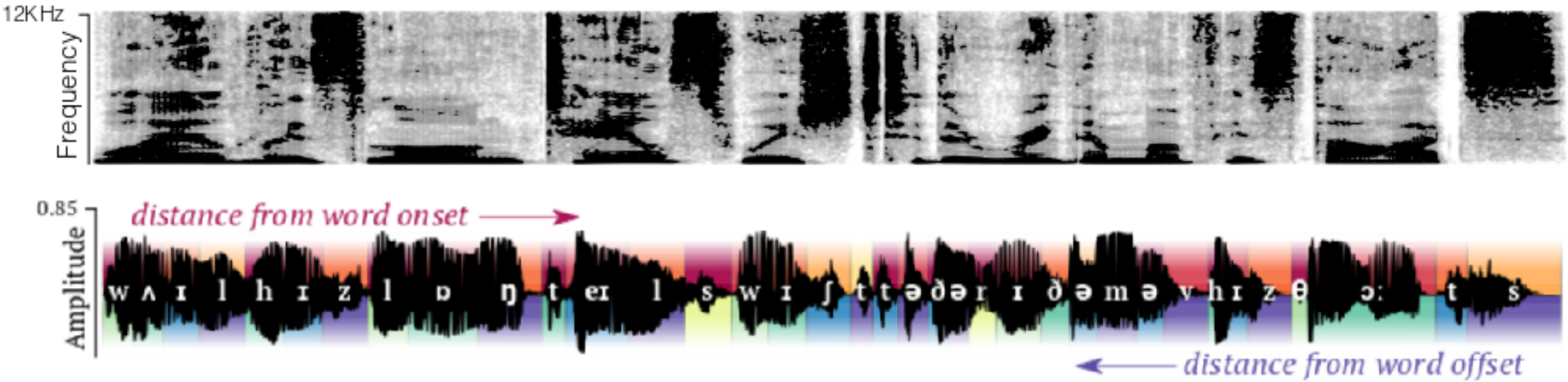
- 21 participants
- Listening to four narrative stories (twice)
- Synthesised voice (MacOS)
  - semi-controlled acoustic input
- 2 x one hour recordings
- KIT 208 channel MEG system
- Engagement task

- ~40,000 phonemes per participant
- ~16,000 words per participant

Gwilliams et al. (2022) *NatComms*

Analysis

"… while his long tail swished to the rhythm of his thoughts…"

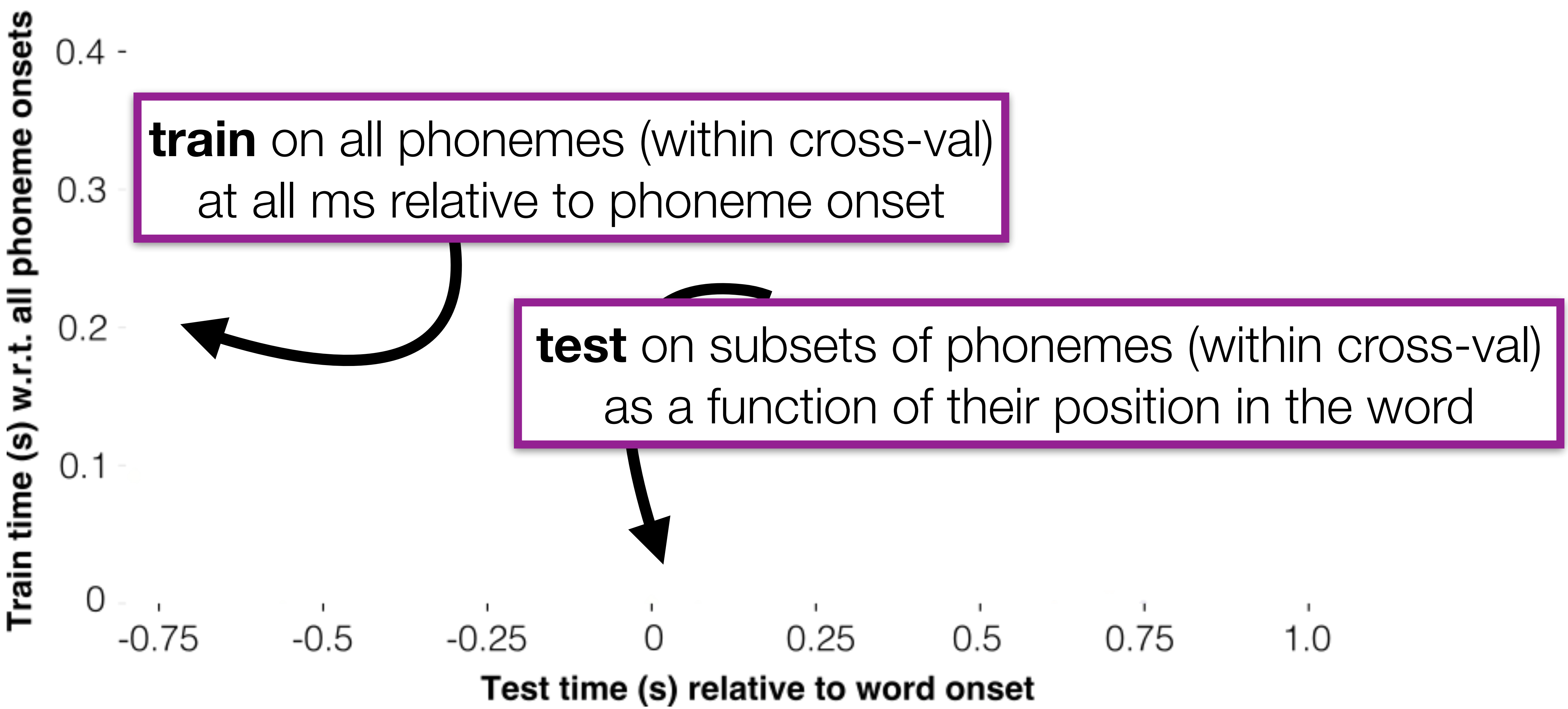


- Decode phonetic features (voicing, manner, place of articulation) as a function of the phoneme's position in the word

Gwilliams et al. (2022) *NatComms*

Transformation of acoustic signal

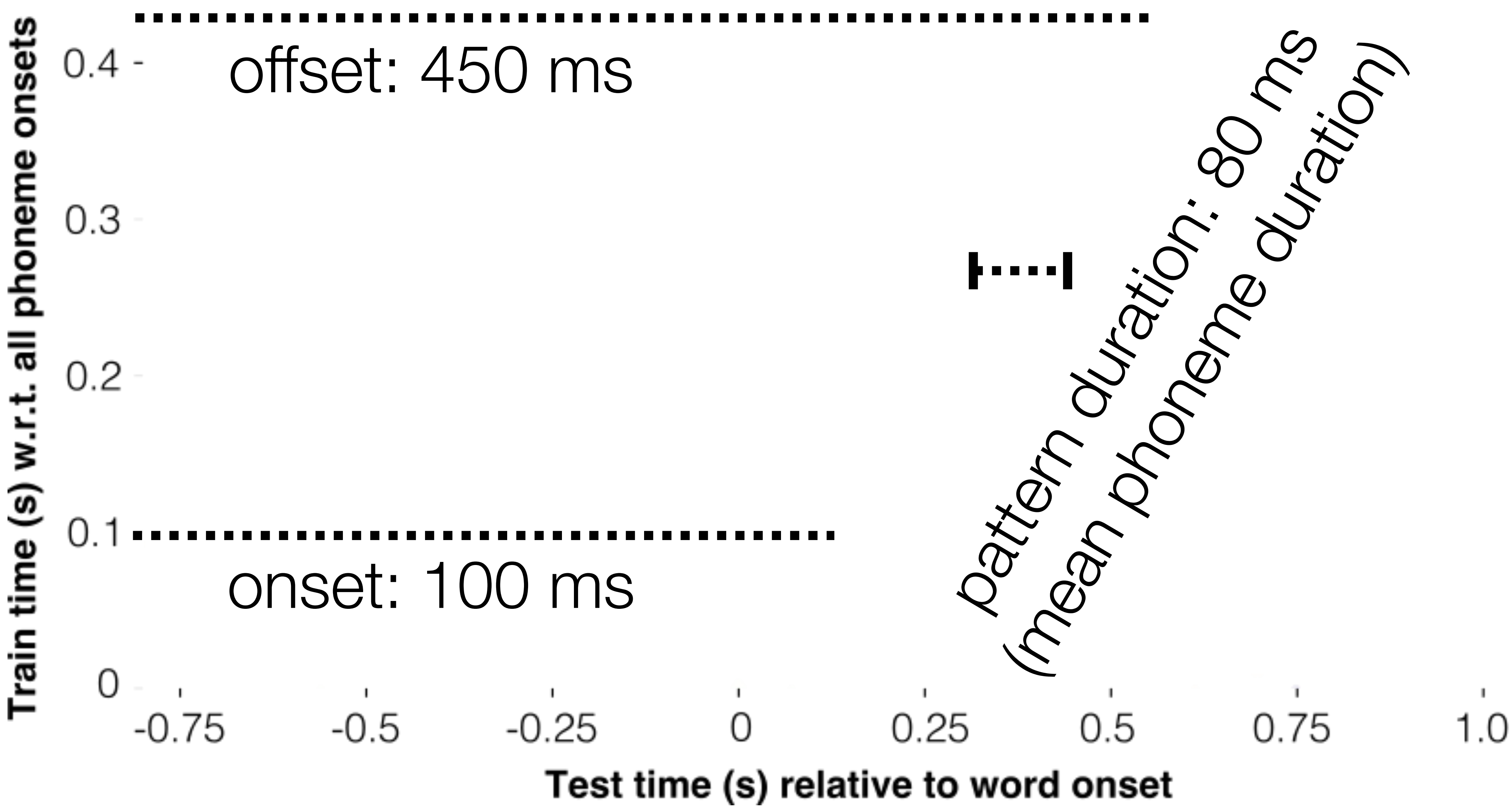
- Apply temporal generalisation to assess processing dynamics



