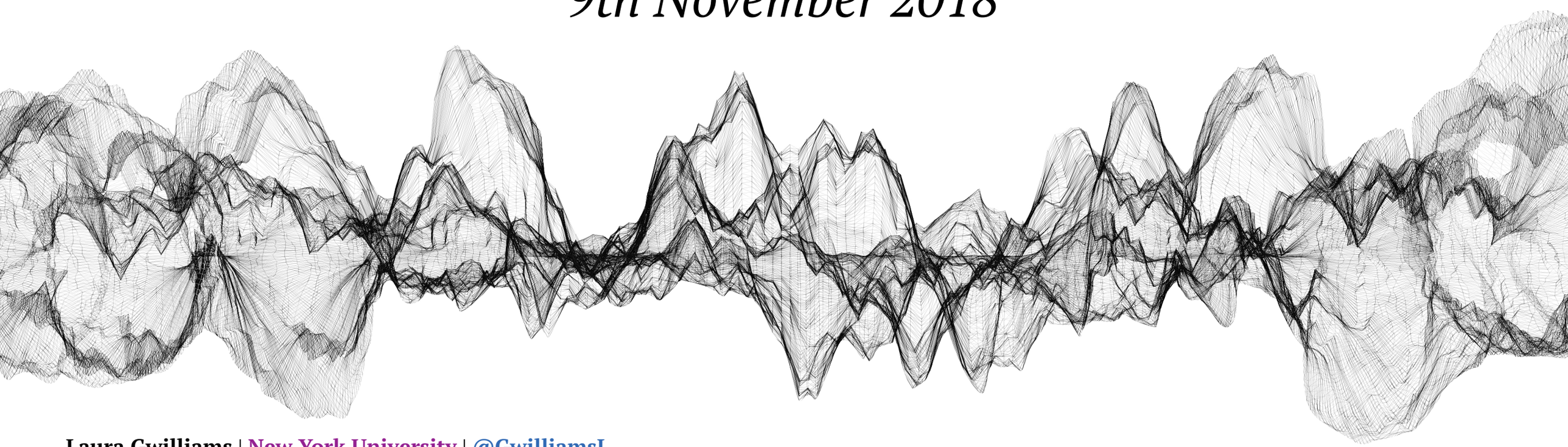
NEW YORK UNIVERSITY

Towards a mechanistic account of speech comprehension

Laura Gwilliams

9th November 2018



Laura Gwilliams | **New York University** | **@GwilliamsL**

Levels of analysis

1. Phonemes within words
   - Responses to phoneme ambiguity, phonetic features and acoustic properties **(bottom-up)**
   - Neural signatures of ambiguity resolution, when provided with lexical information **(top-down)**

2. Words within sentences
   - Which linguistic properties encoded in brain activity?
   - What are the relative time-courses of processing each property?
   - What is the computational architecture?

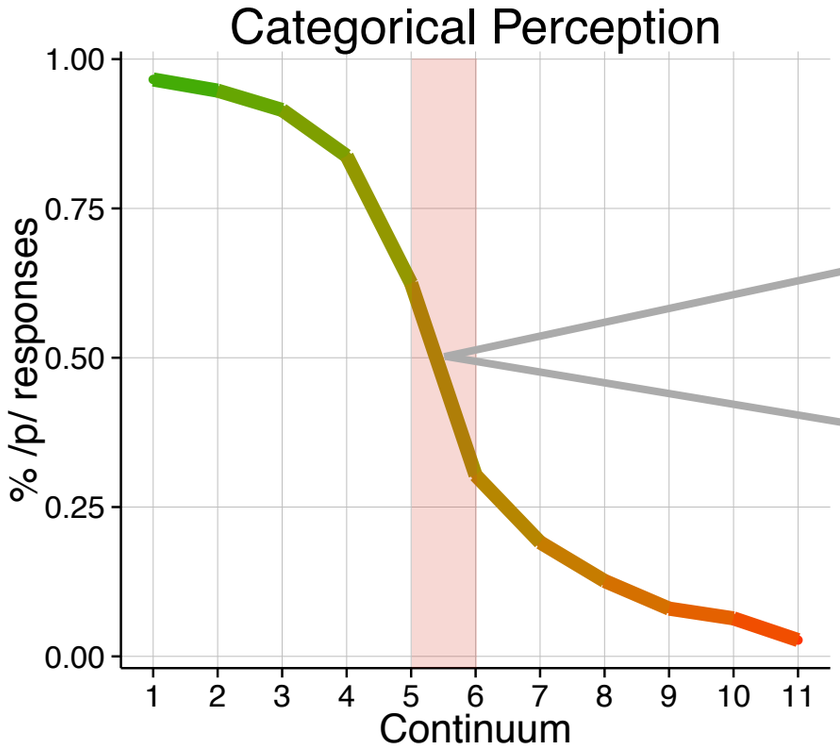
Future Influences on Perception

- Speech is an inherently **noisy and ambiguous** signal

- To fluently derive meaning, listeners must **integrate top-down** contextual information to guide their interpretation

- Top-down input occurring *after* an acoustic signal can be integrated to **affect the perception of earlier sounds**
  (Bicknell et al., submitted; Connine et al., 1991; Samuel, 1981; Szostak & Pitt, 2013; Warren & Sherman, 1974)

Future Influences on Perception



p

b

ee t

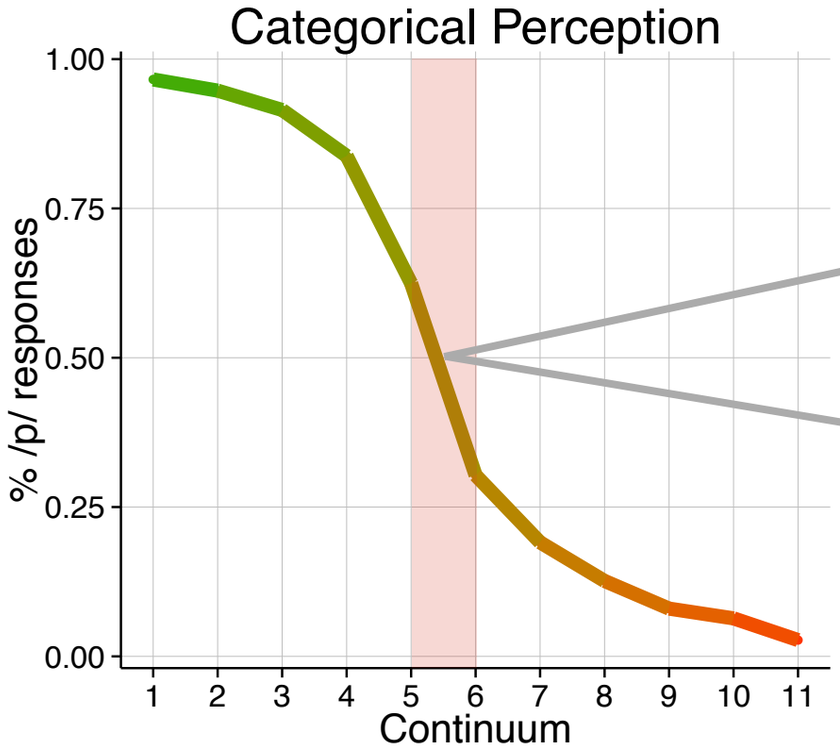
ai d

(this is a parakeet)

(this is a barricade)

“**P**oint **o**f **D**isambiguation” (POD)

Future Influences on Perception



p a r a k **ee** t

b a r a k **ai** d

(this is a parakeet)

(this is a barricade)

"**P**oint **o**f **D**isambiguation" (POD)

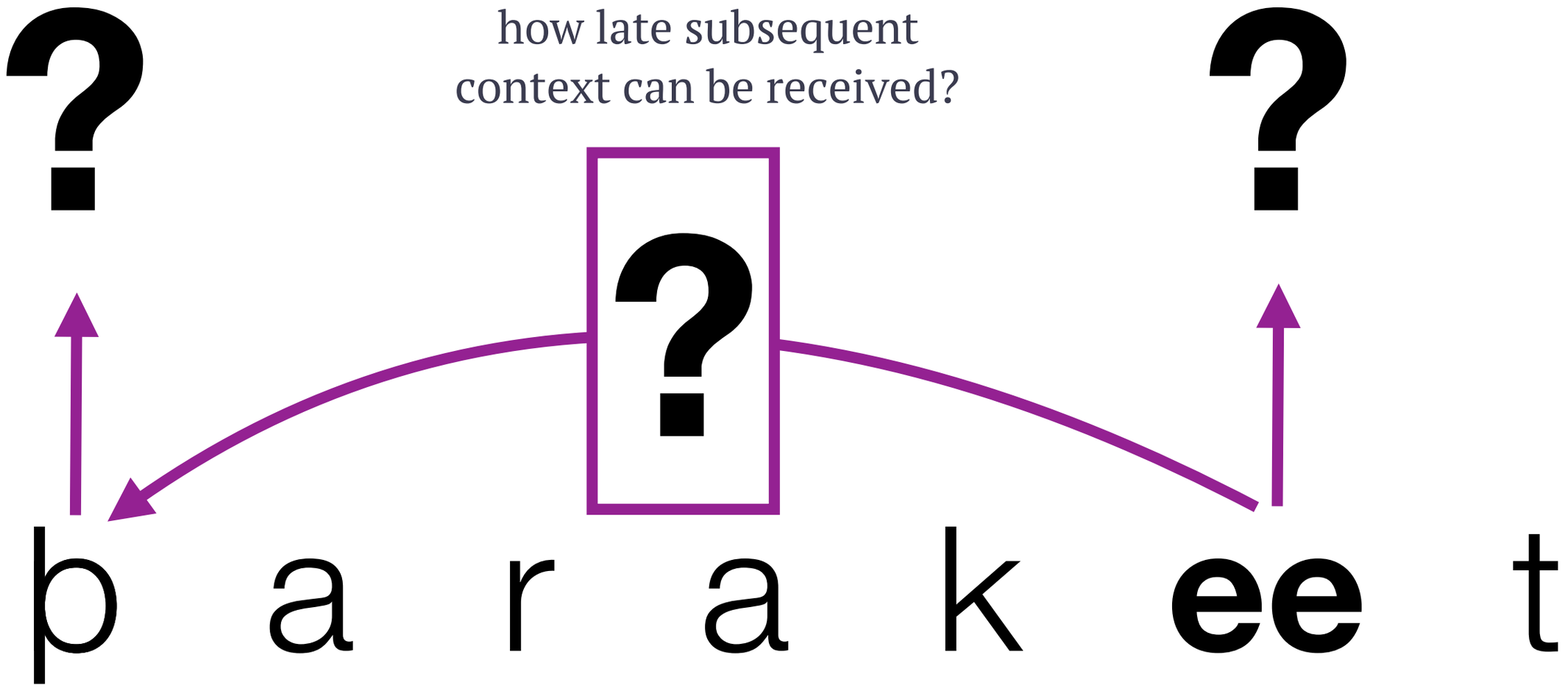
Today's Questions

i could focus on the ambiguity resoldtion part here, rather than the original response to ambiguity. then, tie in the ambiguity response part later, linking it with AI?

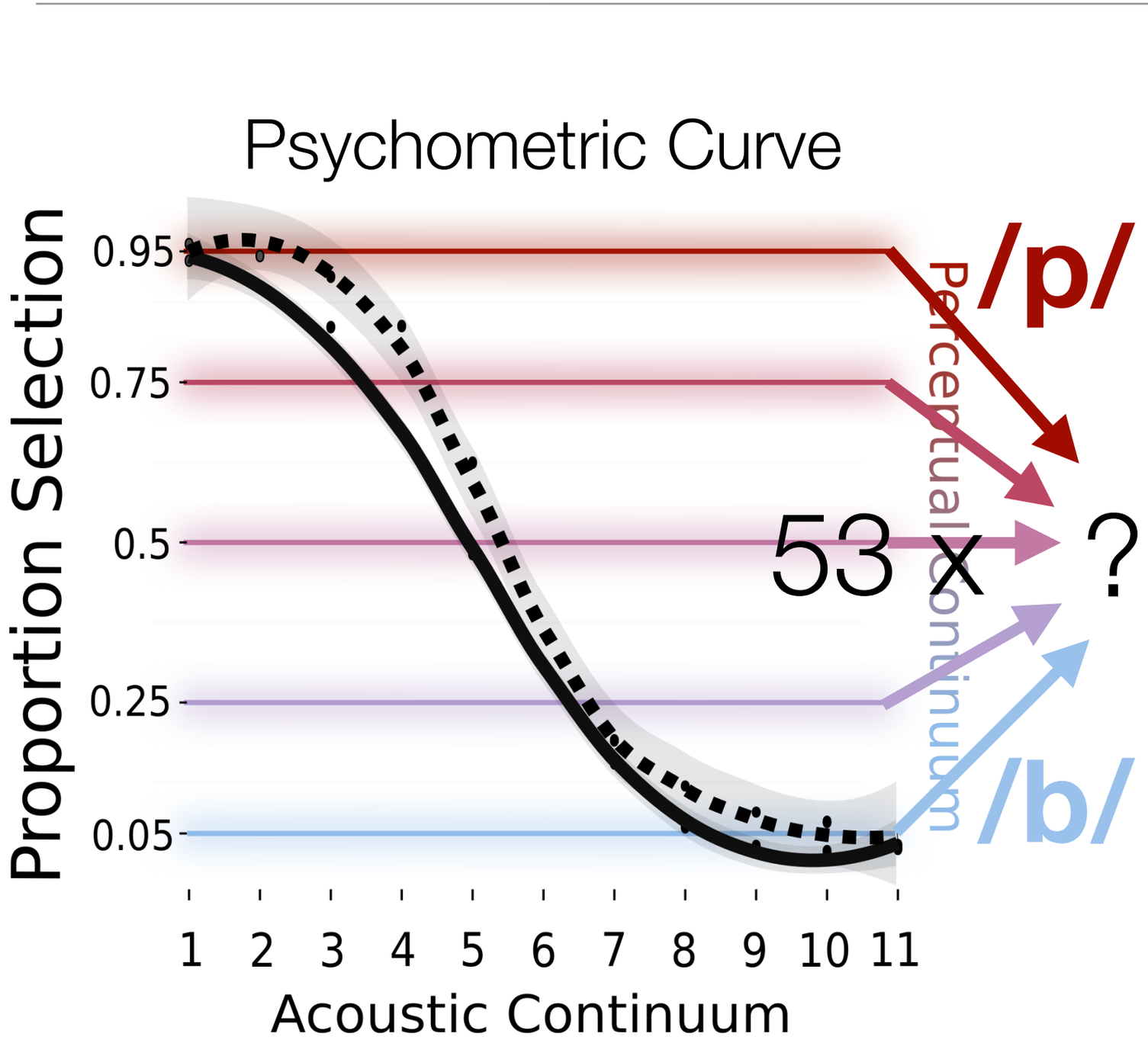
How does the auditory cortex **respond** to phonological ambiguity?

What is the **time-limit** on how late subsequent context can be received?

What are the neural signatures of ambiguity **resolution**?

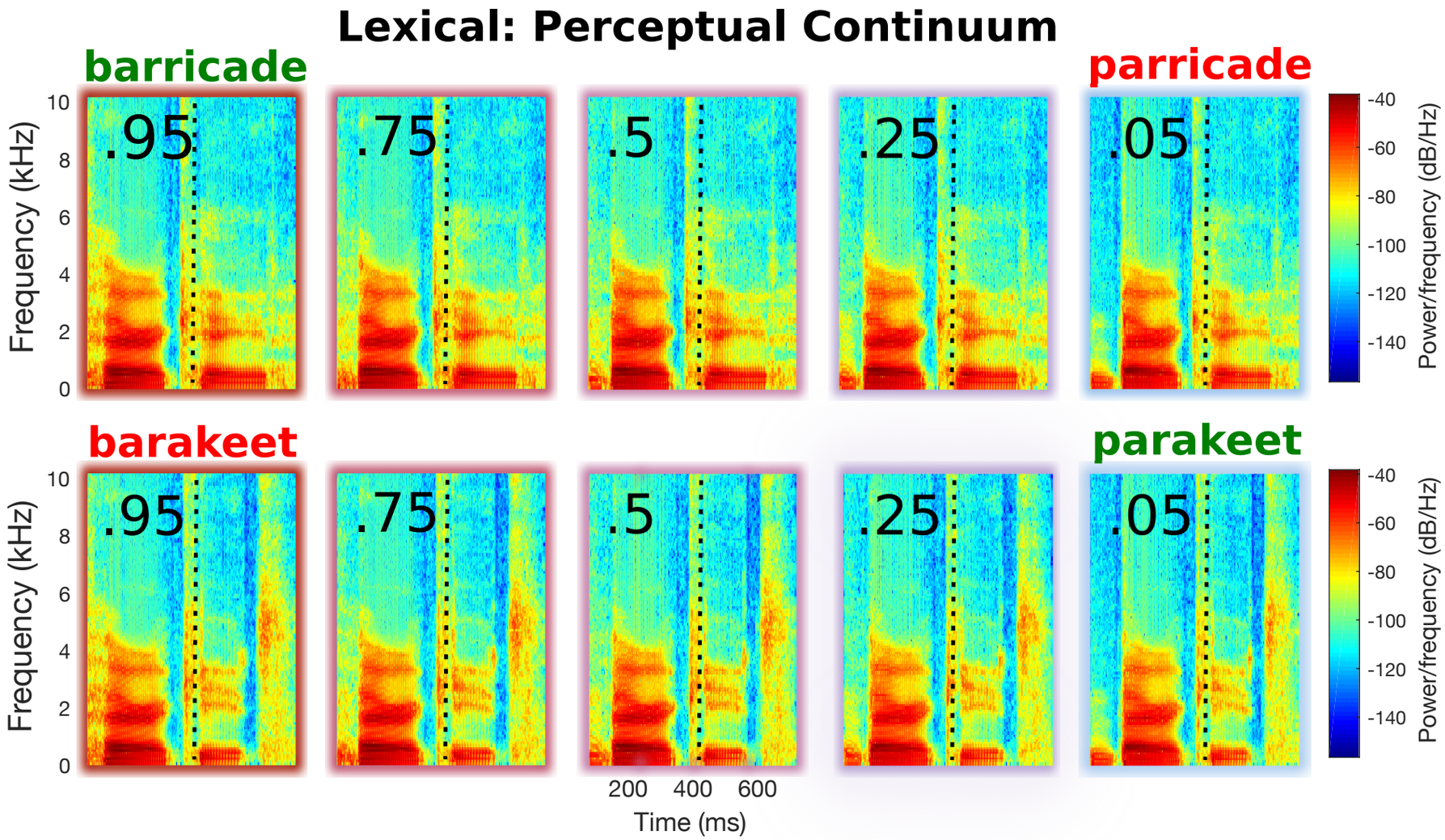


Design & Materials



p p p b b

Design & Materials

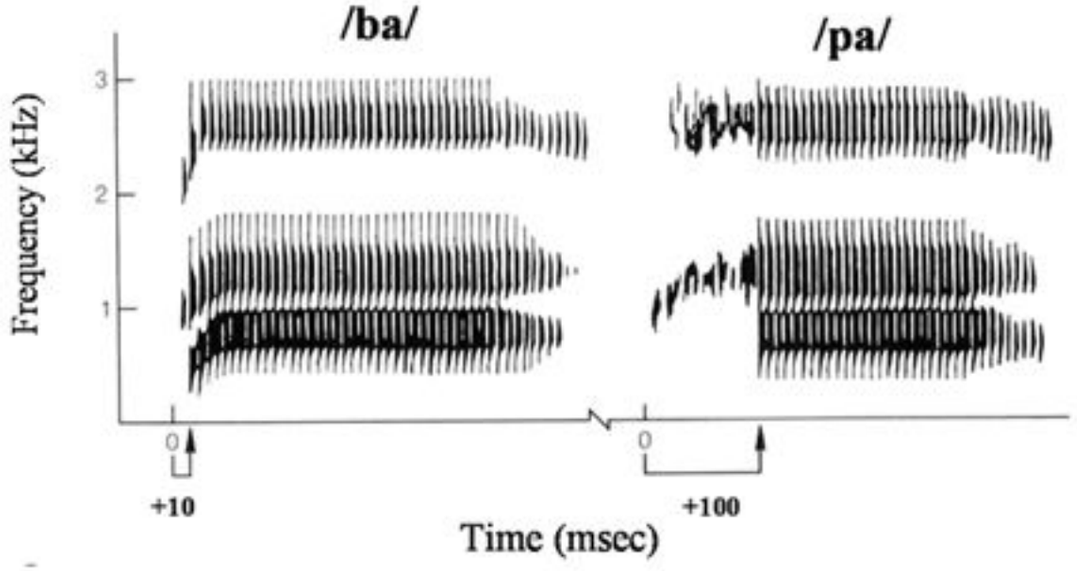


- Point of Disambiguation (POD) ranged 3-8 phonemes / 150-750 ms
- VOT (31 pairs) {p-b, t-d, k-g} and POA (22 pairs) {t-k, p-t}

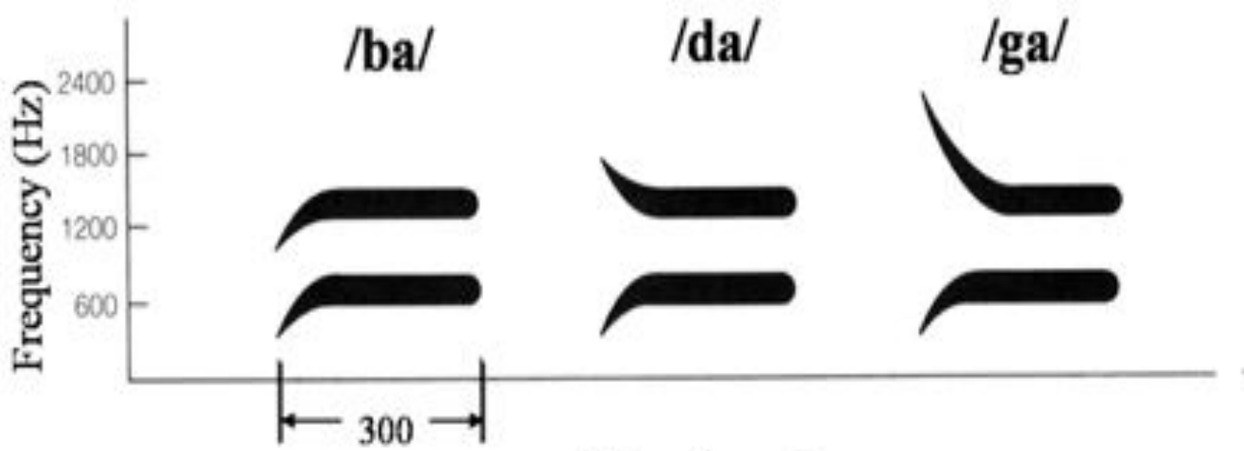
p p p b b

Design & Materials

Contrast: Voice Onset Time (VOT)

