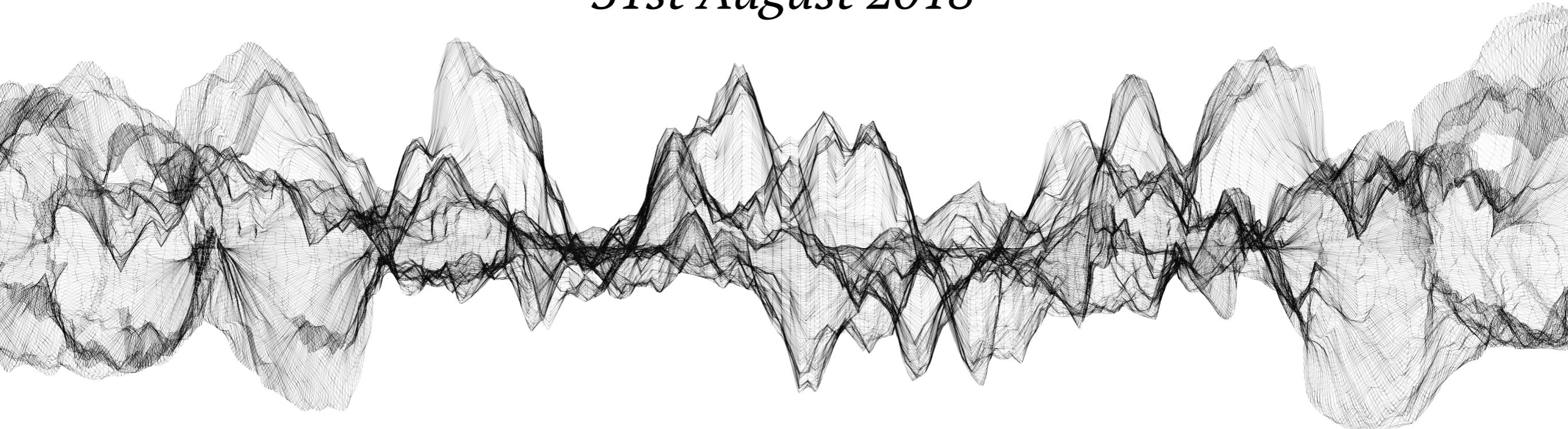




Towards a mechanistic account of speech comprehension

Laura Gwilliams

31st August 2018



Levels of analysis

1. Phonemes within words

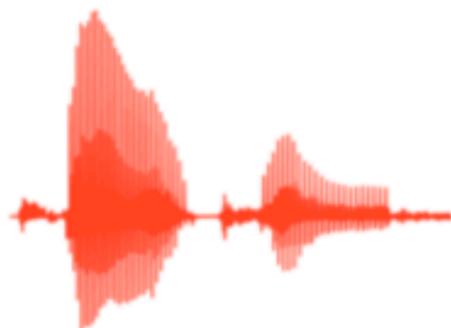
- Responses to phoneme ambiguity, phonetic features and acoustic properties (**bottom-up**)
- Neural signatures of ambiguity resolution, when provided with lexical information (**top-down**)
- What is the computational architecture?

2. Words within sentences

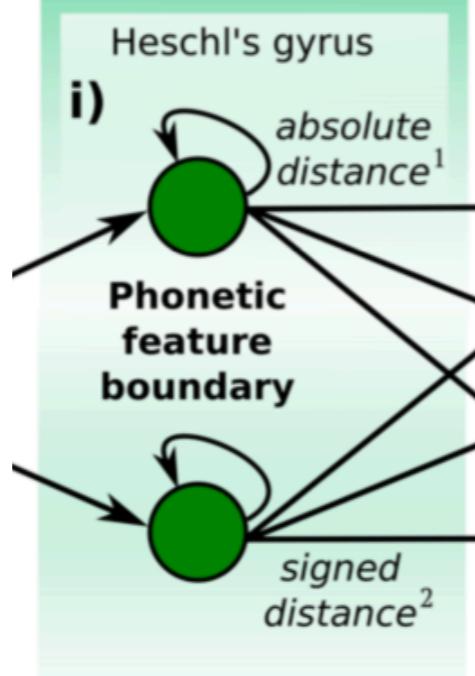
- Which linguistic properties encoded in brain activity?
- What are the relative time-courses of processing each property?
- What is the computational architecture?

Putting together the processing pieces

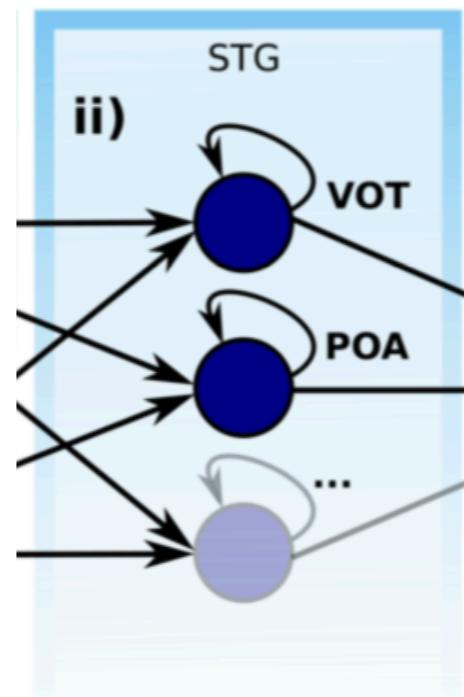
acoustic input



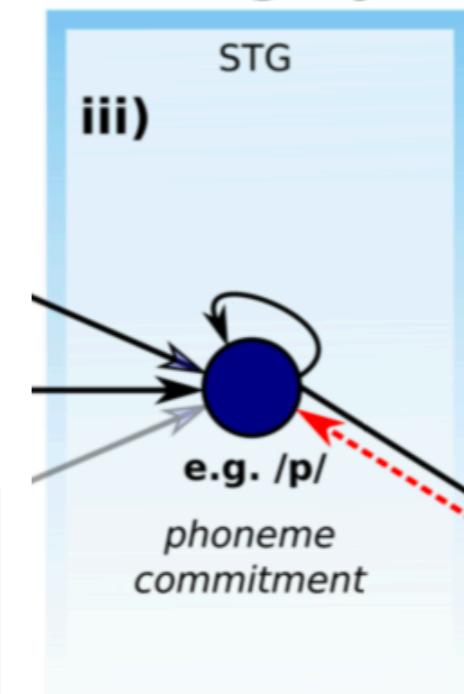
subphonetic acoustics



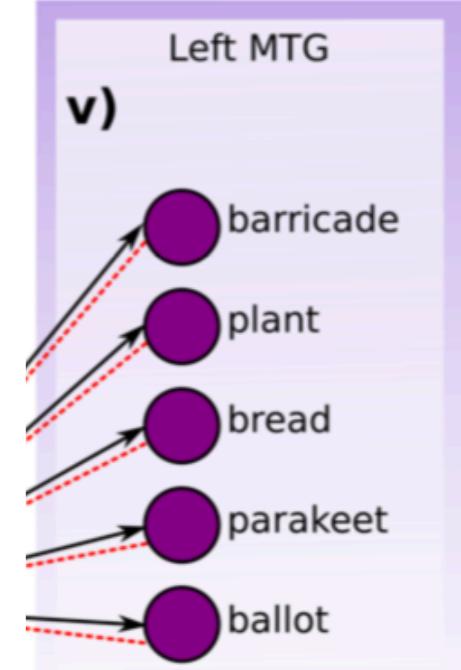
phonetic features



phoneme category



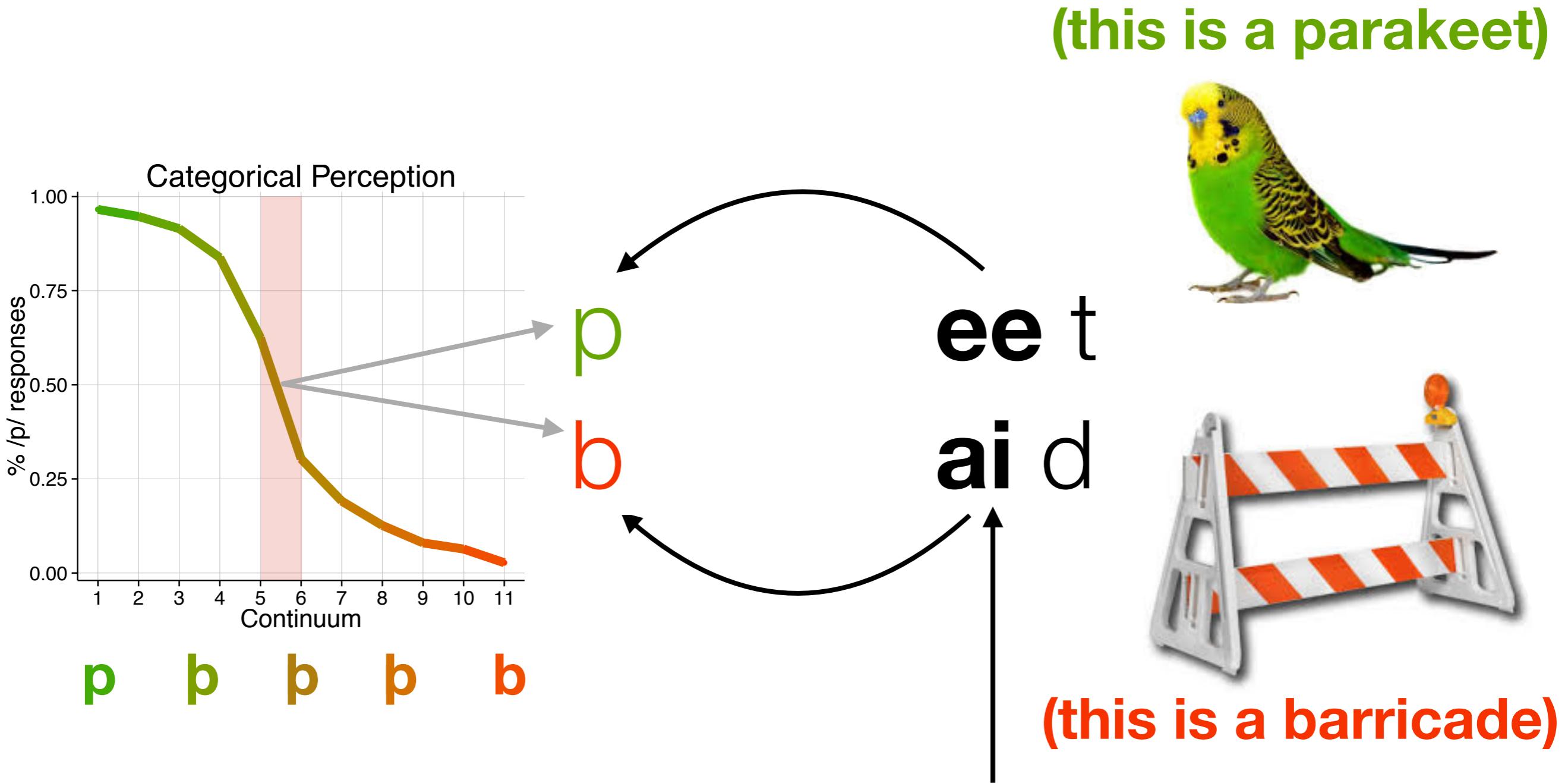
lexical candidates



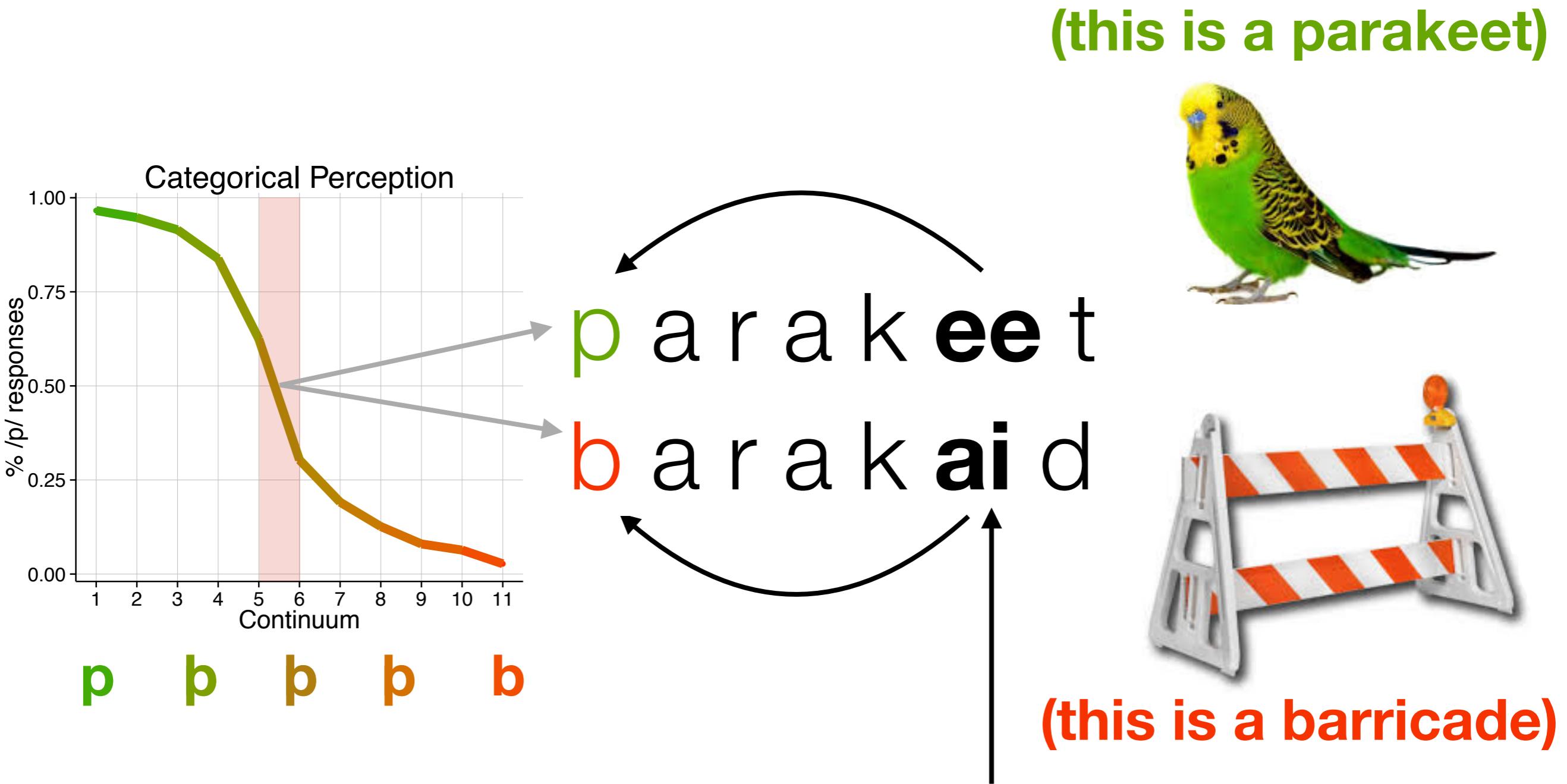
Future Influences on Perception

- 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



Future Influences on Perception



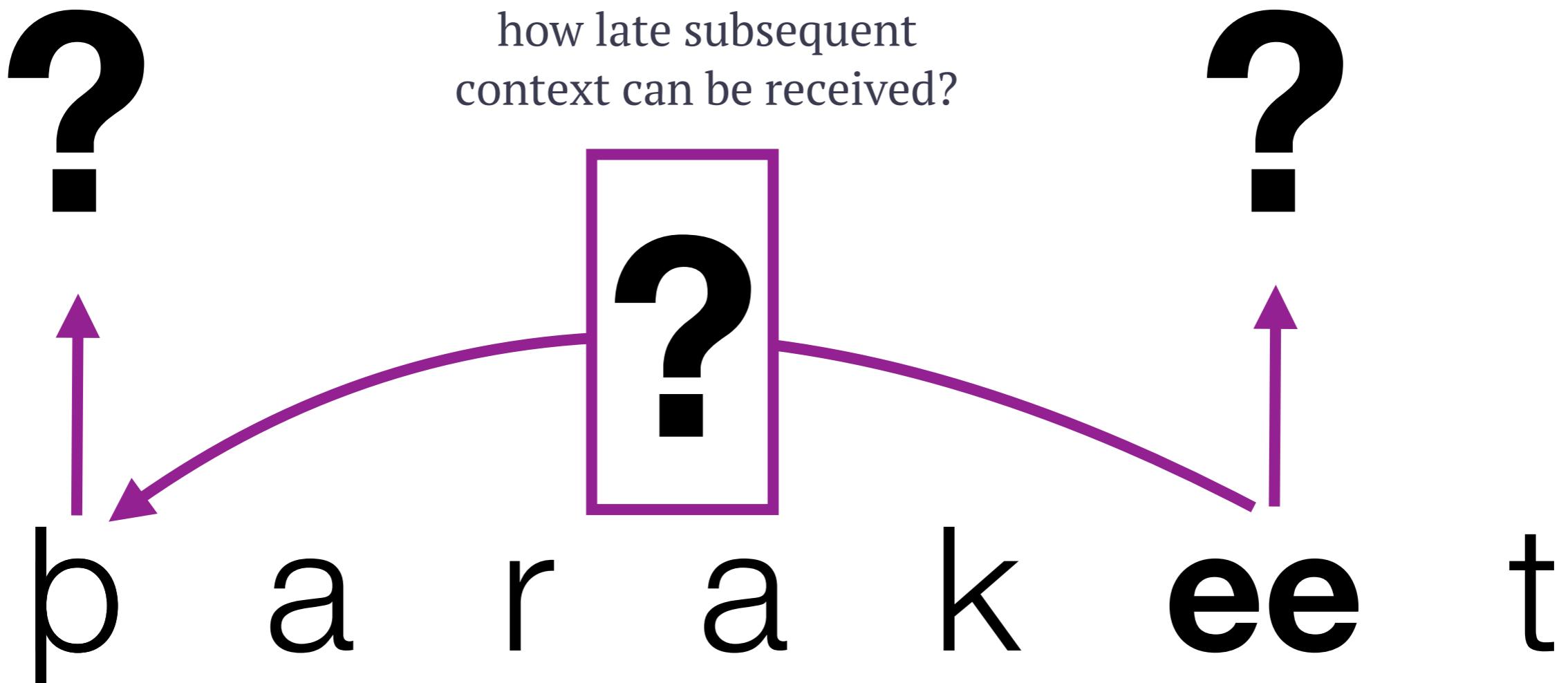
Today's Questions

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

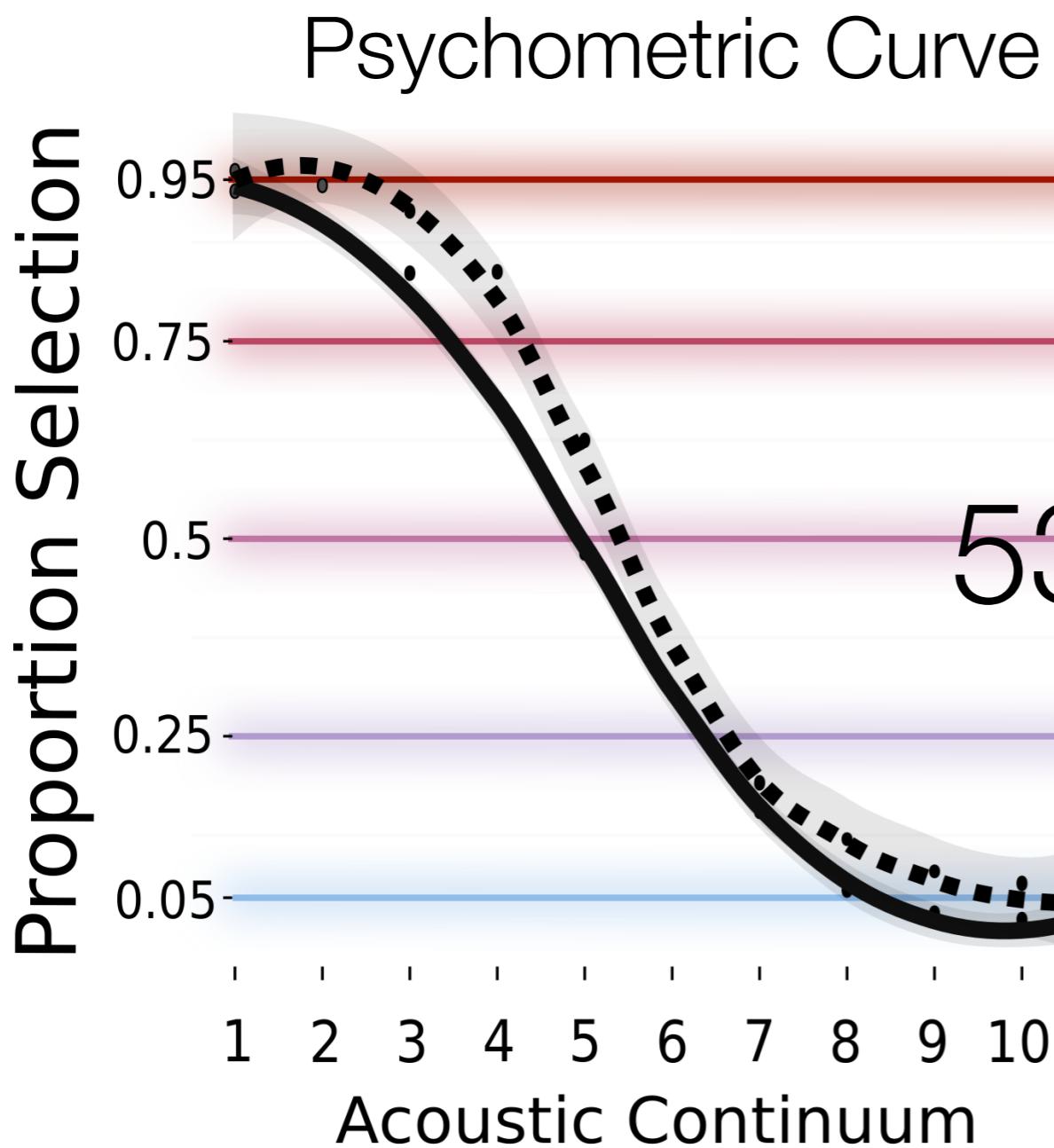
i could focus on the ambiguity resolution part here, rather than the original response to ambiguity. then, tie in the ambiguity response part later, linking it with AI?