Gwilliams et al. (2022) *NatComms*

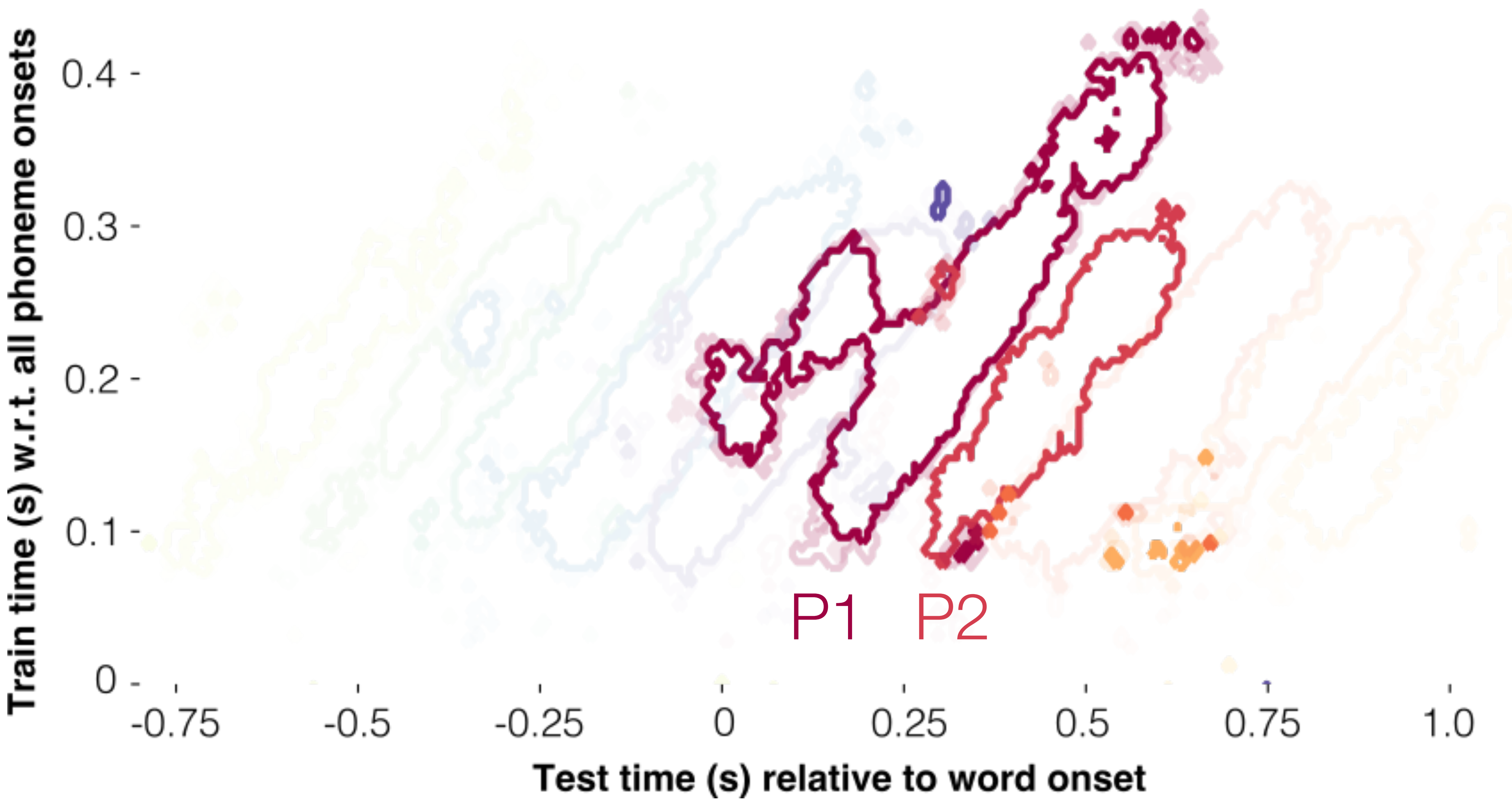
Transformation of acoustic signal

- Apply temporal generalisation to assess processing dynamics



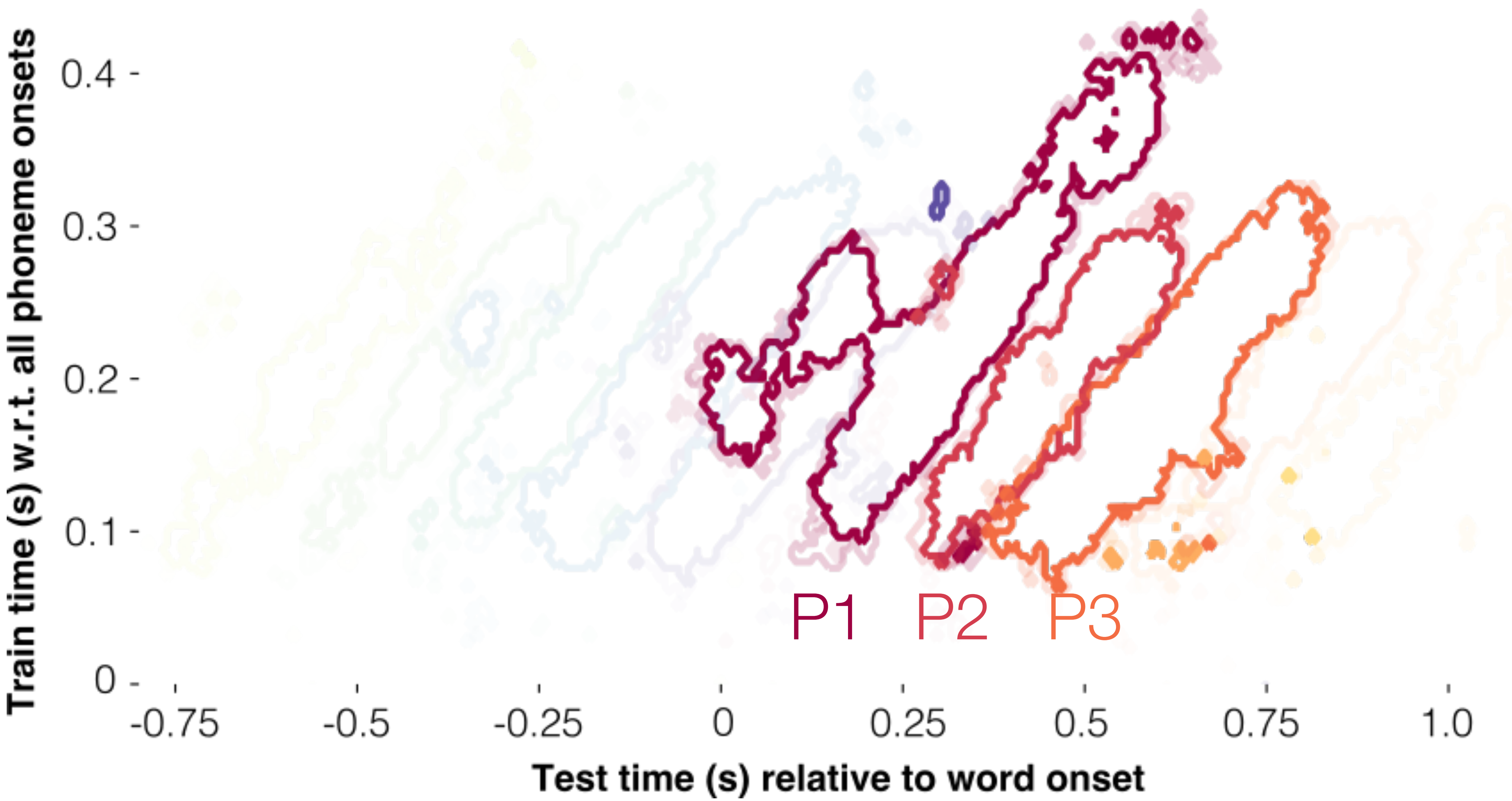
Gwilliams et al. (2022) *NatComms*

Transformation of acoustic signal



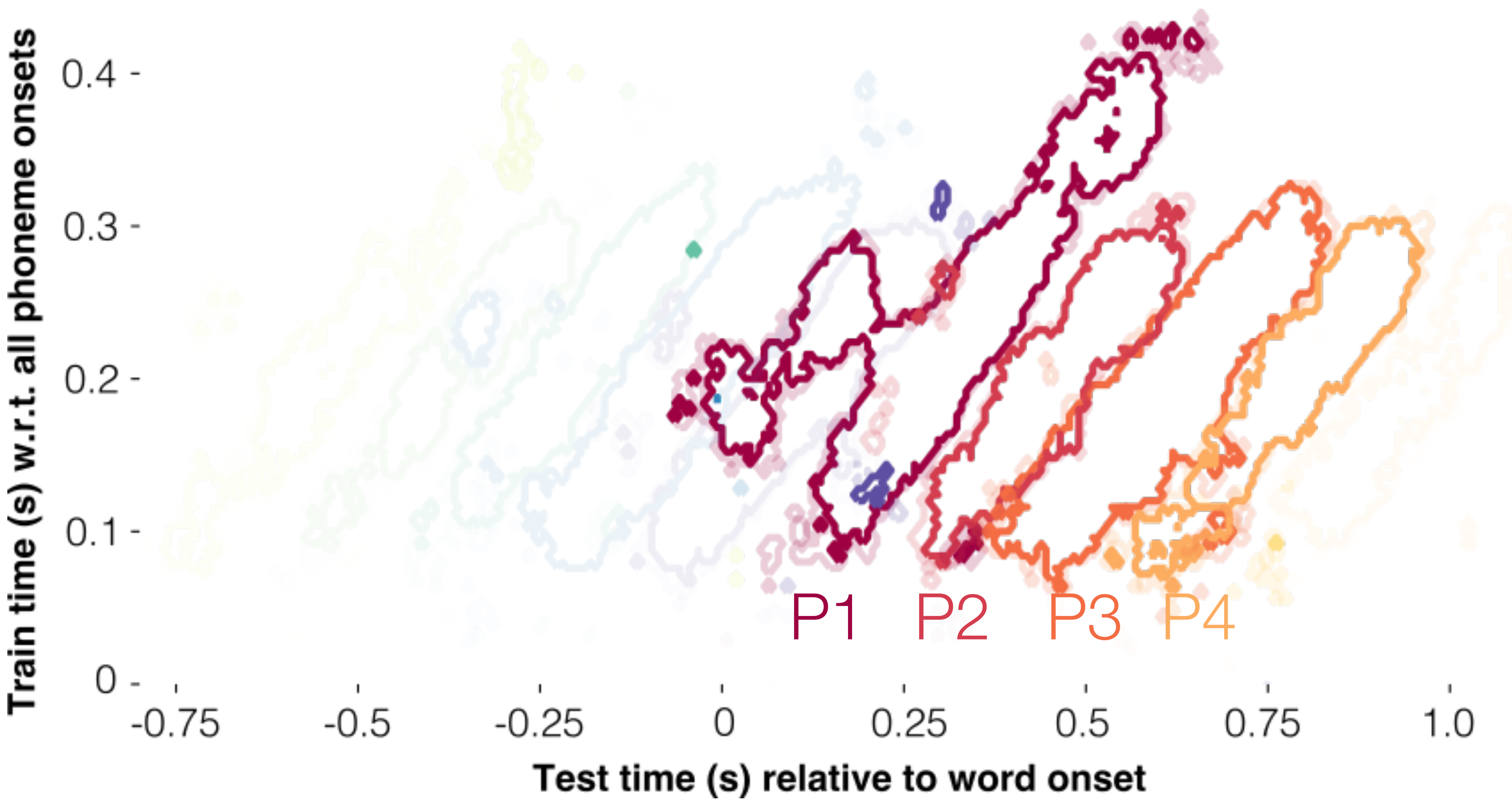
Gwilliams et al. (2022) *NatComms*

Transformation of acoustic signal



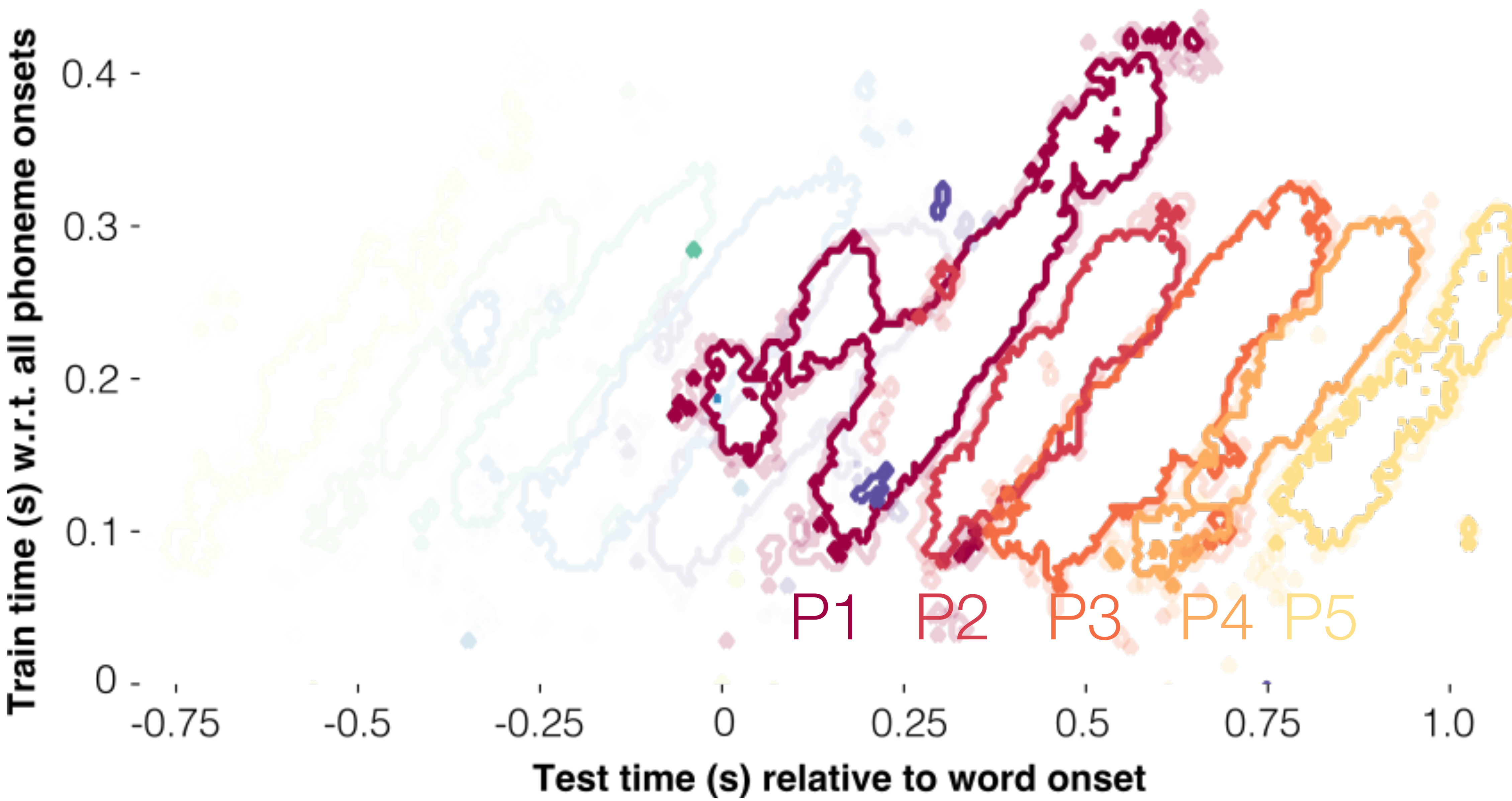
Gwilliams et al. (2022) *NatComms*

Transformation of acoustic signal



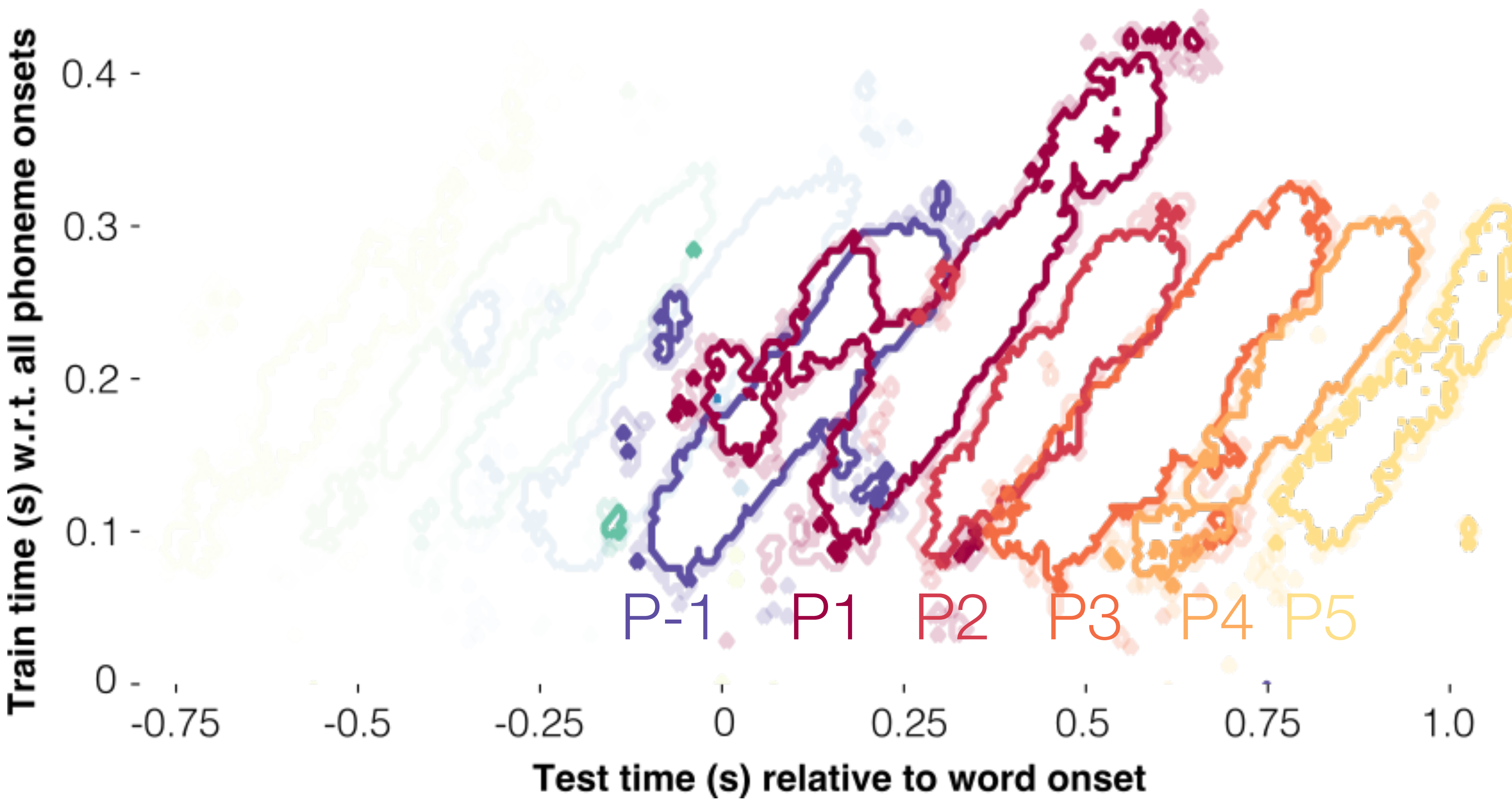