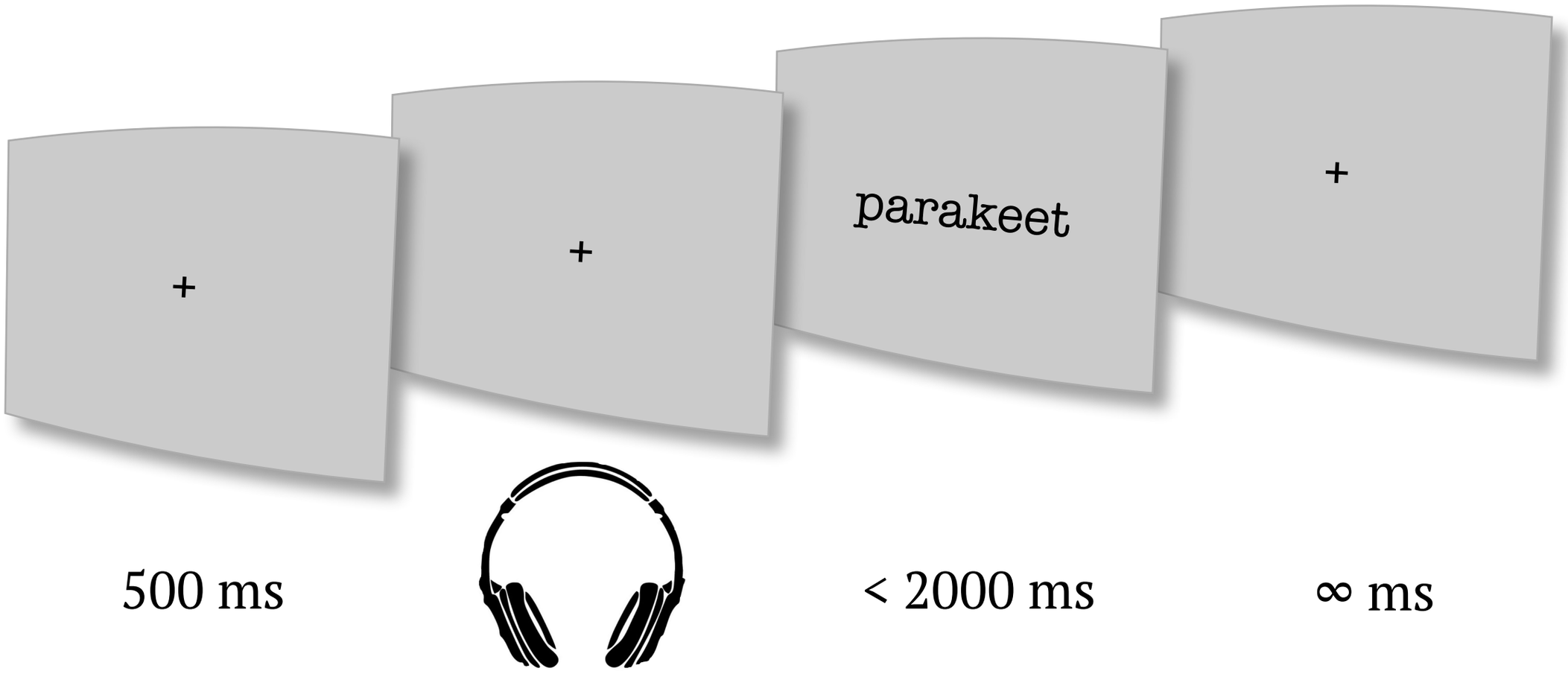


Contrast: Place of Articulation (POA)



- Point of Disambiguation (POD) ranged 3-8 phonemes / 150-750 ms
- VOT (31 pairs) {p-b, t-d, k-g} and POA (22 pairs) {t-k, p-t}

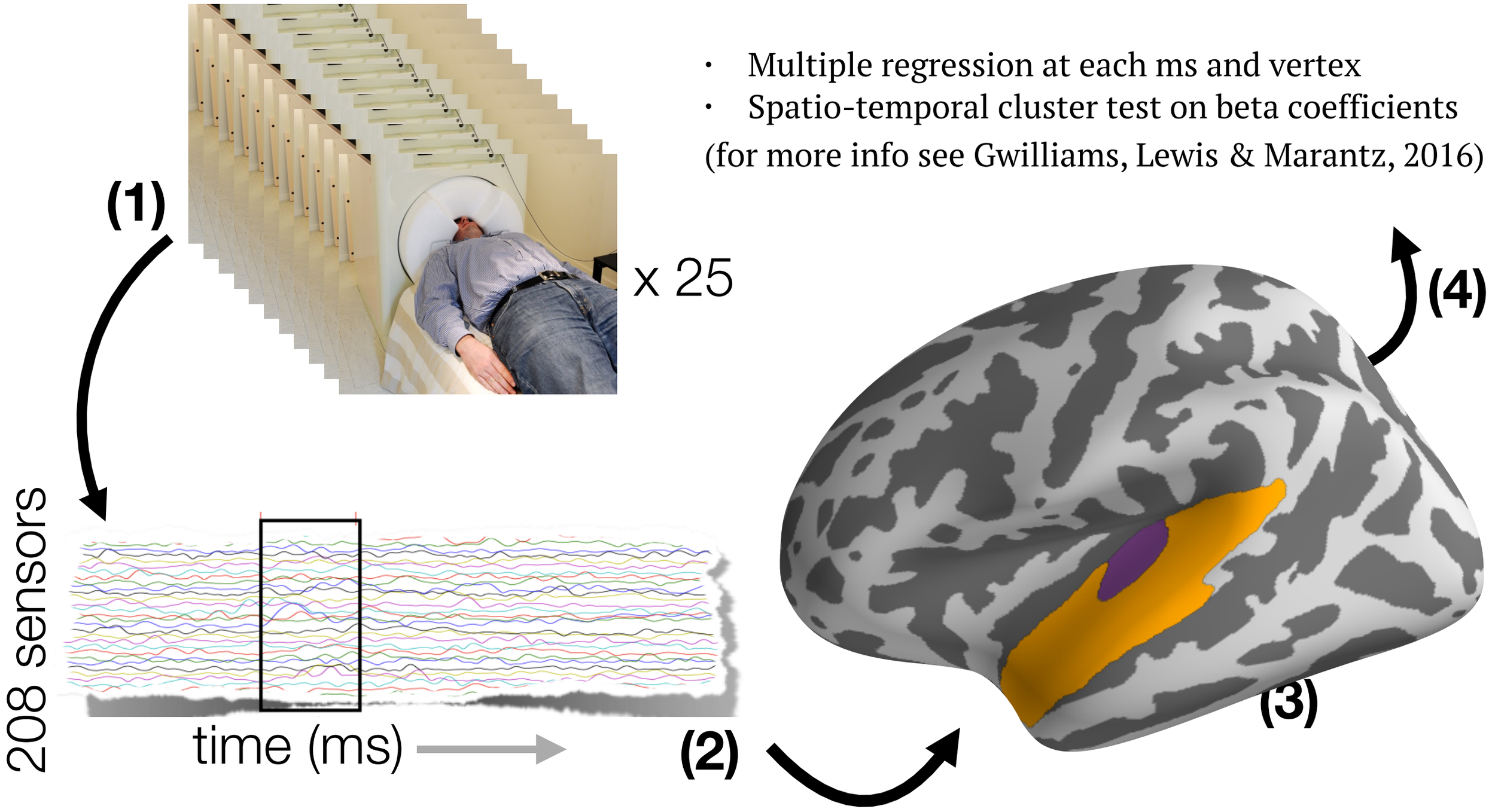
Design & Materials



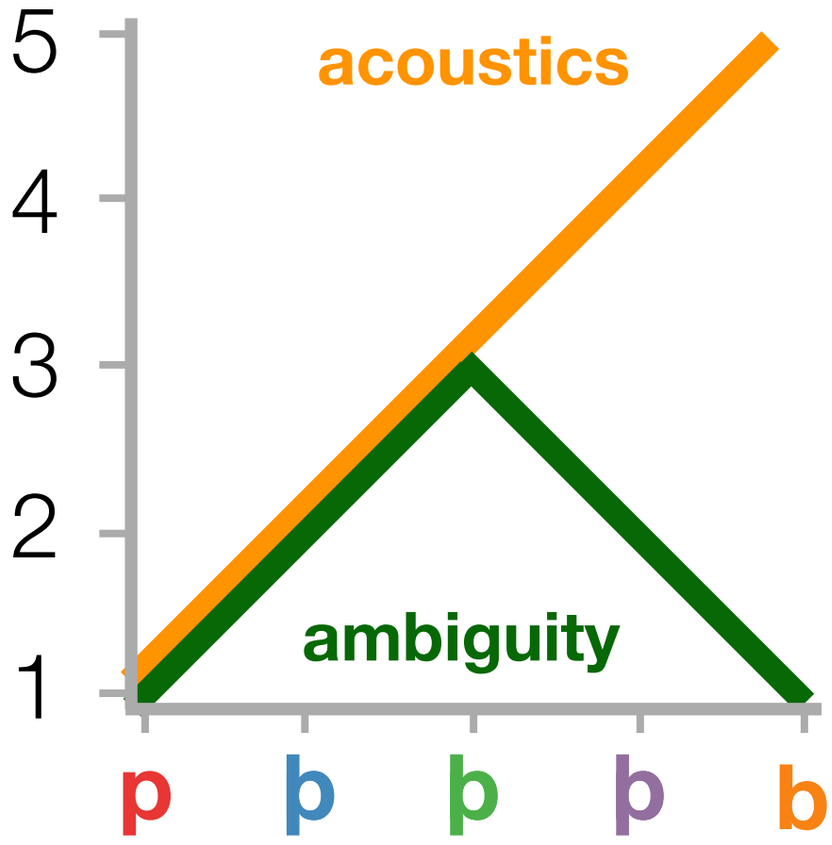
Procedure & Analysis

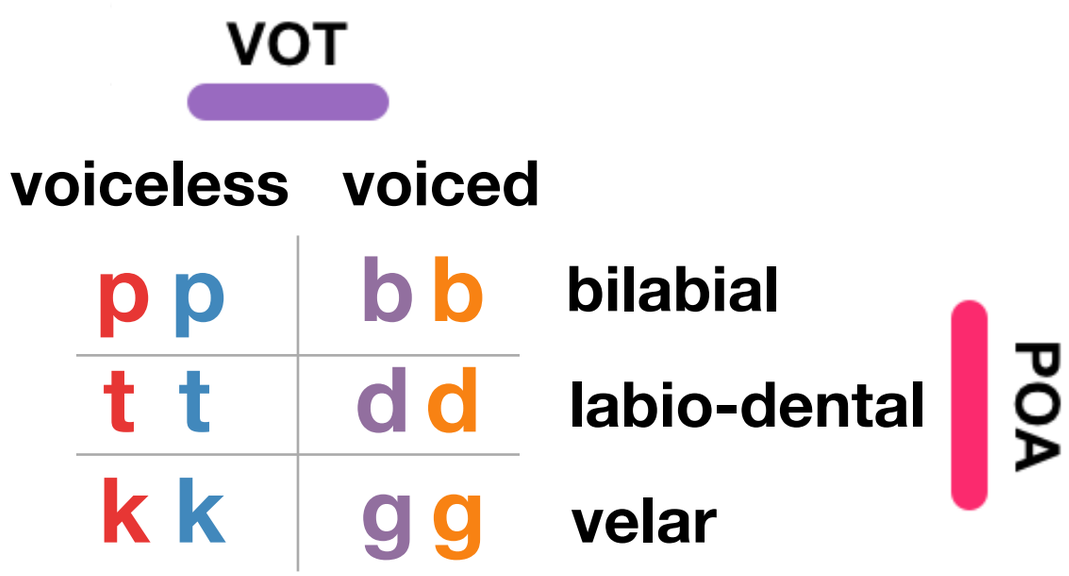
- Multiple regression at each ms and vertex
- Spatio-temporal cluster test on beta coefficients

(for more info see Gwilliams, Lewis & Marantz, 2016)



Four Experimental Variables





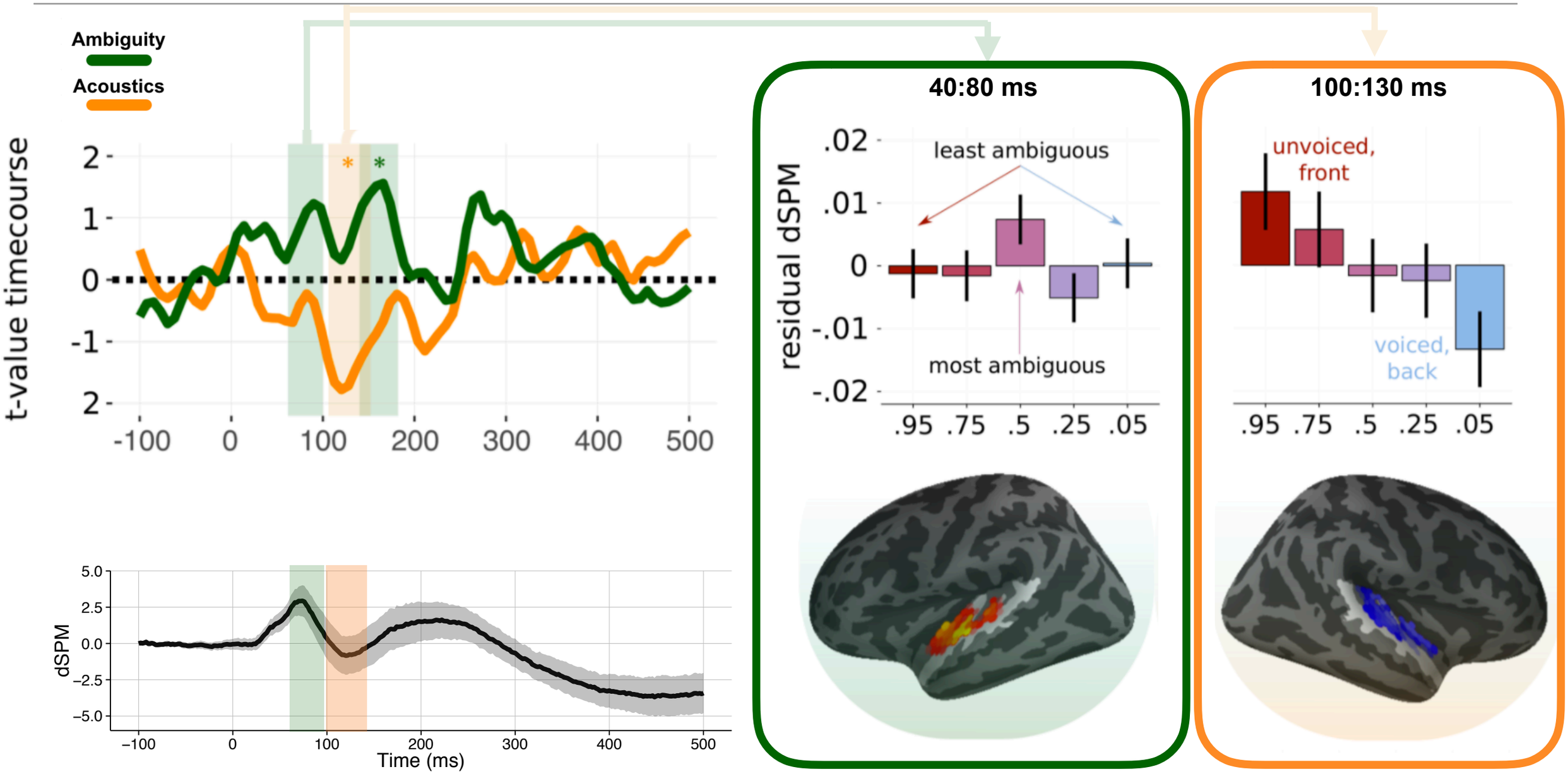
Today's Questions

How does the auditory cortex **respond** to phonological ambiguity?