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

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

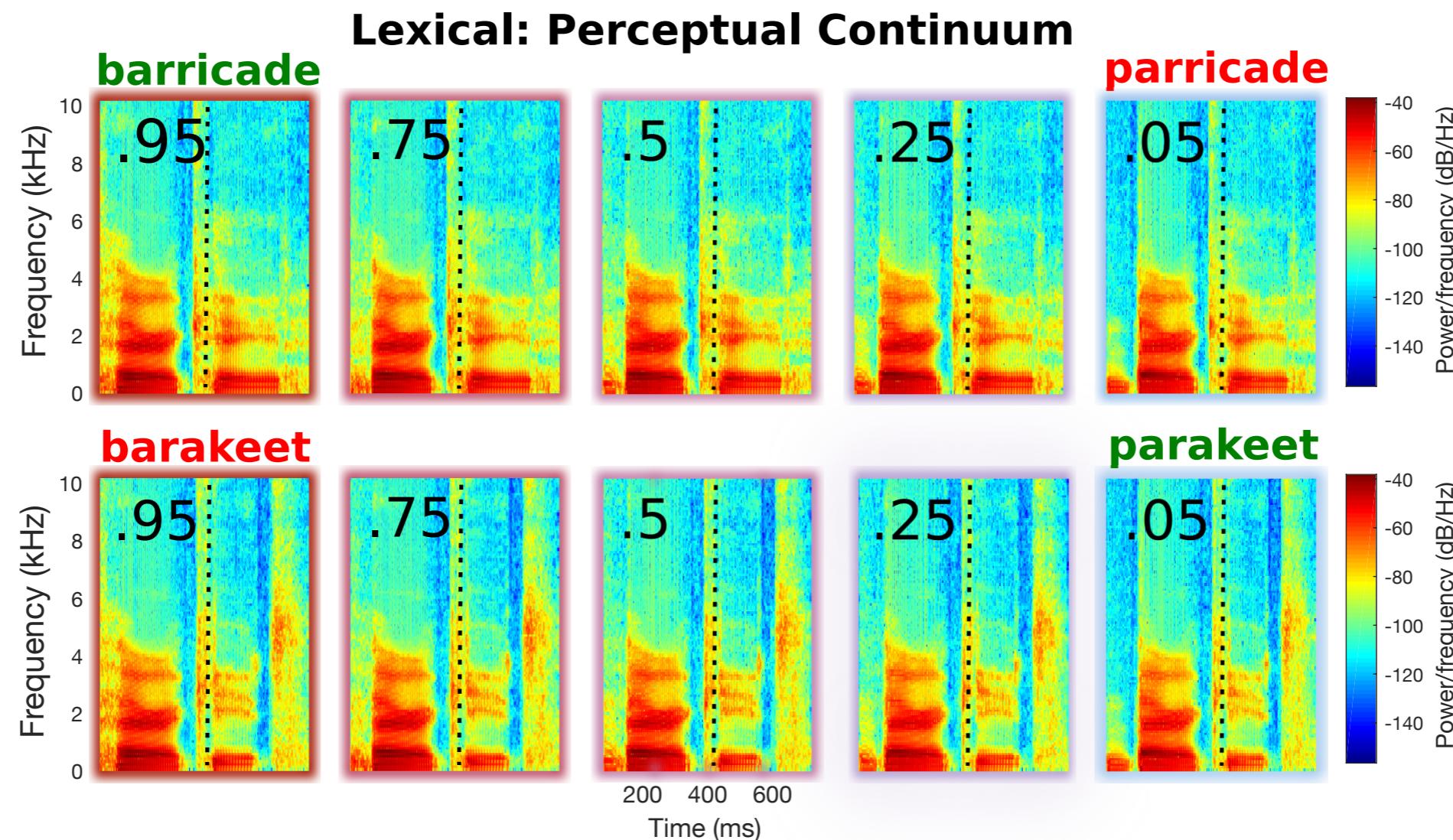


Design & Materials



p p b b b

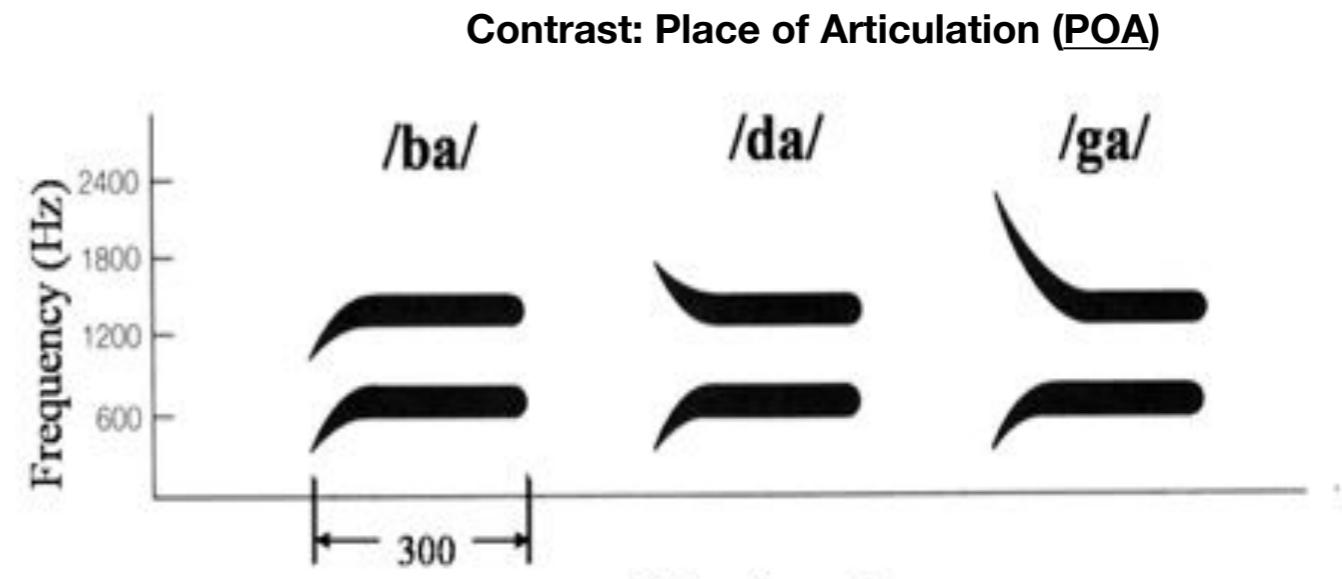
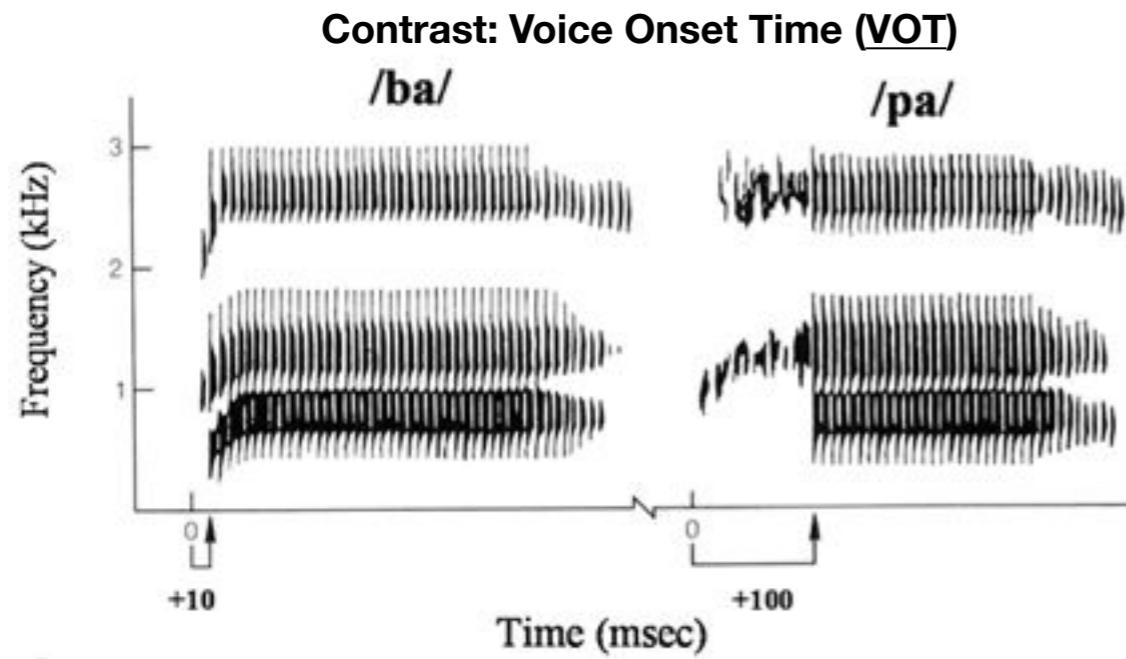
Design & Materials



- 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}

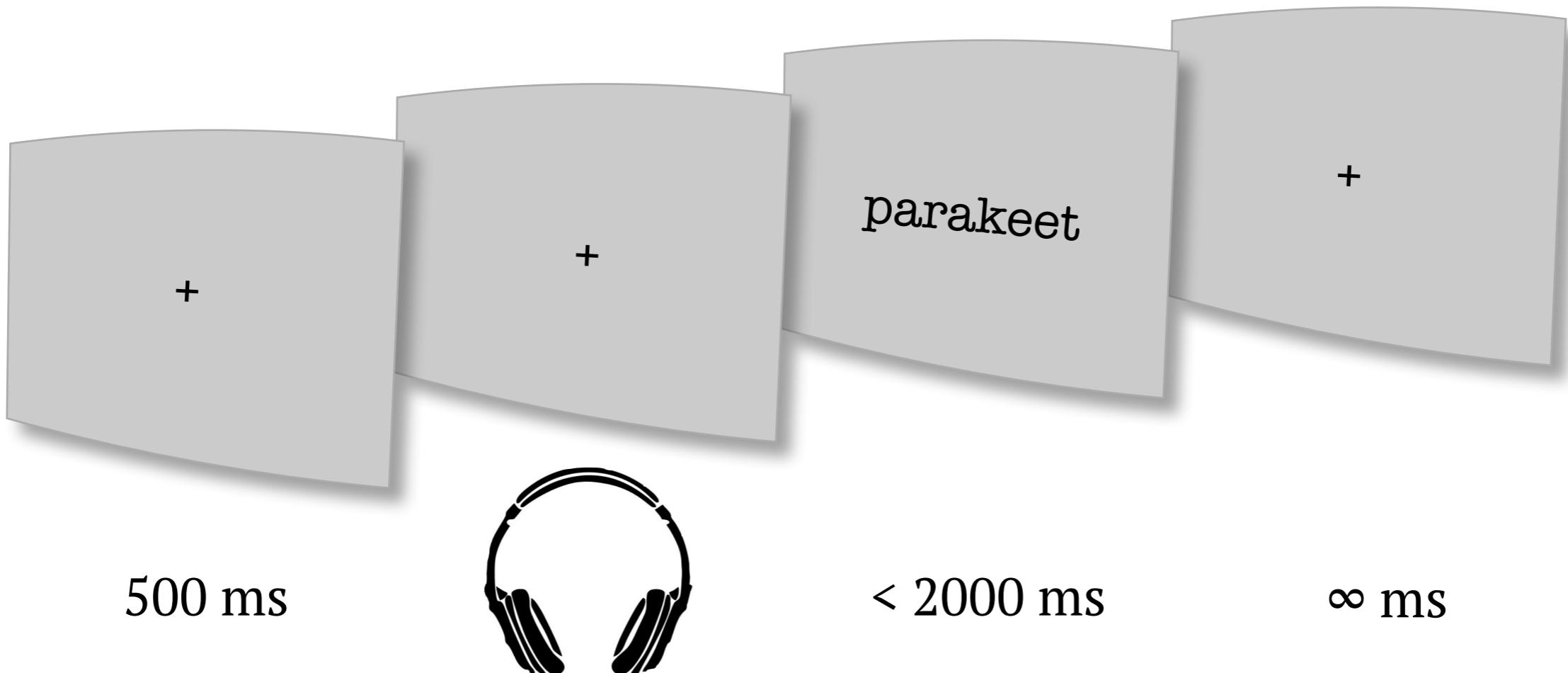
p b b b b

Design & Materials

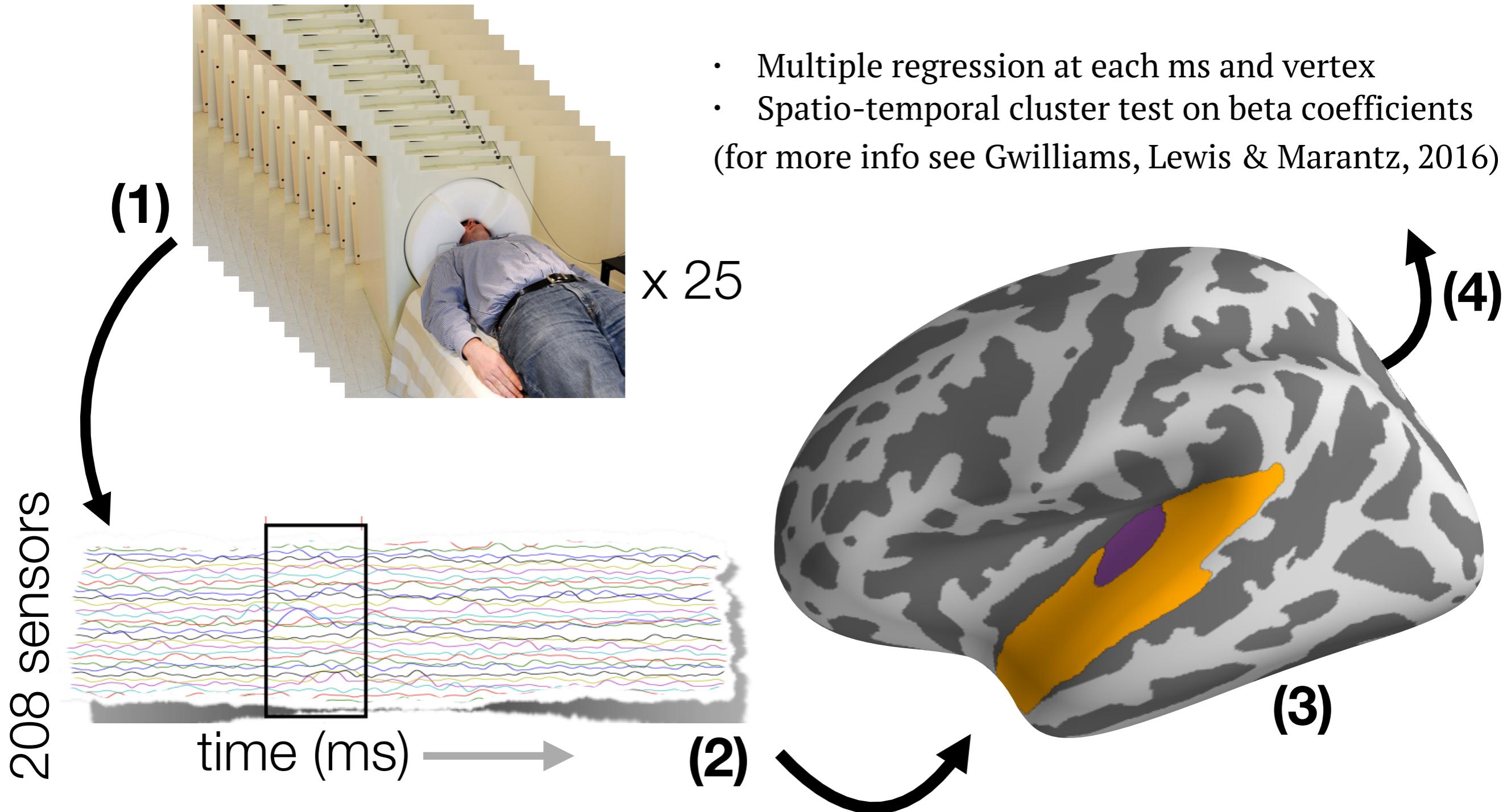


- Point of Disambiguation (POD) ranged 3-8 phonemes / 150-750 ms
- VOT (31 pairs) {p-b, t-d, k-g} and POA (22 pairs) {t-k, p-t}

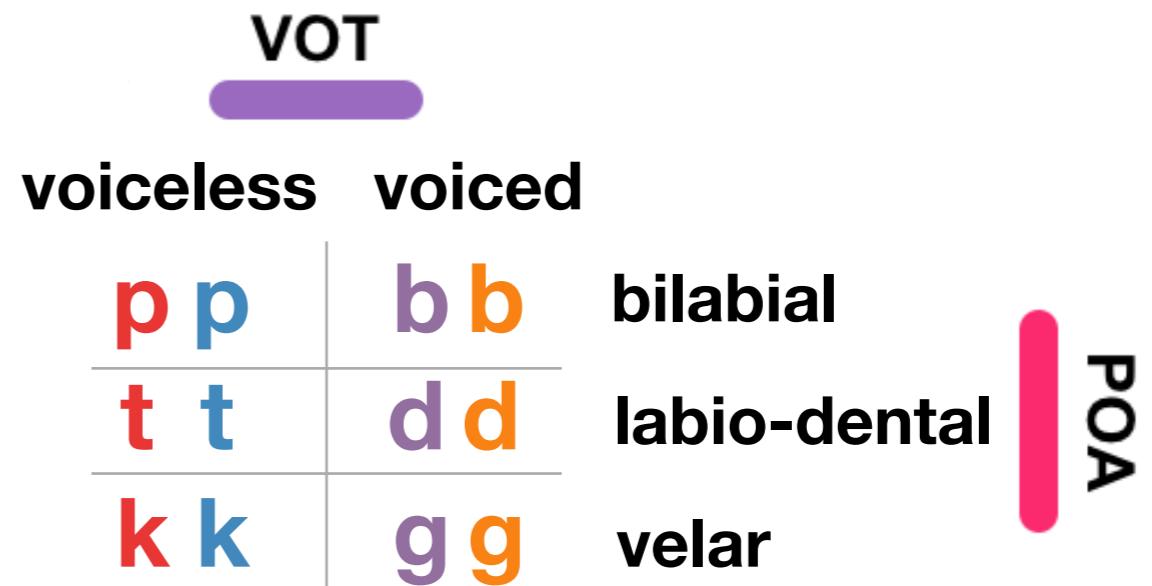
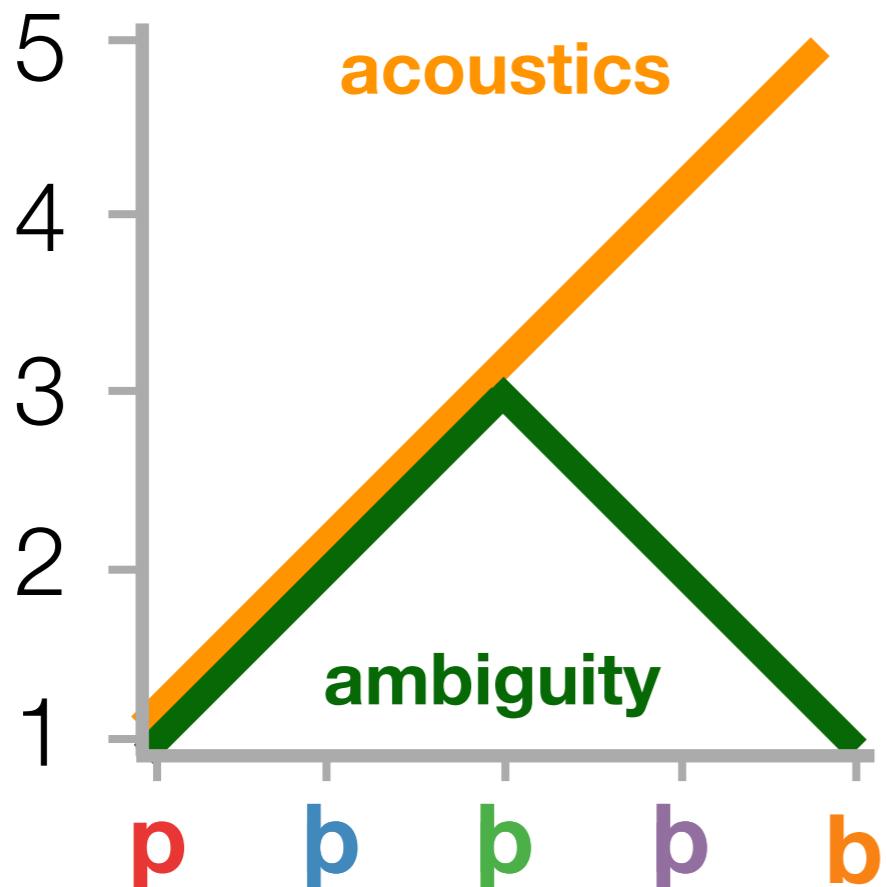
Design & Materials