Gwilliams et al. (2022) *NatComms*

Transformation of acoustic signal

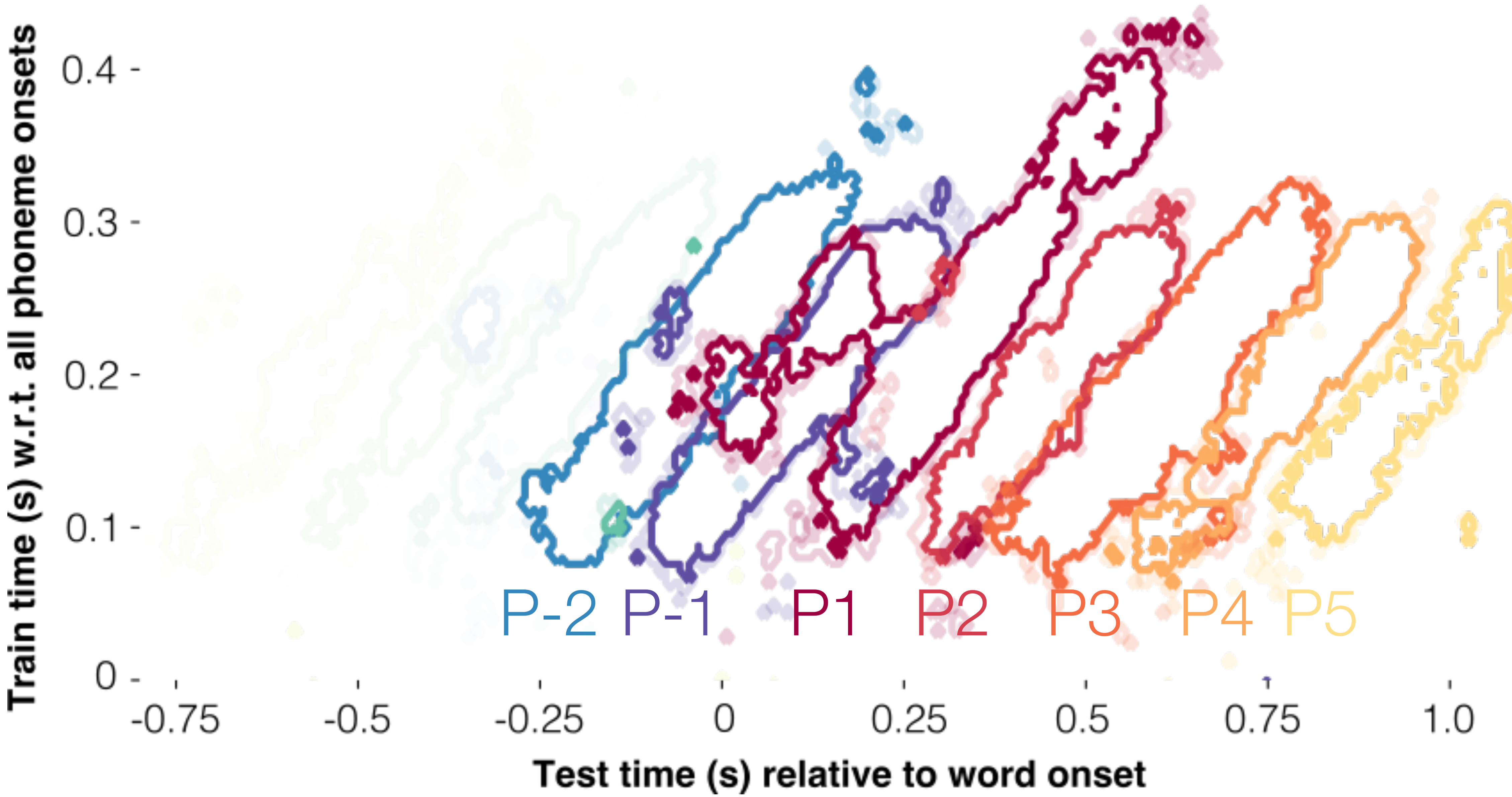


Gwilliams et al. (2022) *NatComms*

Transformation of acoustic signal

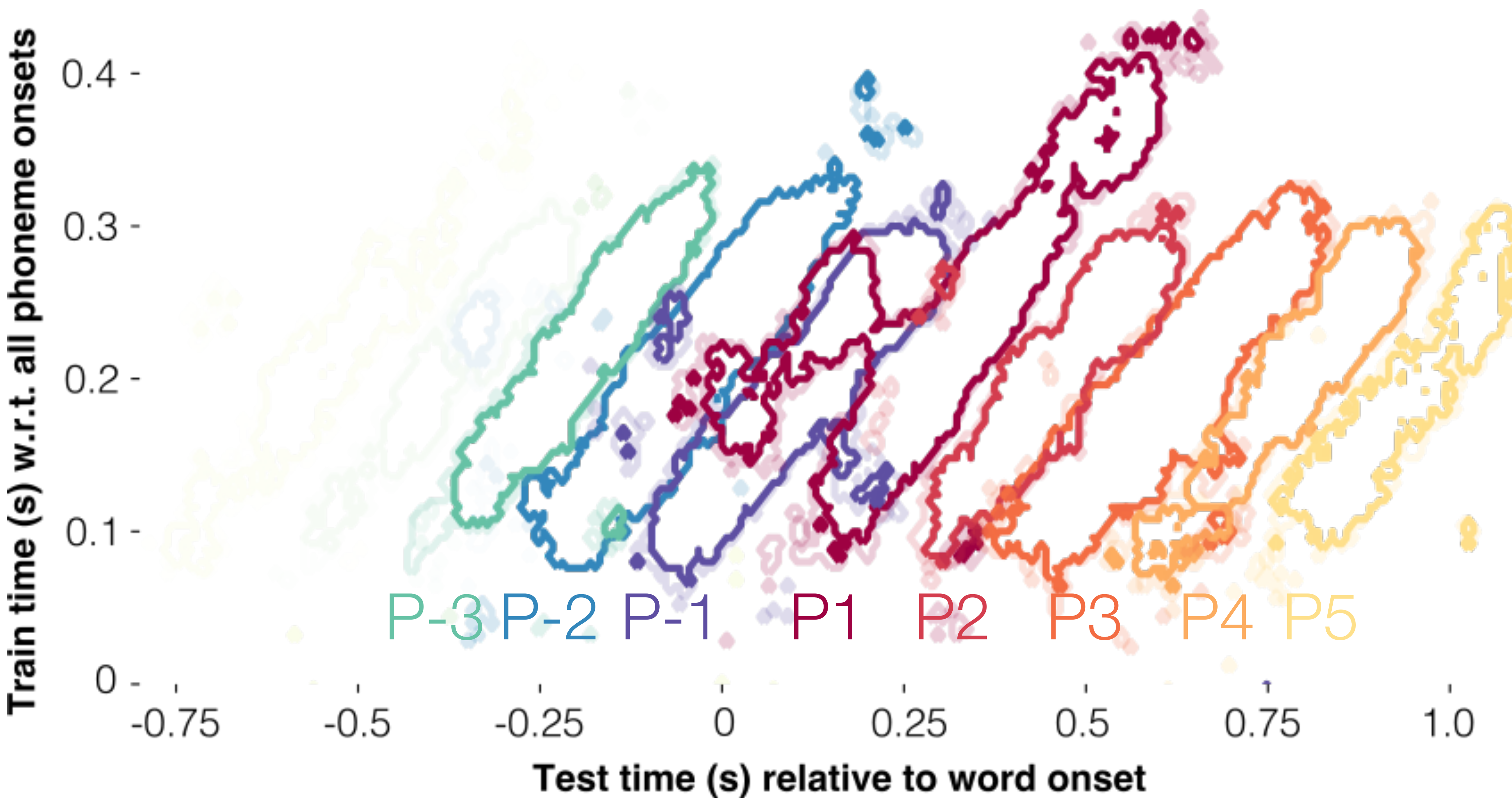


Transformation of acoustic signal



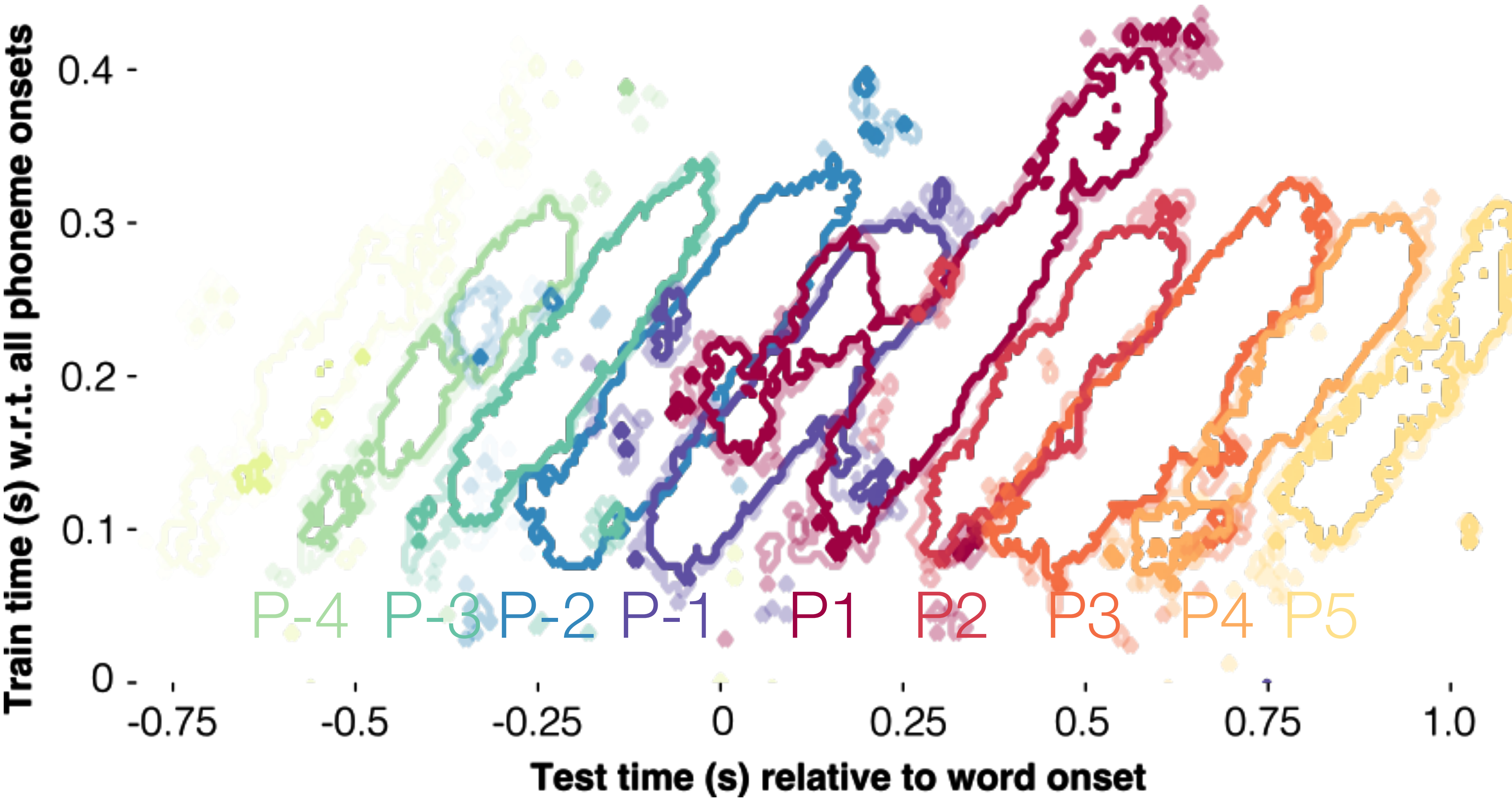
Gwilliams et al. (2022) *NatComms*

Transformation of acoustic signal



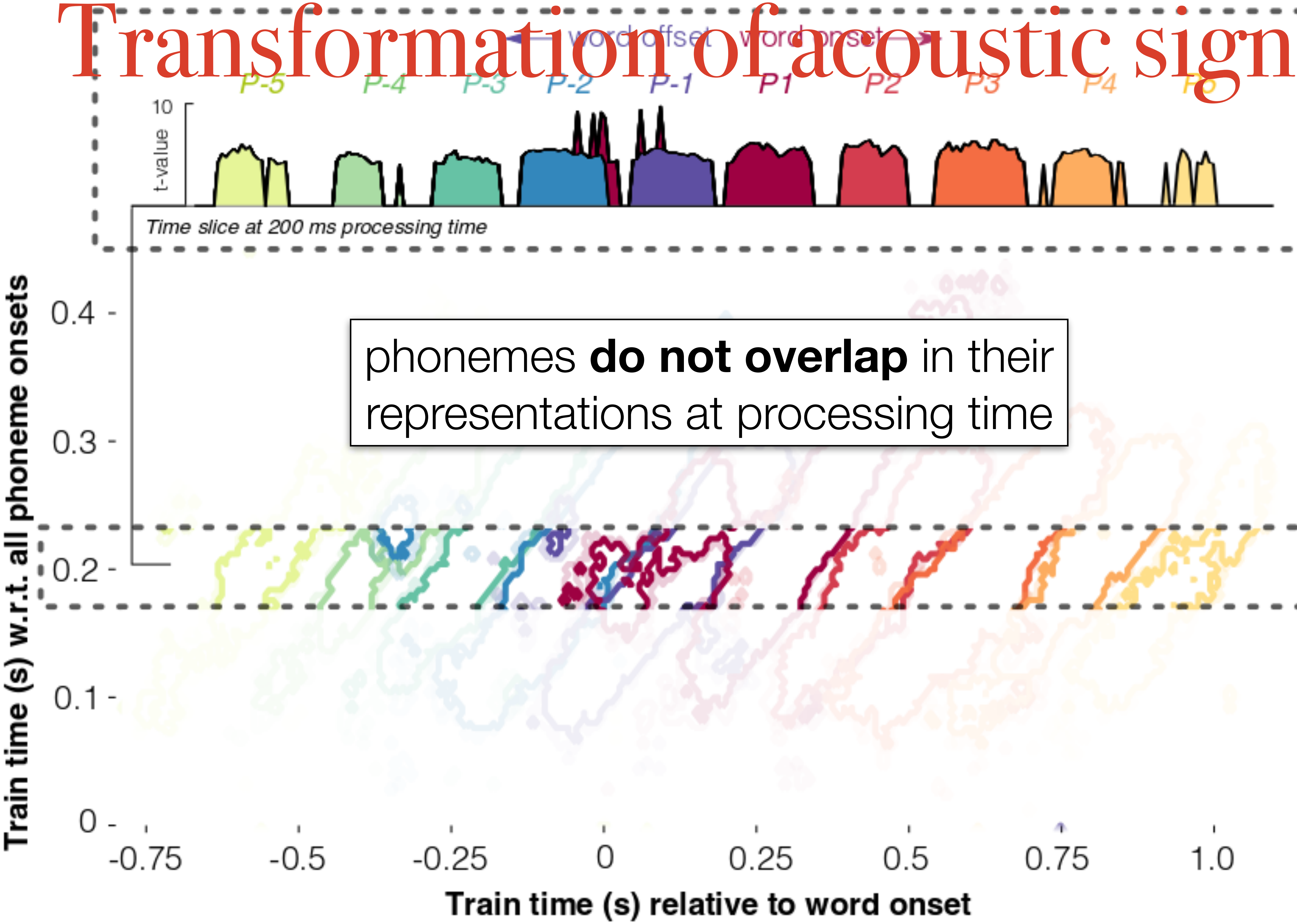
Gwilliams et al. (2022) *NatComms*

Transformation of acoustic signal

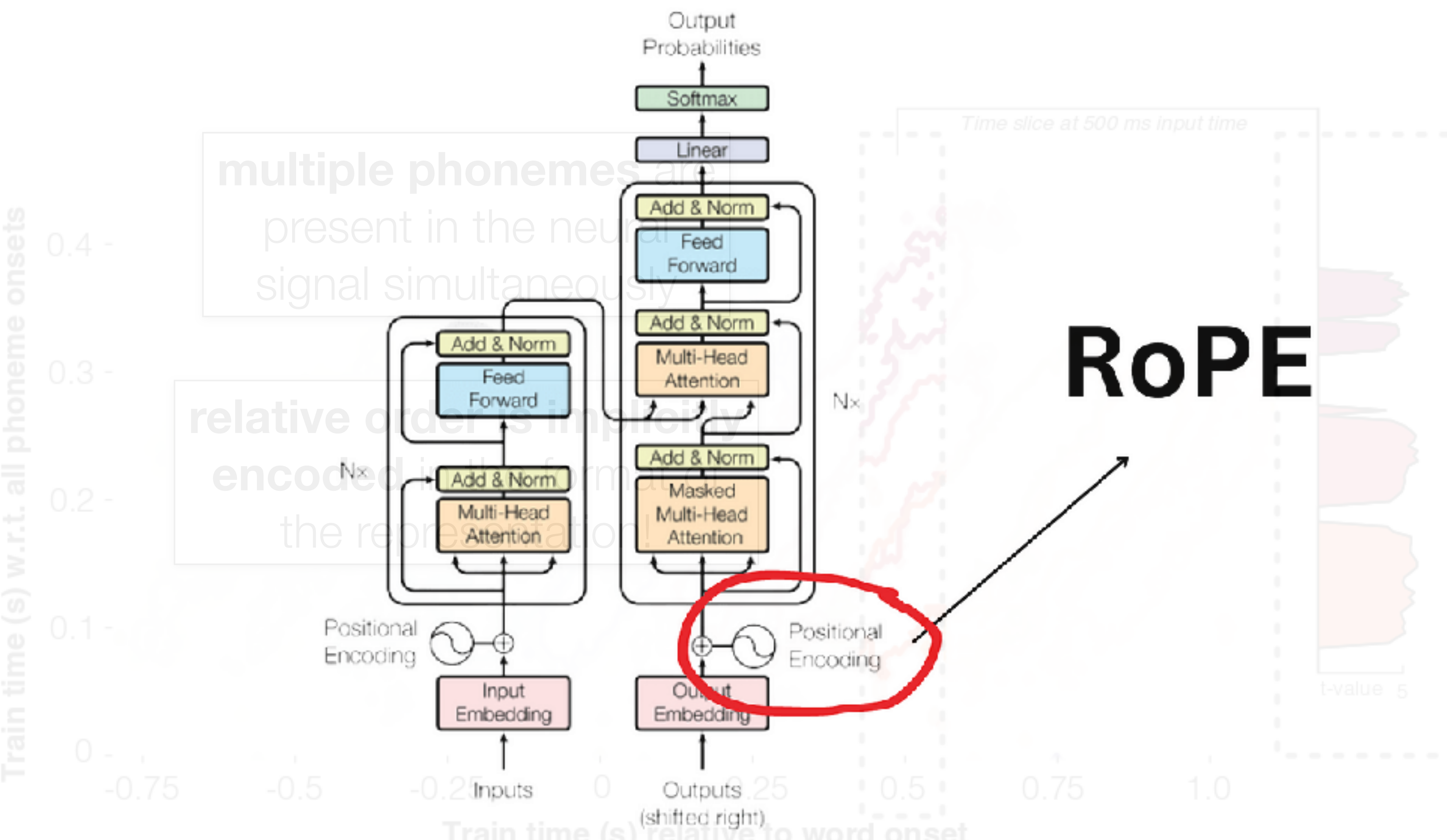