p a r a k **ee** t

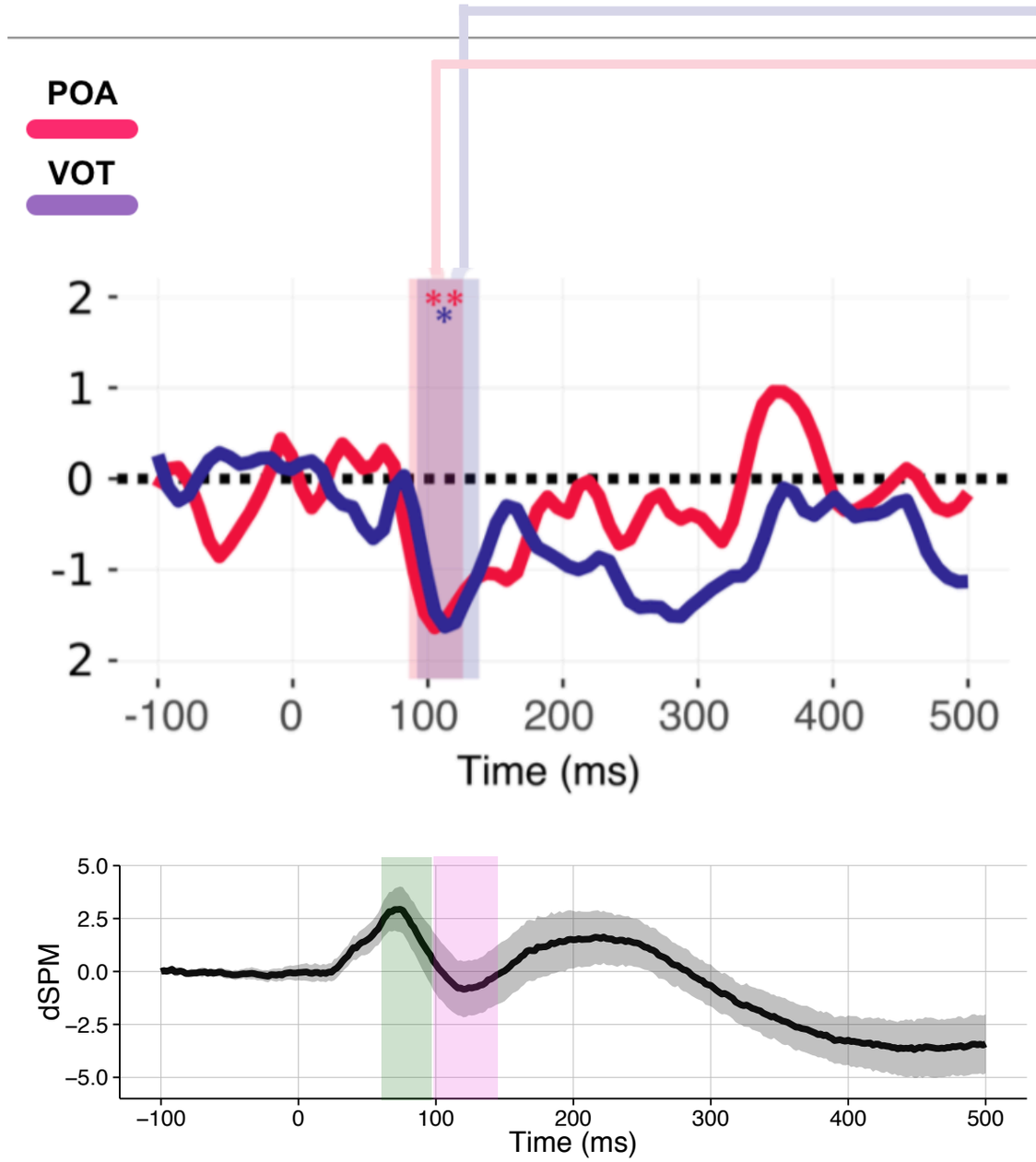
p p b b b

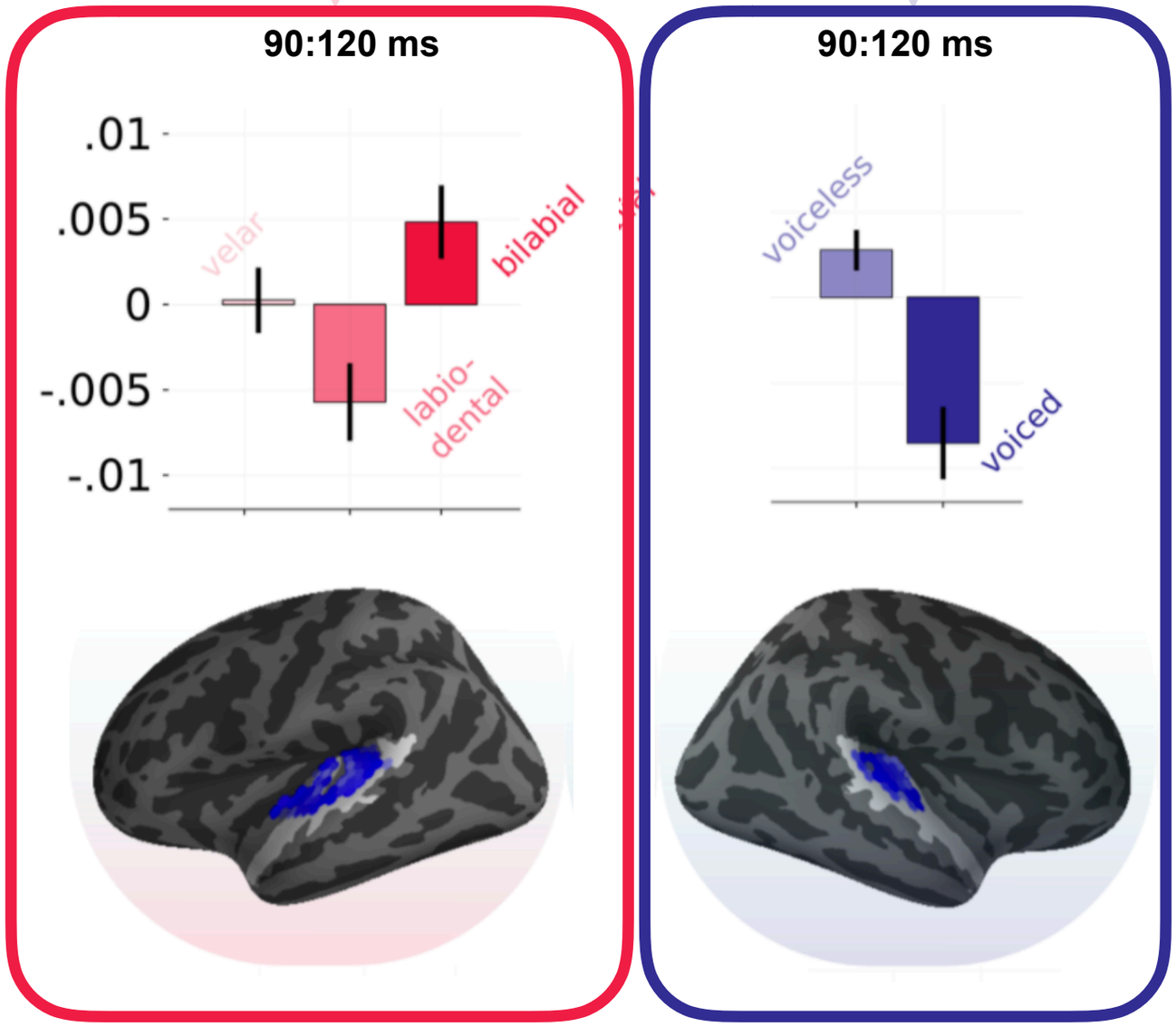
Subphonetics at Onset



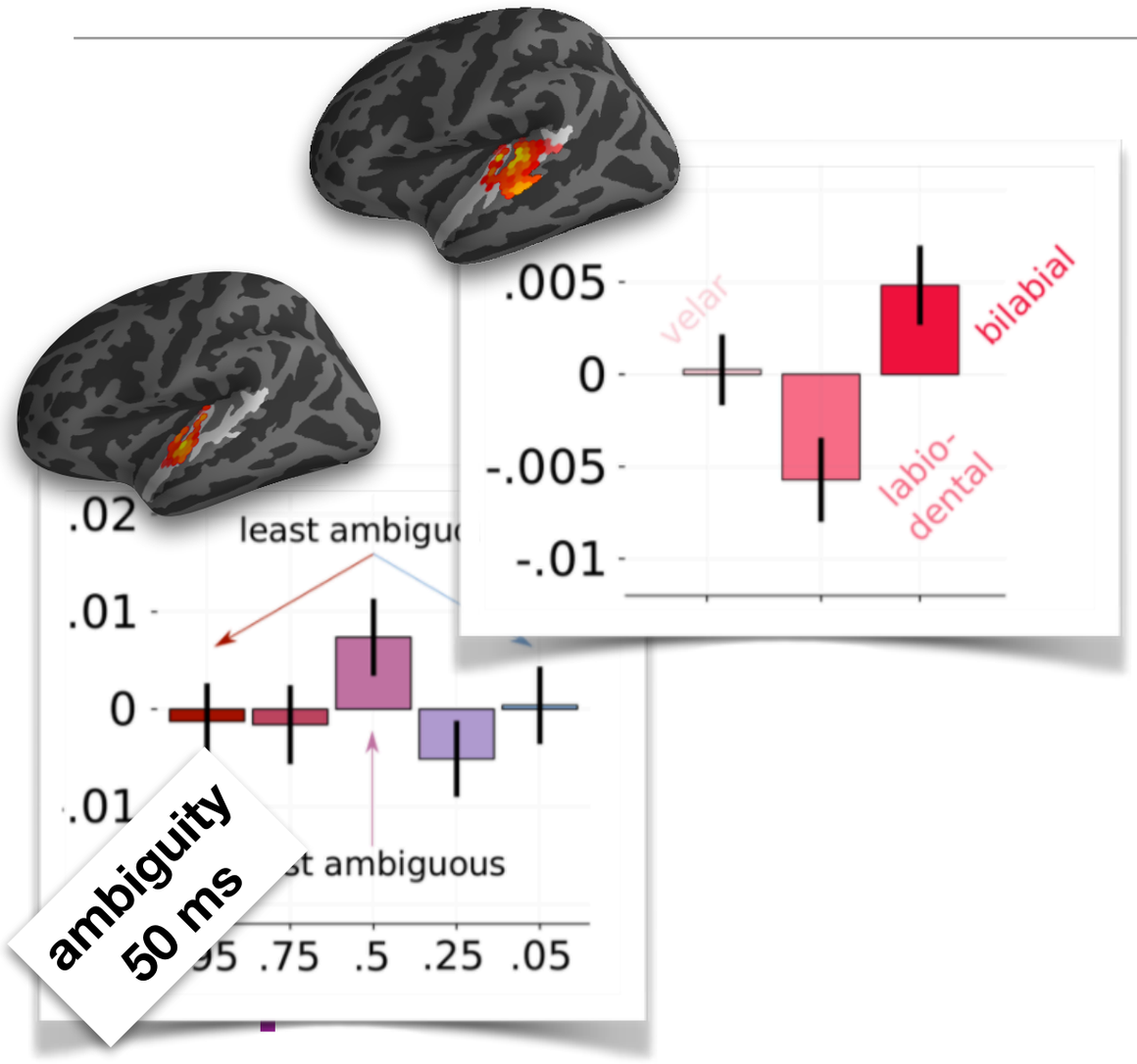
p p p b b

Phonetic Features at Onset





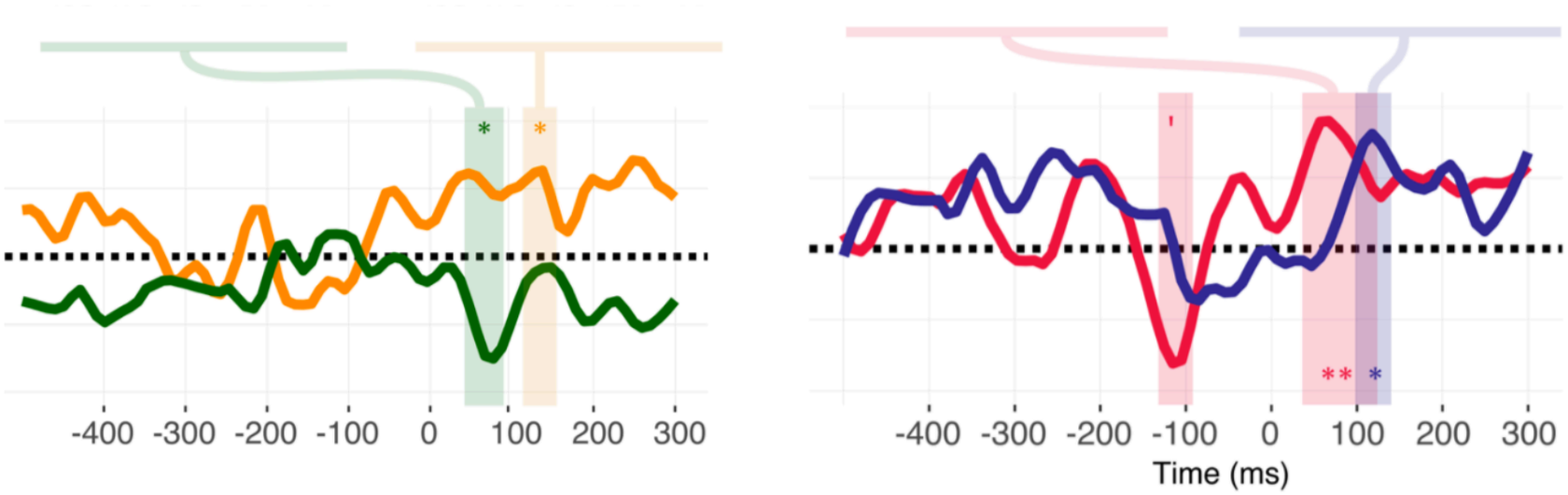
Interim Conclusion



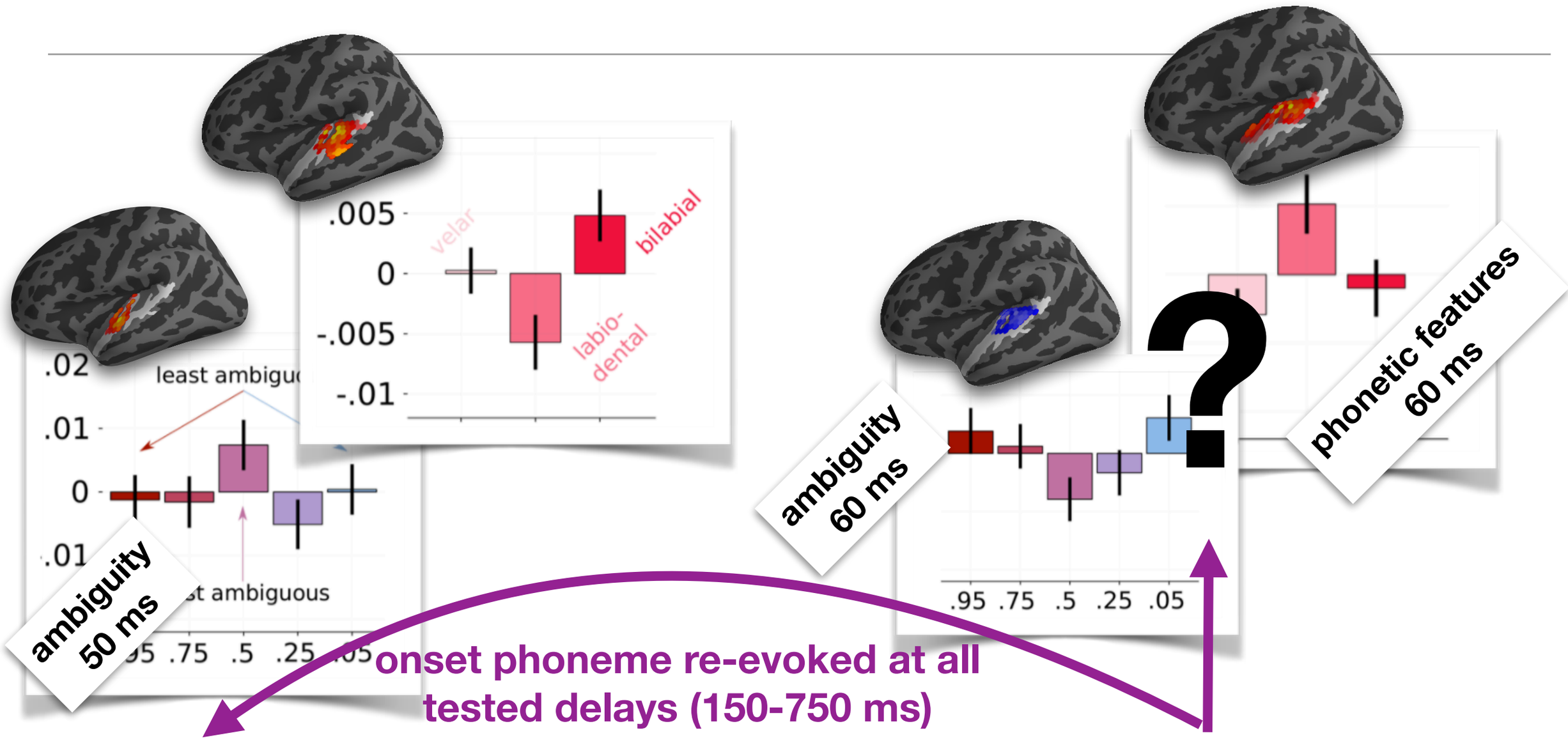
?

þ a r a k **ee** t

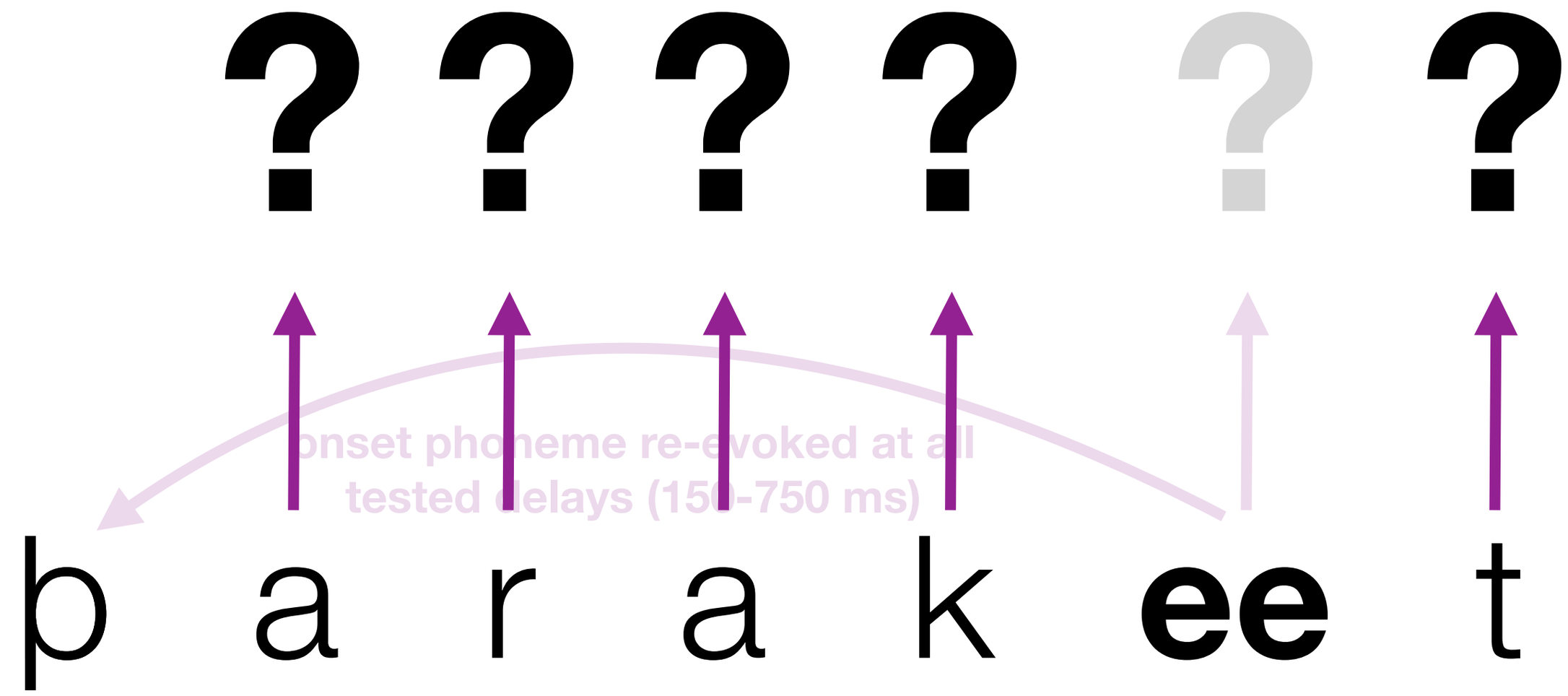
Ambiguity at POD



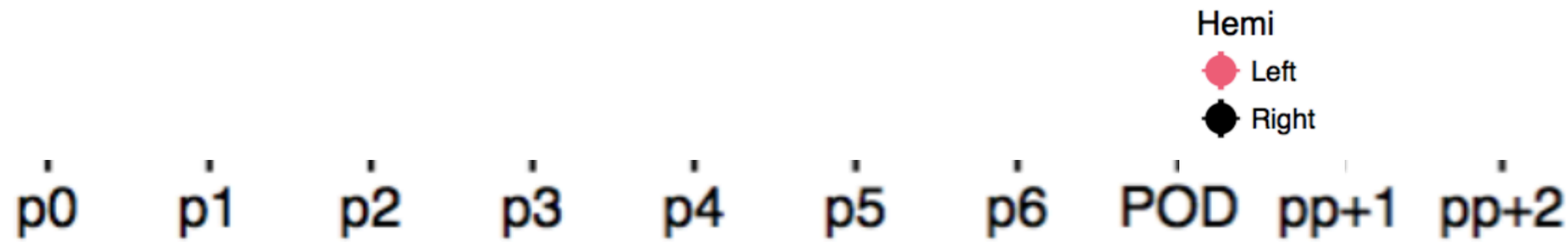
Interim Conclusion



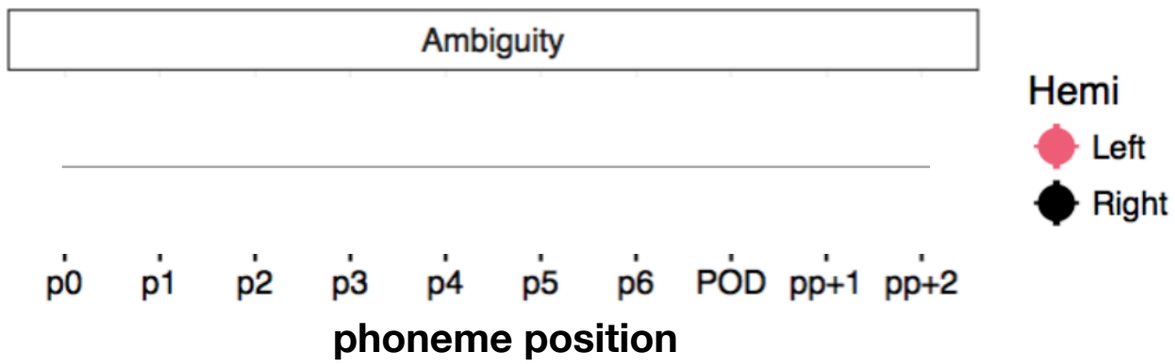
Interim Conclusion



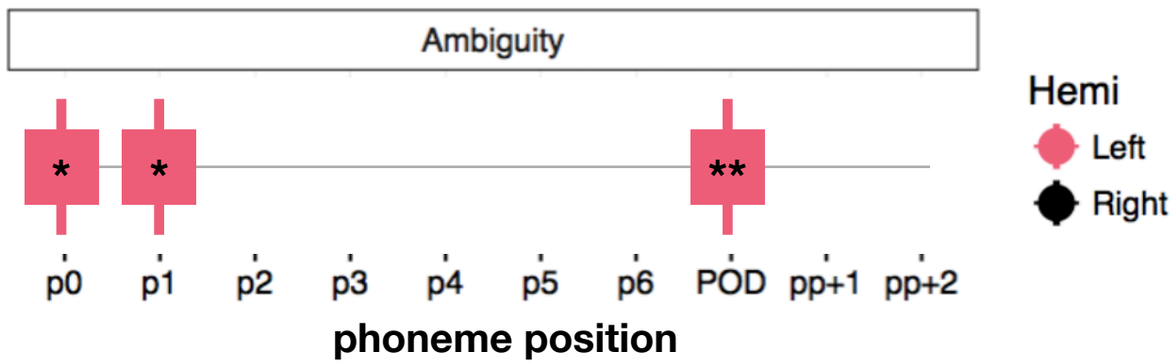
Reactivation in Intermediate Positions



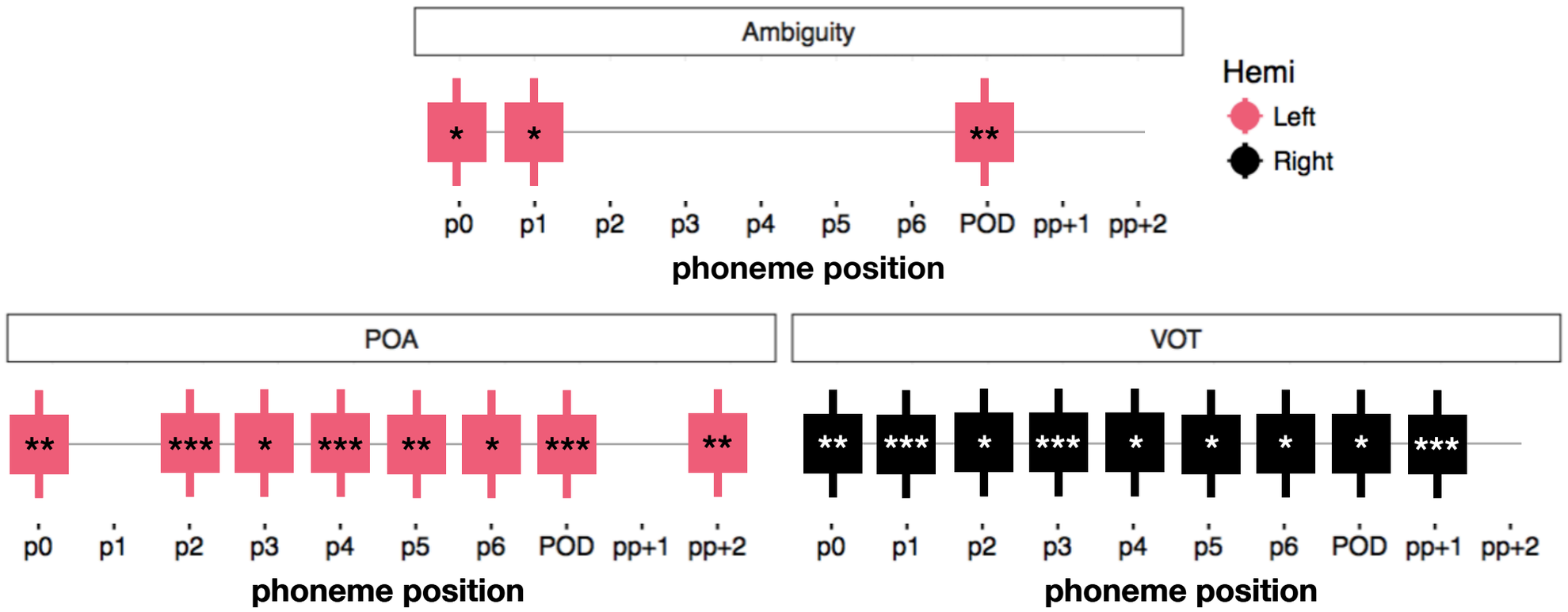
Reactivation in Intermediate Positions



Reactivation in Intermediate Positions



Reactivation in Intermediate Positions



Interim Conclusion

/p/

phonological commitment

p a r a k ee t

acoustic-phonetic maintenance

ballet

bath

palate

prove

bond

book

pin

pacify

beef

b

pants

bind

poke

balance

boast

paddle

panda

Laura Gwilliams | **New York University** | **@GwilliamsL**

ballet

prove

pin

bath

pacify

bond

palate

beef

book

b

pants

balance

bind

paddle

boast

poke

panda

Laura Gwilliams | **New York University** | **@GwilliamsL**

ballet

prove

pin

bath

pacify

bond

palate

beef

book

pants

balance

bind

paddle

boast

poke

panda

Laura Gwilliams | **New York University** | **@GwilliamsL**

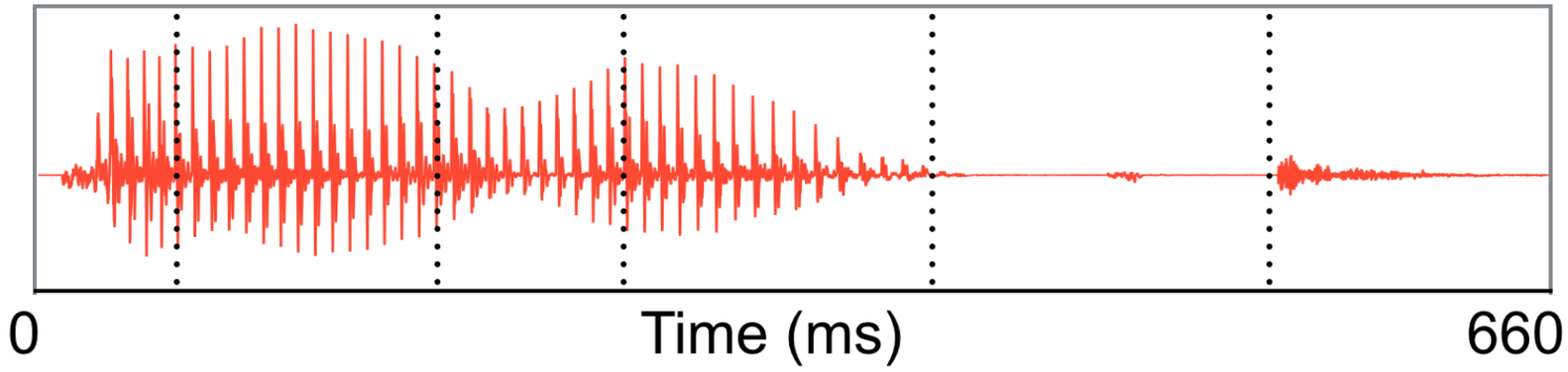
Testing for phonological commitment

- **Surprisal**:

No commitment
Commitment

$$-log_2\left(P(\varphi_a|A)\frac{f(\varphi_a,\varphi_2,\ldots,\varphi_t)}{f(\varphi_a,\varphi_2,\ldots,\varphi_{t-1})}Q_a^t+\right.$$

/d/ /ɪ/ /r/ /ɛ/ /k/ /t/



- **Entropy**:

No commitment
Commitment

$$P(w|C,A) = P(w|C_a)P(\varphi_a|A)+ P(w|C_b)P(\varphi_b|A)$$

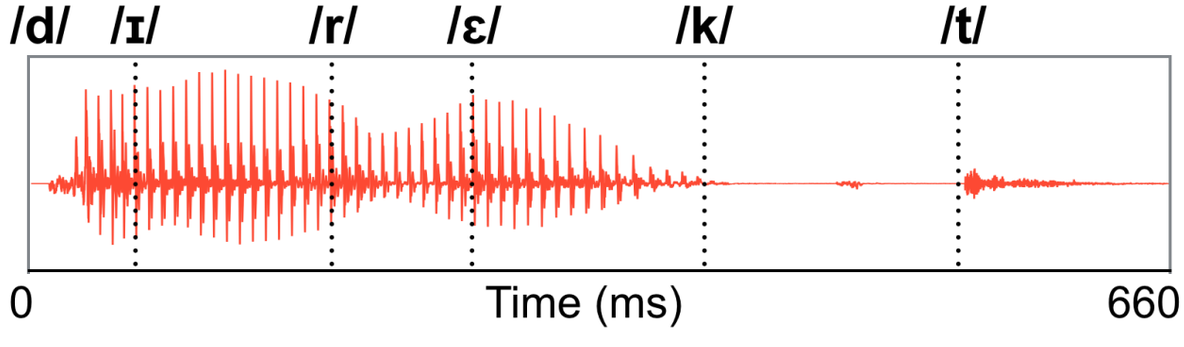
Model Setup

- **Critical variables**:

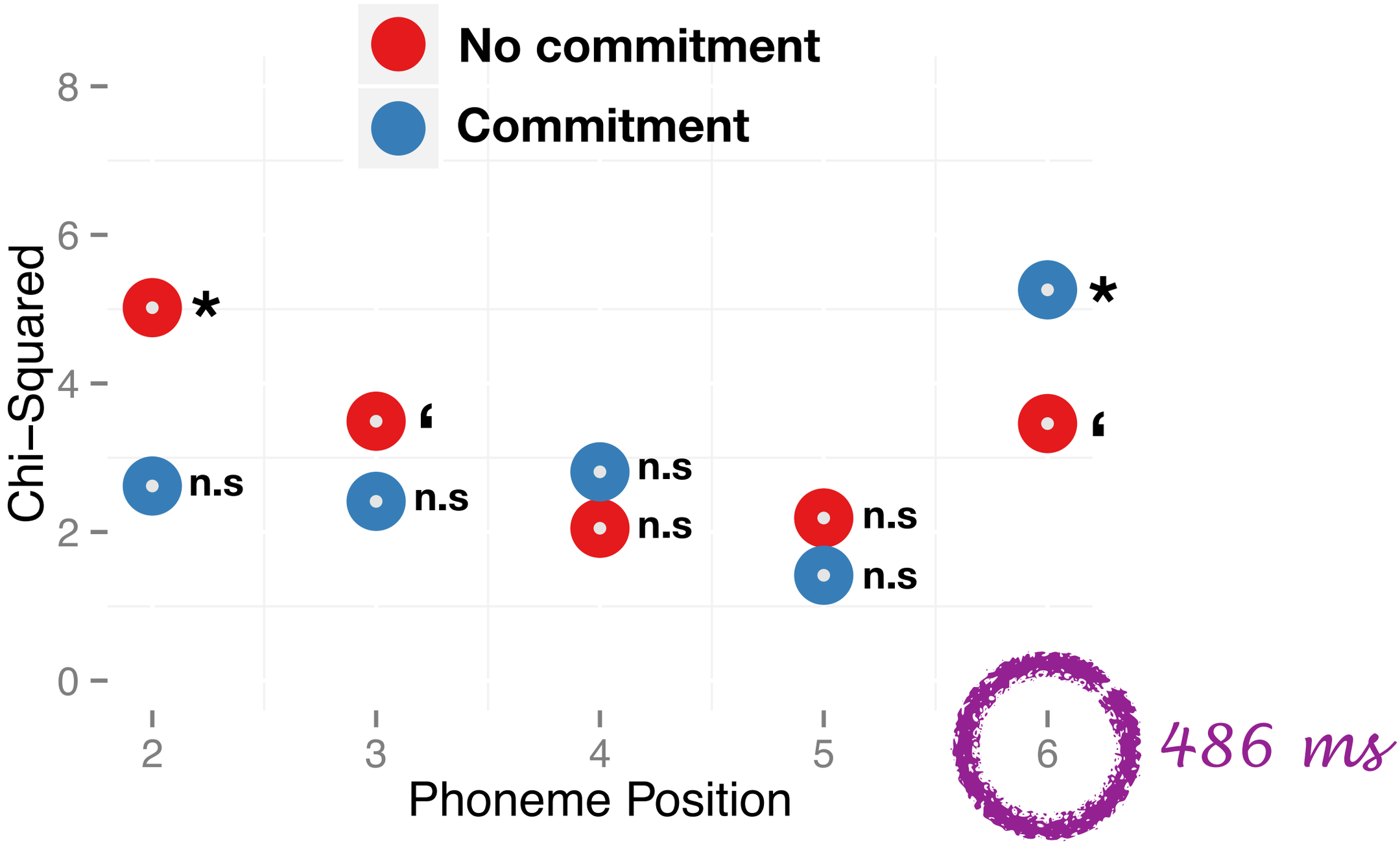
no commitment entropy
no commitment surprisal
commitment entropy
commitment surprisal

- **Control variables**:

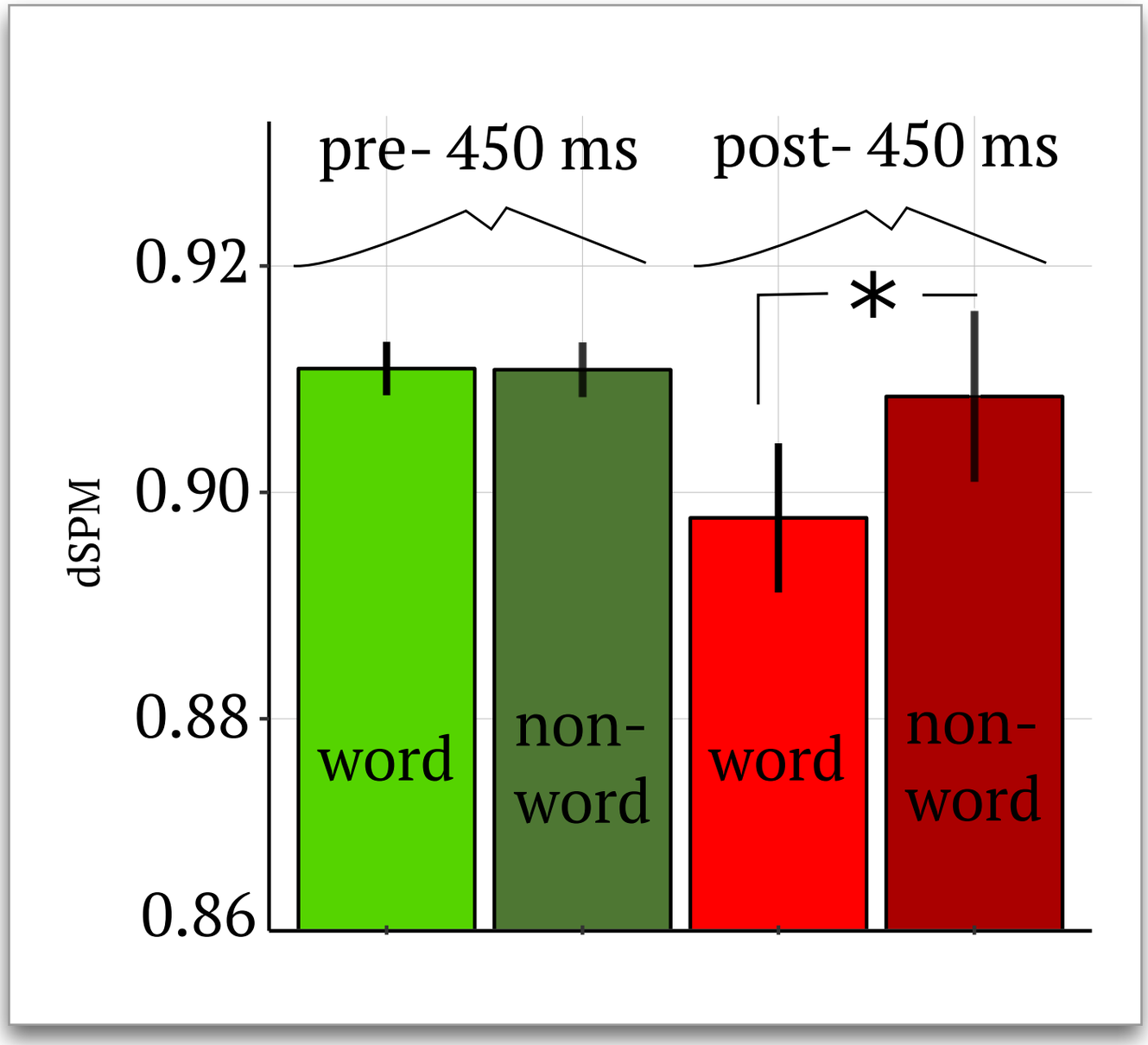
phoneme latency (ms)
phoneme latency (number of phonemes)
trial number
block number
stimulus amplitude
phoneme pair
ambiguity



Results

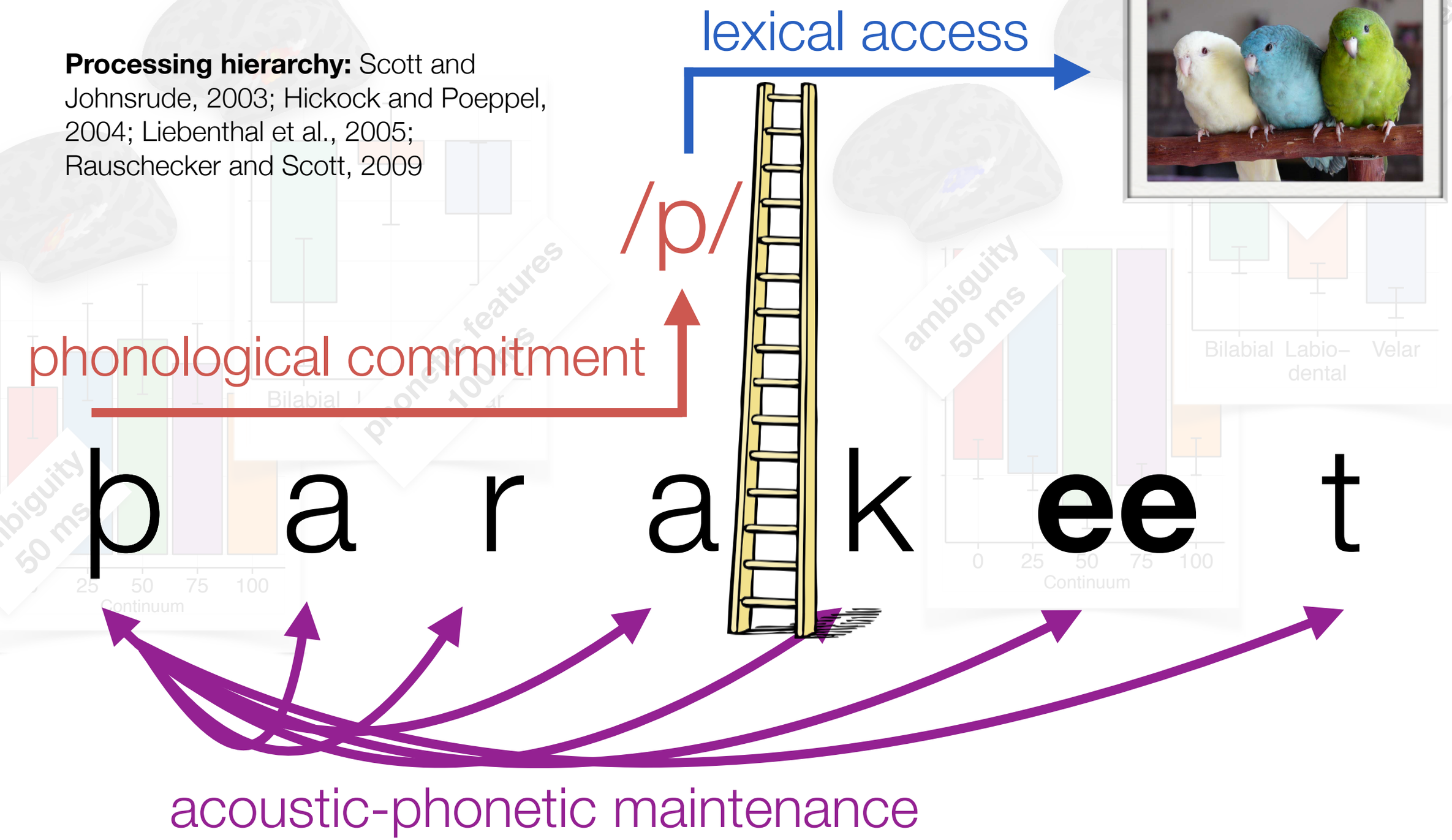

No commitment
Commitment
Chi–Squared
8
6
4
2
0
*
n.s
‘
n.s
n.s
n.s
n.s
*
‘
2
3
4
5
6
Phoneme Position
486 ms

Further test of commitment



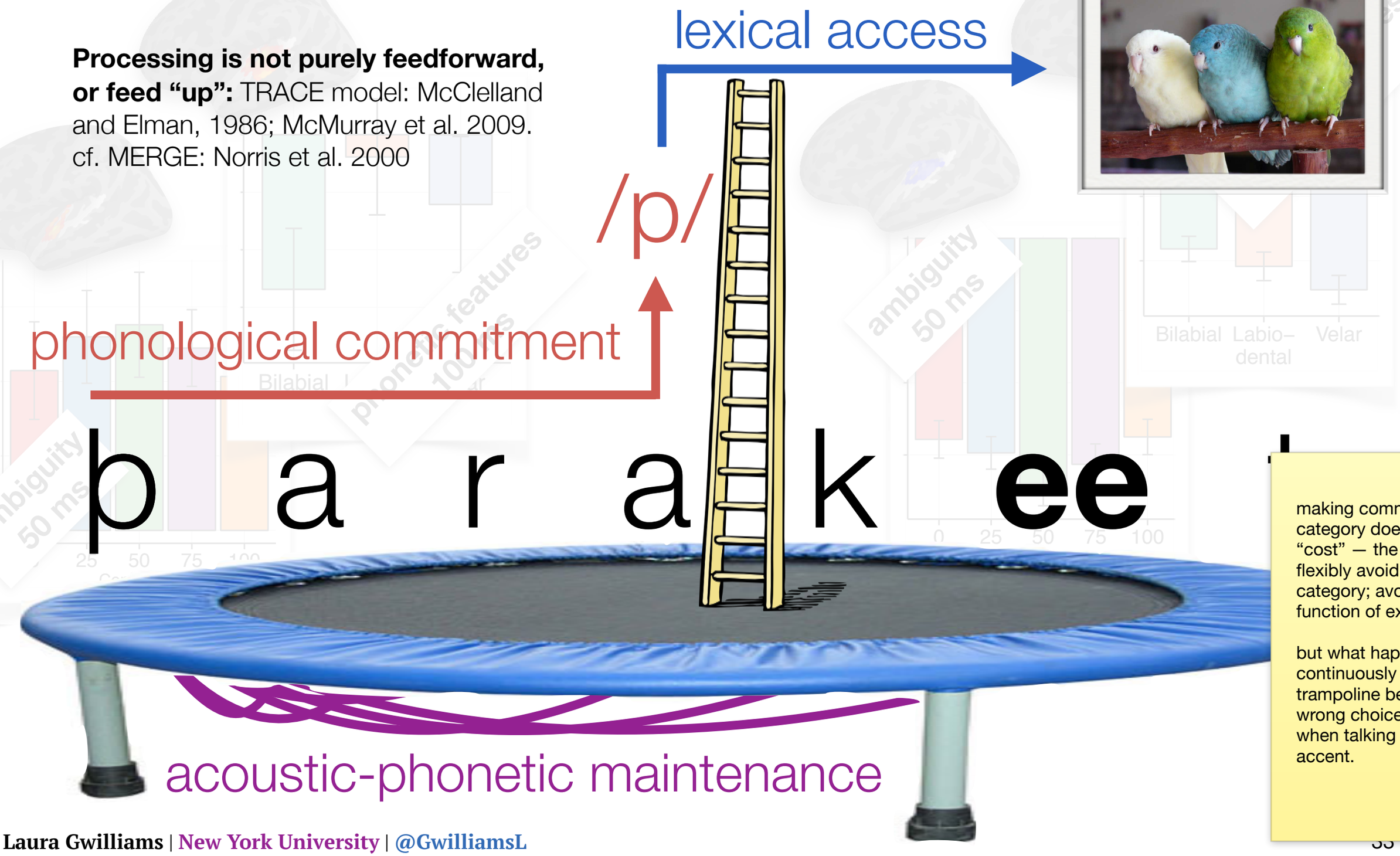
Interpretation

Processing hierarchy: Scott and Johnsrude, 2003; Hickock and Poeppel, 2004; Liebenthal et al., 2005; Rauschecker and Scott, 2009



Interpretation

Processing is not purely feedforward, or feed "up": TRACE model: McClelland and Elman, 1986; McMurray et al. 2009. cf. MERGE: Norris et al. 2000



making commitmen
category does not c
"cost" — the syster
flexibly avoid comm
category; avoiding t
function of exposur

but what happens i
continuously jumpir
trampoline because
wrong choice? well
when talking to som
accent.

acoustic input

subphonetic acoustics

phonetic features

phoneme category

phoneme sequence

lexical candidates

Heschl's gyrus

STG

STG

?

Left MTG

i)

ii)

iii)

iv)

v)

absolute distance[1]

Phonetic feature boundary

signed distance[2]

p0

e.g. þ

VOT

POA

...

e.g. /p/

phoneme commitment

/ˈp/

/ˈpa/

/ˈparəkiːt/

barricade

plant

bread

parakeet

ballot

possum

p1

e.g. a

e.g. /a/

1. Co-linear with phoneme ambiguity
2. Co-linear with perceptual continuum

phonemes

connection types

recurrent

feedforward

feedback

Laura Gwilliams | New York University | @GwilliamsL

Levels of analysis

1. Phonemes within words
   - Responses to phoneme ambiguity **(bottom-up)**
   - Neural signatures of ambiguity resolution, when provided with lexical information **(top-down)**

2. Words within sentences
   - Which linguistic properties encoded in brain activity?
   - What are the relative time-courses of processing each property?
   - What is the computational architecture?