Procedure & Analysis



Four Experimental Variables



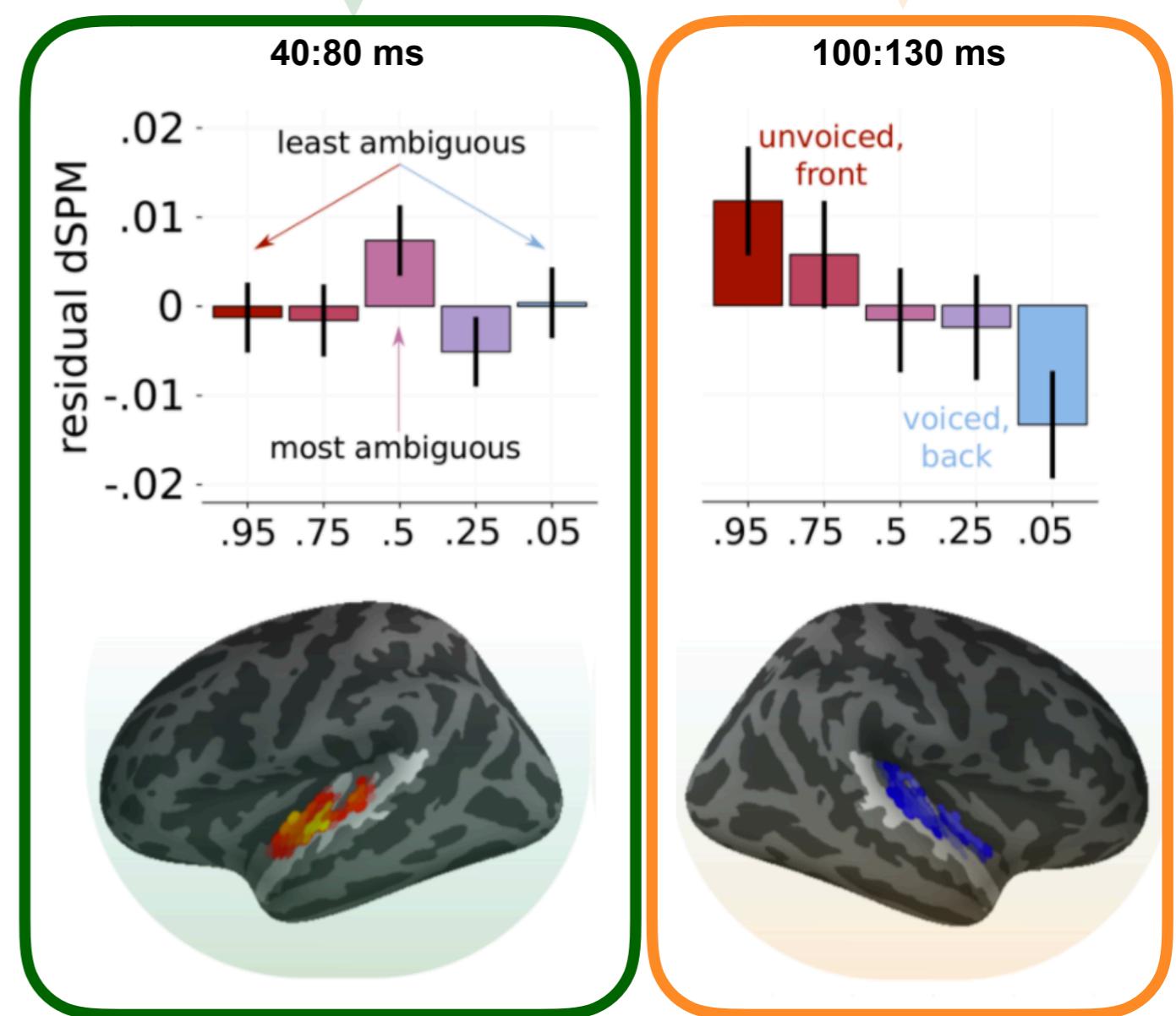
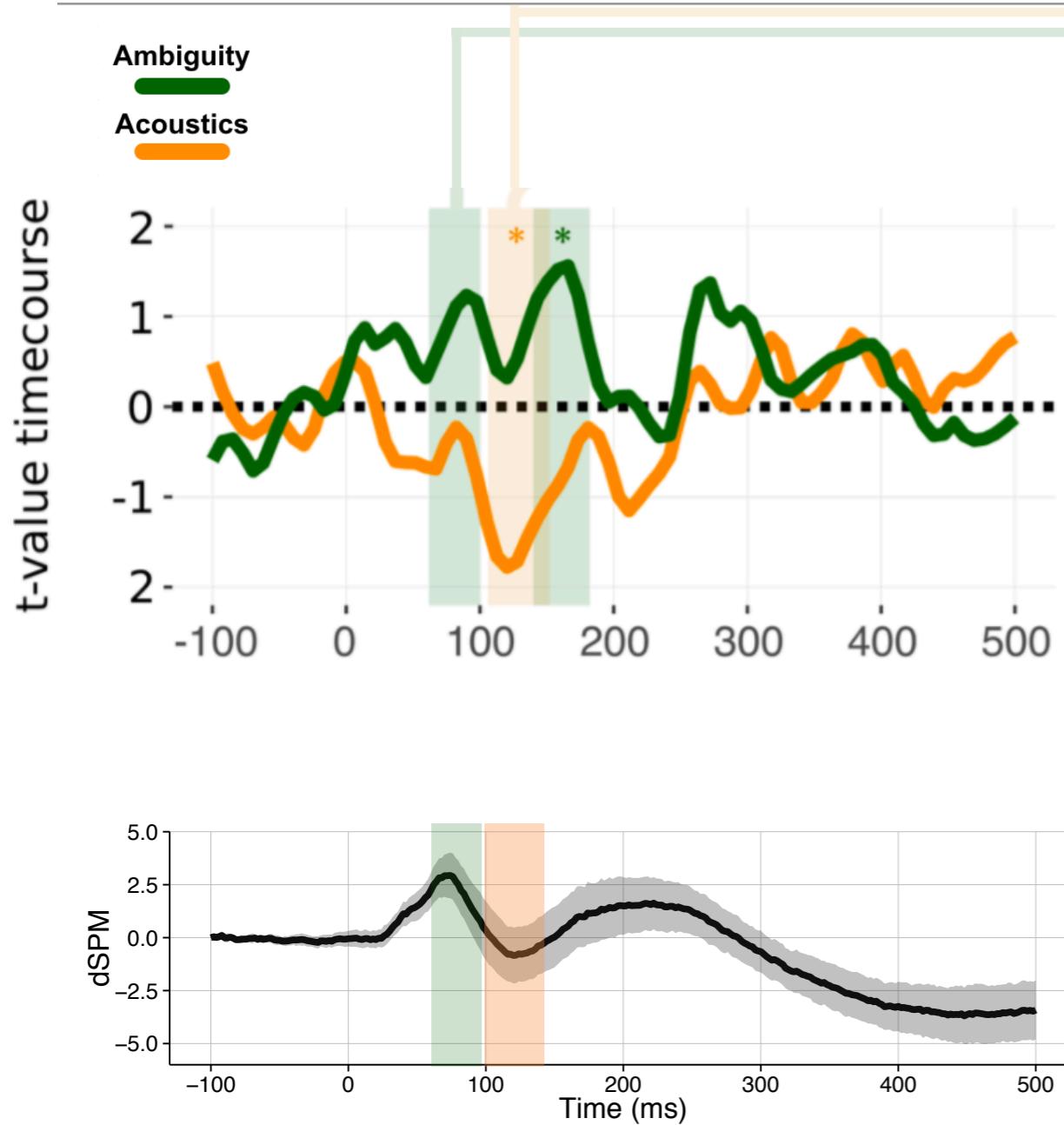
Today's Questions