Gwilliams et al. (2022) *NatComms*

Transformation of acoustic signal



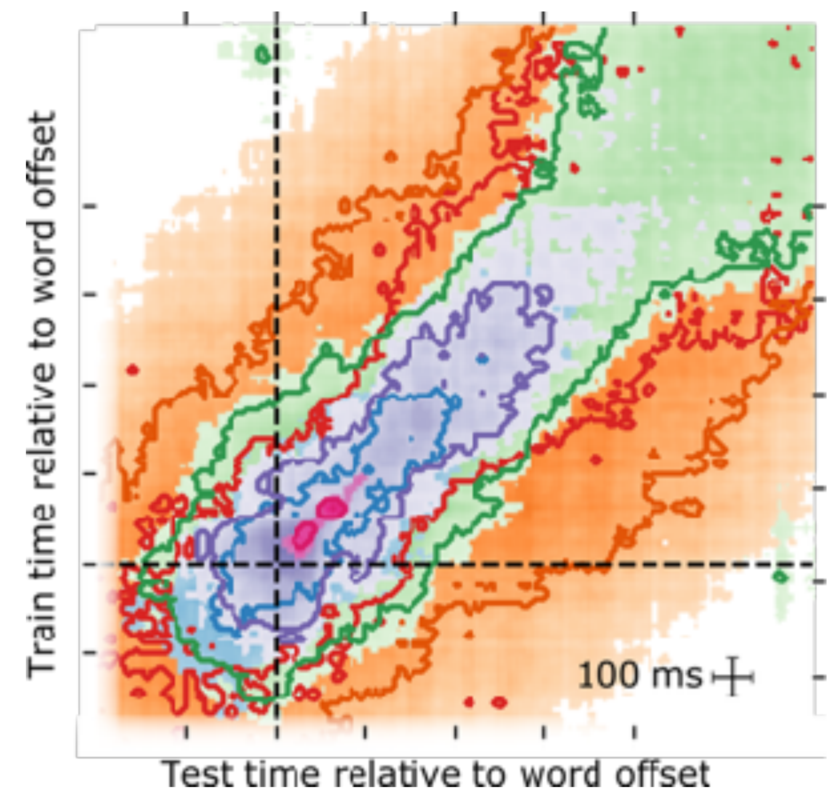
phonemes **do not overlap** in their representations at processing time


Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
Nx
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Nx
Positional Encoding
Positional Encoding
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)

RoPE

Nested hierarchy

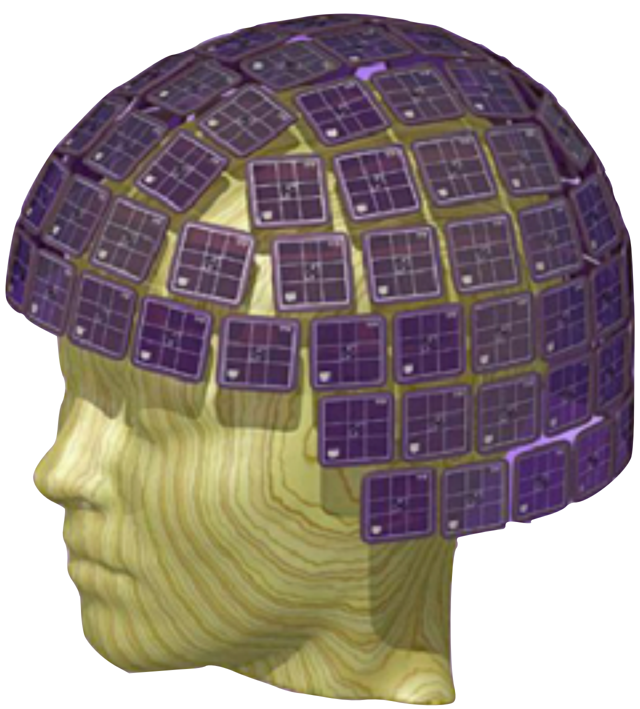
- Dynamic trajectory for all representational levels
- Speed of evolution scales with unit size