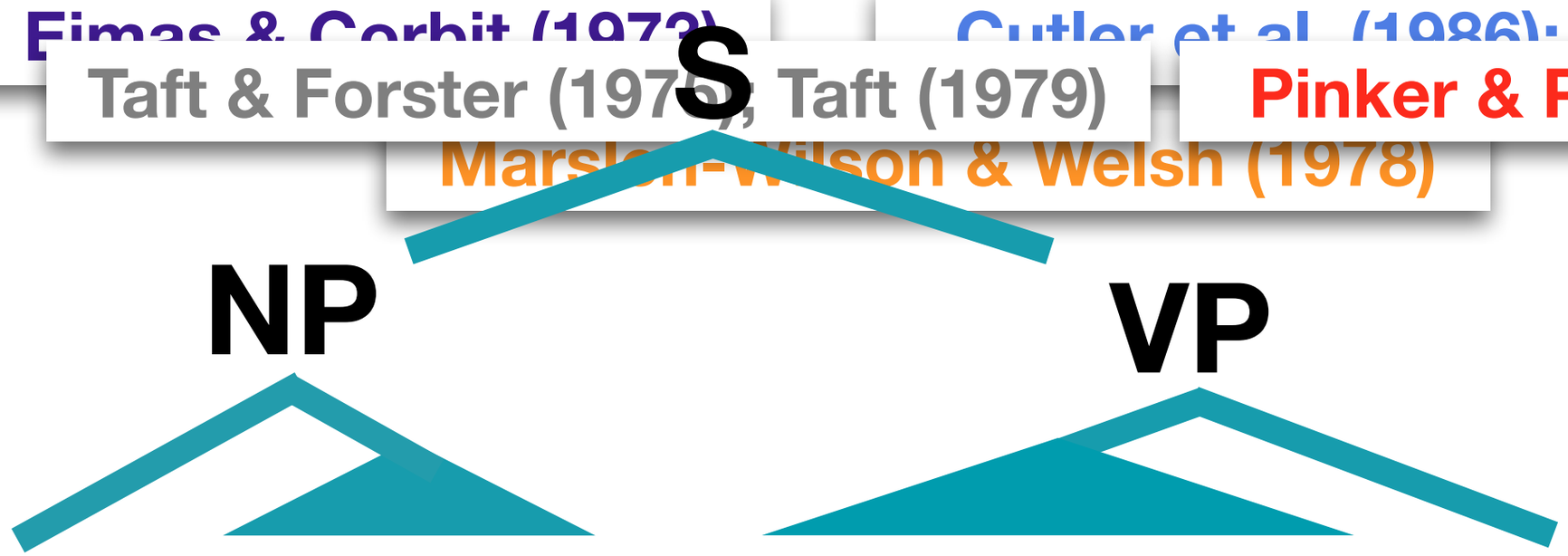
Eimas & Corbit (1973)

Cutler et al. (1986); Barry (1980; 1984)

Taft & Forster (1975); Taft (1979)

Pinker & Prince (1988)

Marslen-Wilson & Welsh (1978)



sentence structure

phrasal structure

the fat cat dis | appear | ed

lemmas

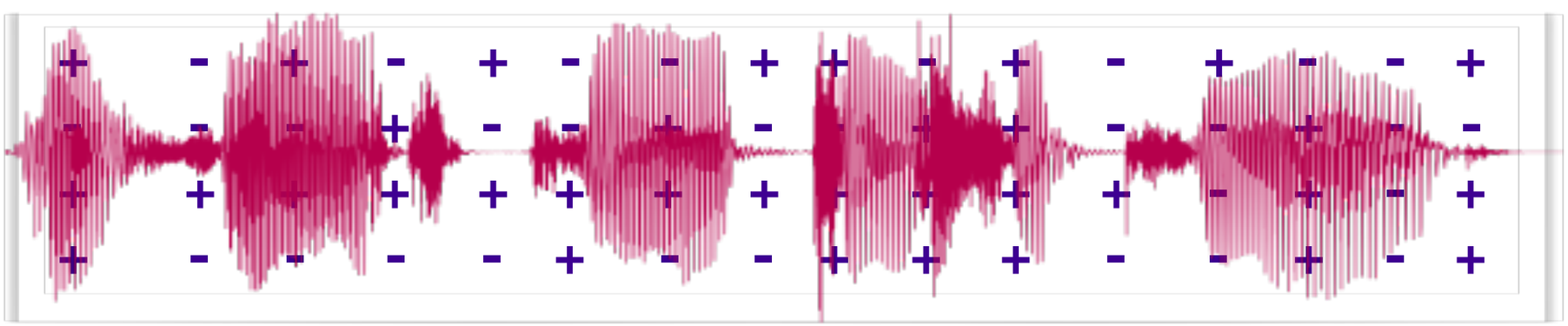
morphemes

dah fat kat dis ah pee ud

syllables

DH AH F AE T K AE T D IH S AH P IH R D

phonemes



phonetic features

acoustics

S

VP

1) which linguistic units are encoded in brain activity?

2) what is the relative time-course?

phrasal structure

lemmas

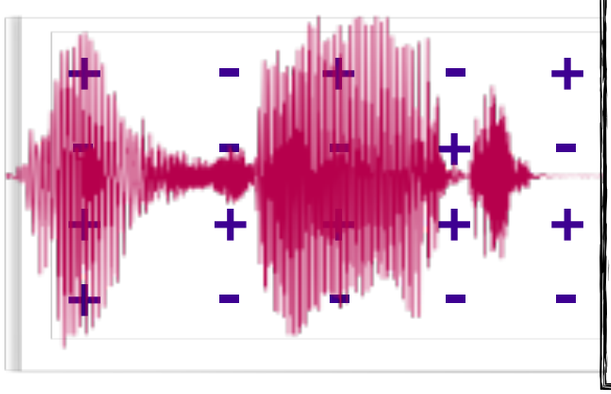
morphemes

th

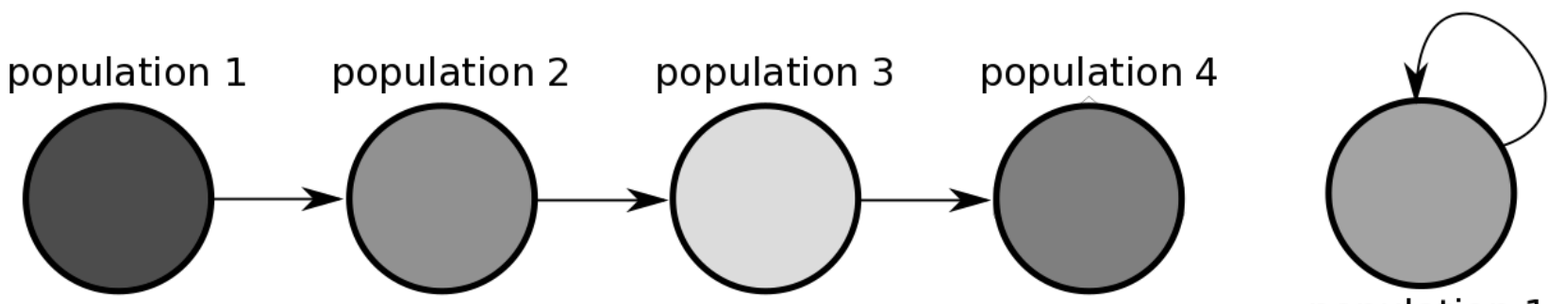
pear | ed

dah fa

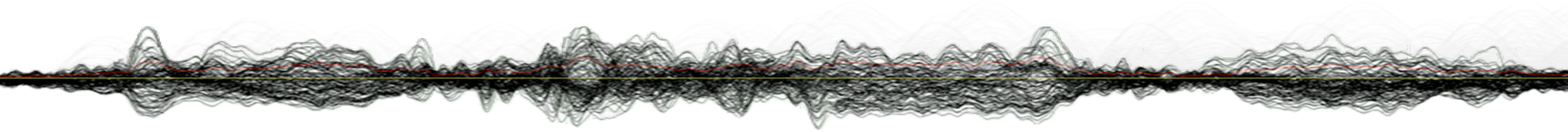
DH AH F AE T



3) what is the computational architecture?



Setup



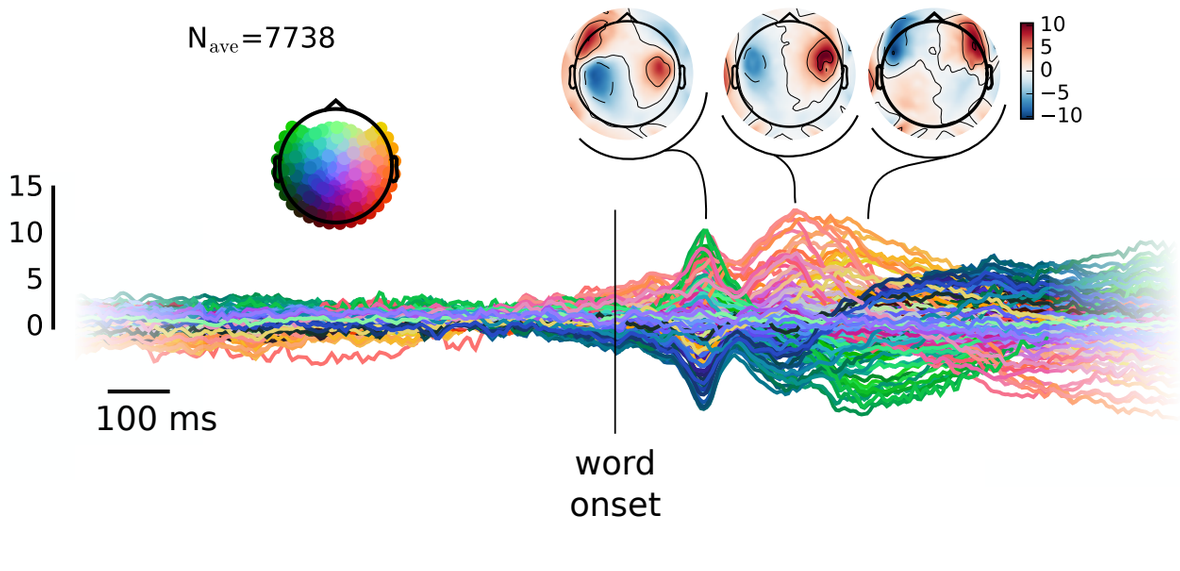
- 18 participants
- Listening to four narrative stories (twice)
- 2 x one hour recordings
- KIT 208 channel MEG system
- Engagement task

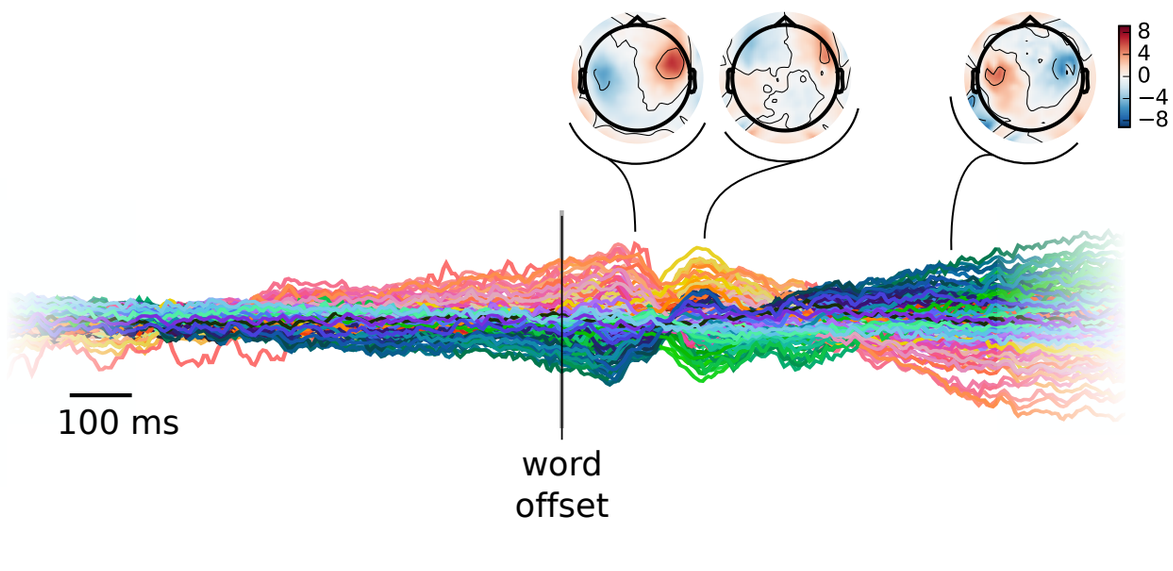
- ~40,000 phonemes per participant
- ~8,000 words per participant

جامعة نيويورك أبوظبي

NYU | ABU DHABI

Event-locked average response

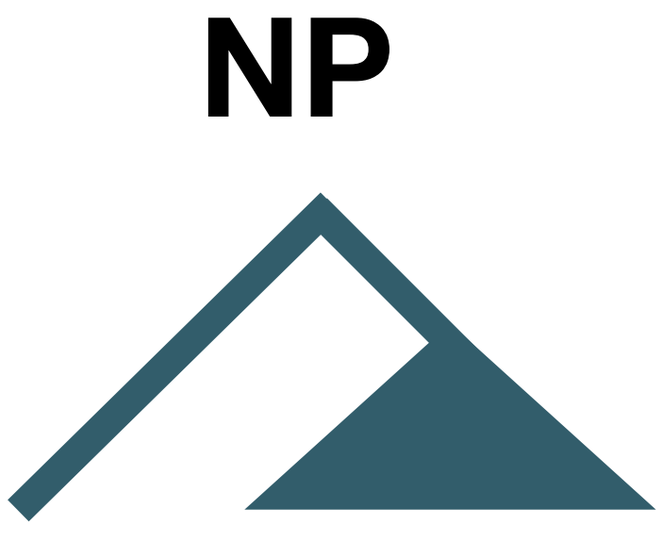




explain this plot/ colour of channels in the topography

Laura Gwilliams | New York University | @GwilliamsL

Stimulus features

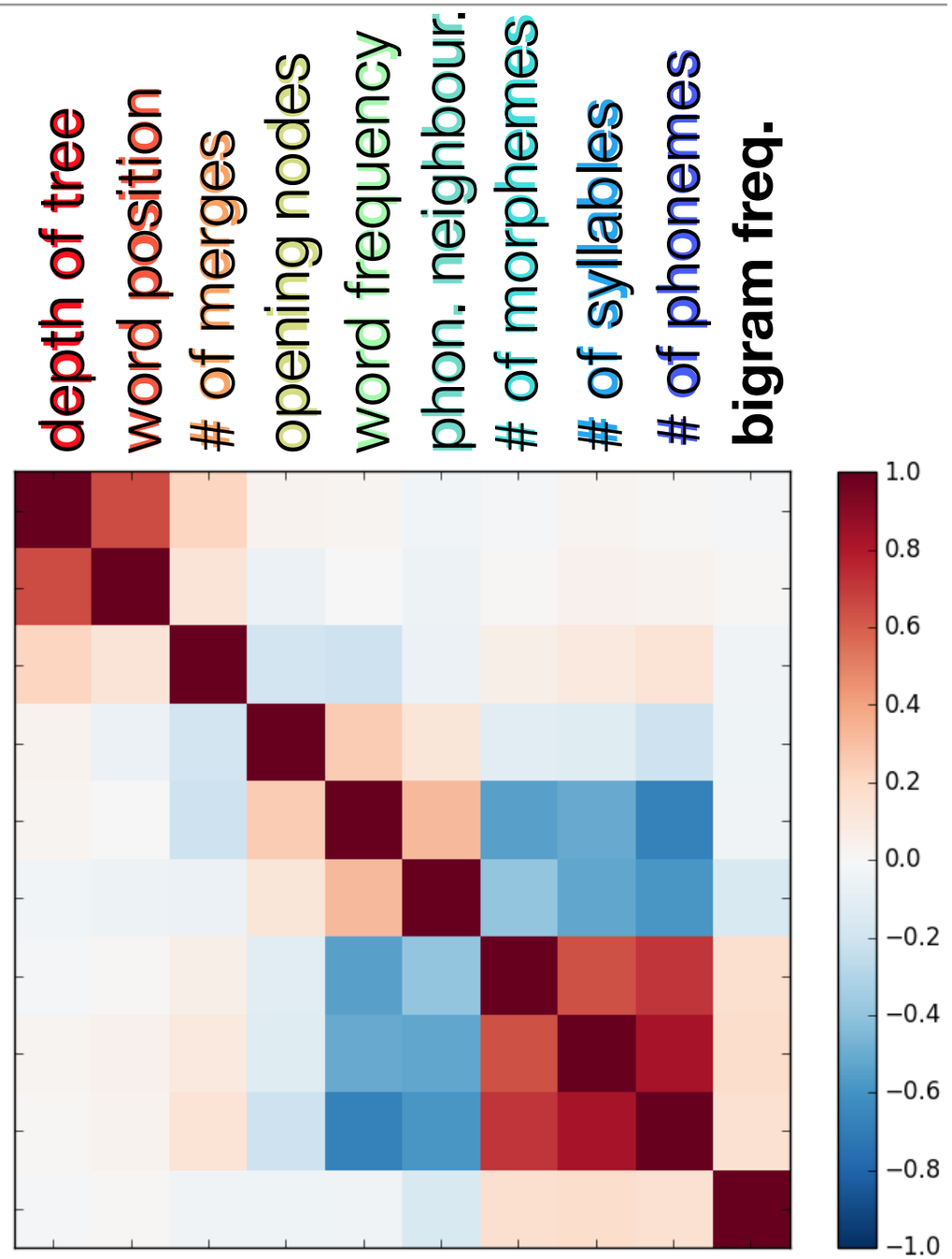


syntactic

word use

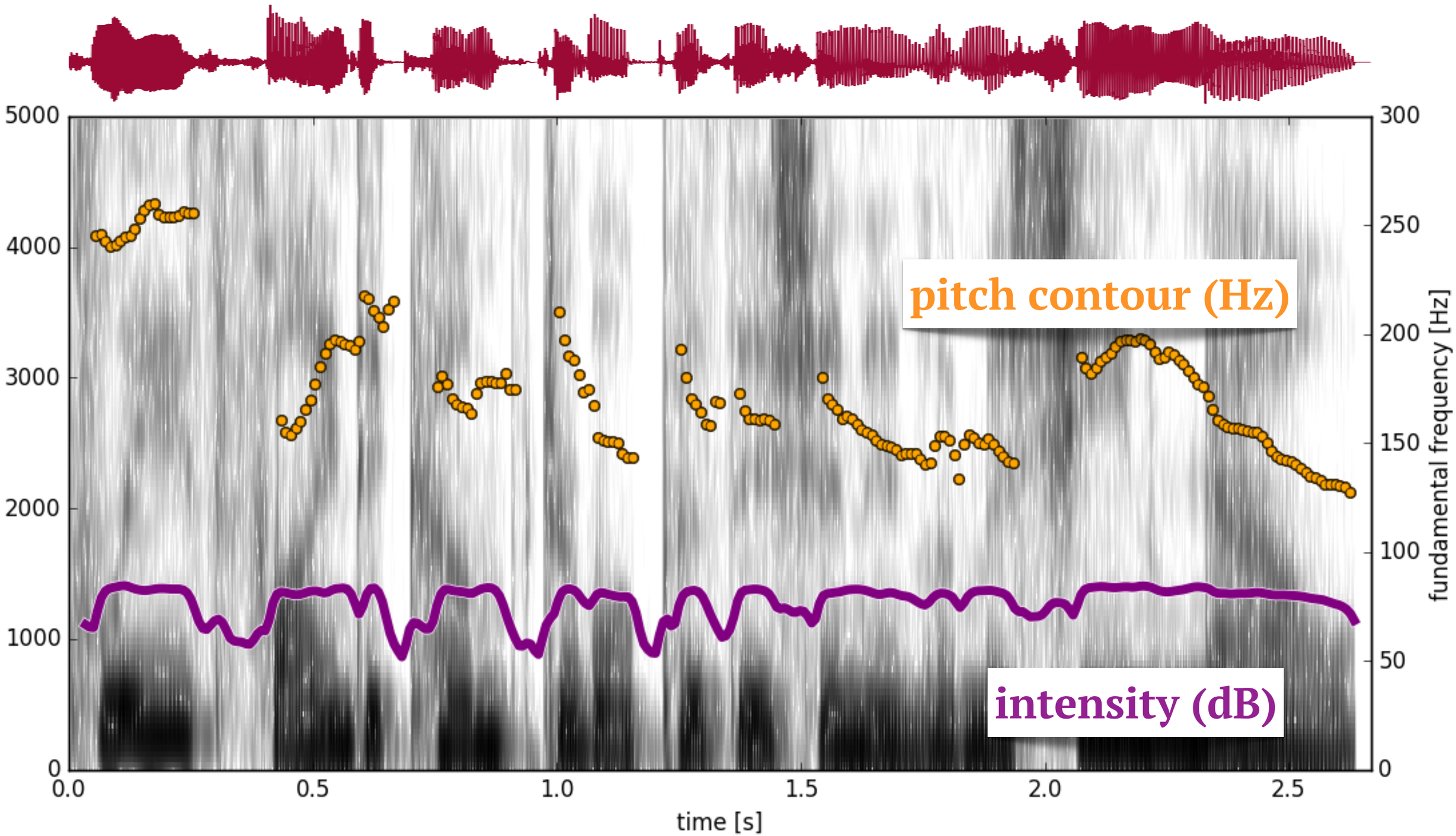
word structure

depth of tree
word position
# of merges
opening nodes
word frequency
phon. neighbourhood
# of morphemes
# of syllables
# of phonemes
bigram frequency