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



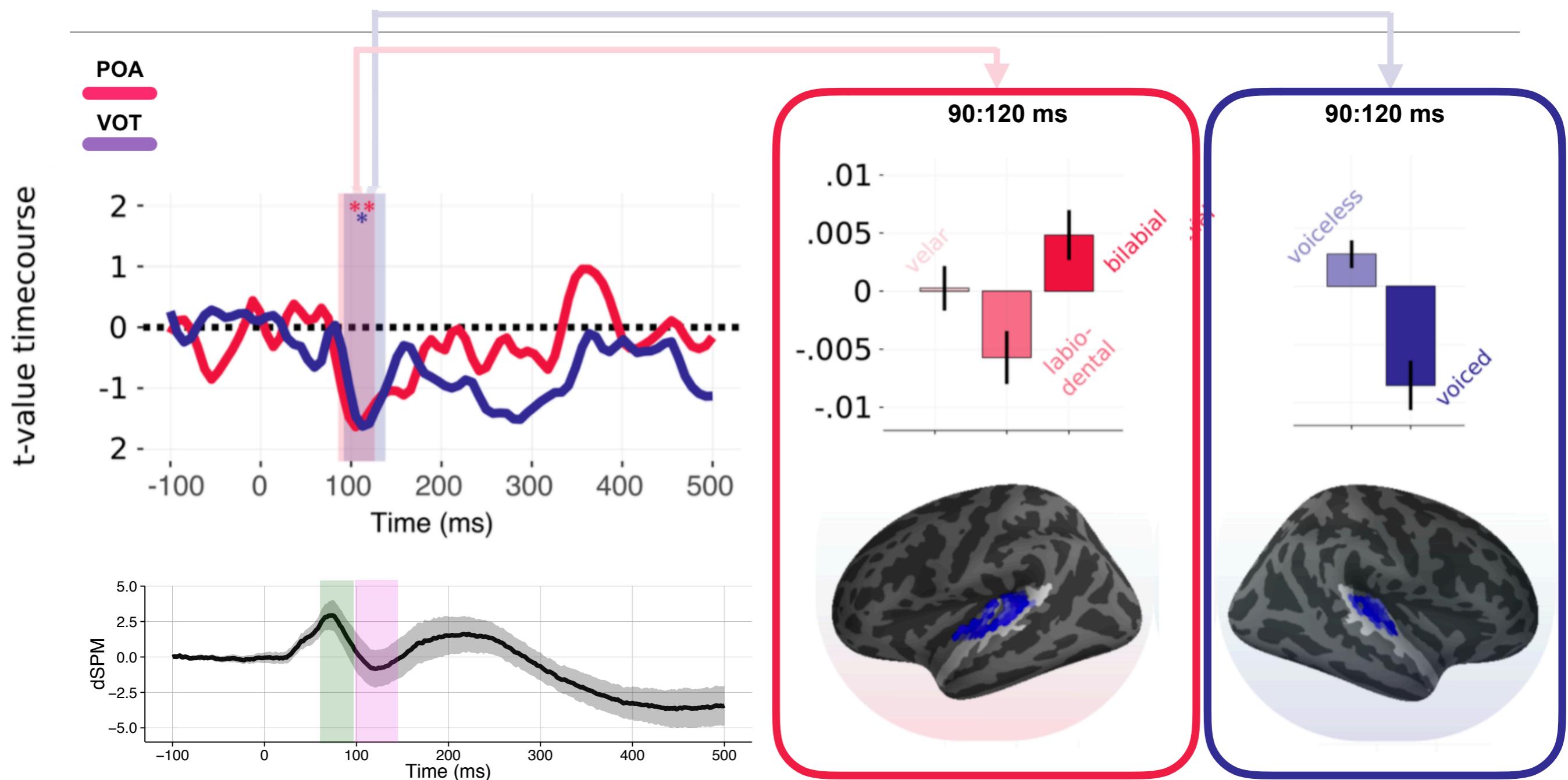
p p b b b

Subphonetics at Onset

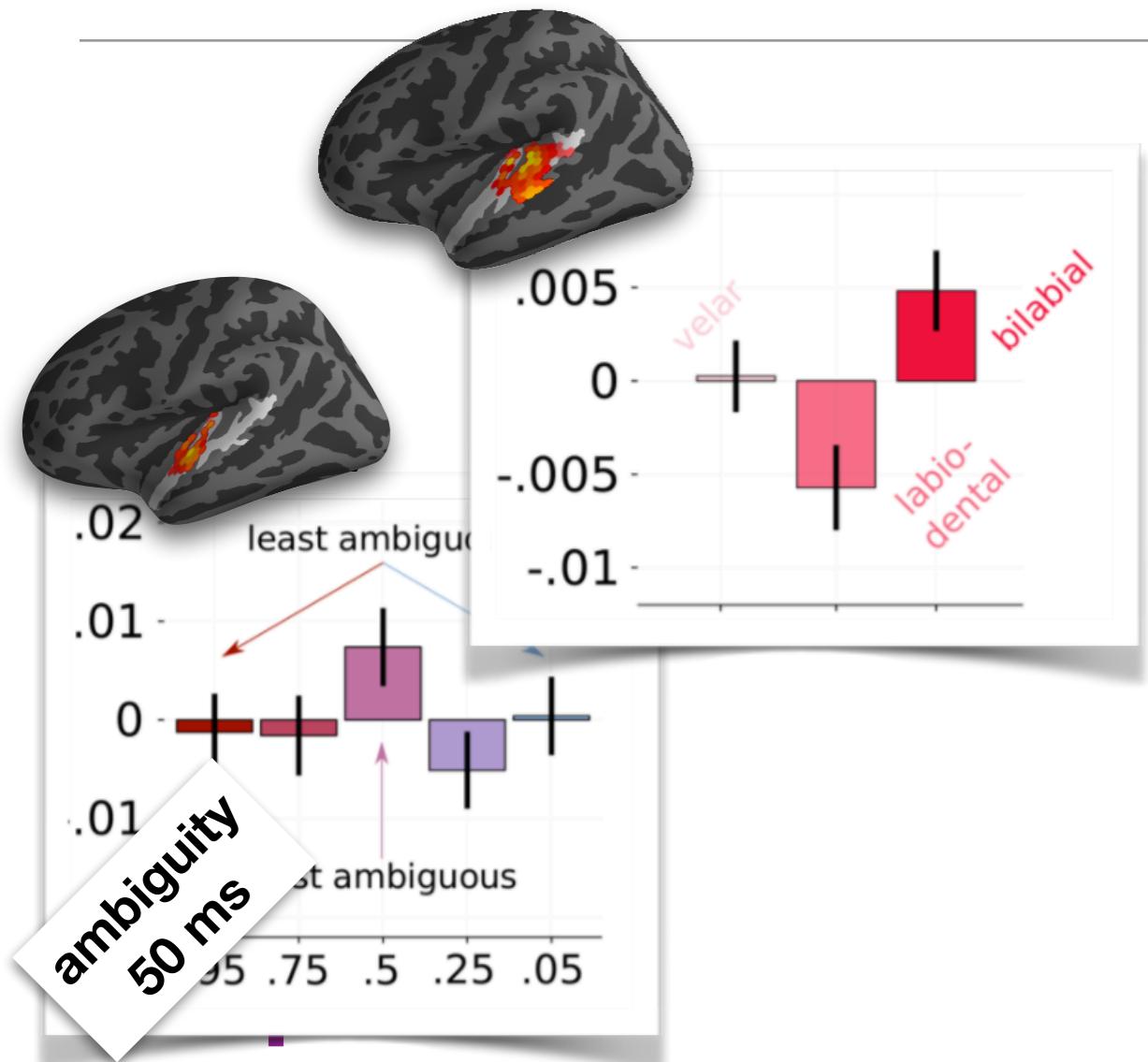


p b b b b

Phonetic Features at Onset

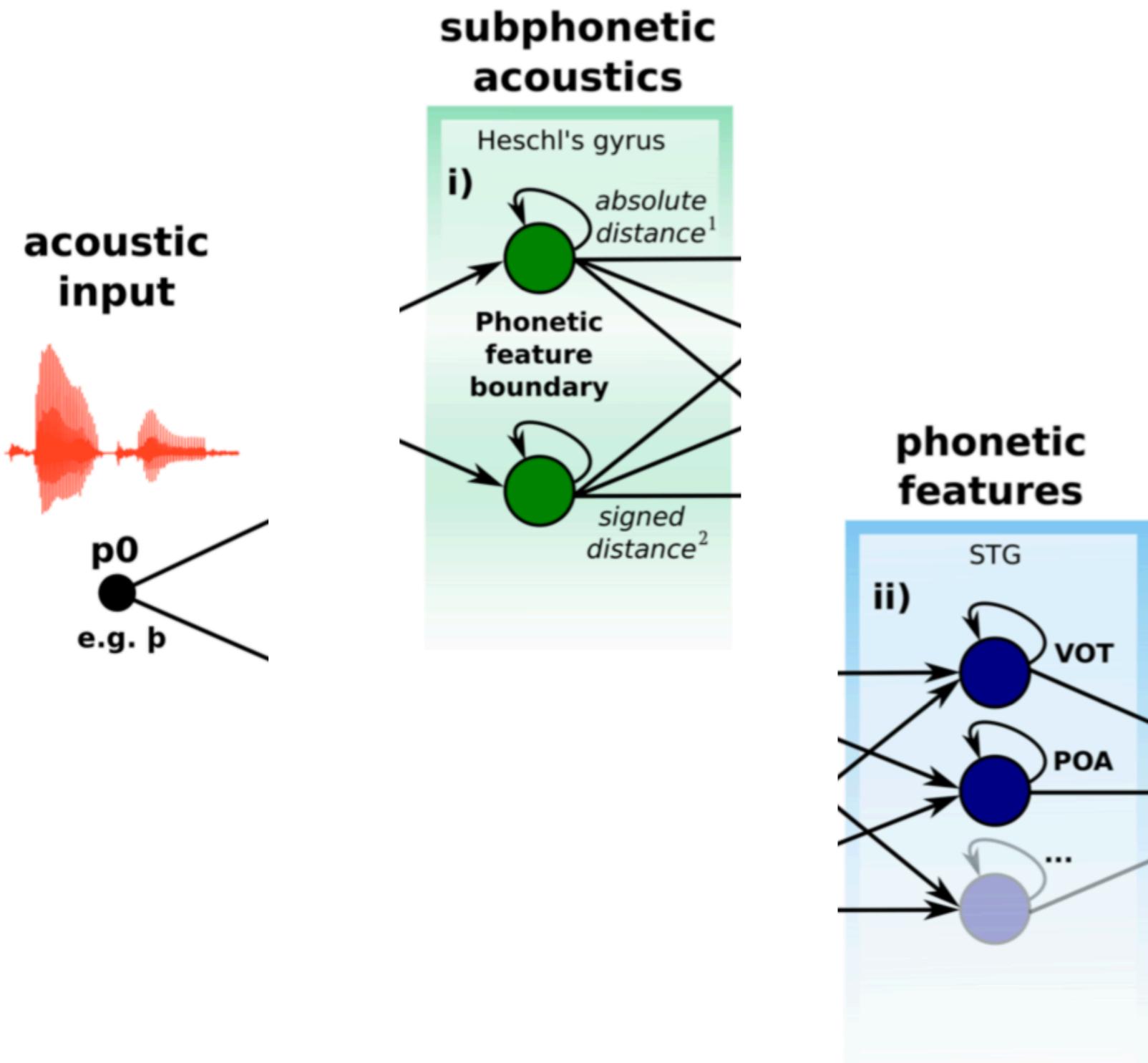


Interim Conclusion

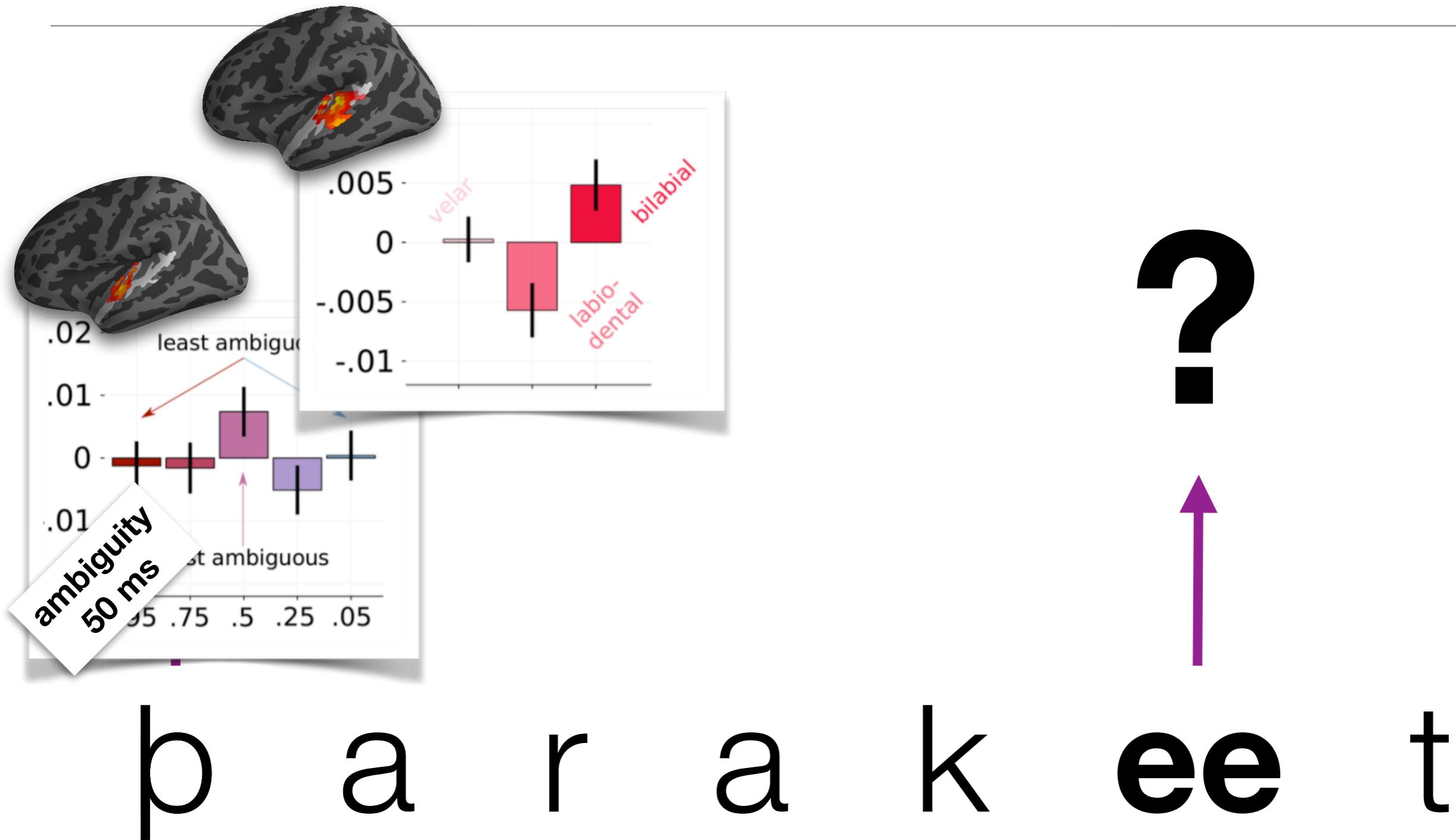


b a r a k ee t

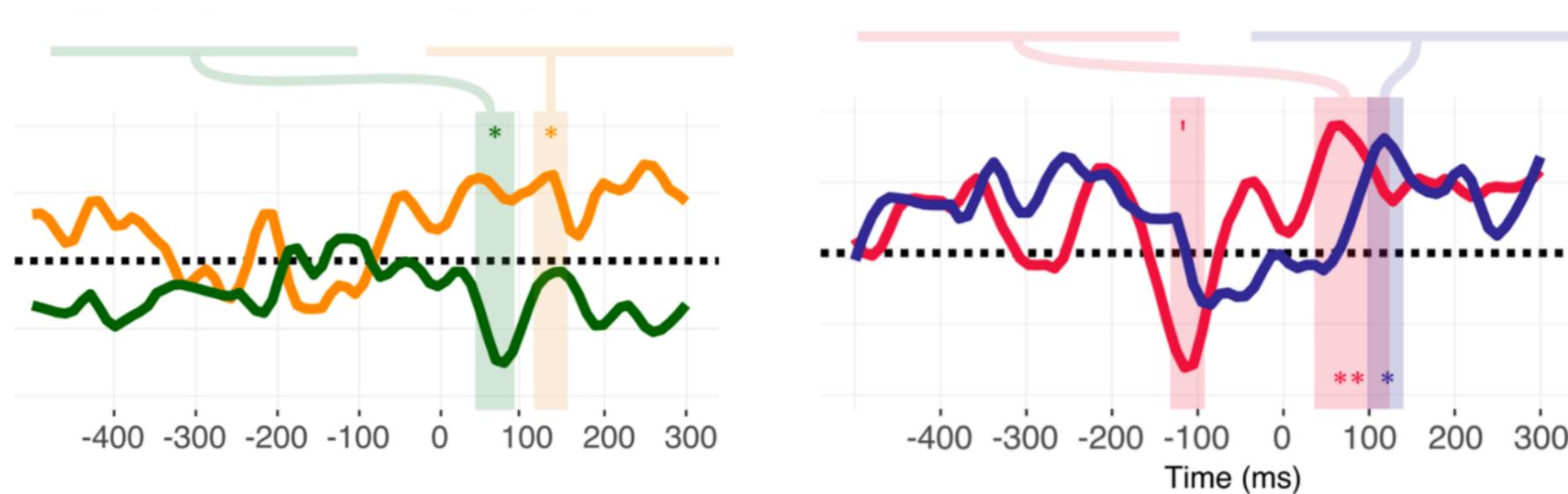
Putting together the processing pieces



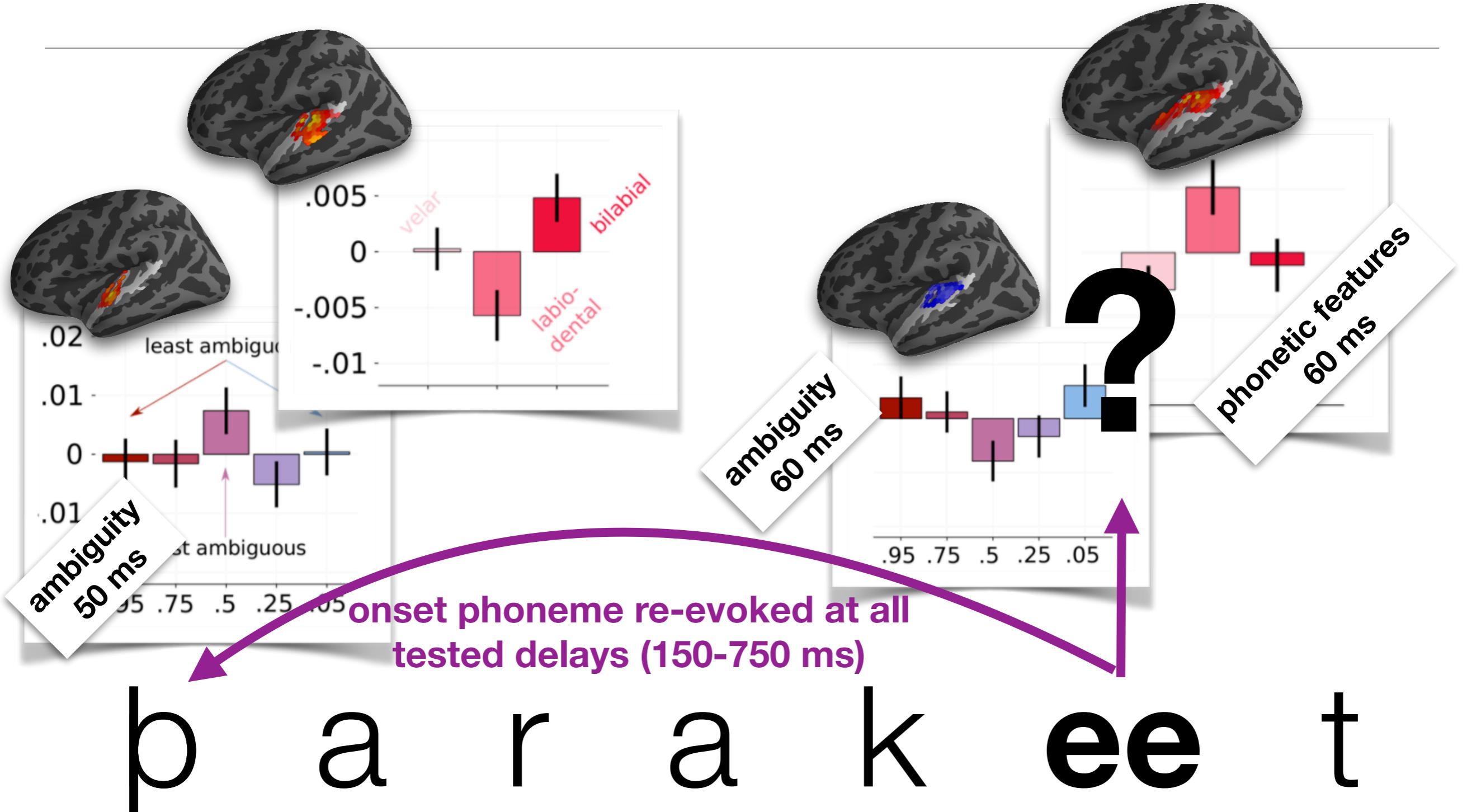
Interim Conclusion



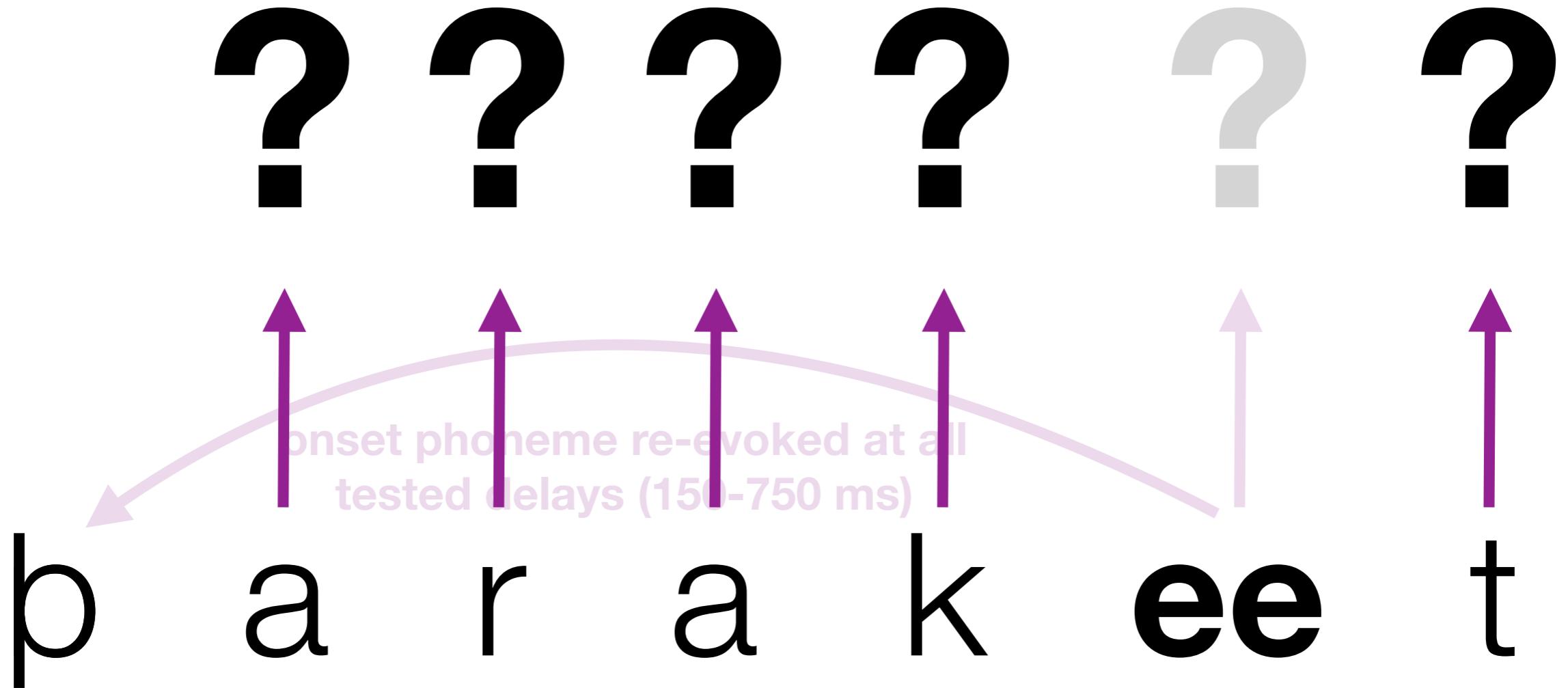
Ambiguity at POD



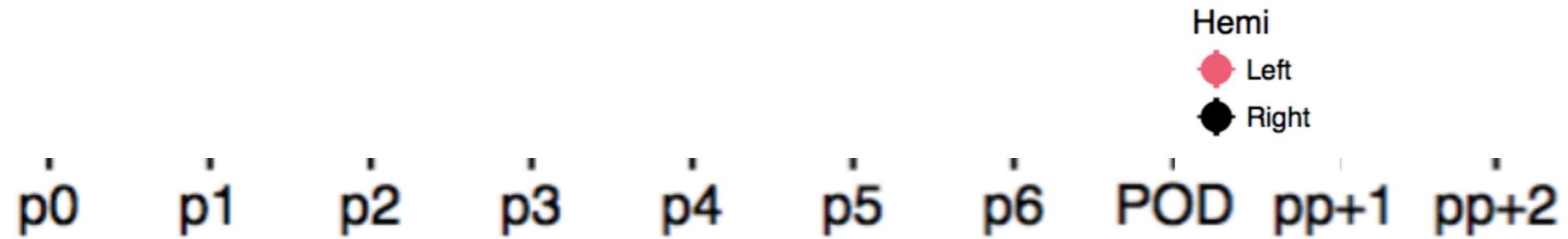
Interim Conclusion



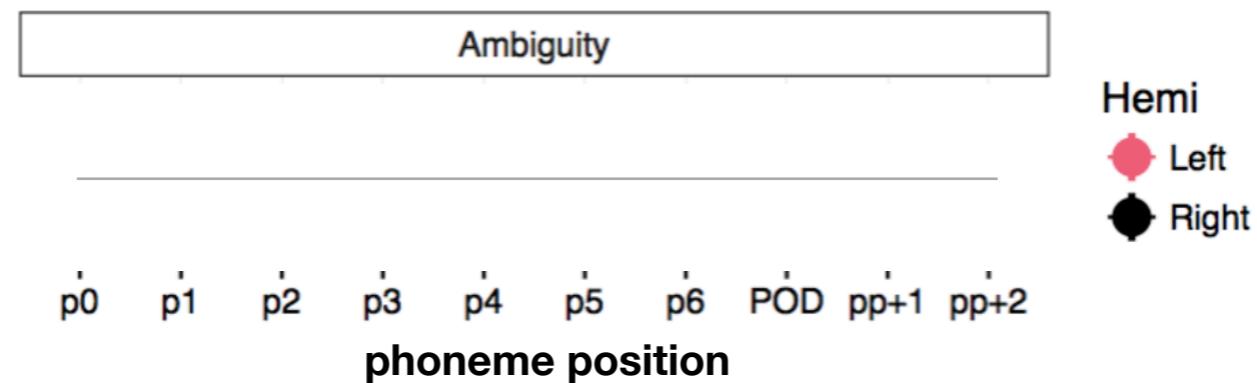
Interim Conclusion



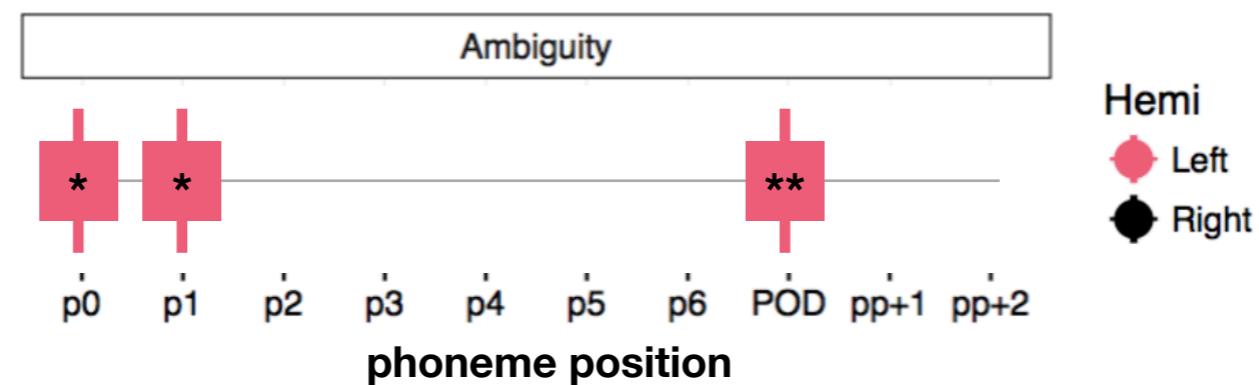
Reactivation in Intermediate Positions



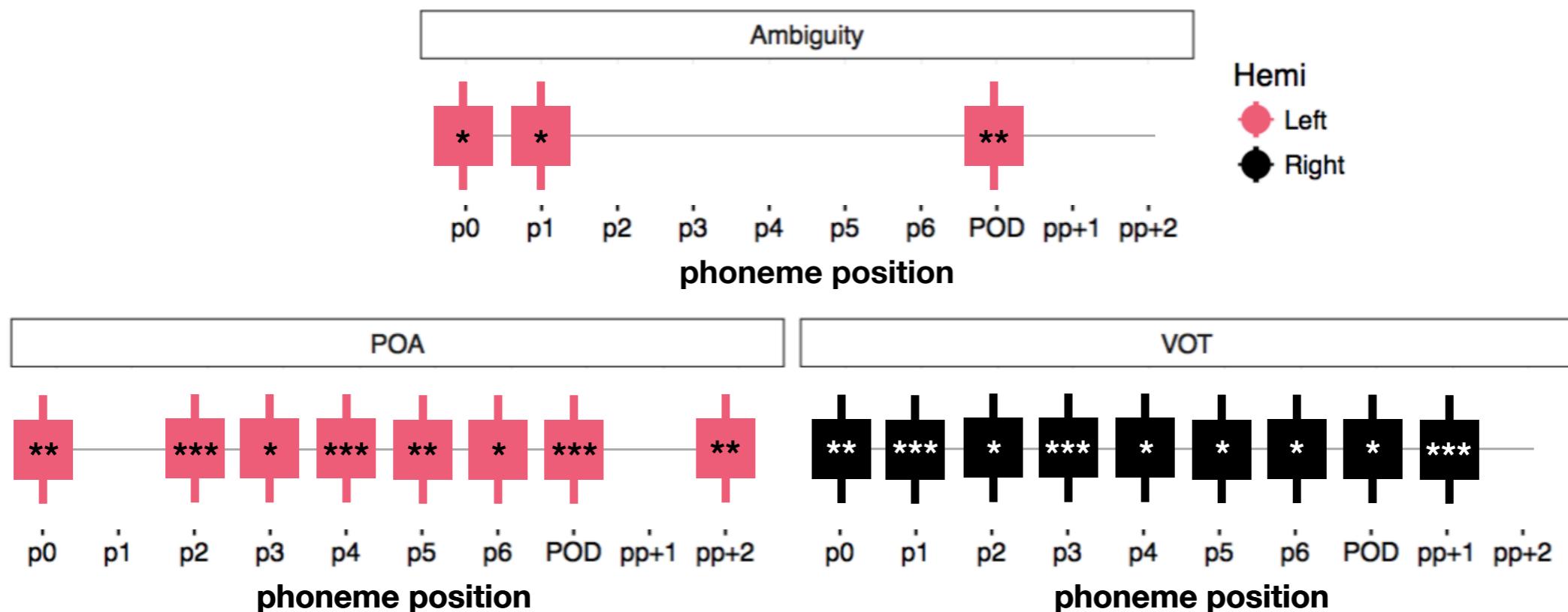