Gwilliams et al. (2024) *Submitted*

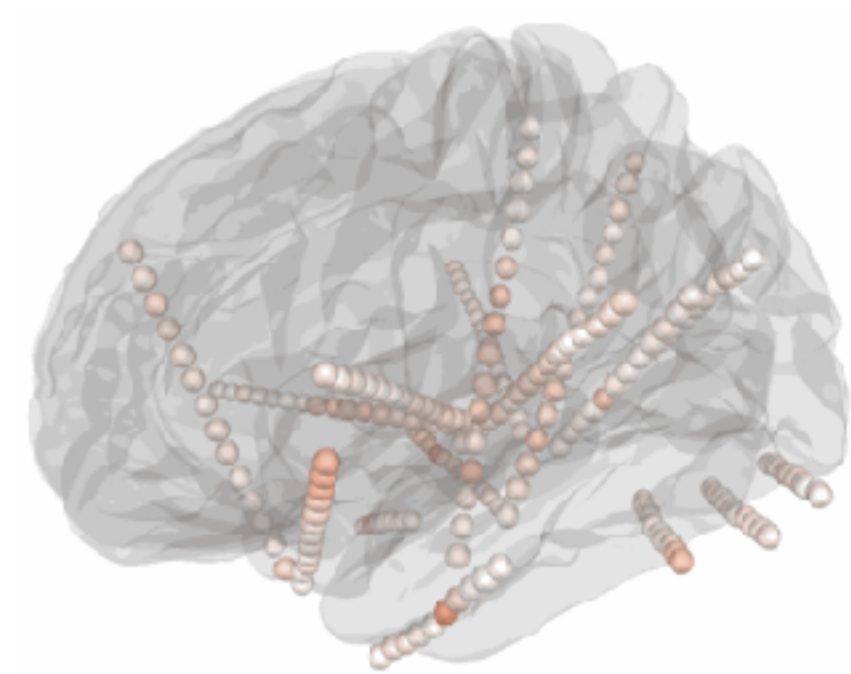
implementation in single neurons

Across neuronal scales



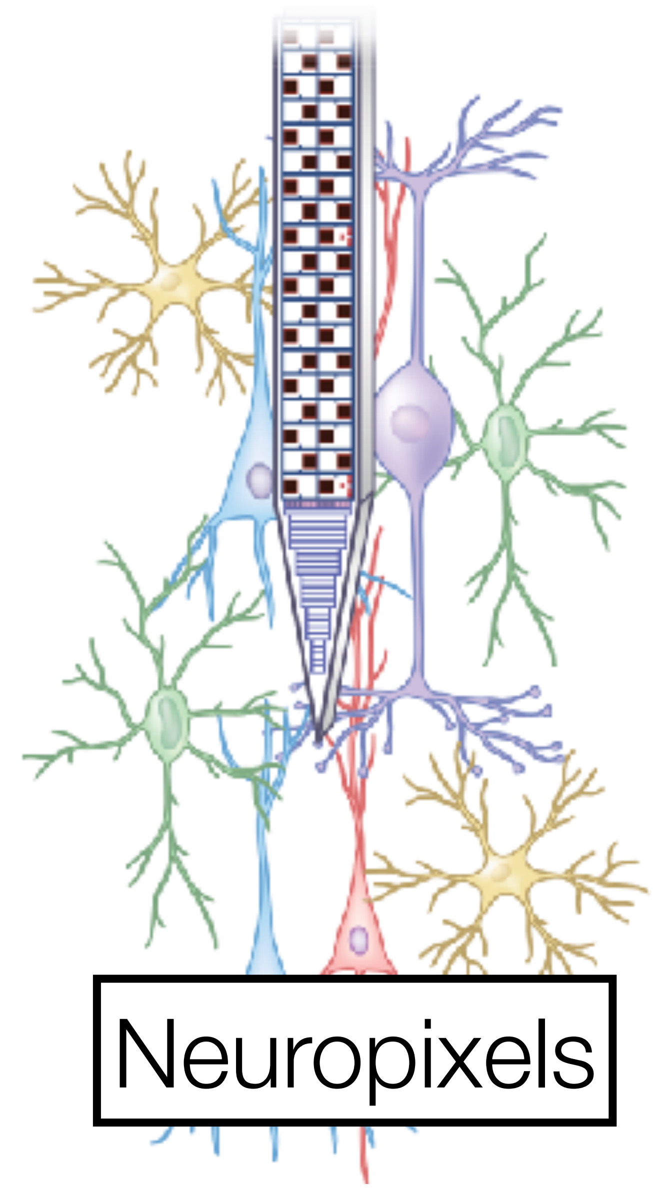
MEG / EEG

~100,000 neurons



iEEG

~10,000 neurons



Neuropixels

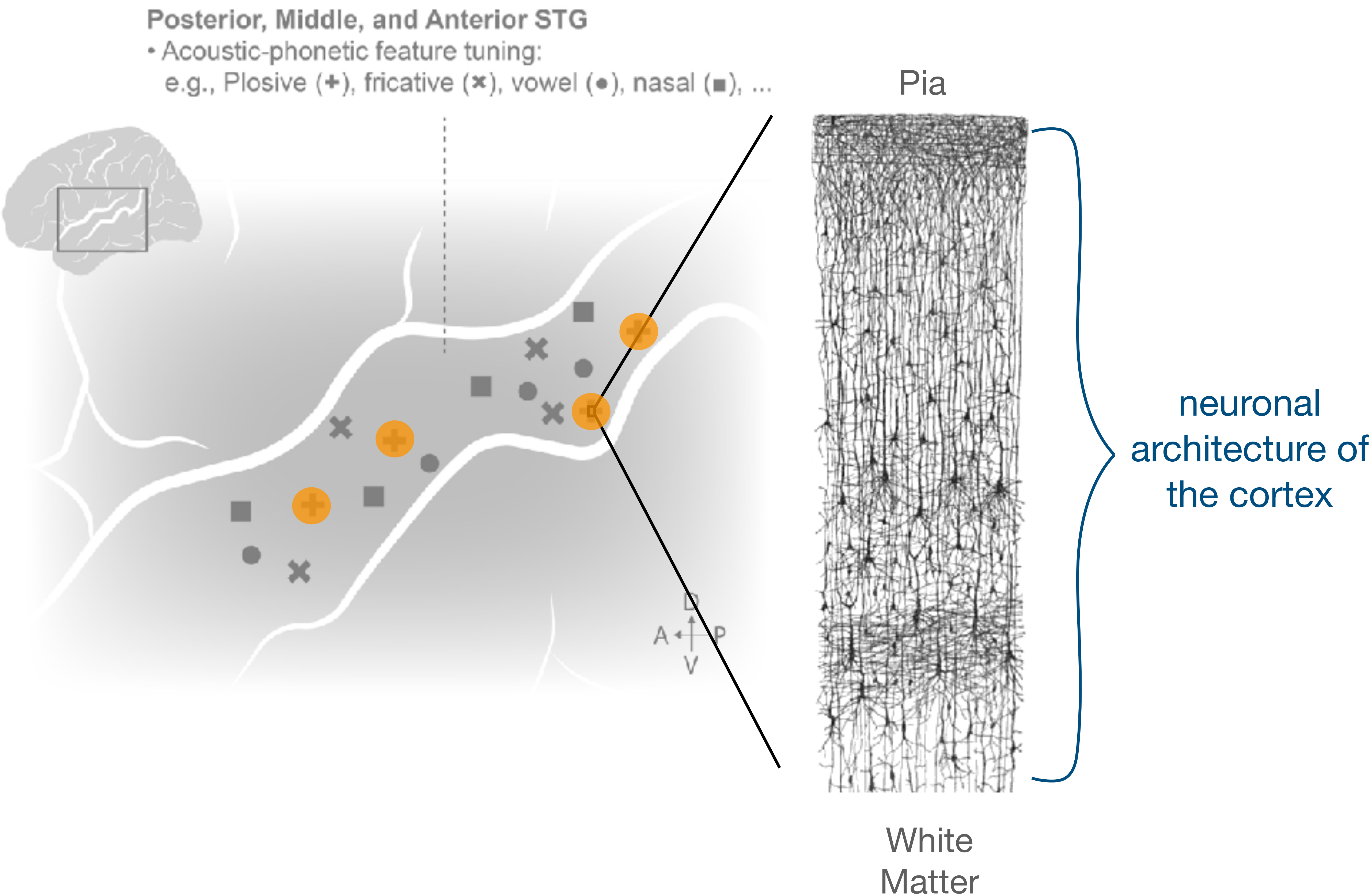
~1 neurons

Superior temporal gyrus (STG) is a core area for speech perception in the human brain

STG neural populations encode **acoustic-phonetic features**

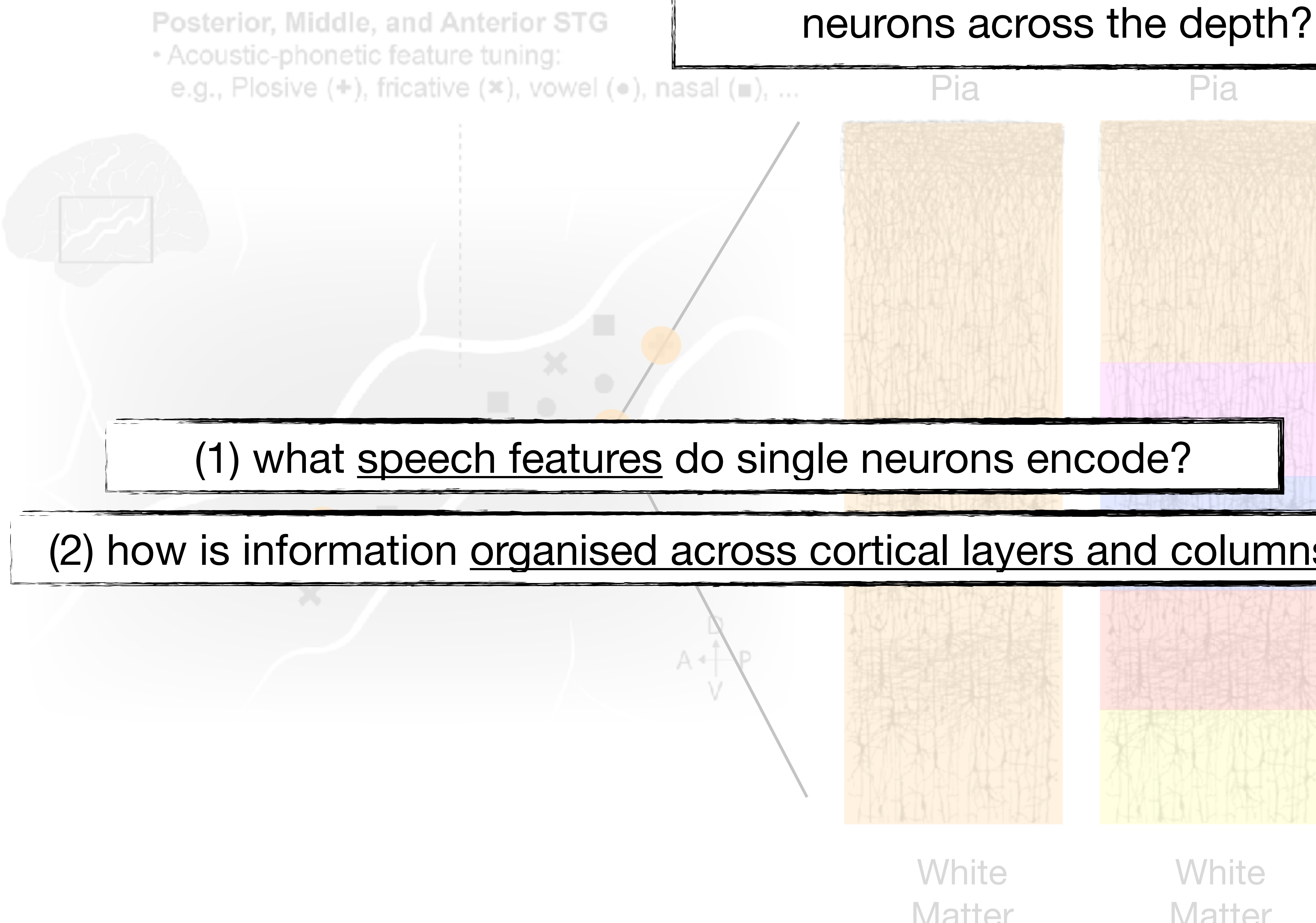
Broader spatial organization for **prosodic features** (onset, envelope, pitch)

Gwilliams*, Leonard* et al. *(2023)* Nature



Gwilliams*, Leonard* et al. *(2023)* Nature

(0) … how do you record single neurons across the depth?



(1) what speech features do single neurons encode?

(2) how is information organised across cortical layers and columns?