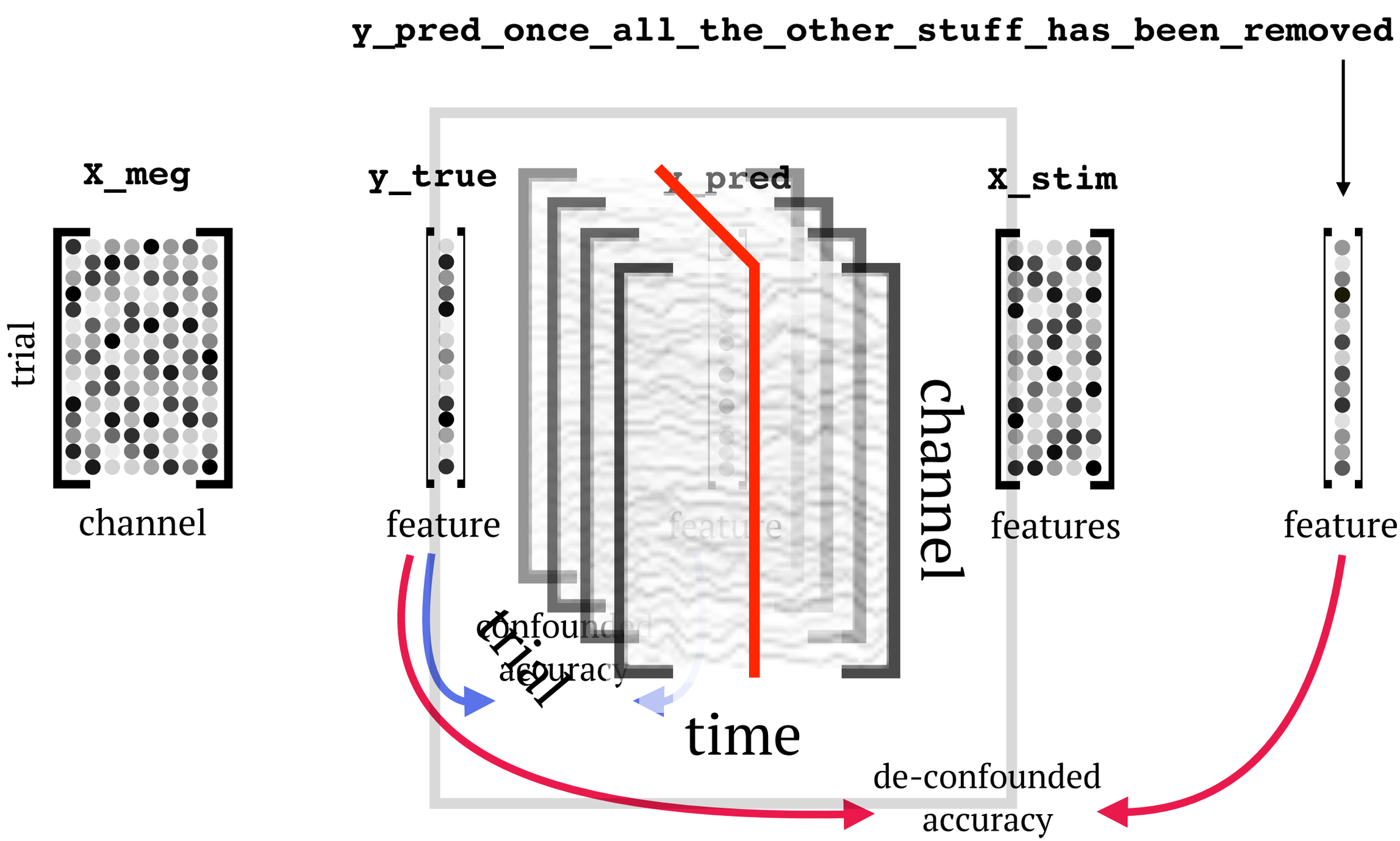


stimulus similarity matrix

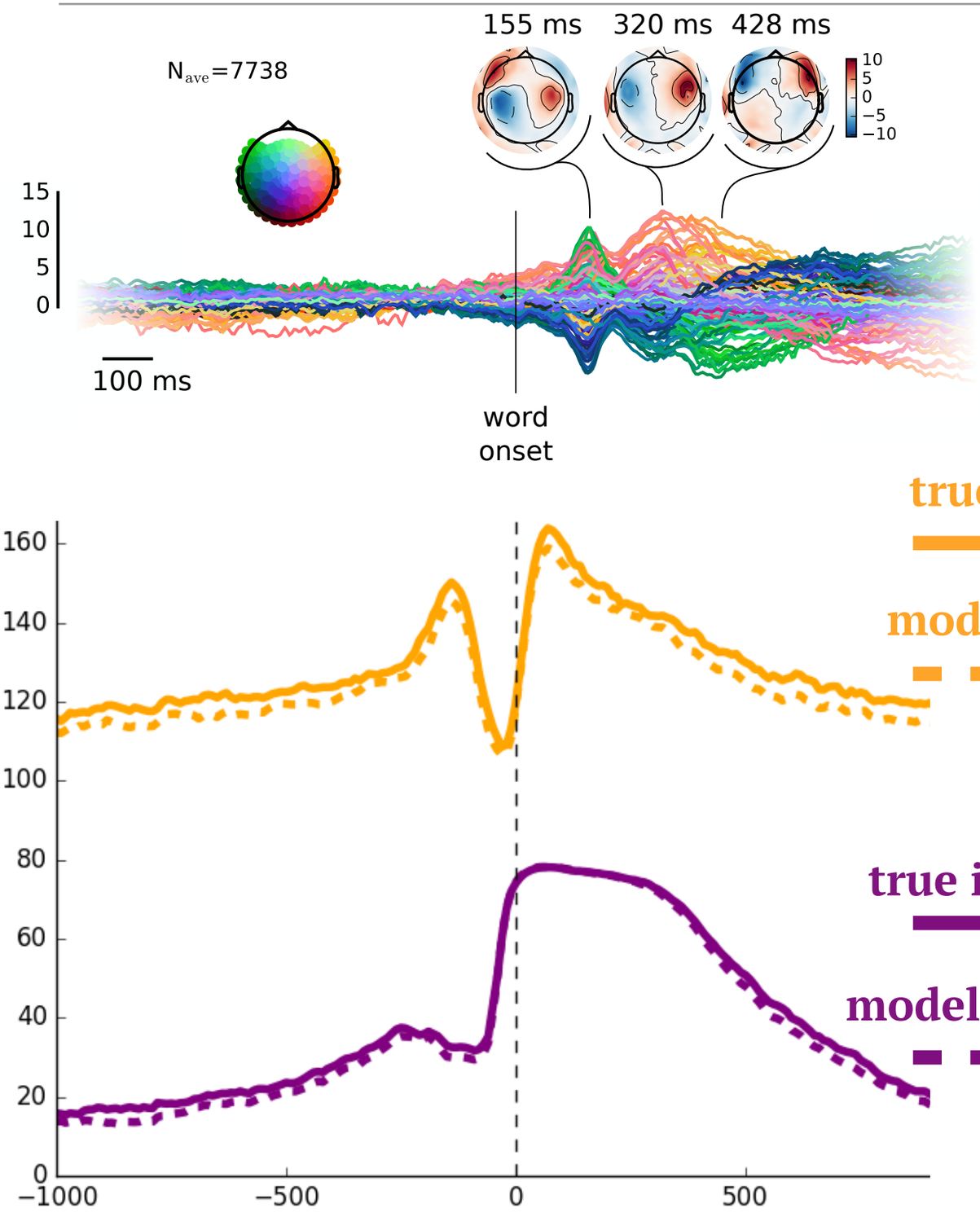
Pitch and intensity in the acoustic signal



Analysis technique



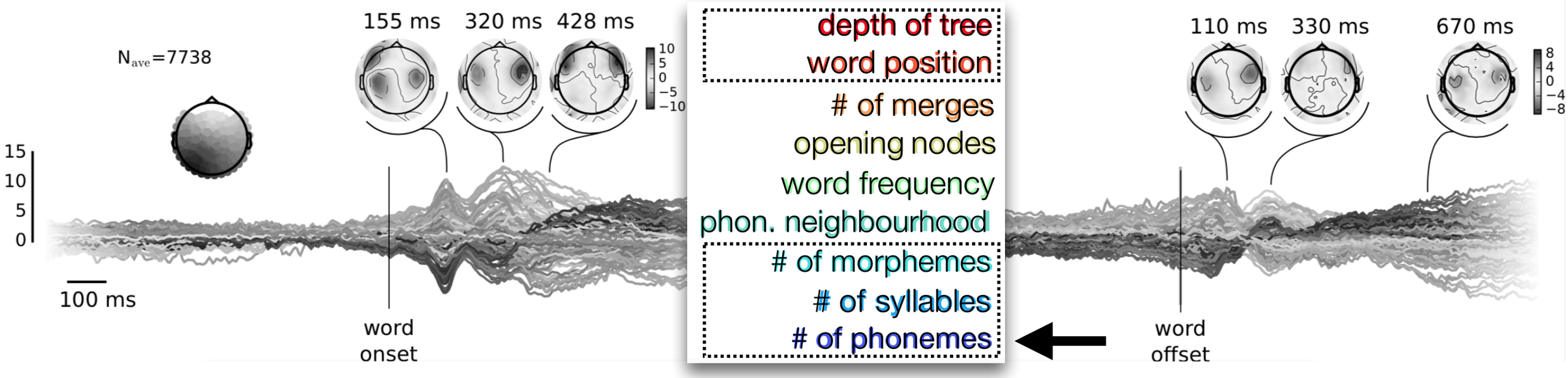
Pitch and intensity

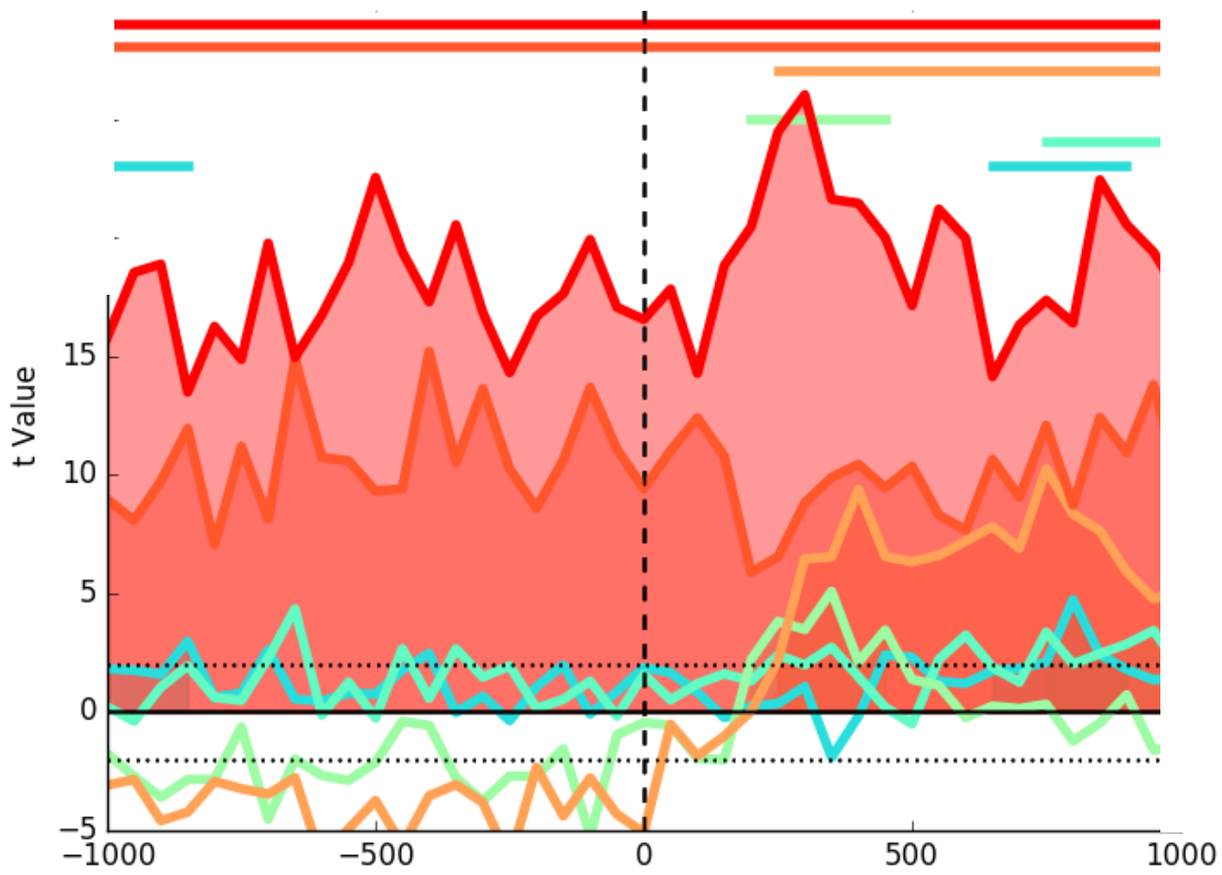


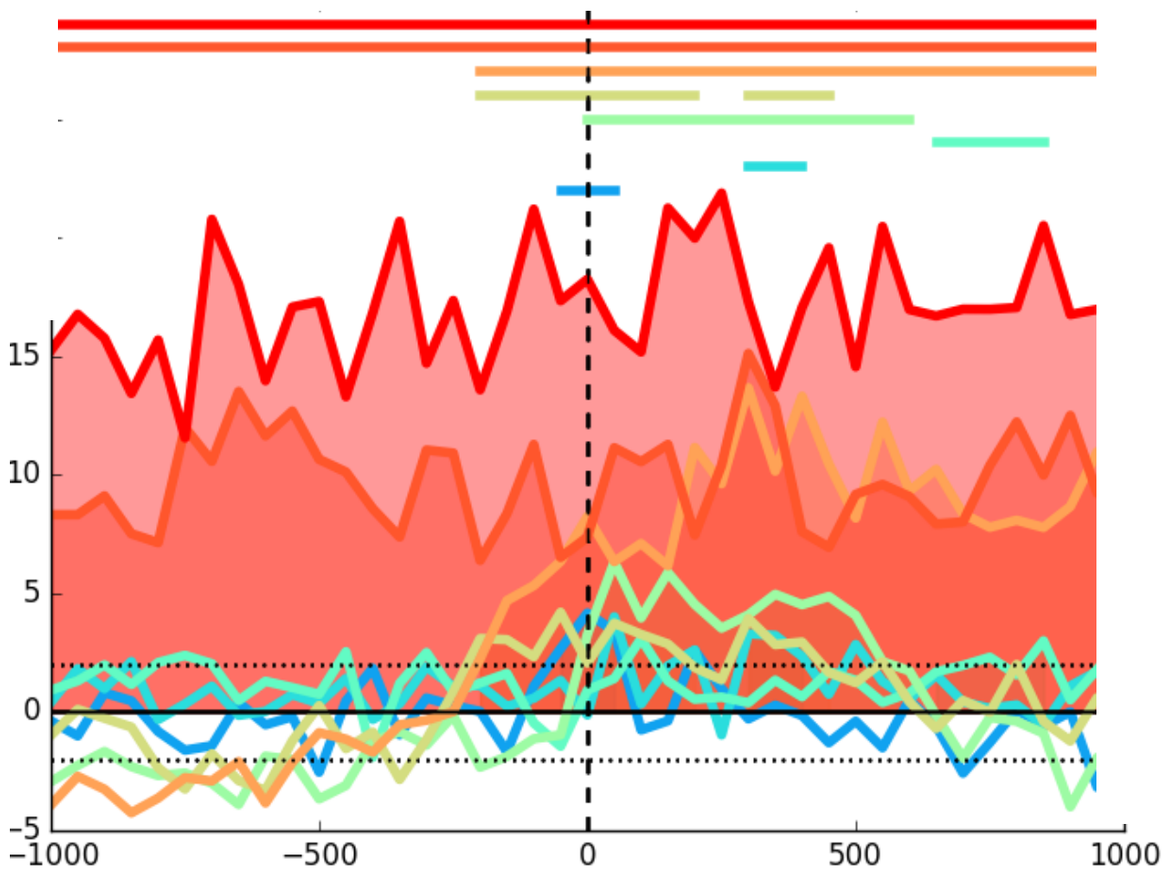
which linguistic units are encoded in brain activity?

Laura Gwilliams | New York University | @GwilliamsL

Deconfounded decoding accuracy



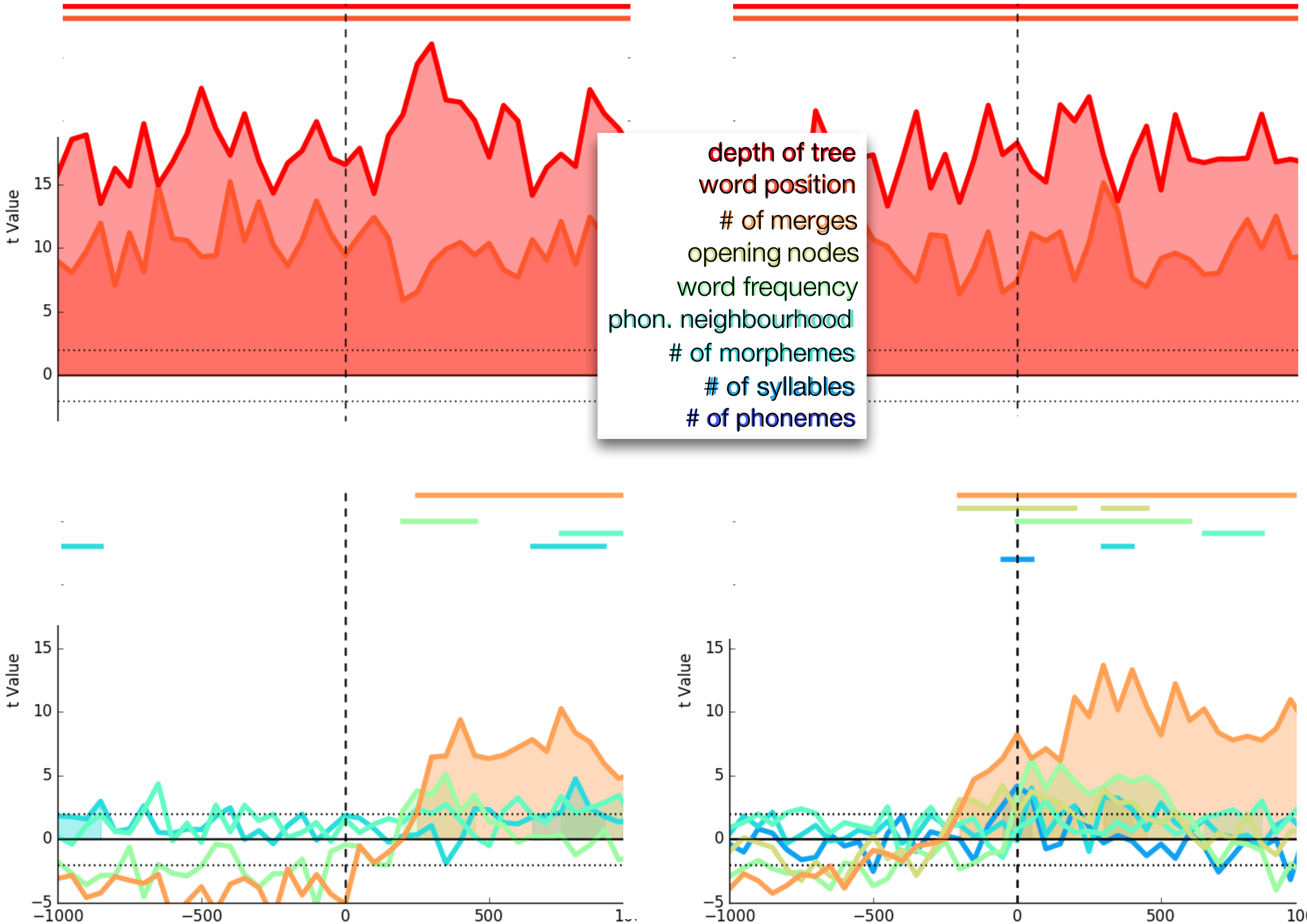




what are the relative time-courses?

Laura Gwilliams | New York University | @GwilliamsL

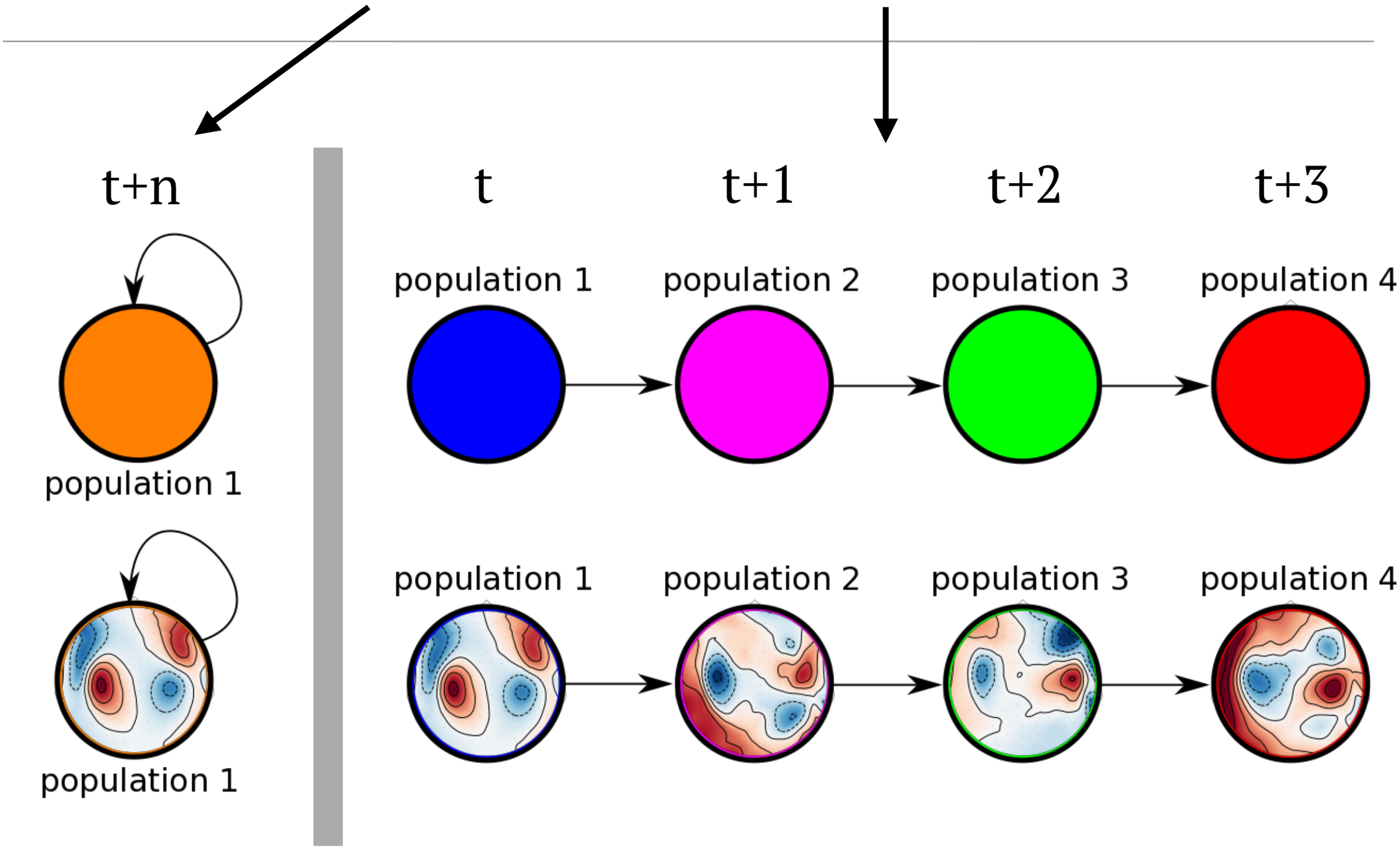
Deconfounded decoding accuracy



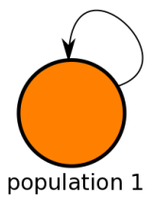
what is the computational architecture?