Reactivation in Intermediate Positions



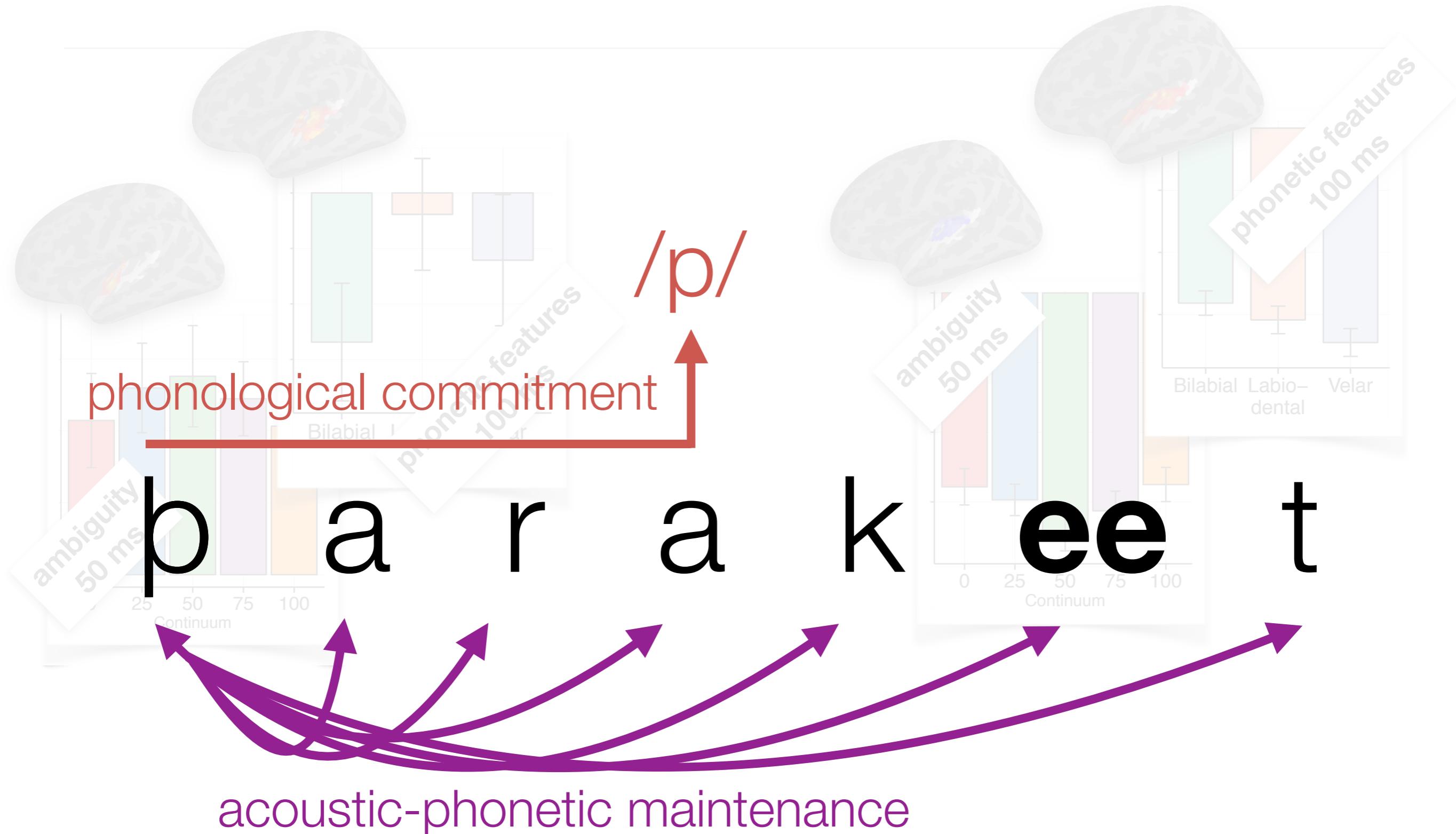
Reactivation in Intermediate Positions



Reactivation in Intermediate Positions



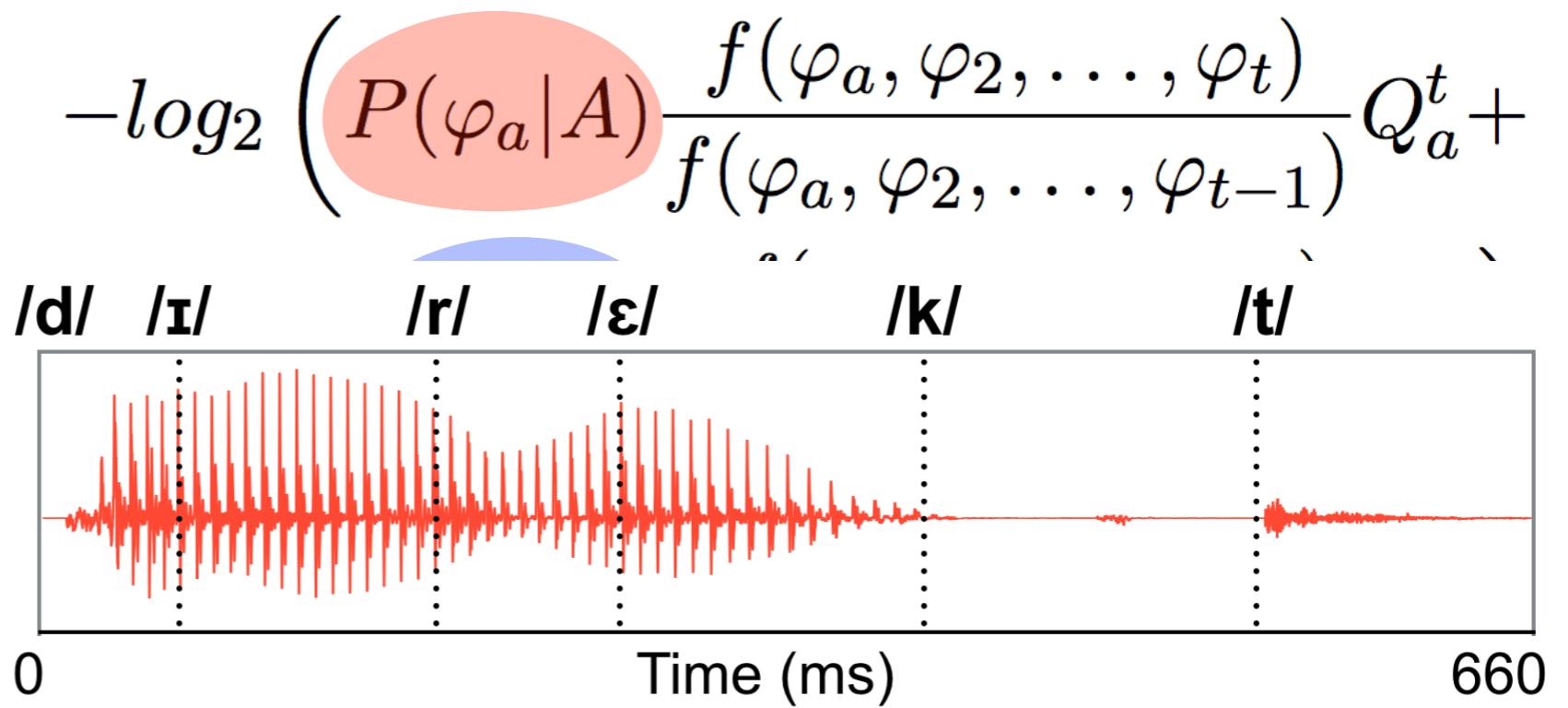
Interim Conclusion



Testing for phonological commitment

- **Surprisal:**

No commitment
Commitment



- **Entropy:**

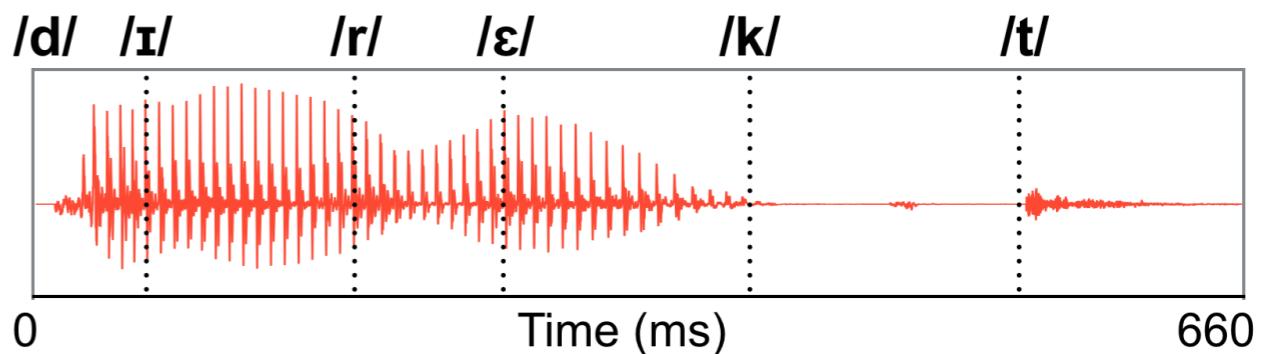
No commitment
Commitment

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

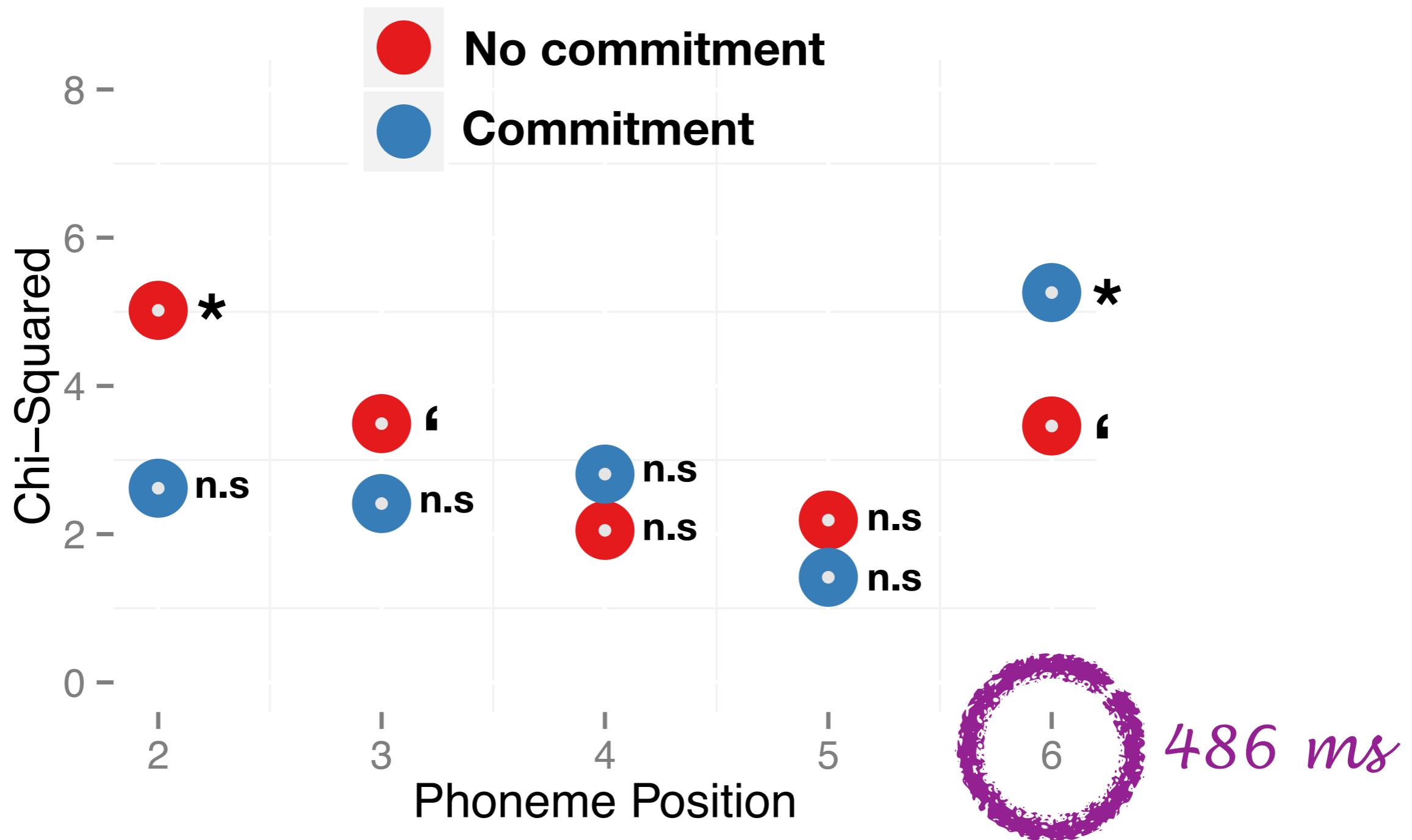
Model Setup

- **Critical variables:**
no commitment entropy
no commitment surprisal
commitment entropy
commitment surprisal

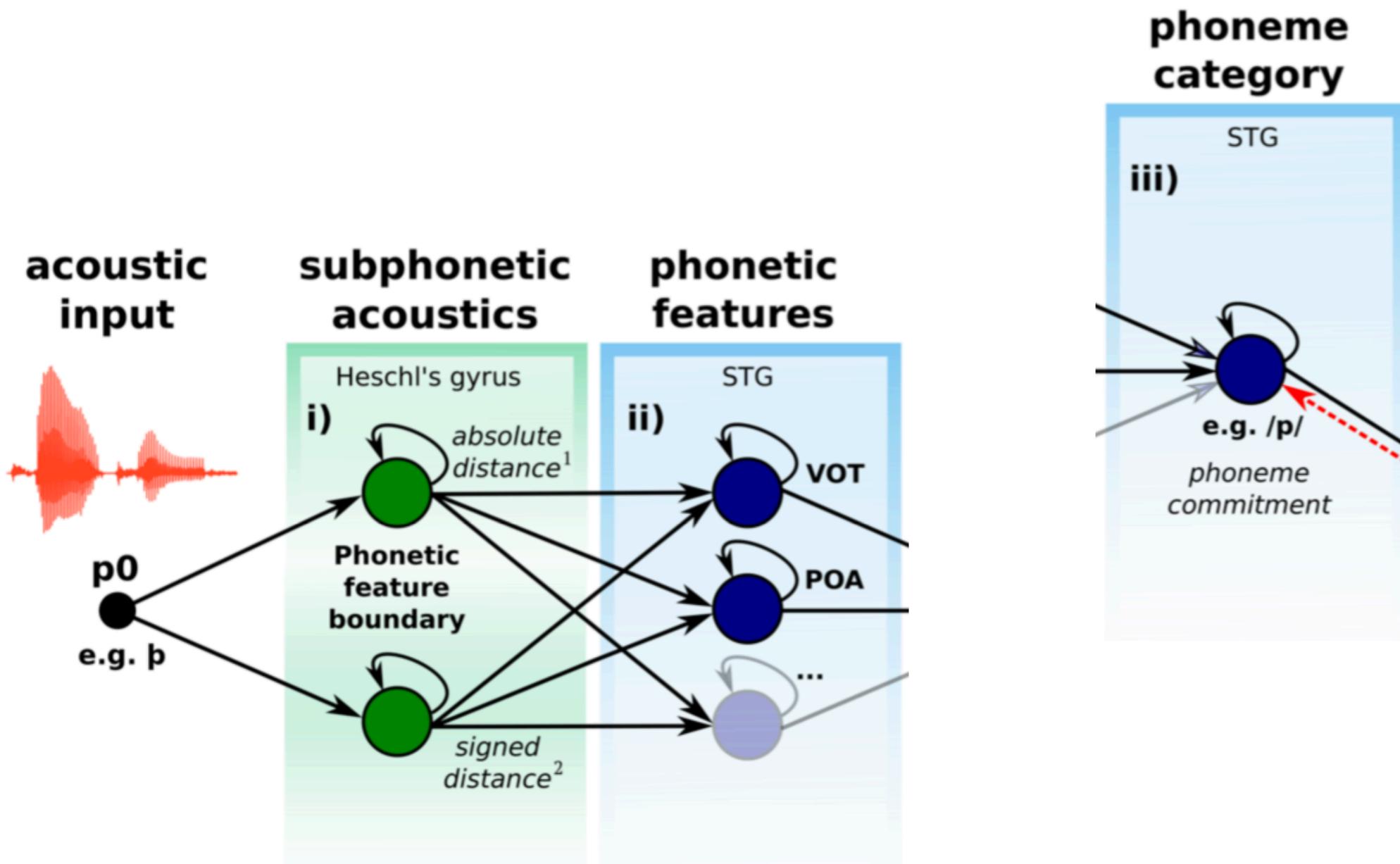
- **Control variables:**
phoneme latency (ms)
phoneme latency (number of phonemes)
trial number
block number
stimulus amplitude
phoneme pair
ambiguity