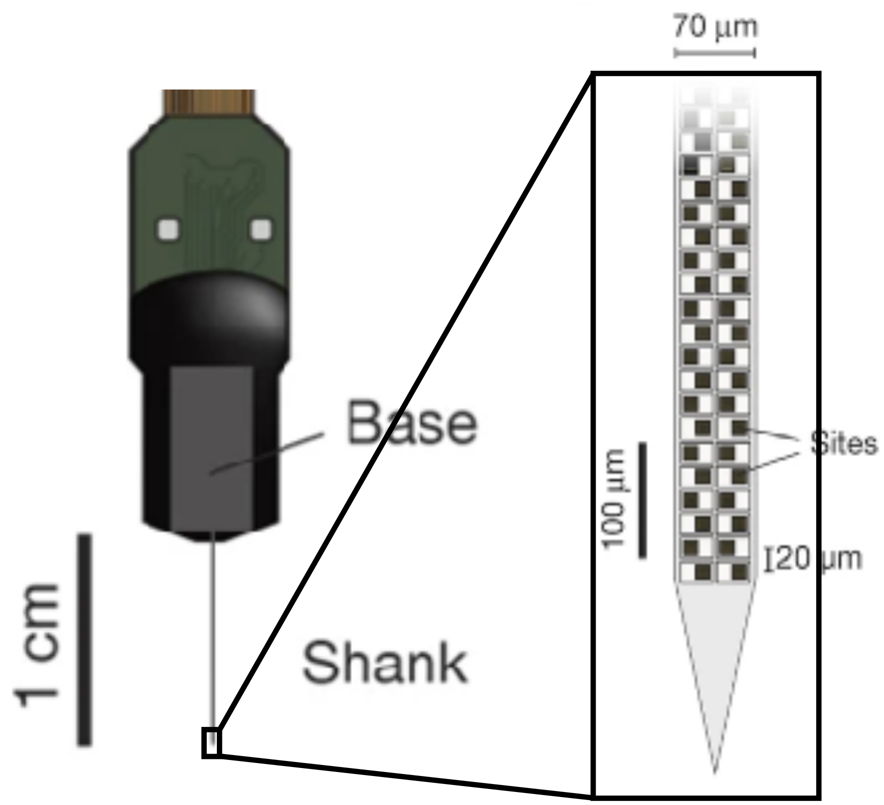
Gwilliams*, Leonard* et al. *(2023)* Nature

Neuropixels

960 electrodes
(384 simultaneous recording)

(0) … how do you record single neurons across the depth?

Revolutionary leap in recording capability for animal models

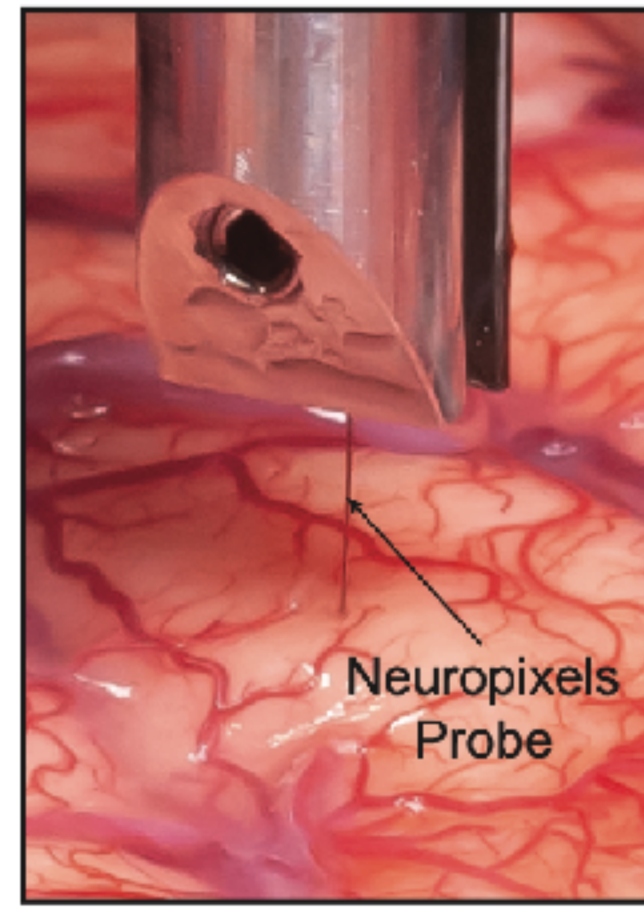


Jun *et al*., 2017, *Nature*

Simultaneous recording of 100-1000s of neurons

Recently translated into humans

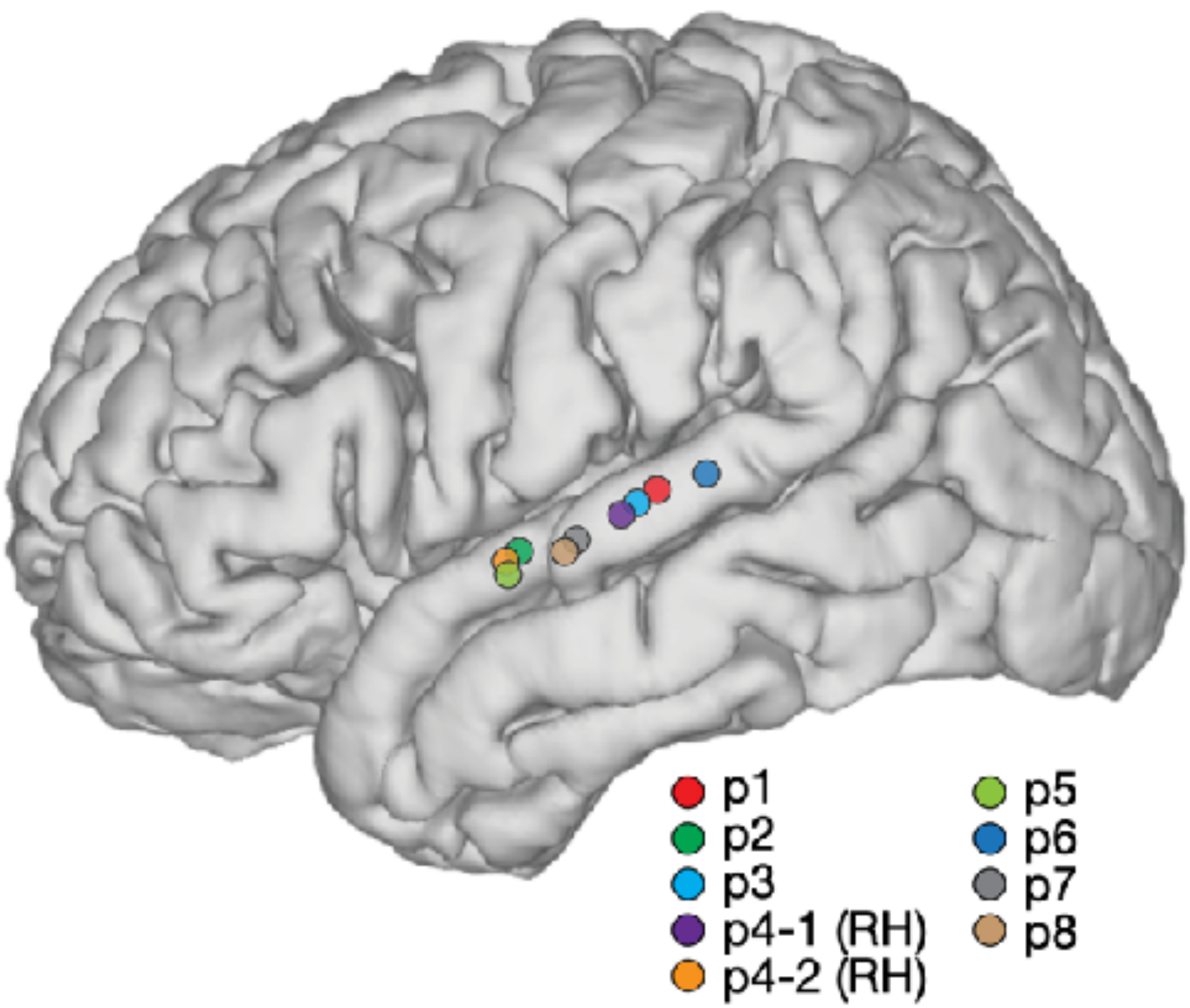
Chung*, Sellers*, **Gwilliams** *et al*., 2022, *Neuron*
Paulk *et al*., 2022, *Nature Neuroscience*