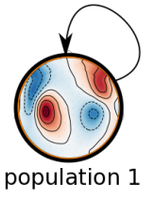
Laura Gwilliams | **New York University** | **@GwilliamsL**

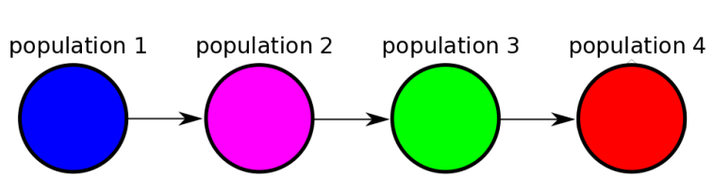
Recurrent vs. Feedforward

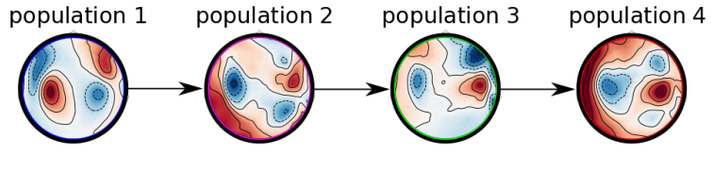


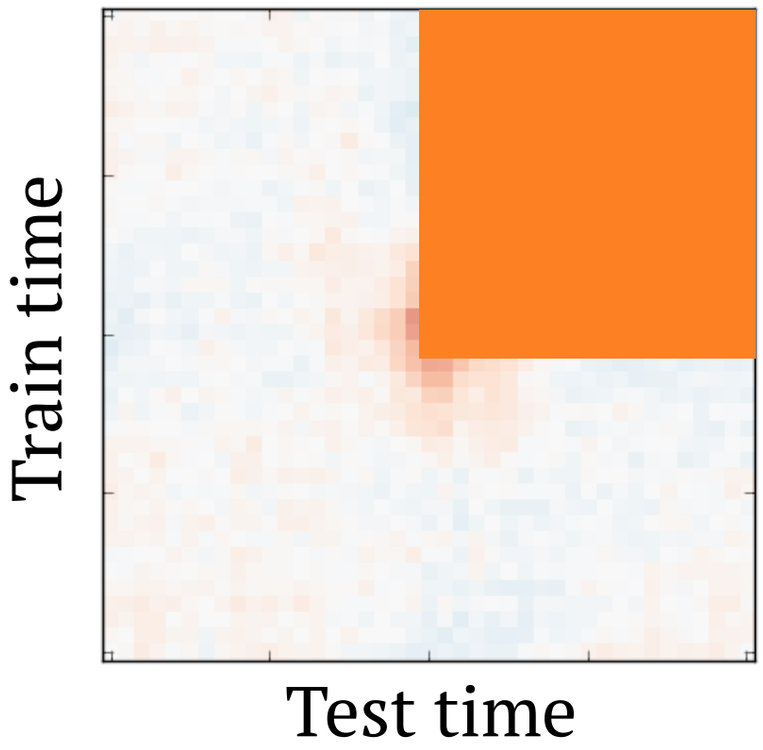
Recurrent vs. Feedforward

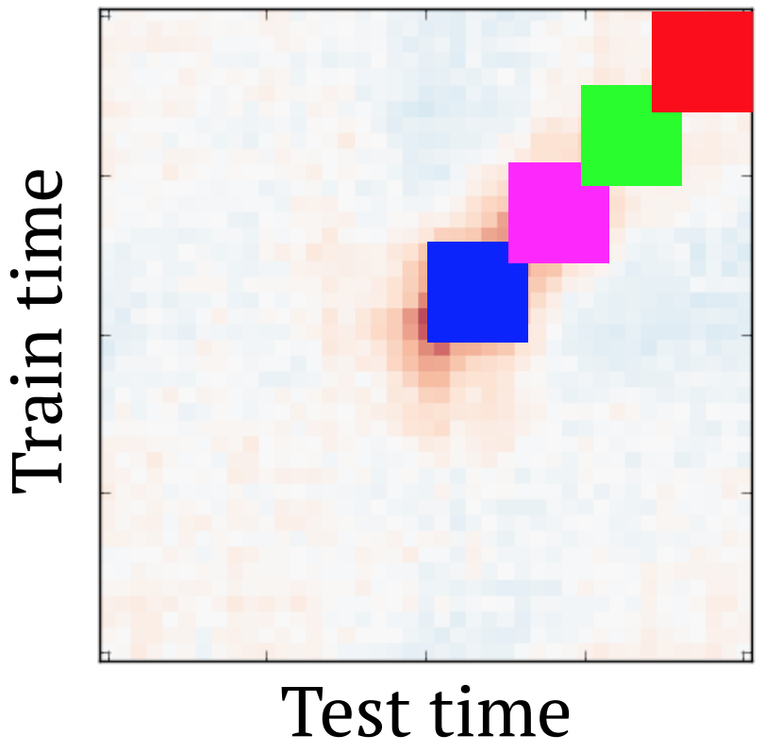






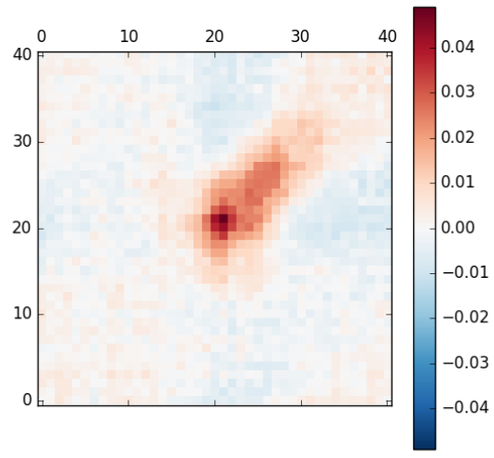




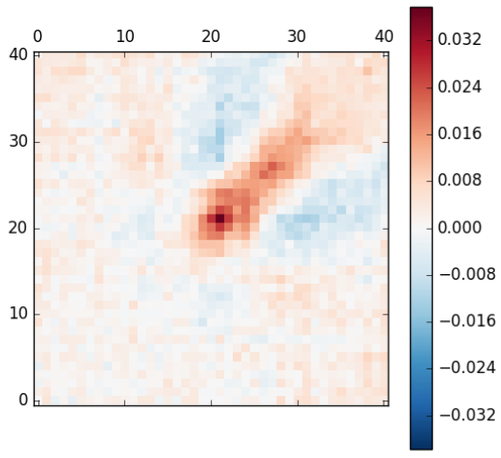


Recurrent vs. Feedforward

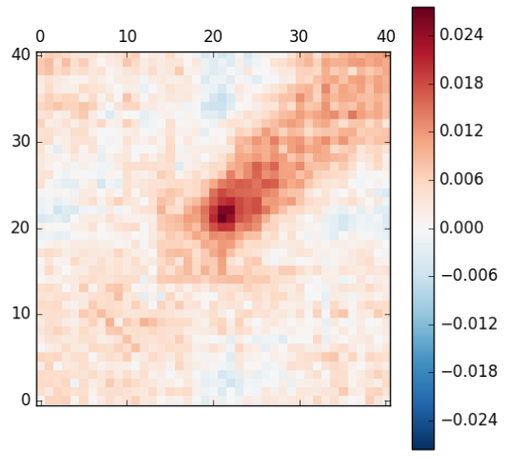
of syllables



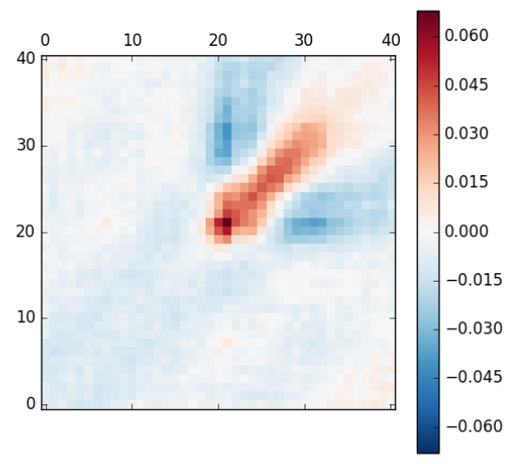
# of morphemes



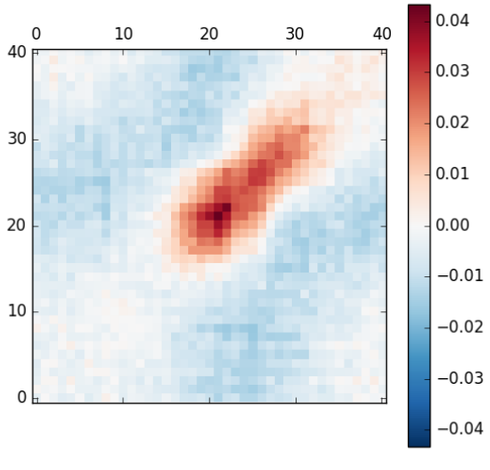
phon. neighbourhood



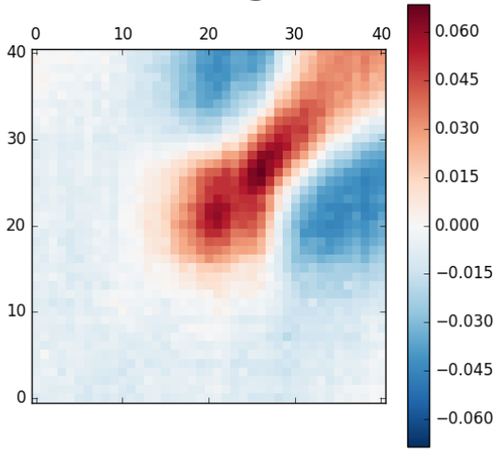
word frequency



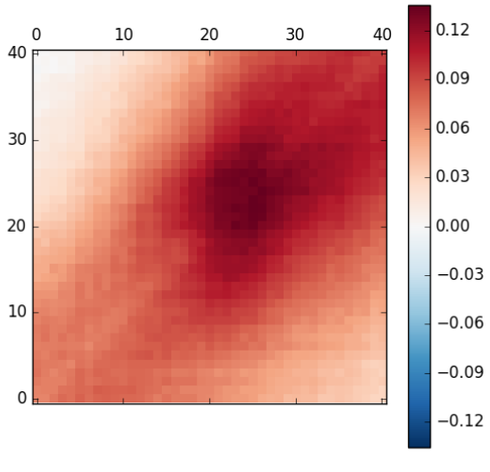
opening nodes



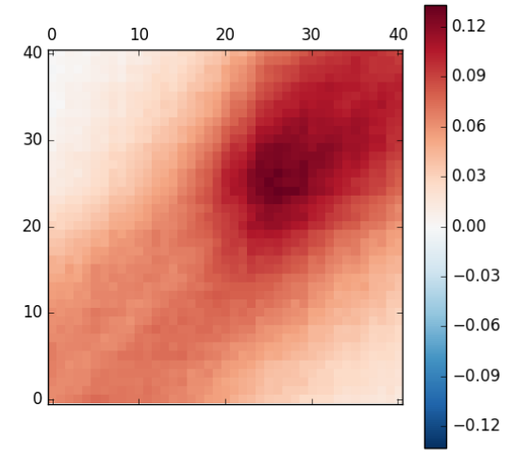
# of merges



word number



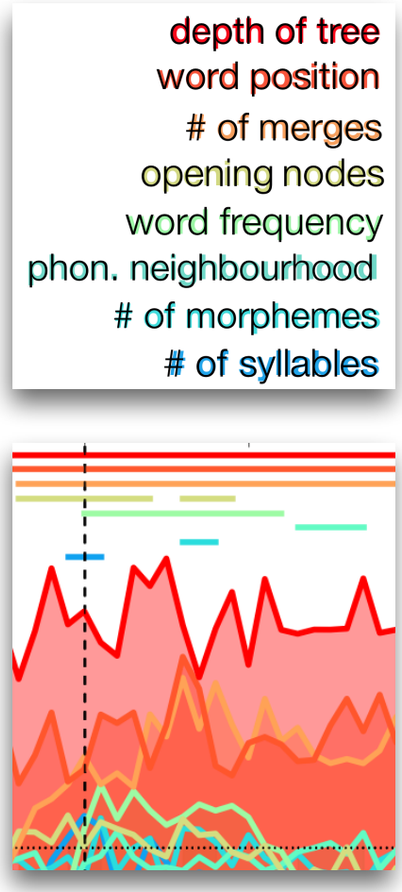
depth



Discussion

(1) Which linguistic units are encoded?

- Multiple features, **spanning the hierarchy**
- Including # of **syllables**; # of **morphemes**

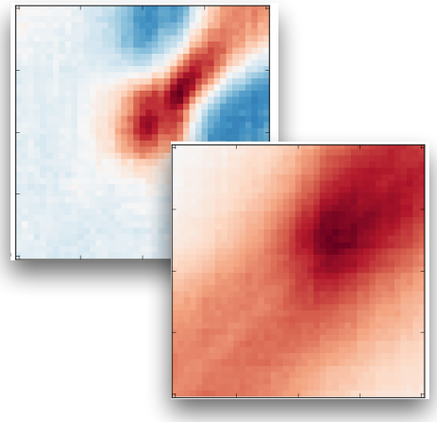


(2) What is the relative time-course?

- Overall a highly **parallel** architecture

(3) What is the computational architecture?

- Both **feedforward** and **recurrent** computations, depending on the linguistic property



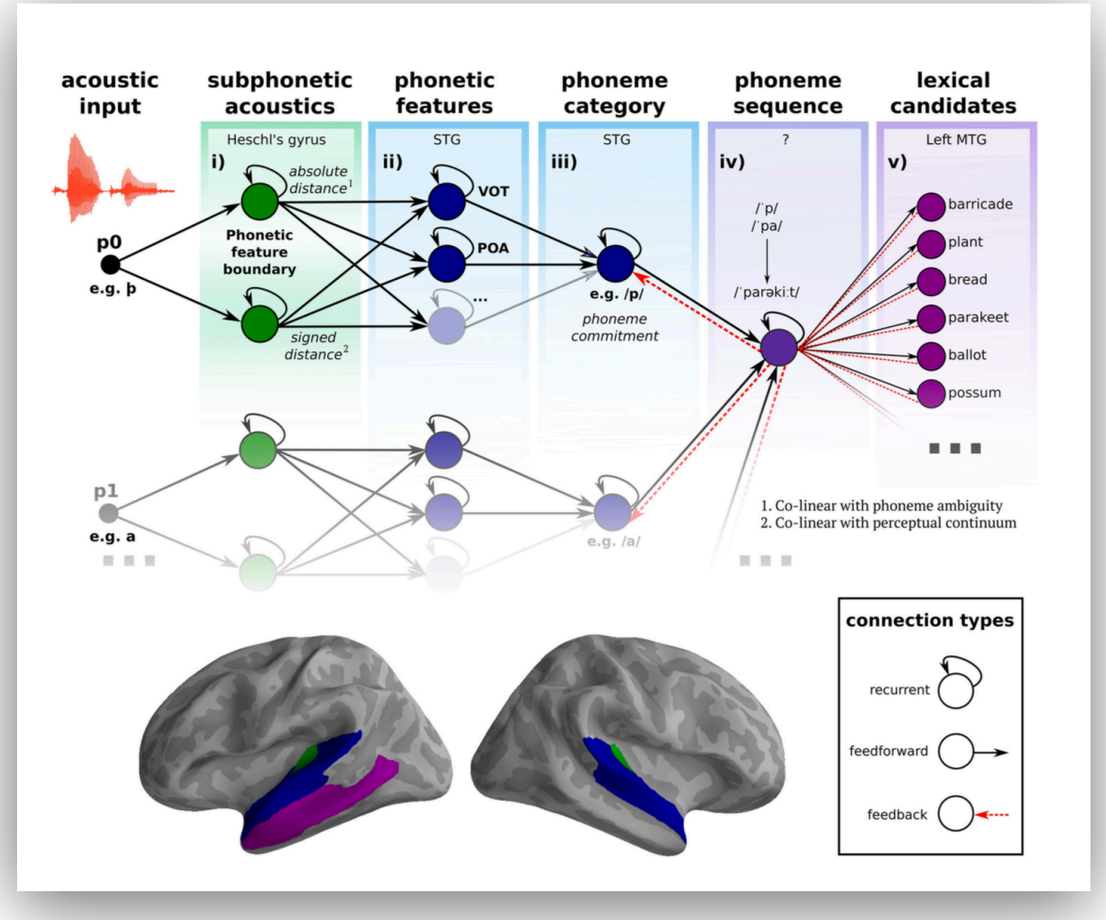
Levels of analysis

1. Phonemes within words
   - Responses to phoneme ambiguity, phonetic features and acoustic properties **(bottom-up)**
   - Neural signatures of ambiguity resolution, when provided with lexical information **(top-down)**

2. Words within sentences
   - Which linguistic properties encoded in brain activity?
   - What are the relative time-courses of processing each property?
   - What is the computational architecture?

Take home message

- **Feedforward, recurrent**, and **highly parallel processes** may be a general architecture supporting speech comprehension
  - words within sentences
  - phonemes within words



laura.gwilliams@nyu.edu
@GwilliamsL

With big thanks to:

- My supervisors, **Alec Marantz** and **David Poeppel**, as well as everyone in the **Neuroscience of Language Lab** and **Poeppel Lab**!

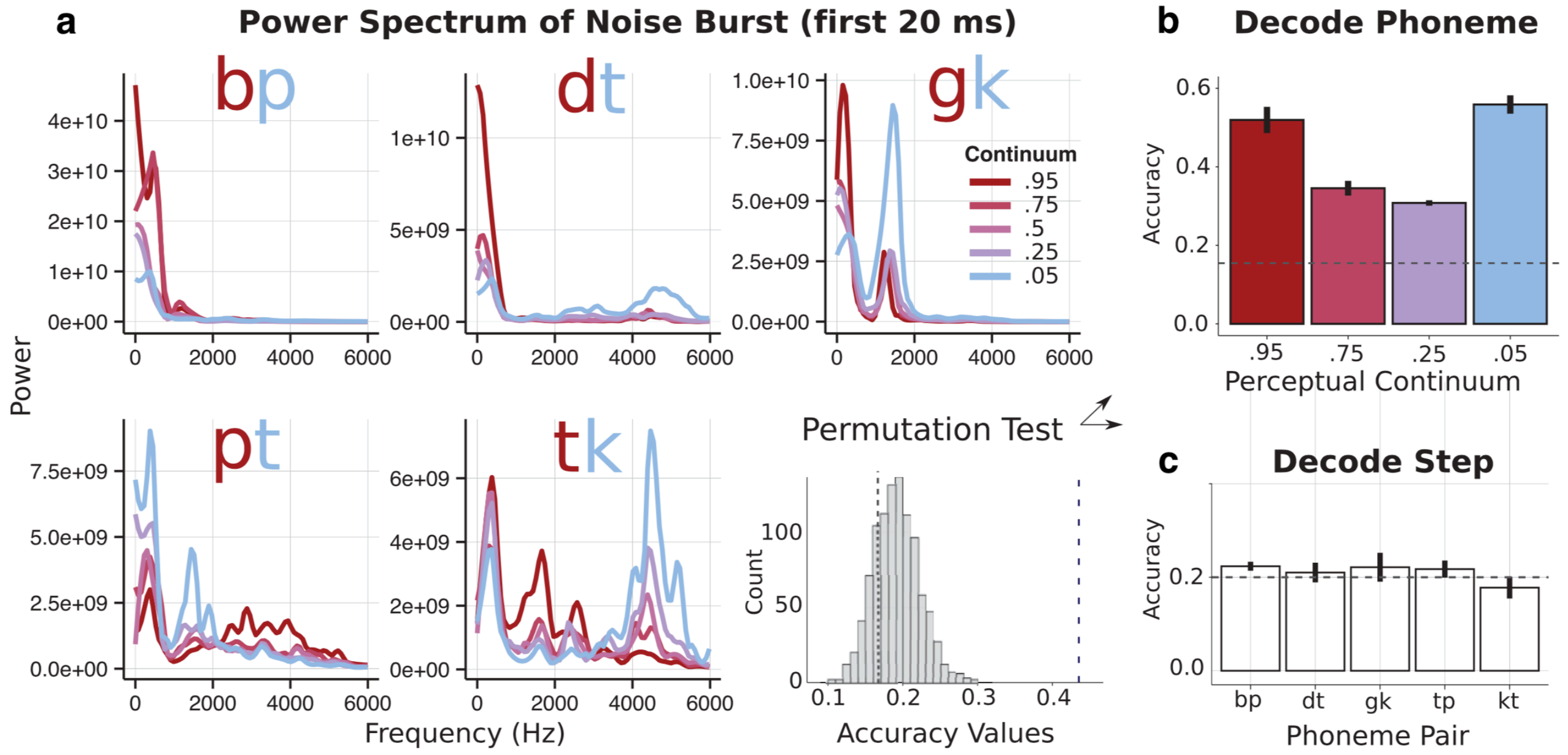
Funding: G1001 Abu Dhabi Institute

NEW YORK UNIVERSITY

laura.gwilliams@nyu.edu
@GwilliamsL

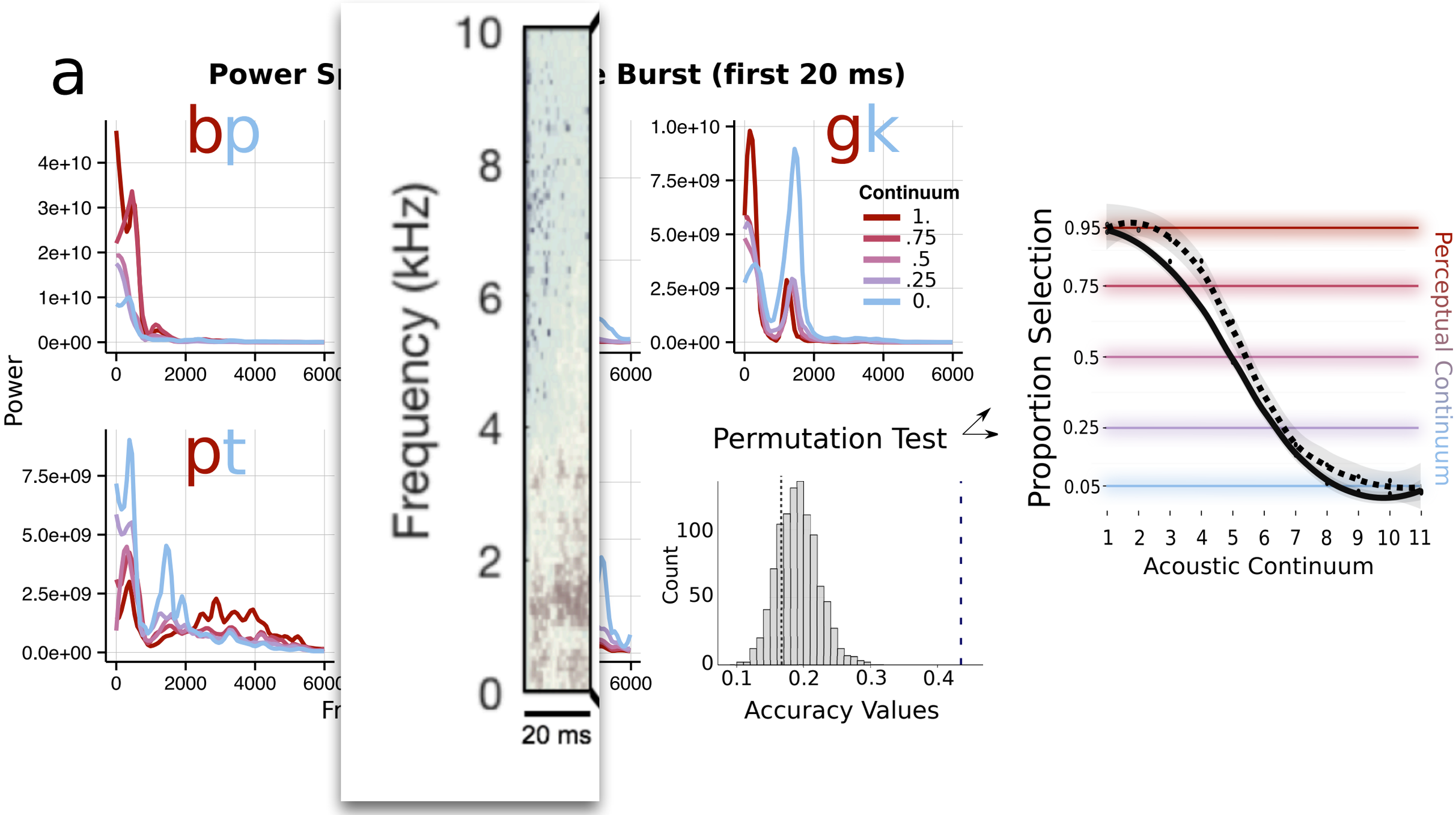
Thank you!