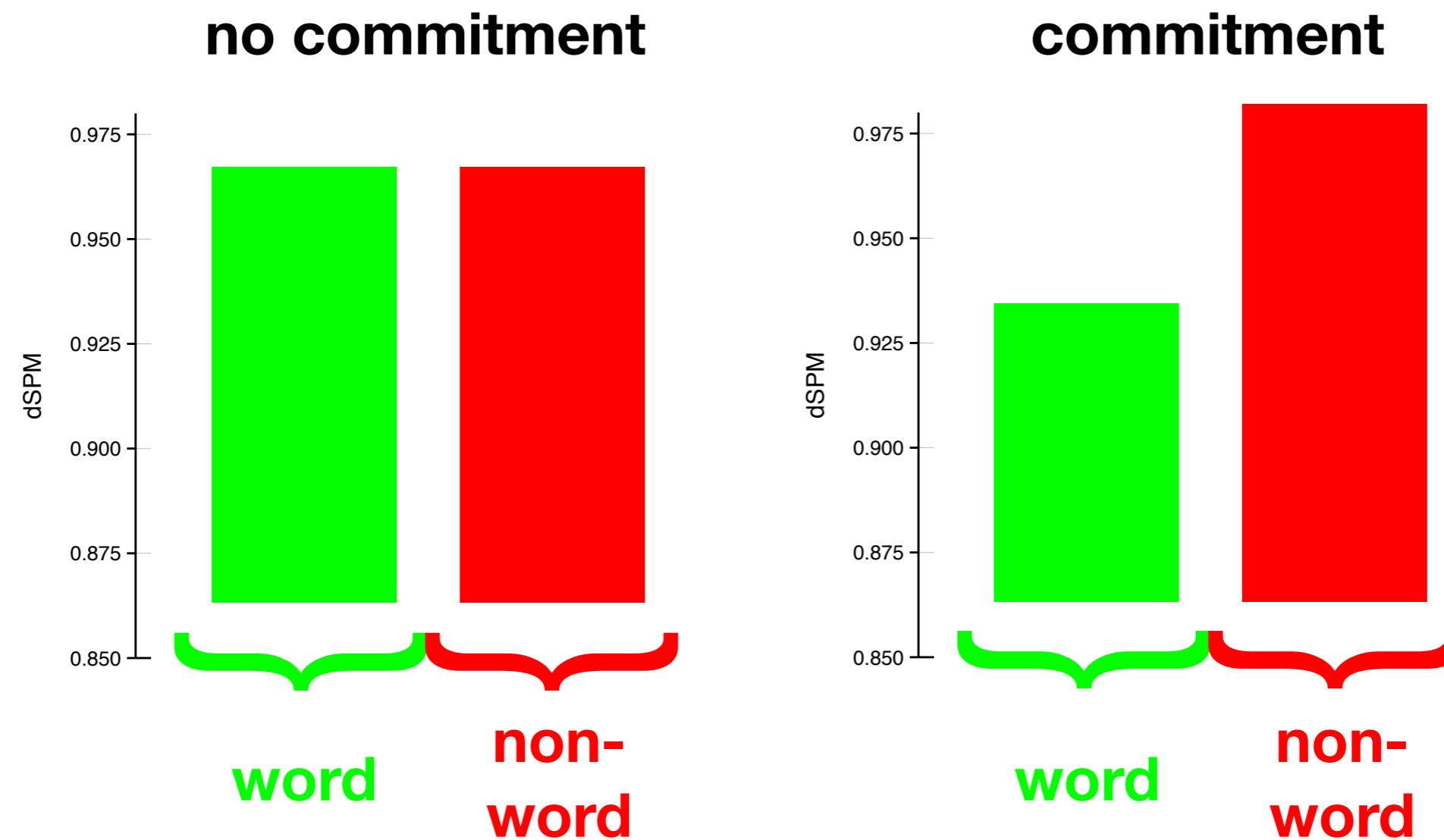
Results



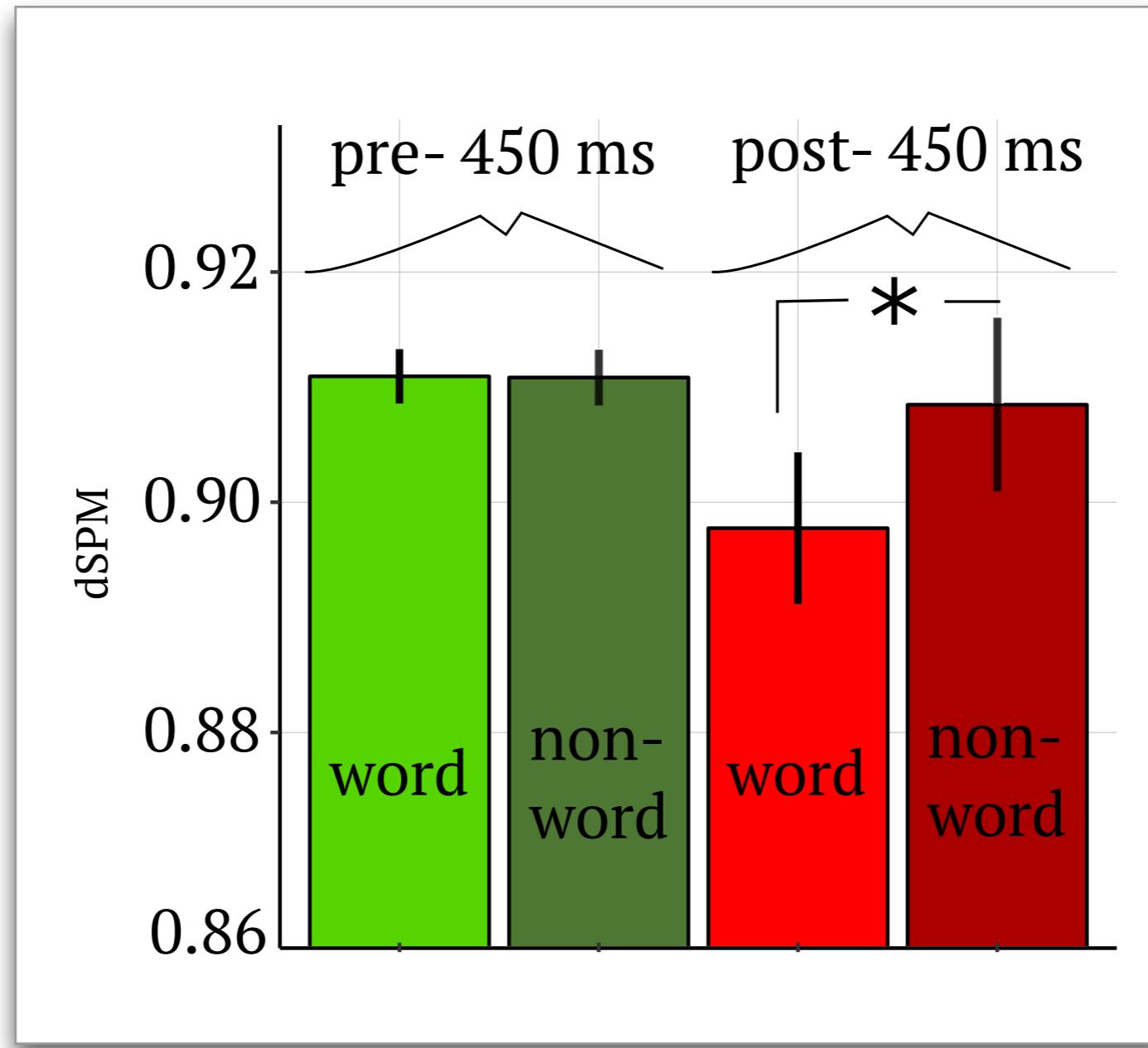
Putting together the processing pieces



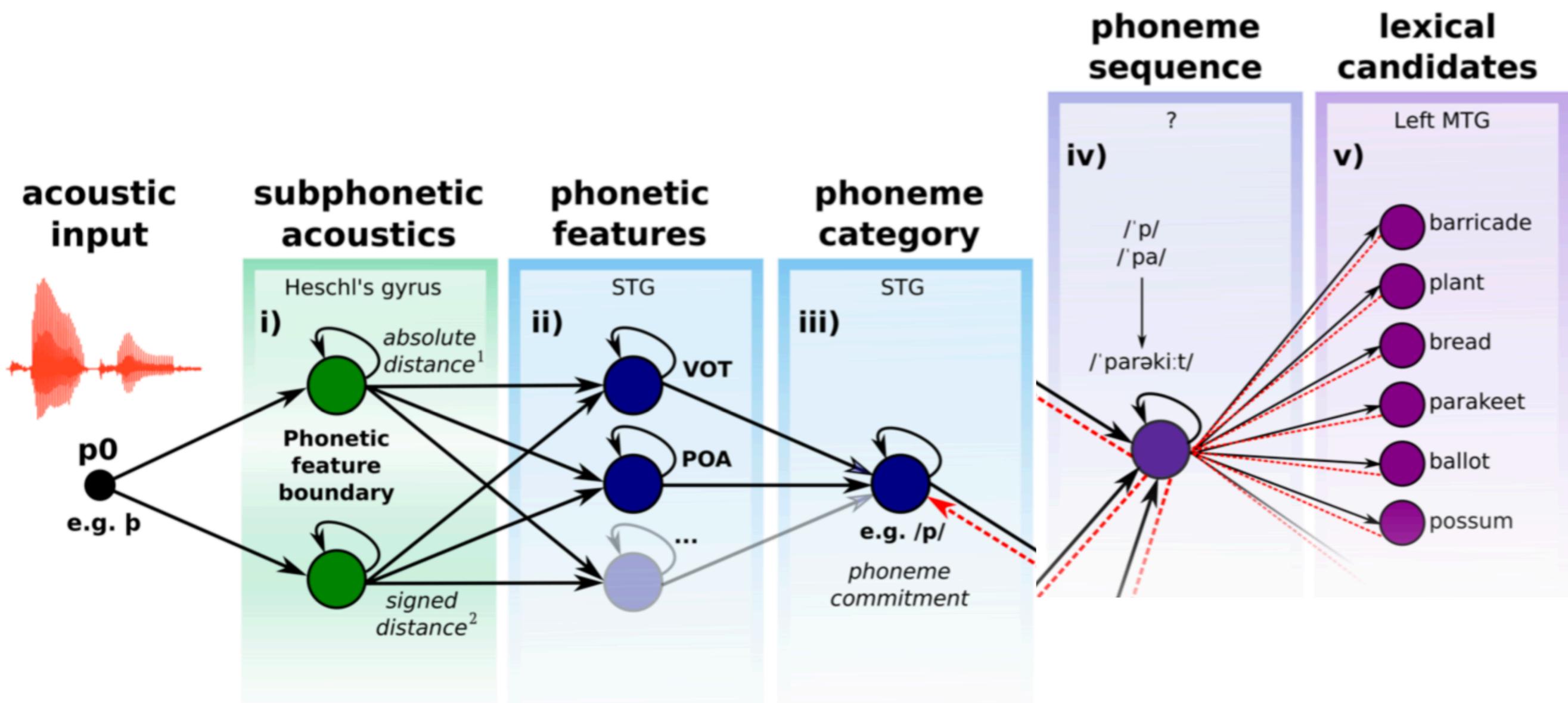
Further test of commitment



Further test of commitment

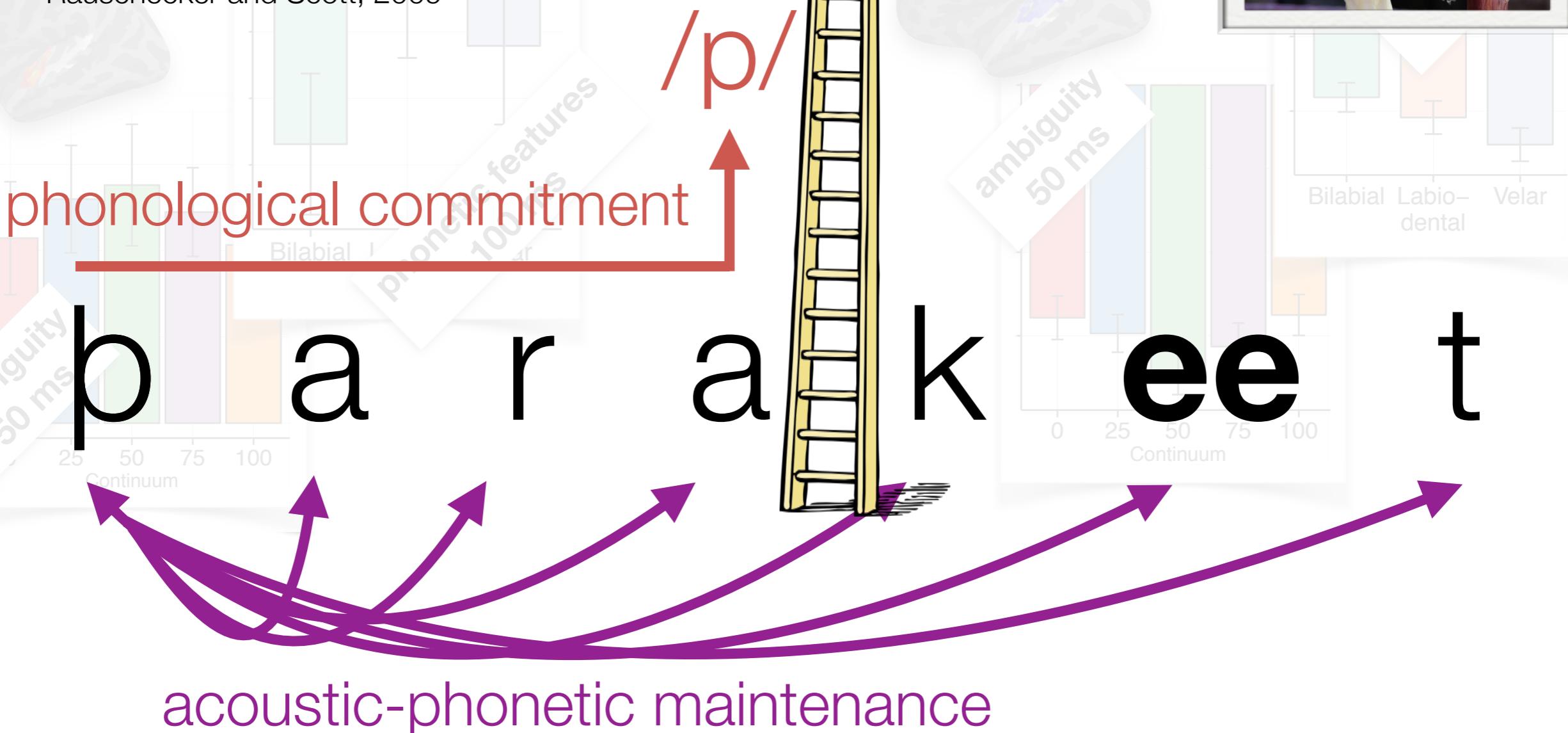


Putting together the processing pieces



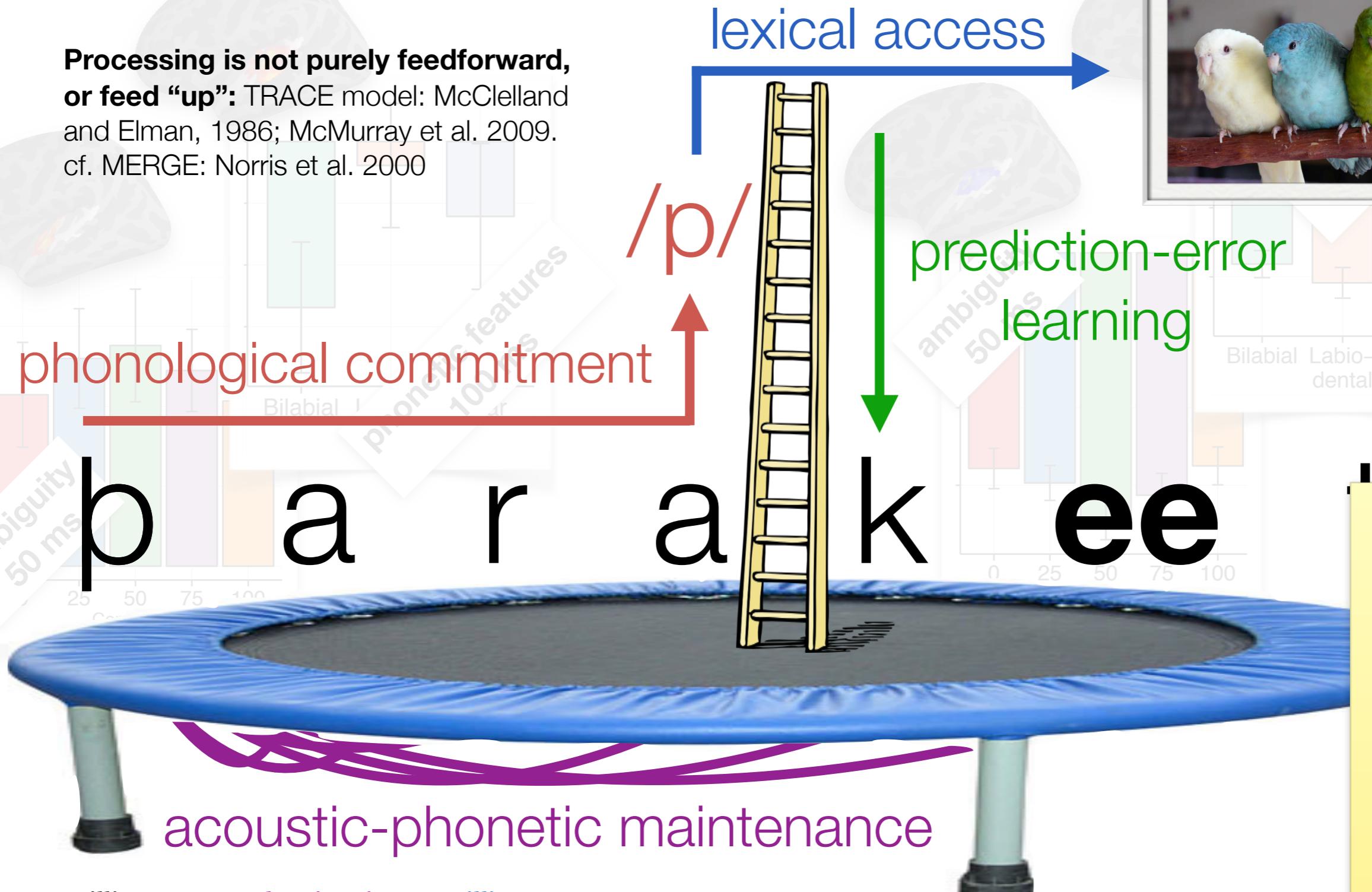
Interpretation

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



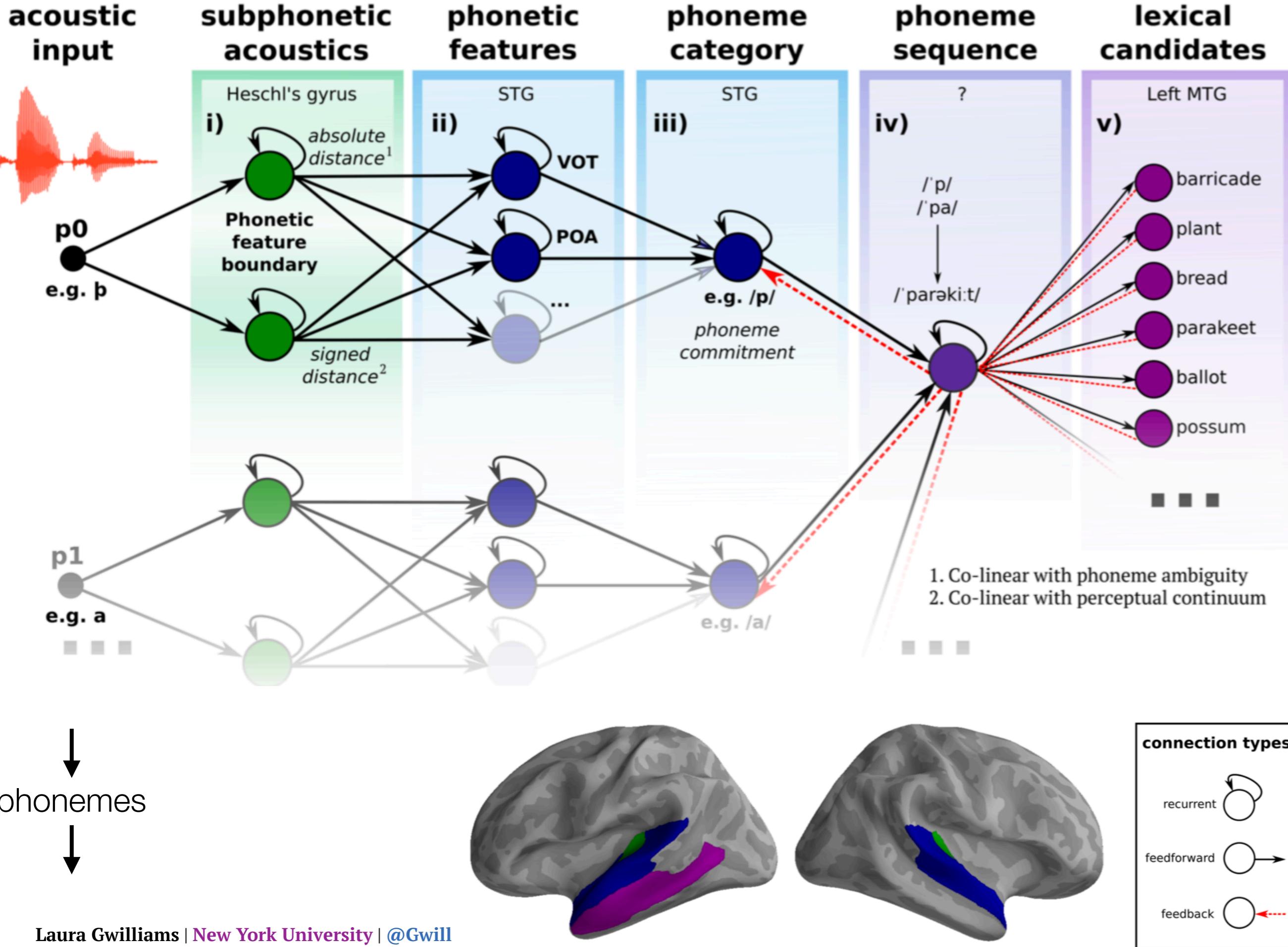
Interpretation

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 commitment to category does not cost — the system can flexibly avoid committing to category; avoiding the function of exposure

but what happens if the system continuously jumps on the trampoline because of a wrong choice? well, when talking to someone with an accent.



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**)
- What is the computational architecture?

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?

Famas & Corbit (1972)

Taft & Forster (1976), Taft (1979)

Marslen-Wilson & Welsh (1978)

Cutler et al. (1986), Barry (1980-1984)

Pinker & Prince (sentence
structure)

NP

VP

phrasal
structure

the fat cat dis | appear | ed

lemmas

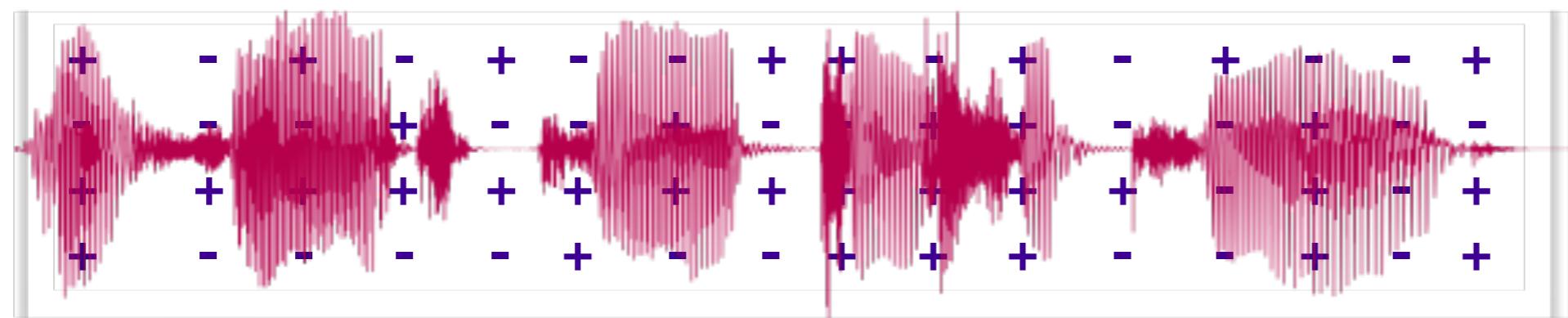
morphemes

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

1)

which linguistic units are encoded in brain activity?