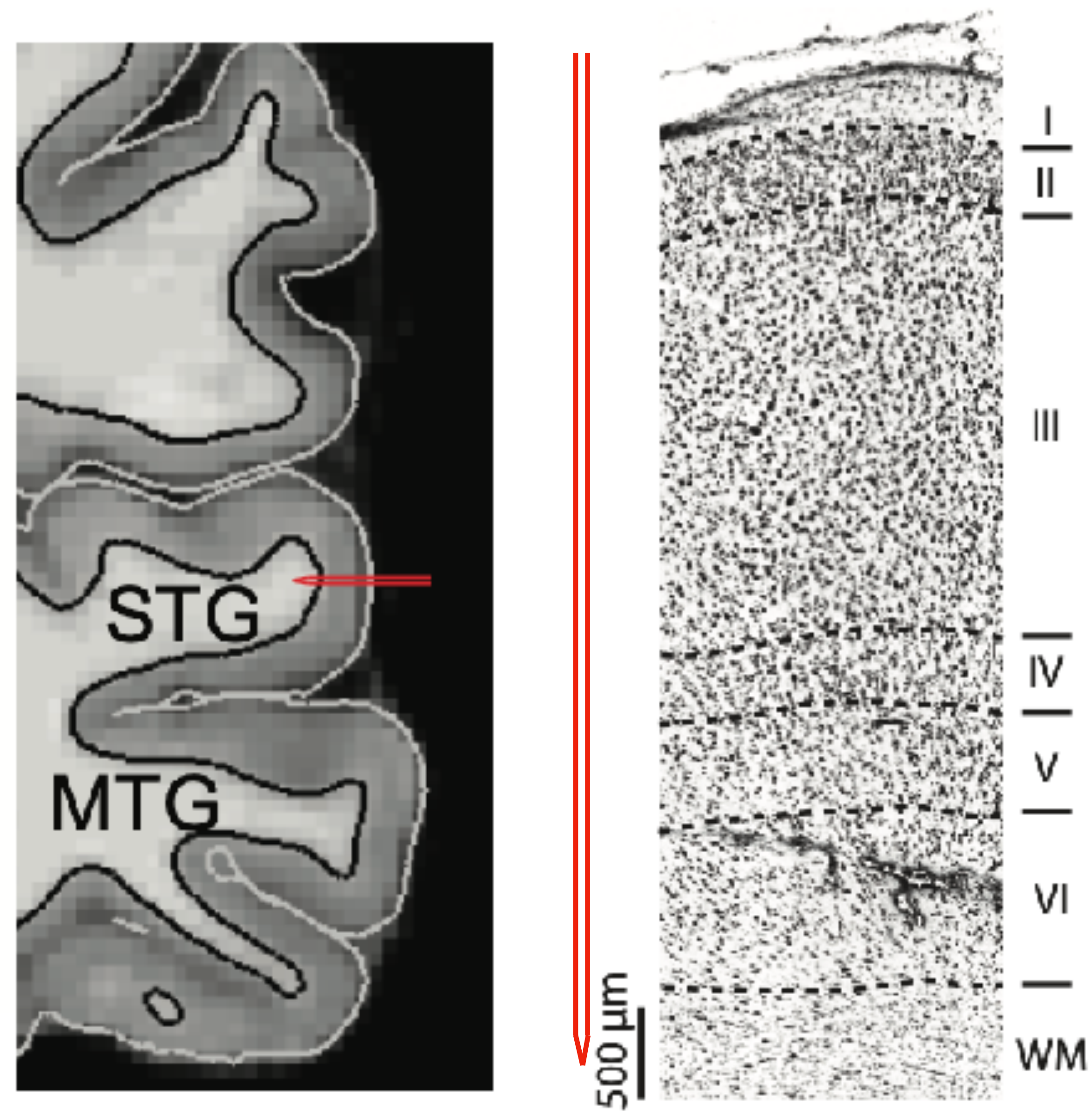


Gwilliams*, Leonard*, *et al*., 2023, *Nature*

~3 years of work to do this safely and efficiently in humans

Simultaneous recording from all cortical layers



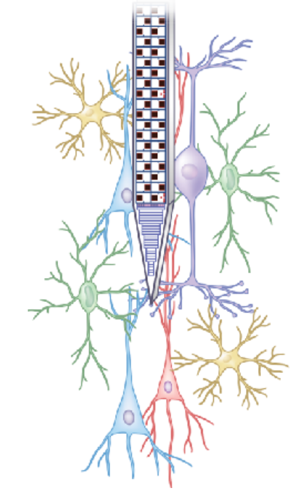


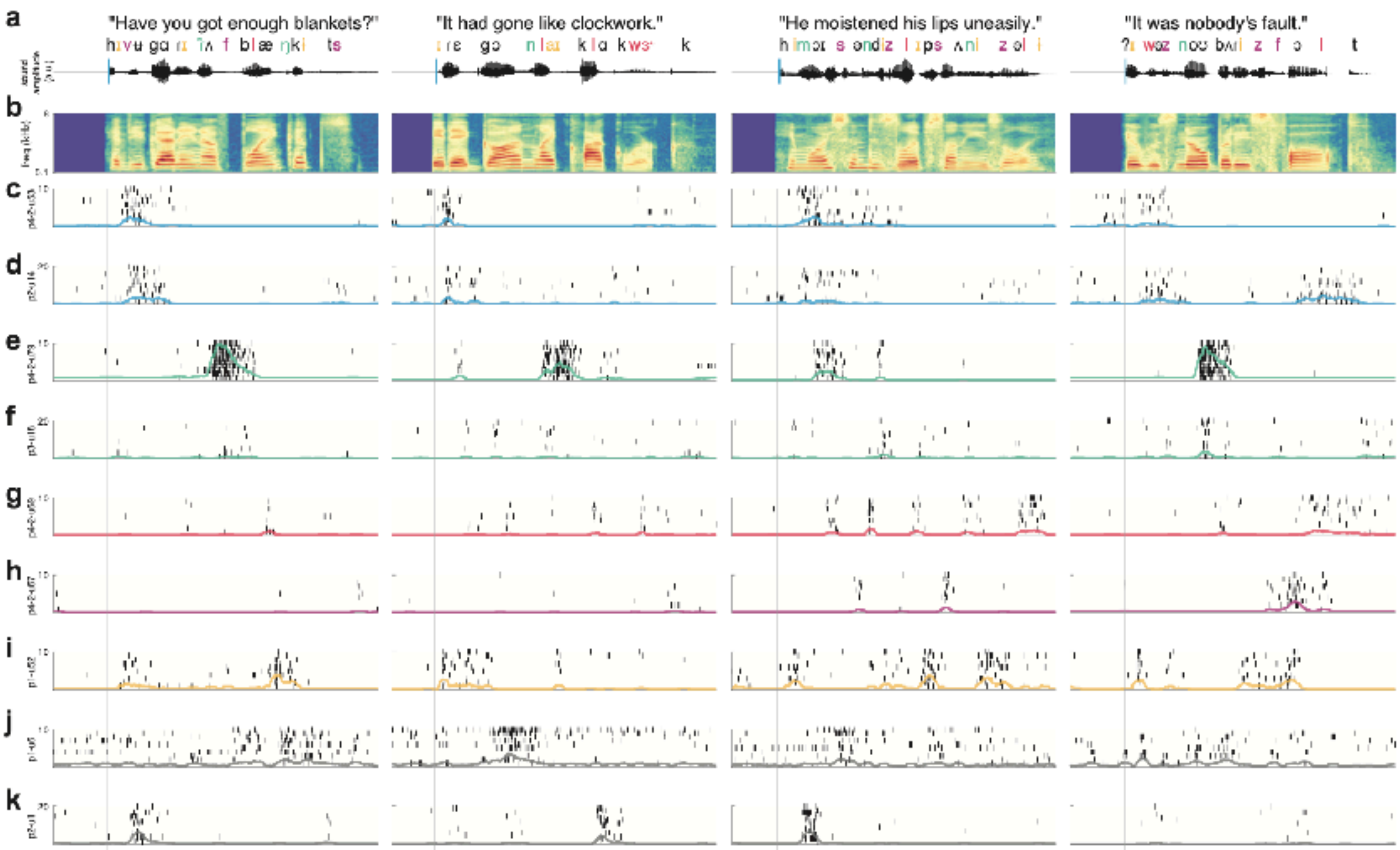
it was nobody's fault

junior what on earth's thematter with you

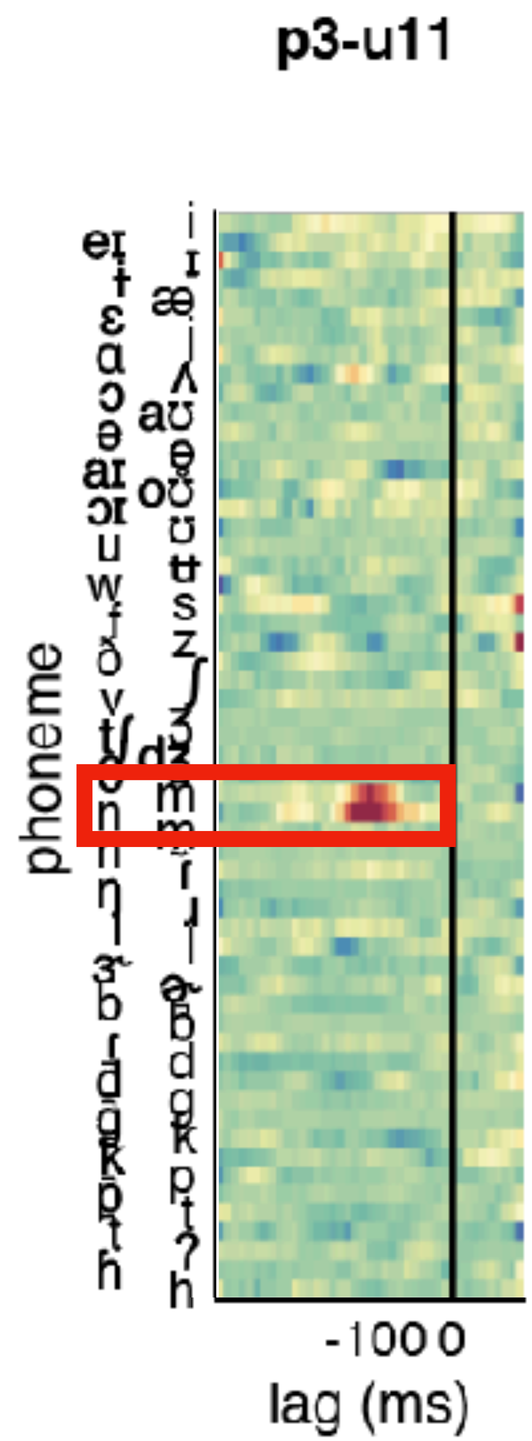
Article

Large-scale single neuron encoding of speech sounds across the depth of human cortex



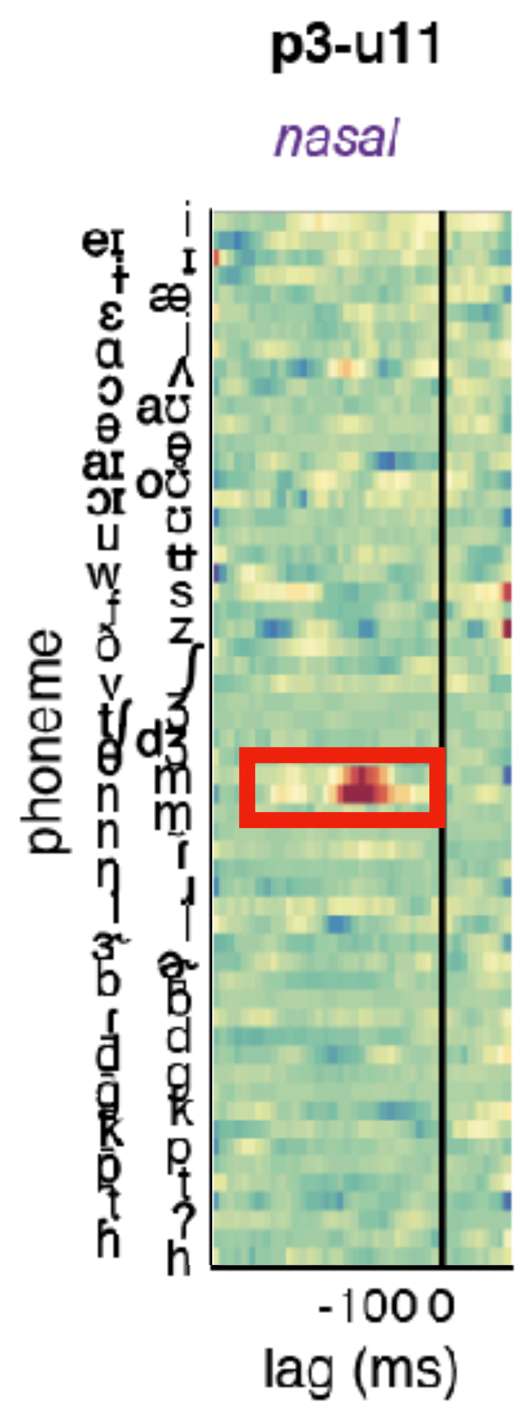


(1) what speech features do single neurons encode?



Linear temporal receptive field (TRF) encoding model weights

(1) what speech features do single neurons encode?