Laura Gwilliams | **New York University** | **@GwilliamsL**

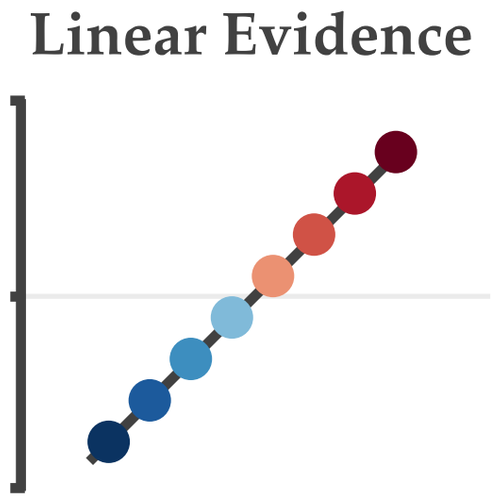


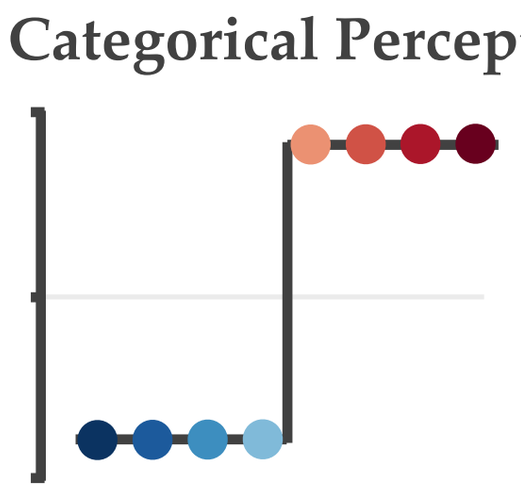
Laura Gwilliams | New York University | @GwilliamsL

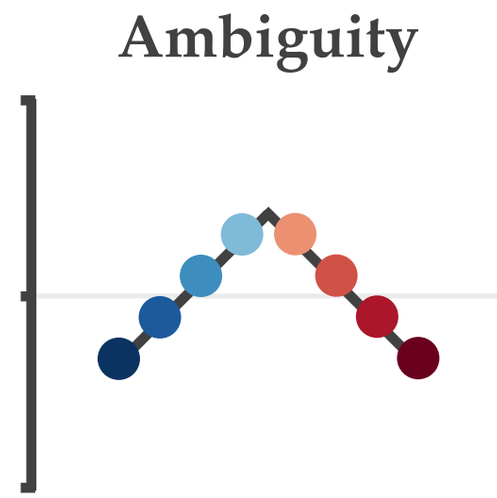
Is ambiguity correlated with acoustic properties?



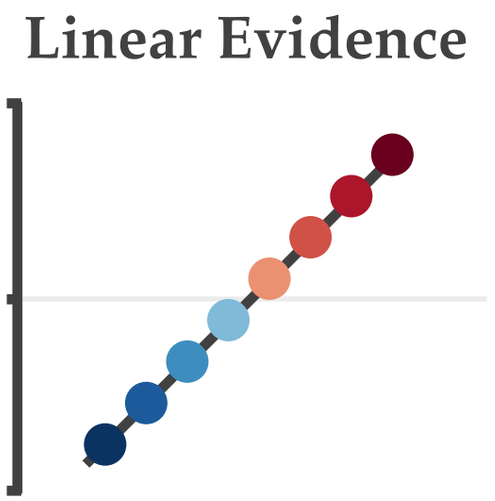
Predictive Coding

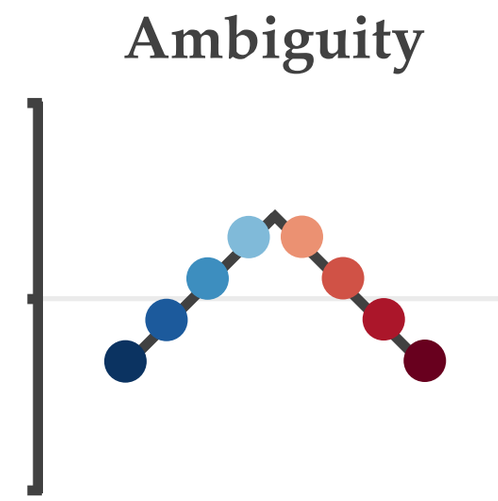


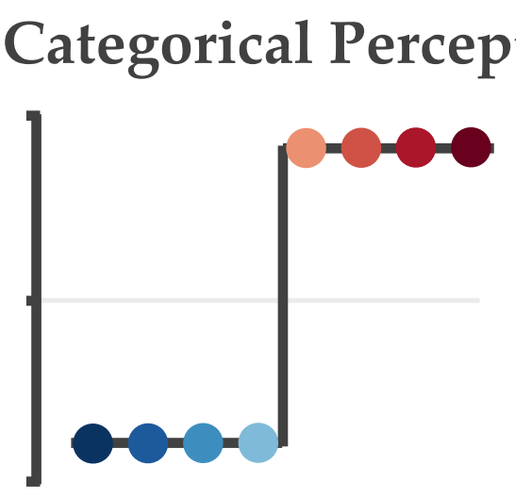




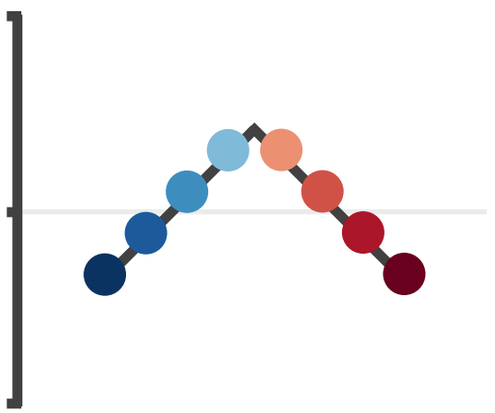
Neutralisation

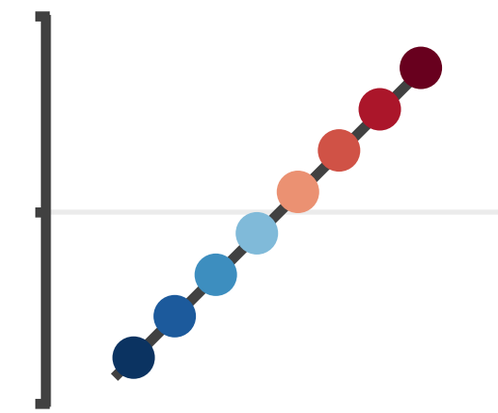


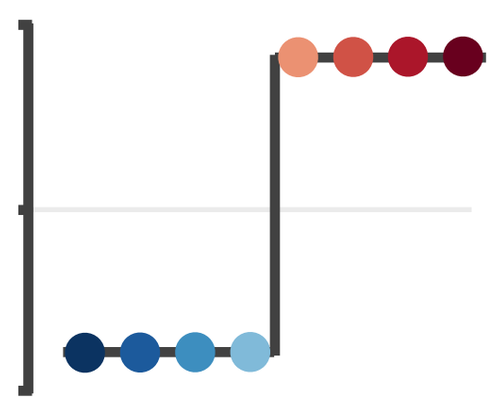




Cut-through connection

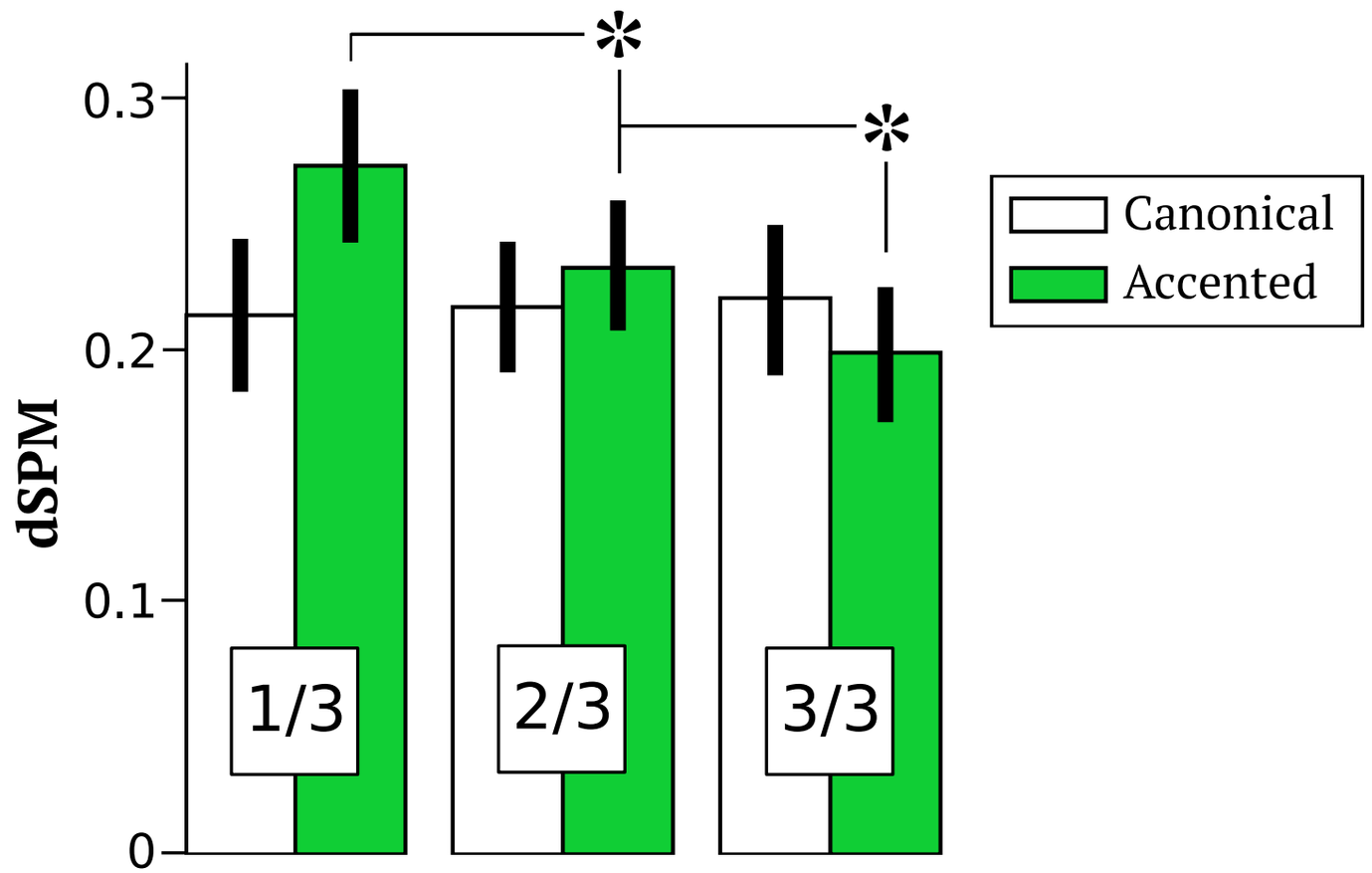




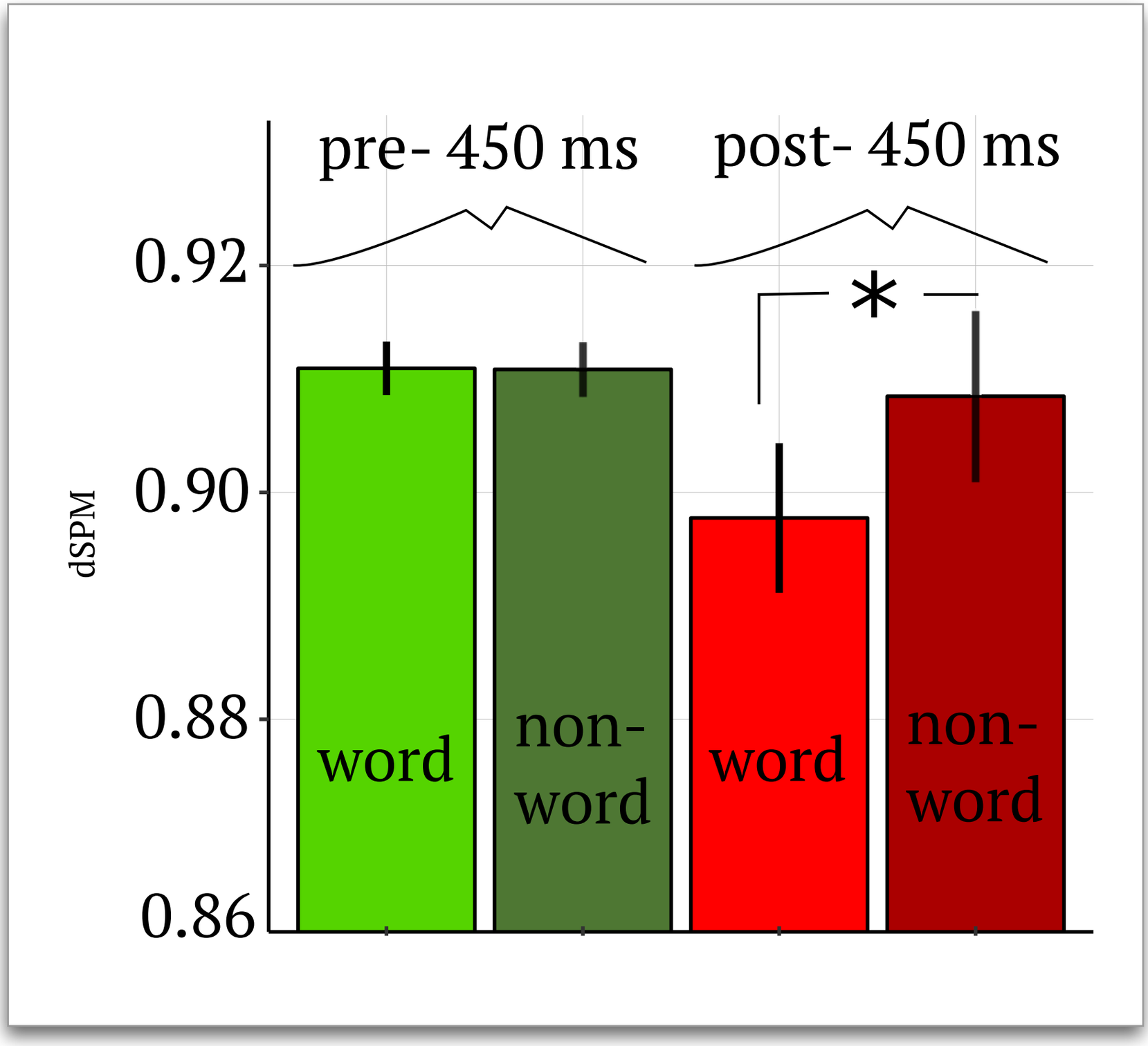


Laura Gwilliams | New York University | @GwilliamsL

Interpretation



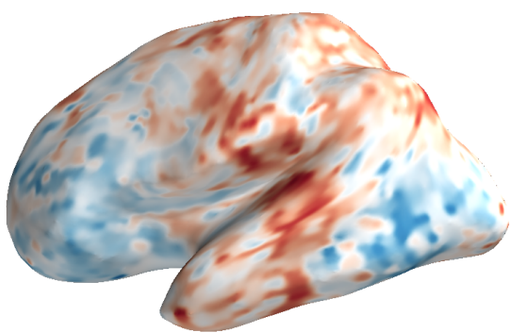
- Attunement is proposed to involve **re-tuning perceptual boundaries** between phonological categories (Norris et al., 2003; Kraljic and Samuel, 2005, 2006, 2007; Maye et al., 2008; see Samuel & Kraljic, 2009 for a review)



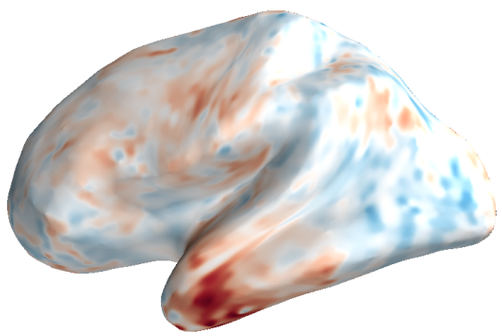
Laura Gwilliams | New York University | @GwilliamsL

(Preliminary) Localisation of effects

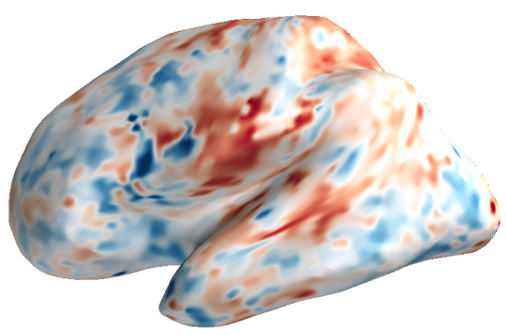
#syllables



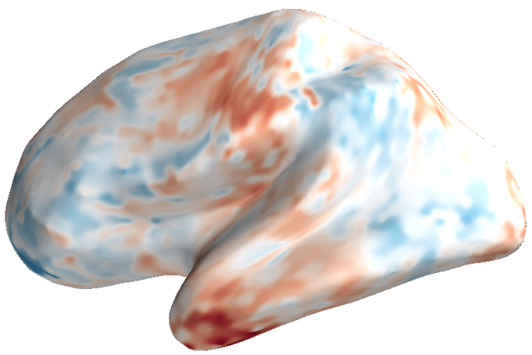
#morphemes



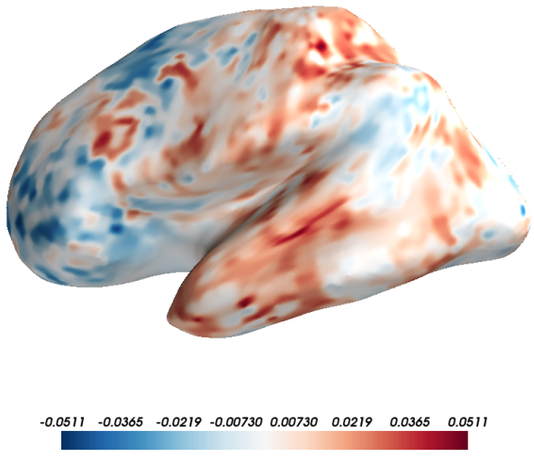
phon. neighborhood



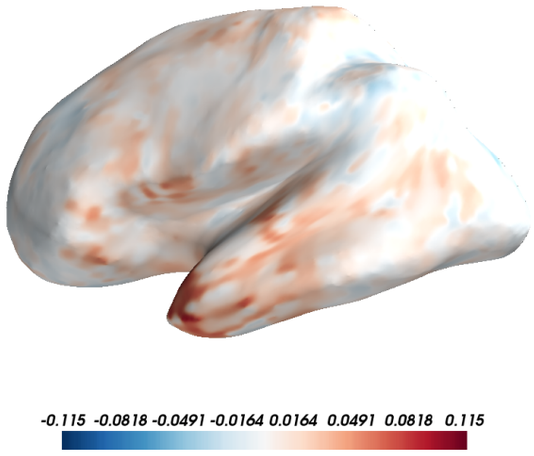
word frequency



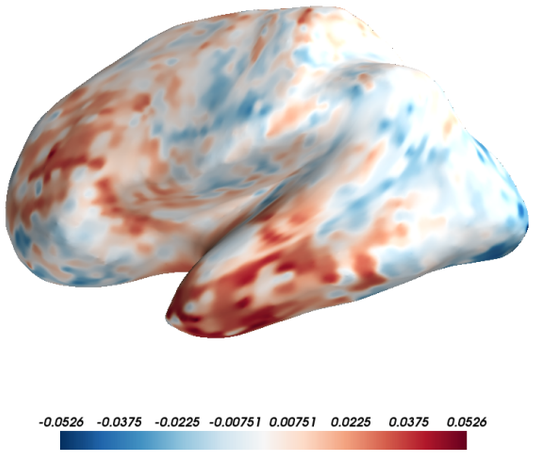
#opening nodes



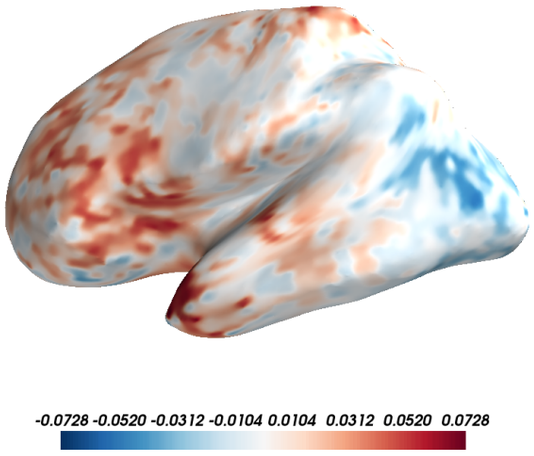
#merge



word number



depth



Confounded decoding accuracy