2) what is the relative time-course?

th



VP

pear | ed

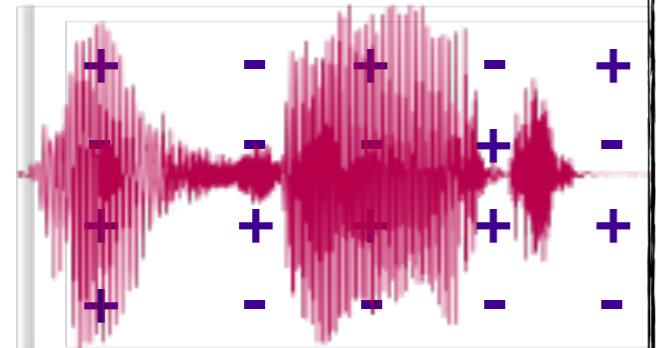
phrasal structure

lemmas

morphemes

dah fa

DH AH F AE T



3) what is the computational architecture?

population 1

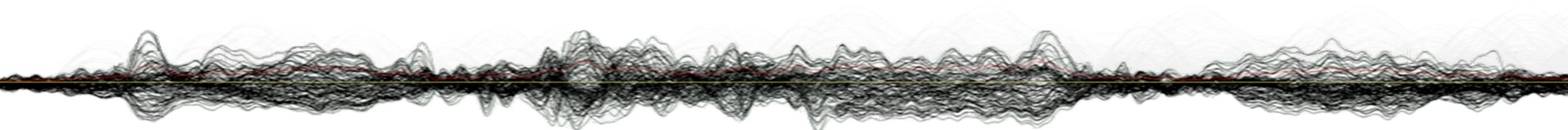
population 2

population 3

population 4

population 1

Setup

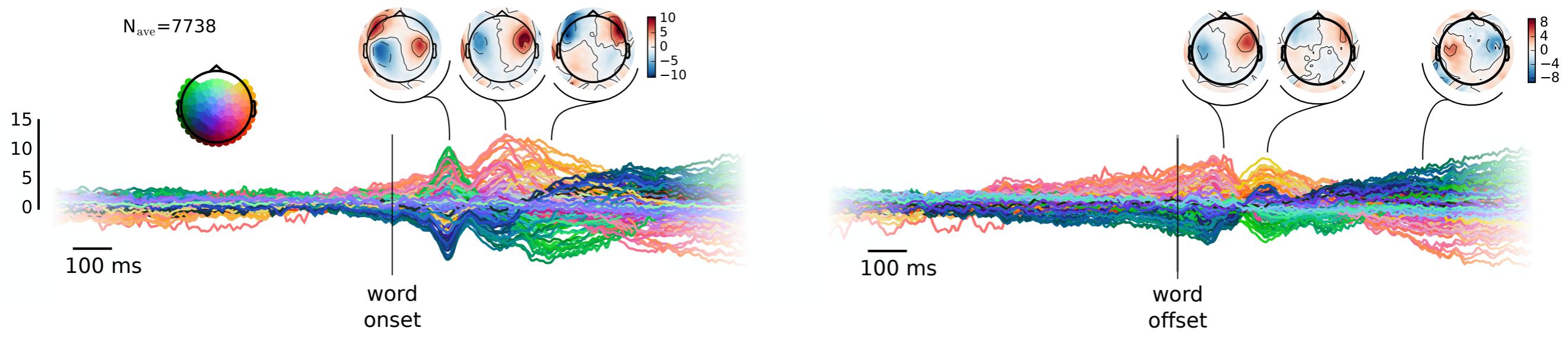


- 18 participants
 - Listening to four narrative stories (twice)
 - 2 x one hour recordings
 - KIT 208 channel MEG system
 - Engagement task
-
- ~40,000 phonemes per participant
 - ~8,000 words per participant



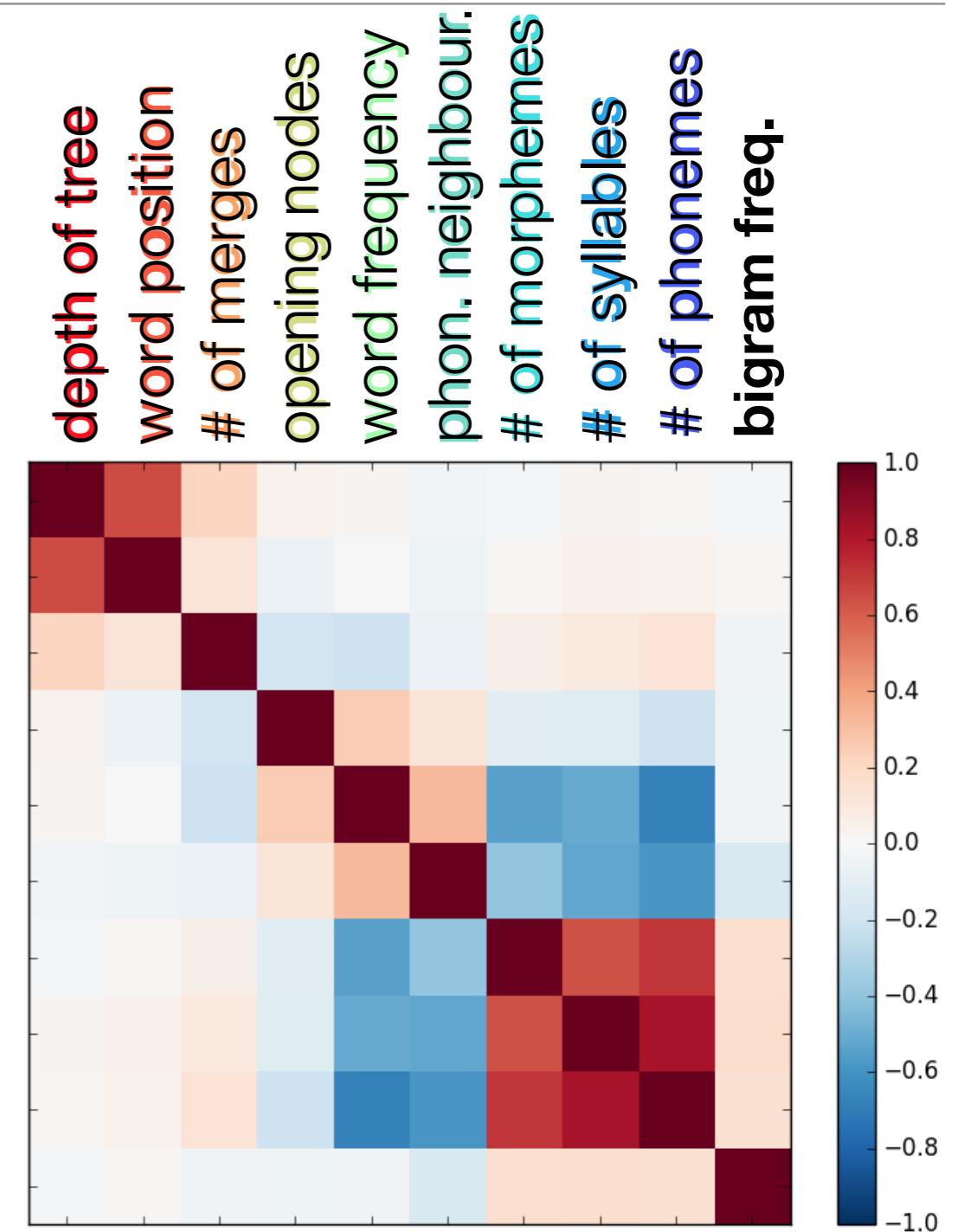
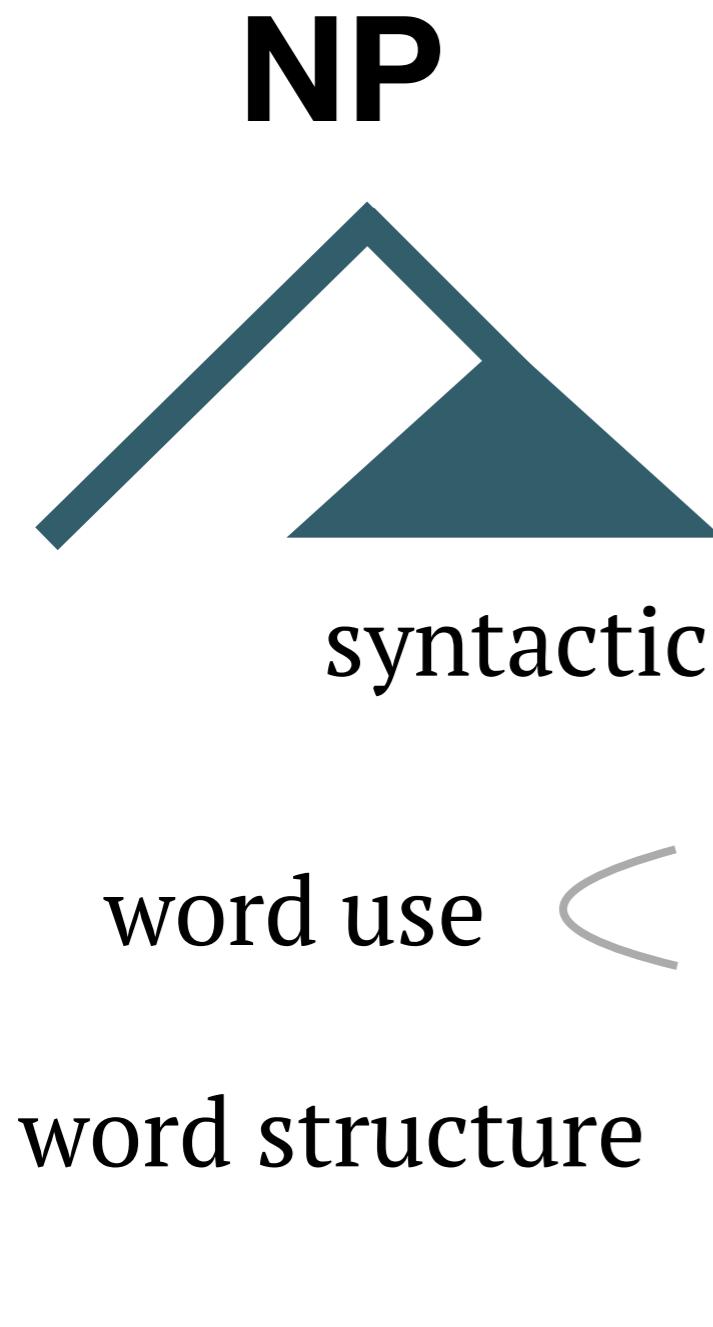
جامعة نيويورك أبوظبي
NYU | ABU DHABI

Event-locked average response

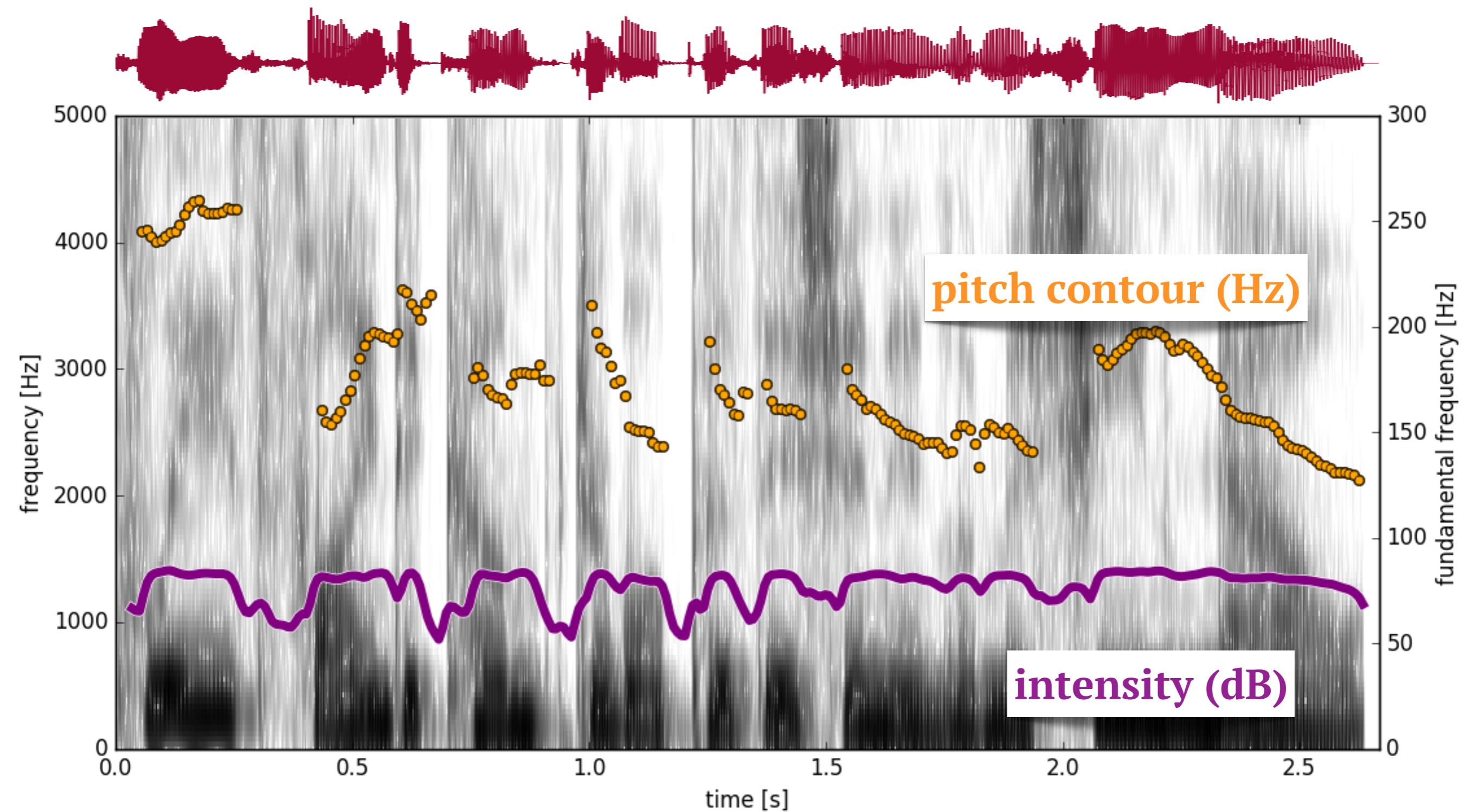


explain this plot/ colour of channels in the topography

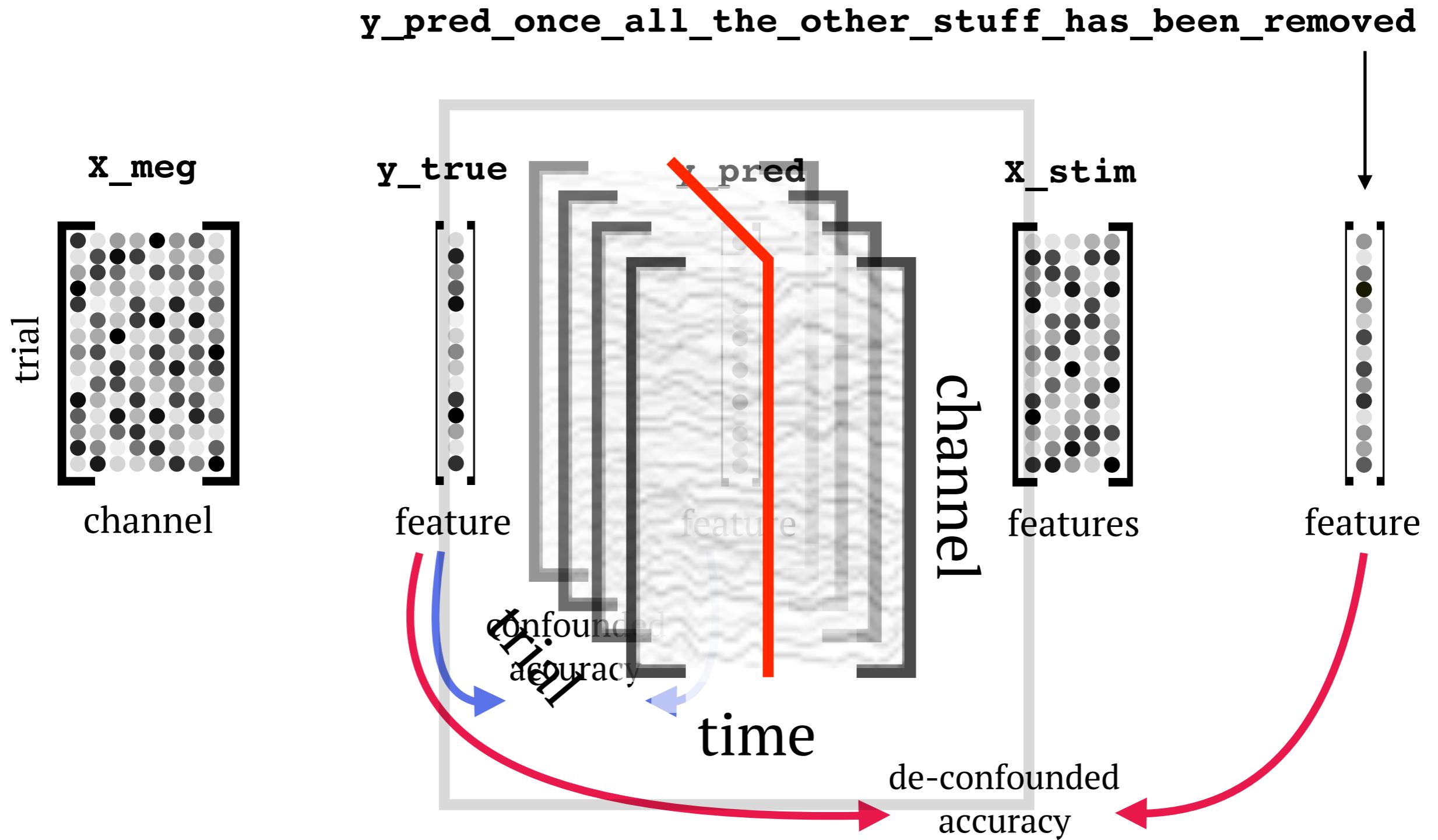
Stimulus features



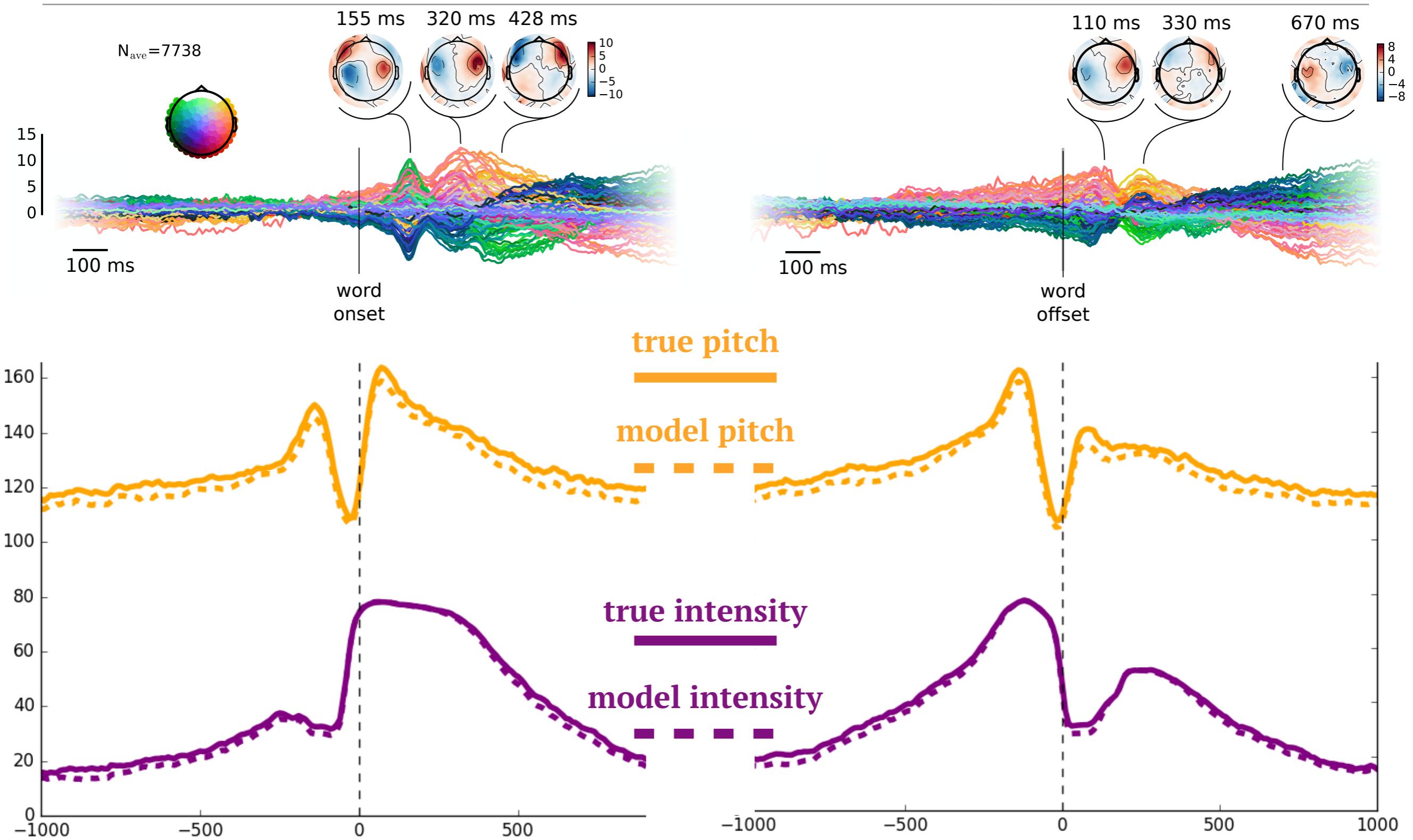
Pitch and intensity in the acoustic signal