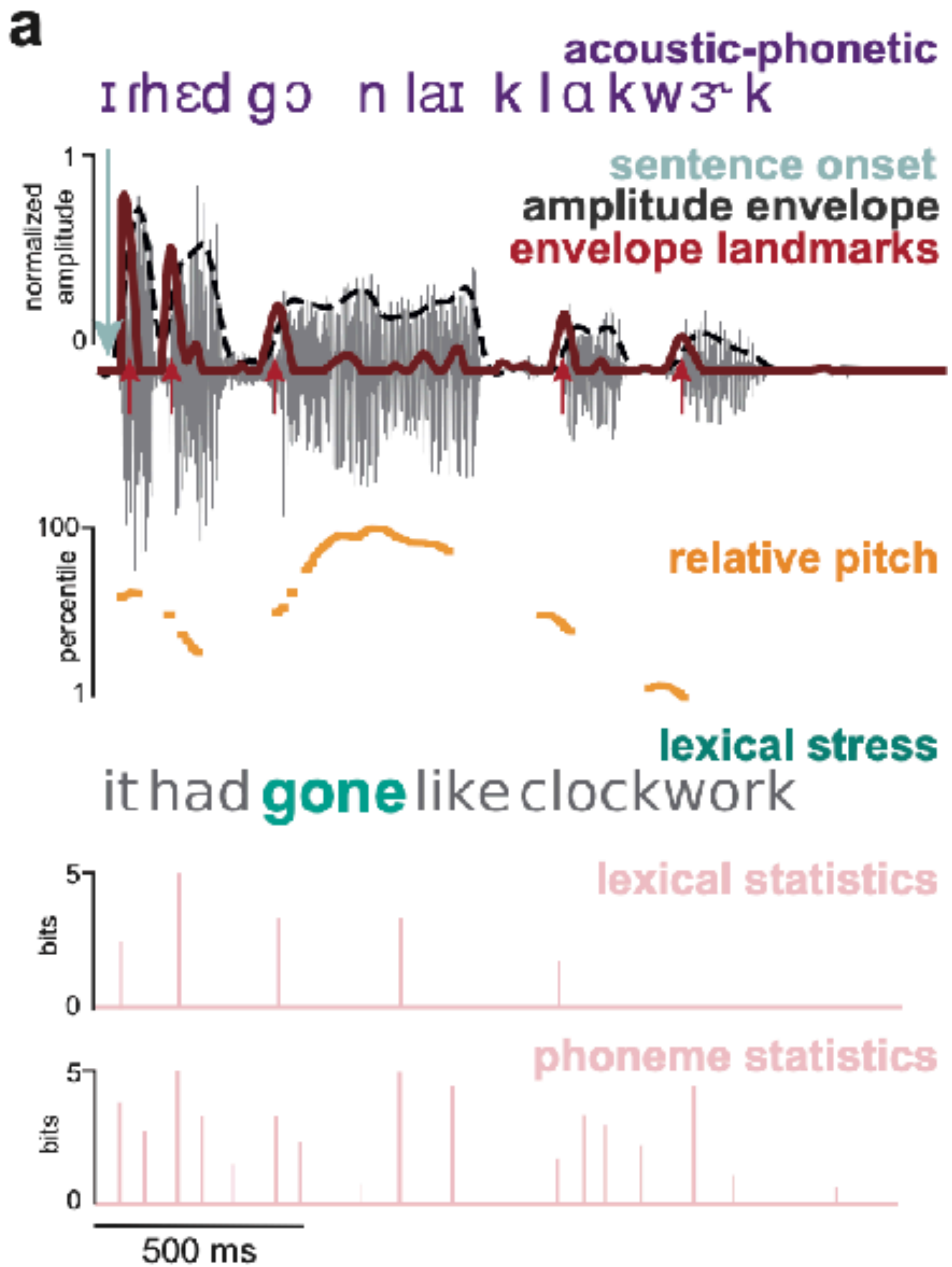
Tuning to phonetic features

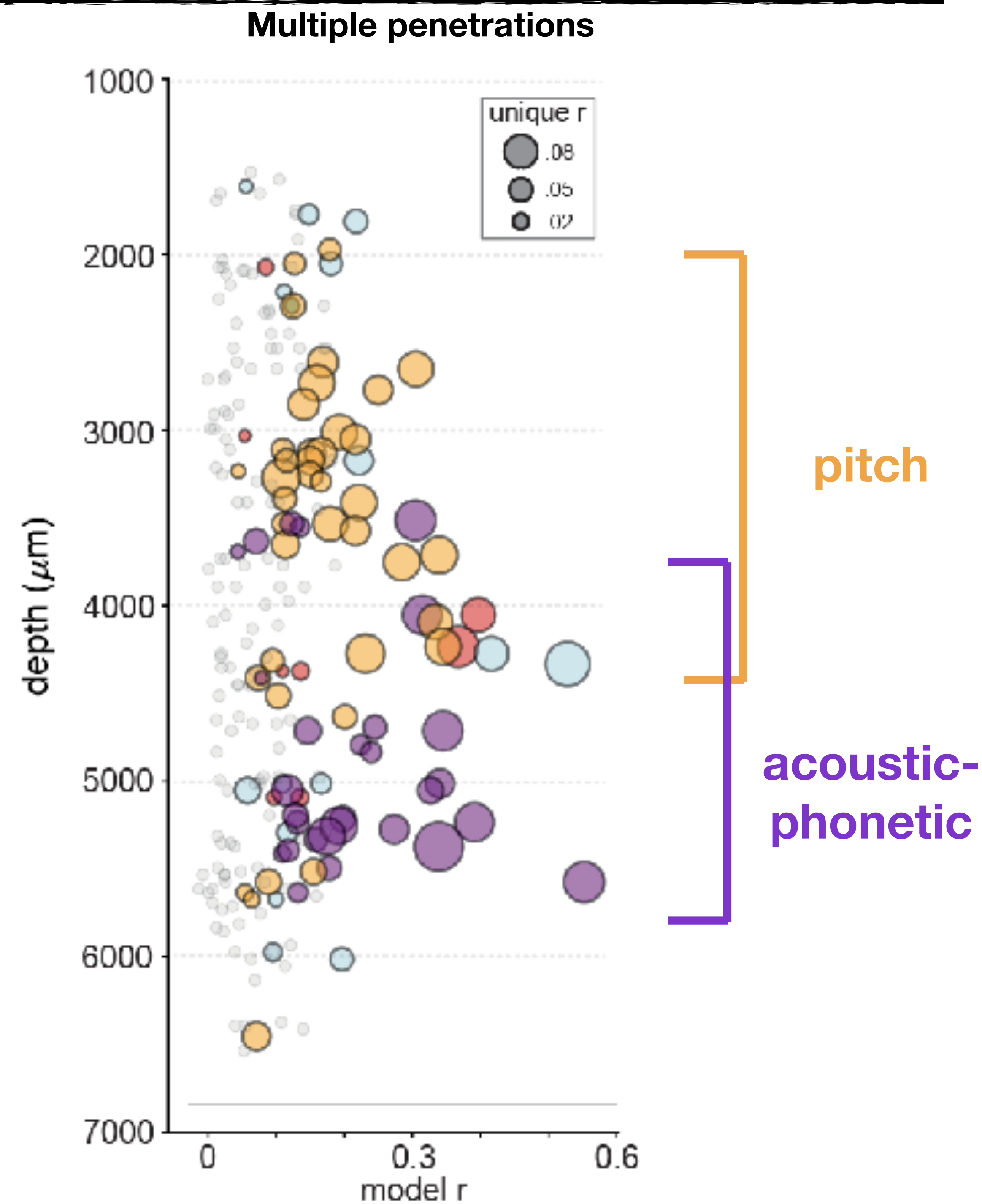
No tuning to individual phonemes

Linear temporal receptive field (TRF) encoding model weights

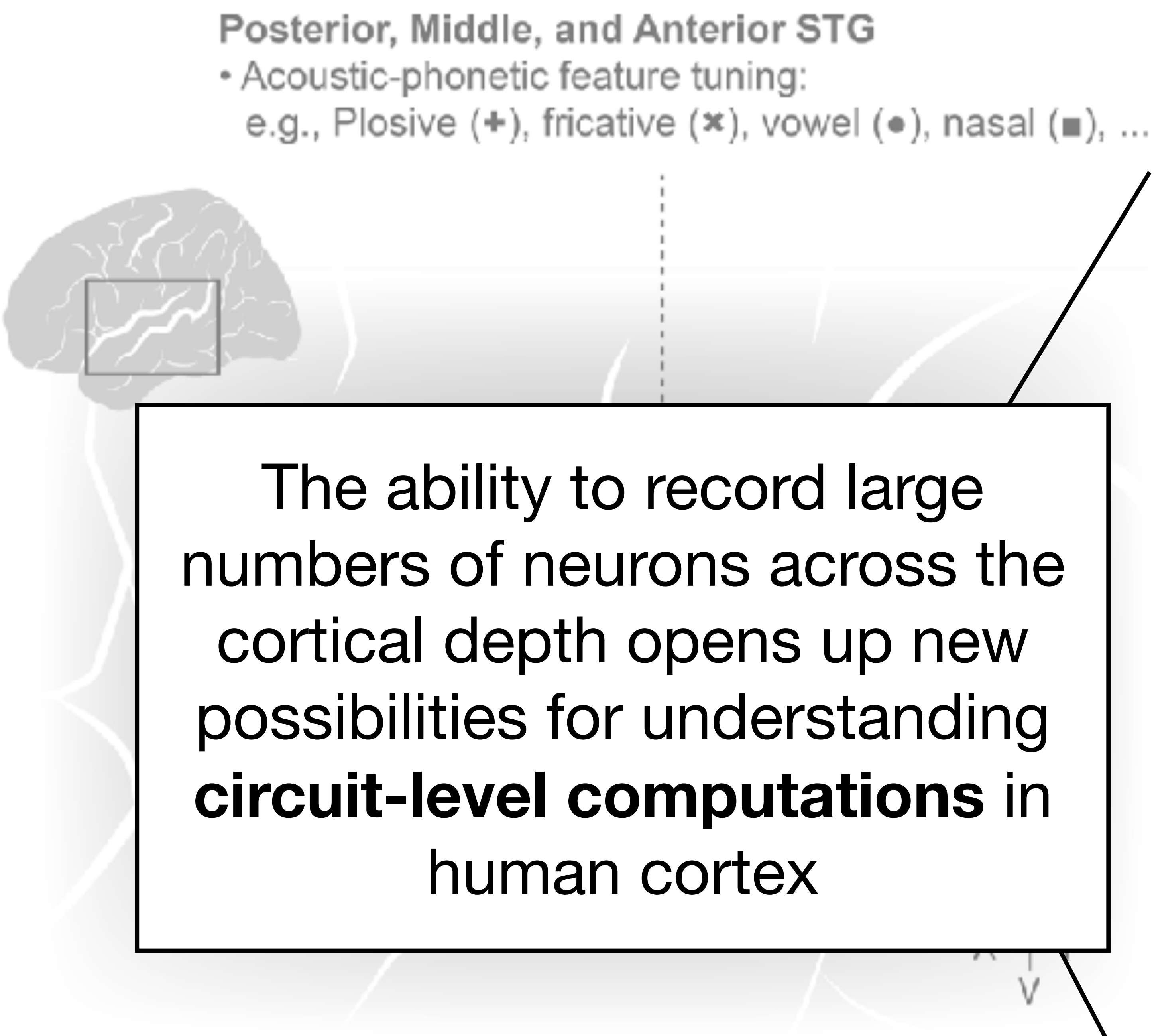
(1) what speech features do single neurons encode?

(2) how is information organised across cortical **layers**?

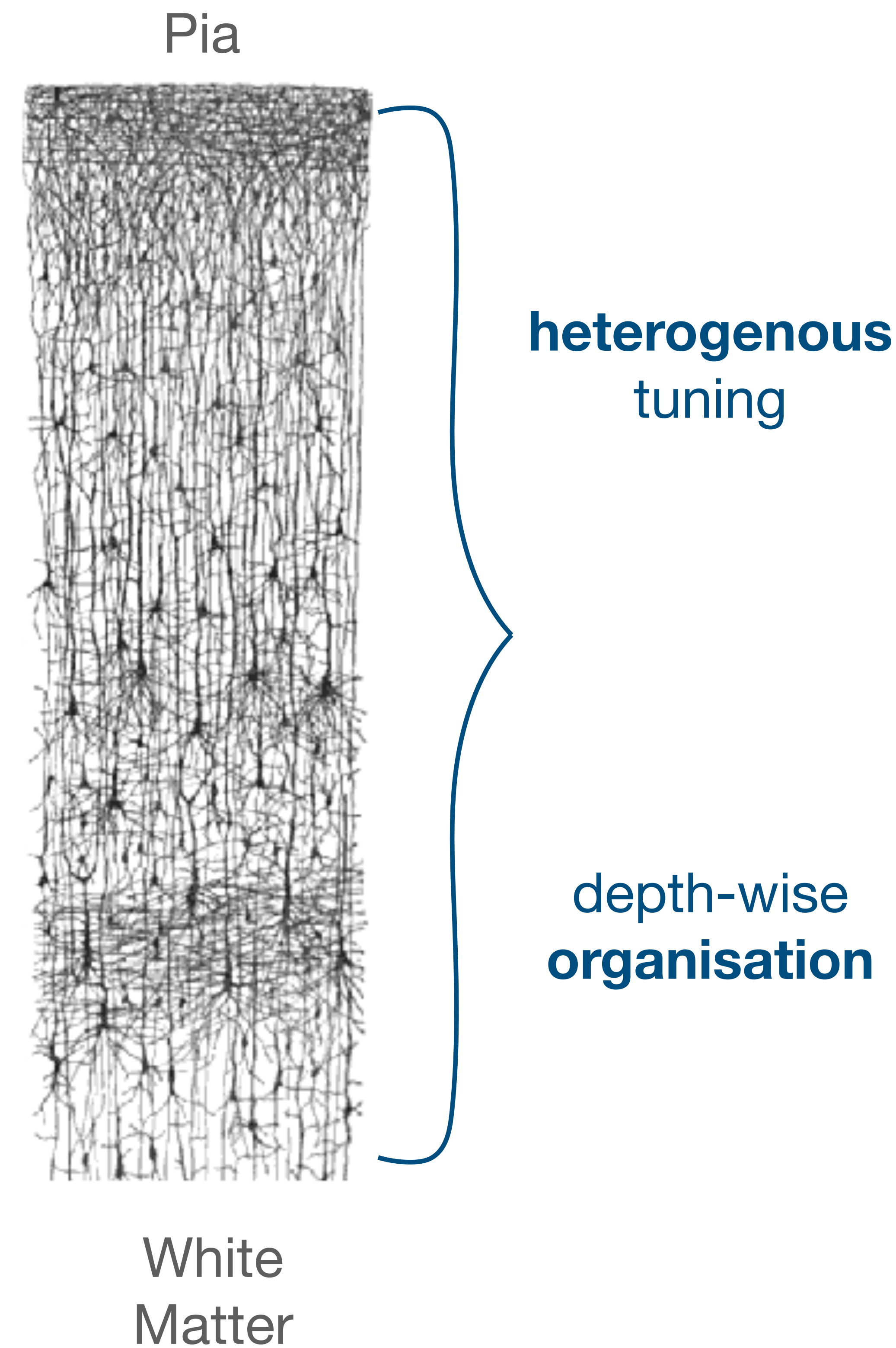




Gwilliams*, Leonard* et al. *(2023)* Nature



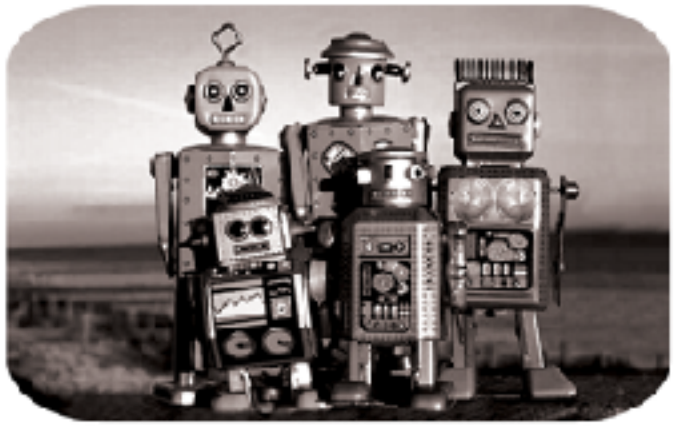
The ability to record large numbers of neurons across the cortical depth opens up new possibilities for understanding **circuit-level computations** in human cortex



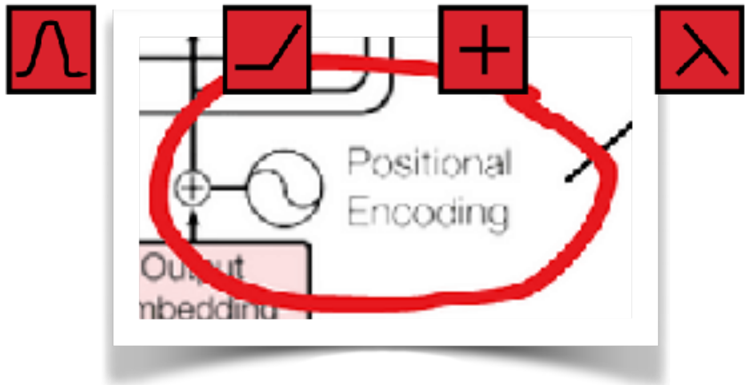
Gwilliams*, Leonard* et al. *(2023)* Nature

Ongoing Work

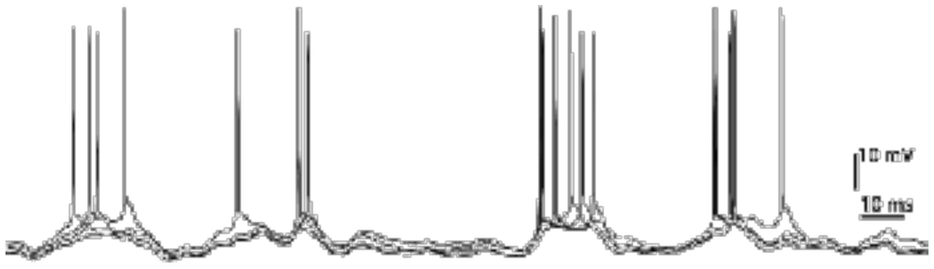
- Isolate and **interpret** symbolic representations across models



- Leverage LLMs to better understand underlying **computations**



- Use empirical spiking properties and receptive fields to build **biologically inspired speech models**



laura.gwilliams@stanford.edu
@GwilliamsL

Funding:

the brain foundation
A-0741551370

WHITEHALL FOUNDATION
2024-08-043

Thank you!