Analysis technique

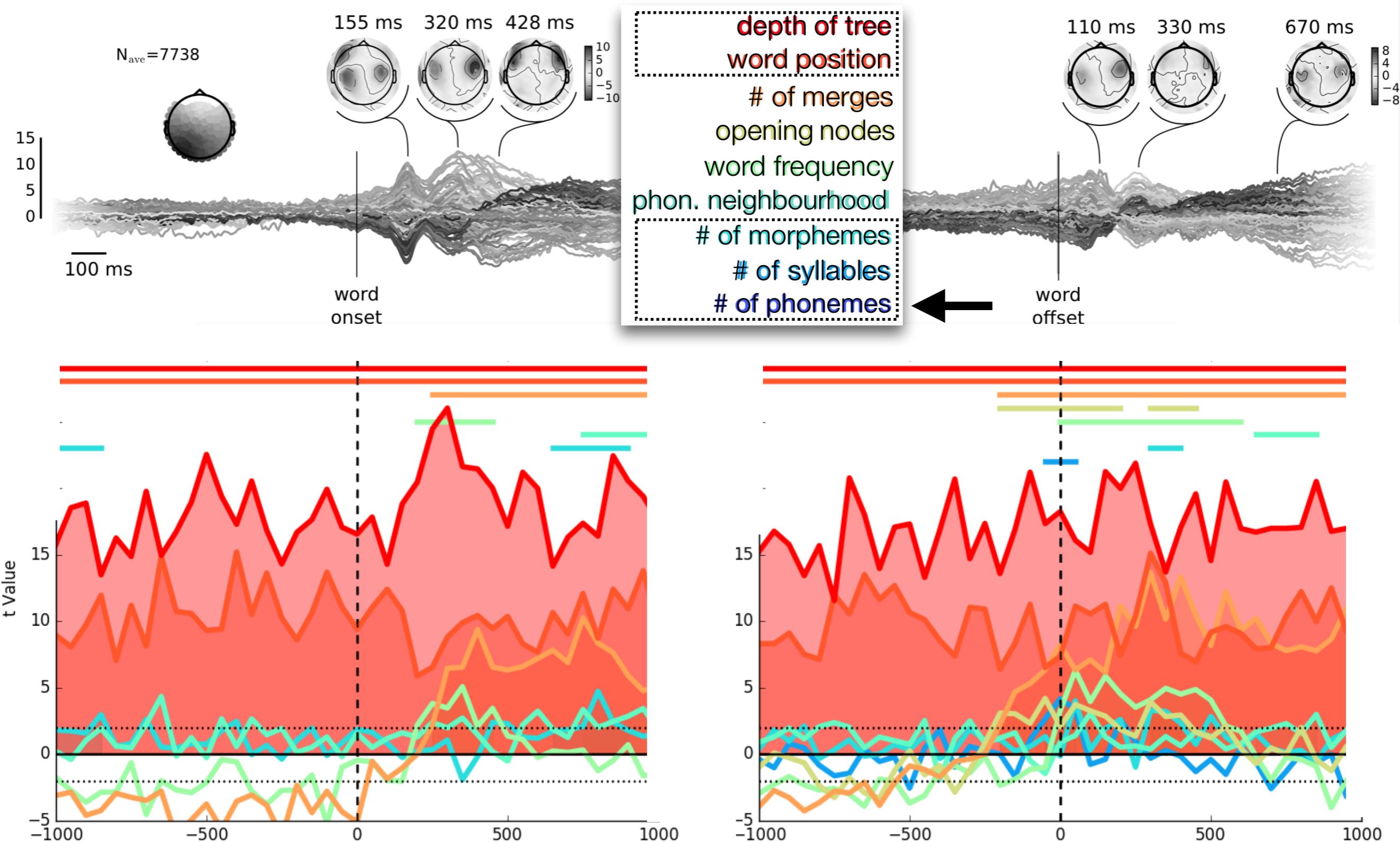


Pitch and intensity



which linguistic units are
encoded in brain activity?

Deconfounded decoding accuracy



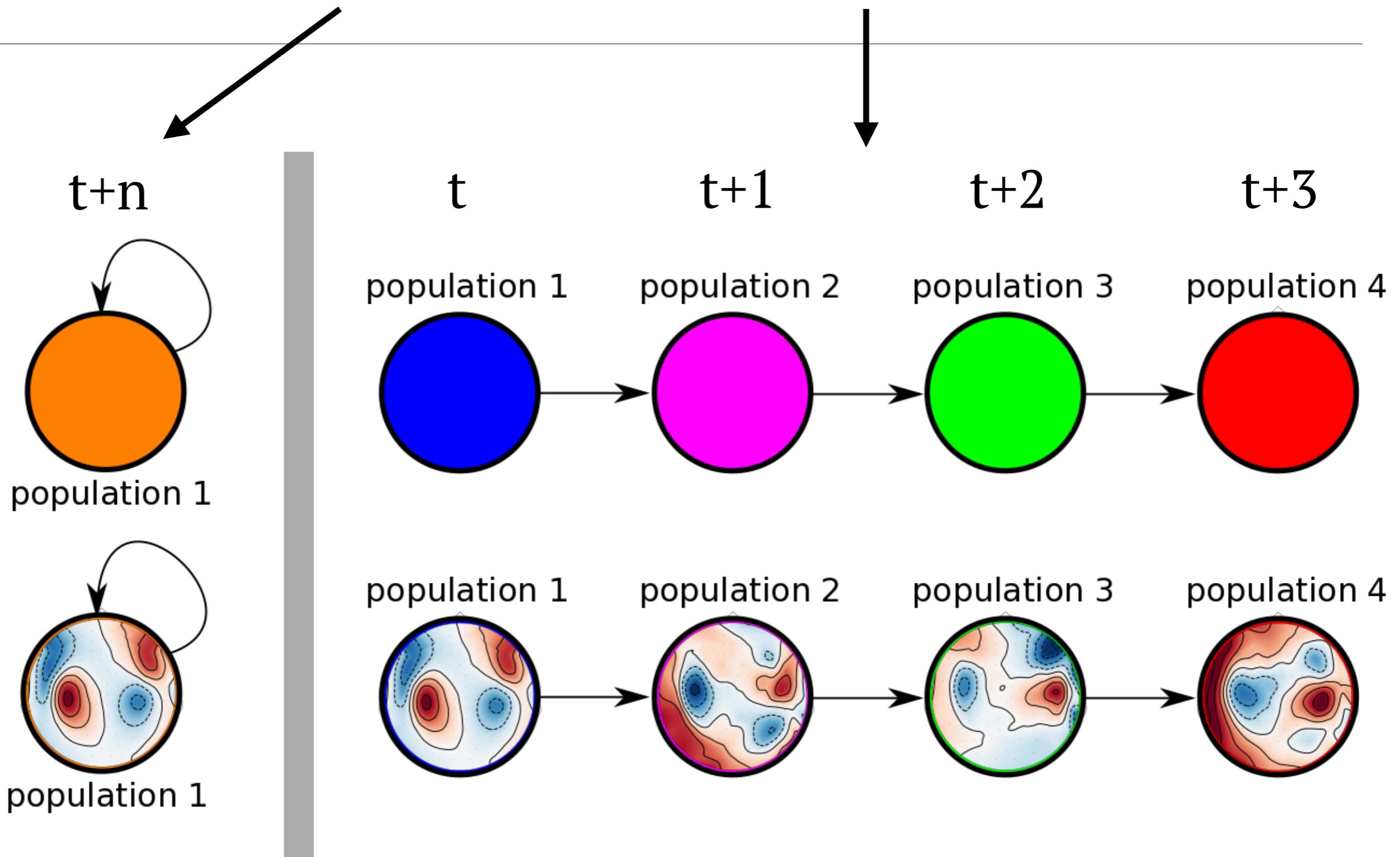
**what are the relative
time-courses?**

Deconfounded decoding accuracy

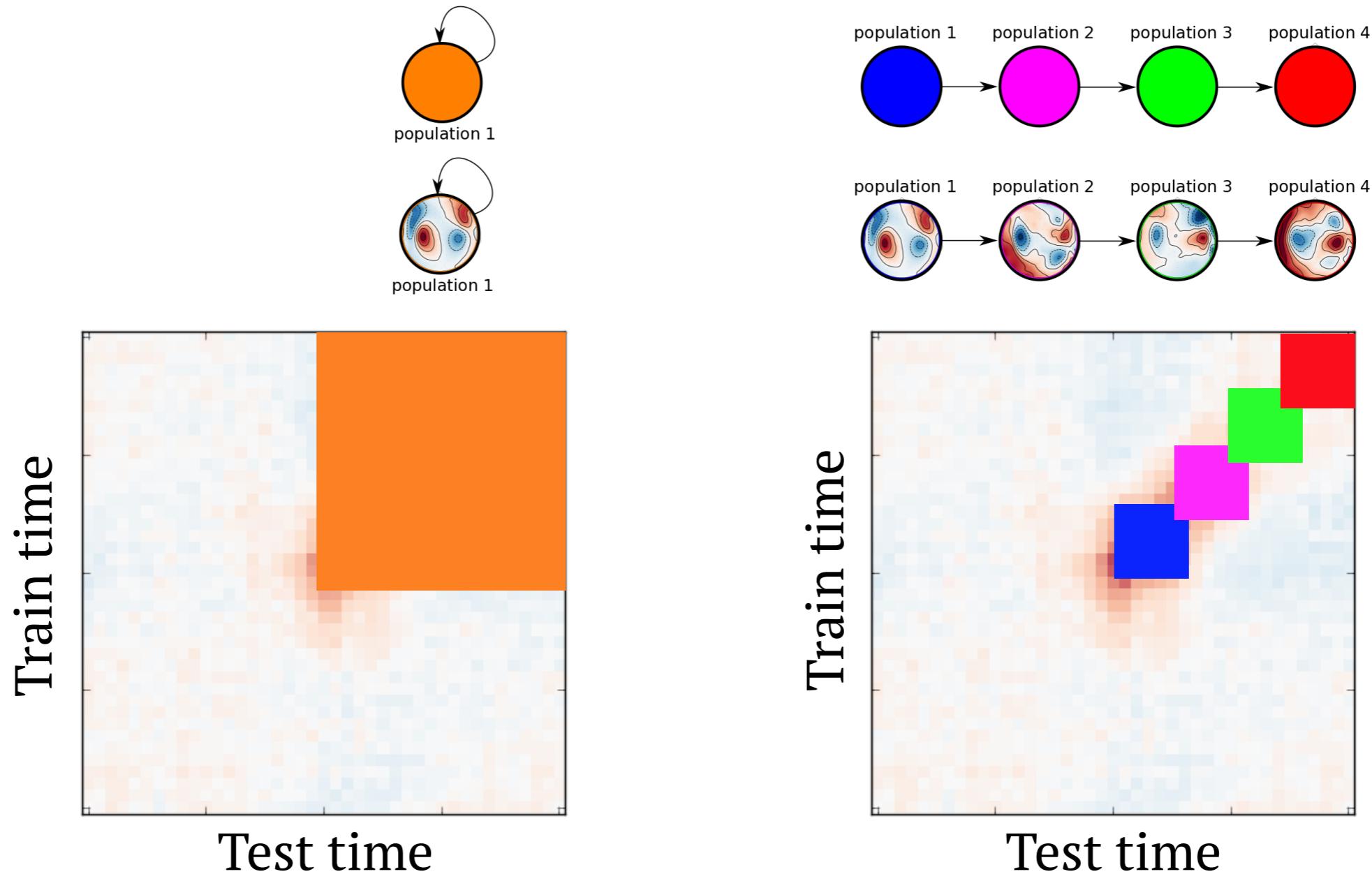


what is the computational architecture?

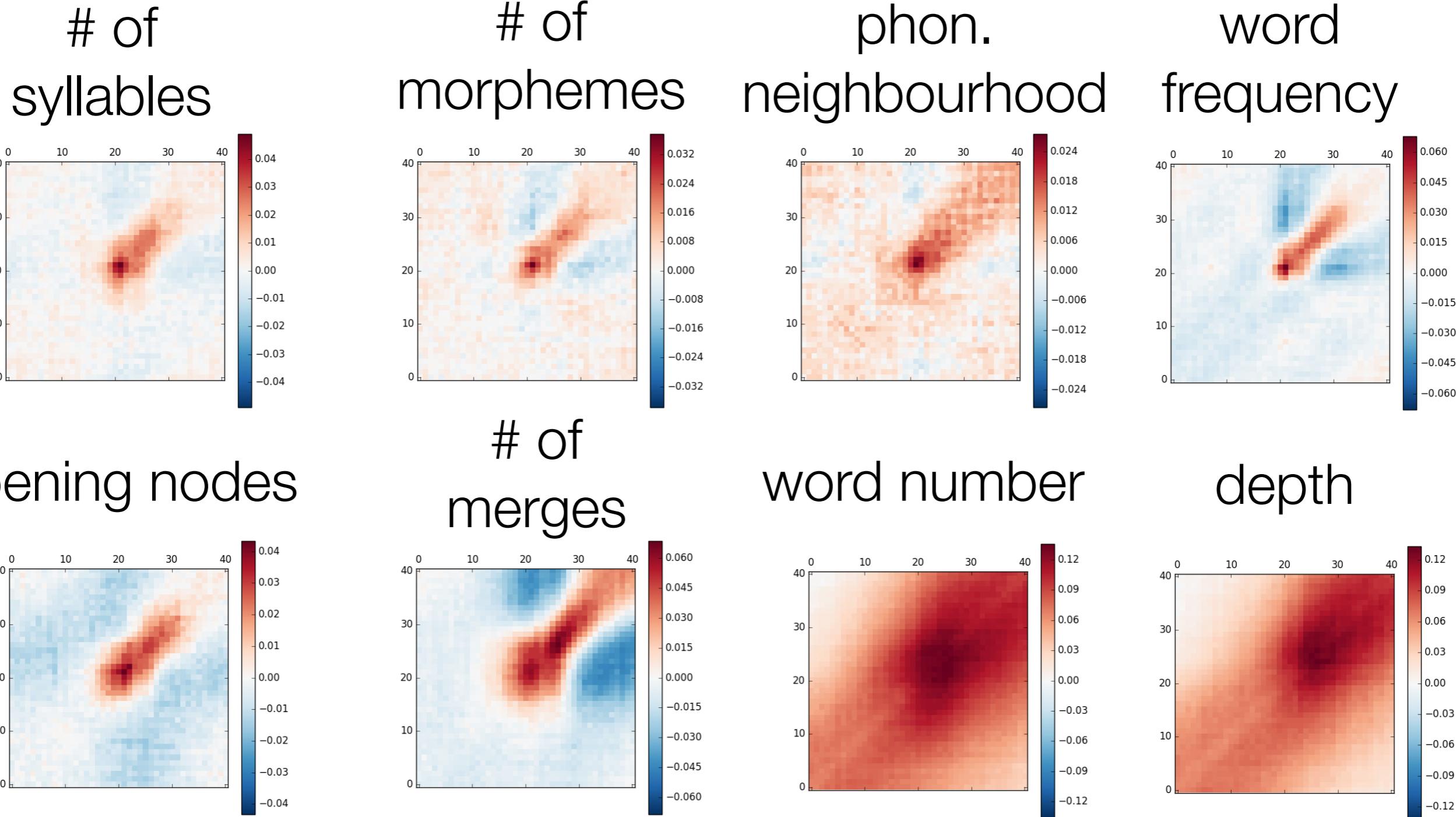
Recurrent vs. Feedforward



Recurrent vs. Feedforward



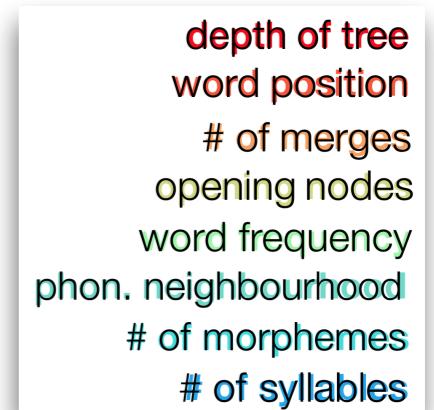
Recurrent vs. Feedforward



Discussion

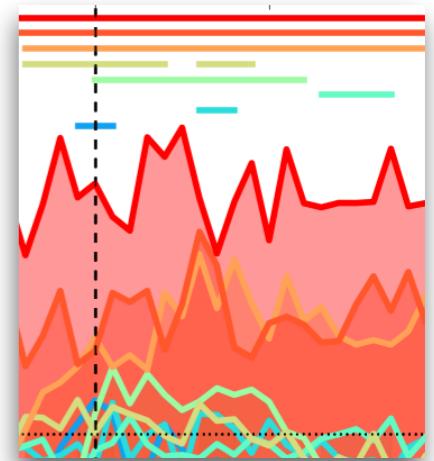
(1) Which linguistic units are encoded?

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



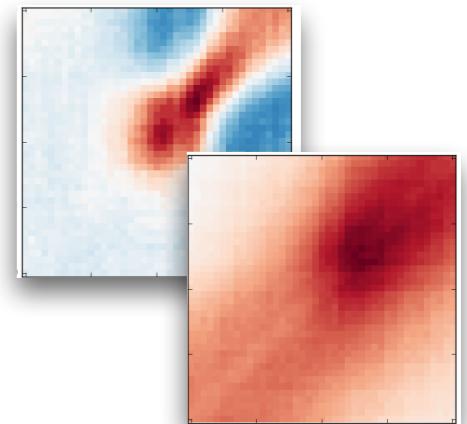
(2) What is the relative time-course?

- 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

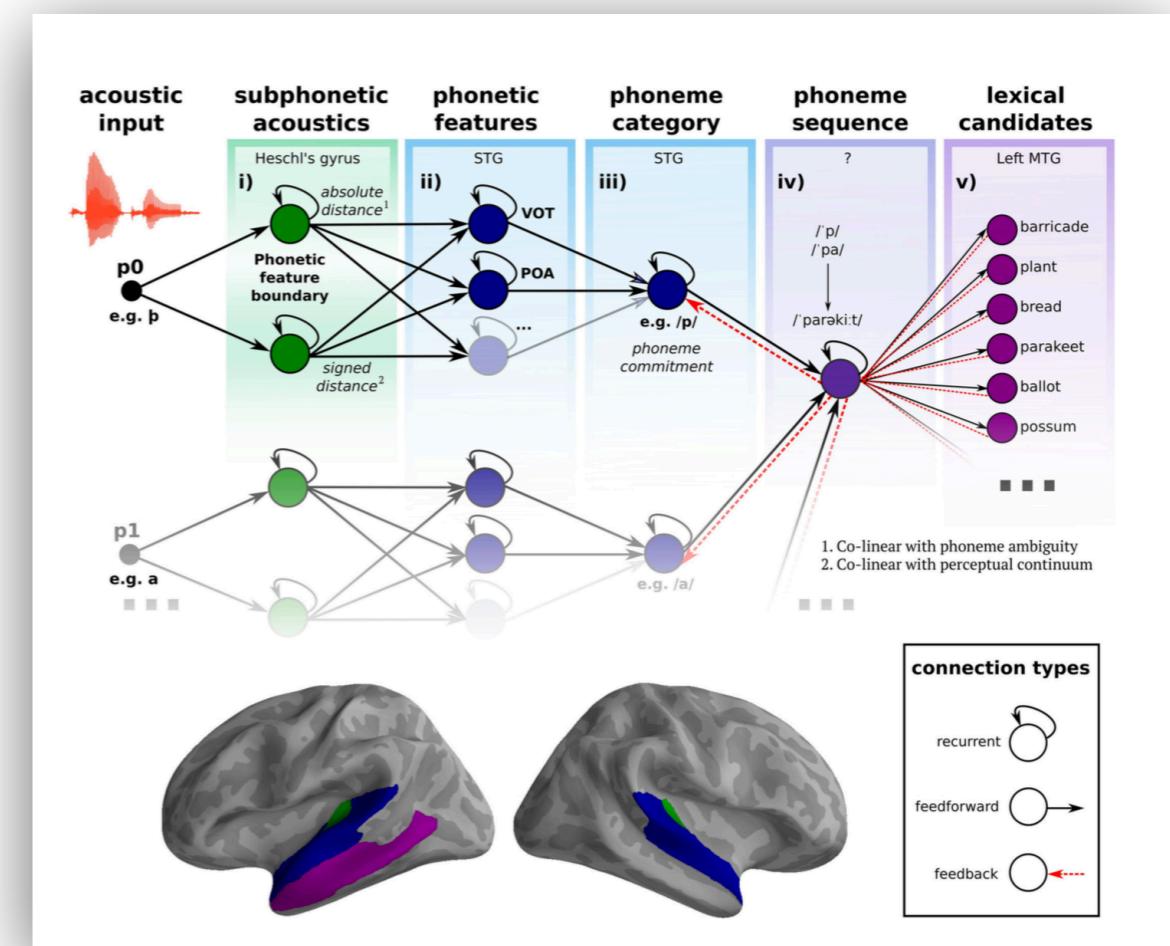
- Responses to phoneme ambiguity (bottom-up)
- Neural signatures of ambiguity resolution, when provided with lexical information (top-down)
- **What is the computational architecture?**

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?**

Conclusion

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



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
Laura Gwilliams | New York University | @GwilliamsL





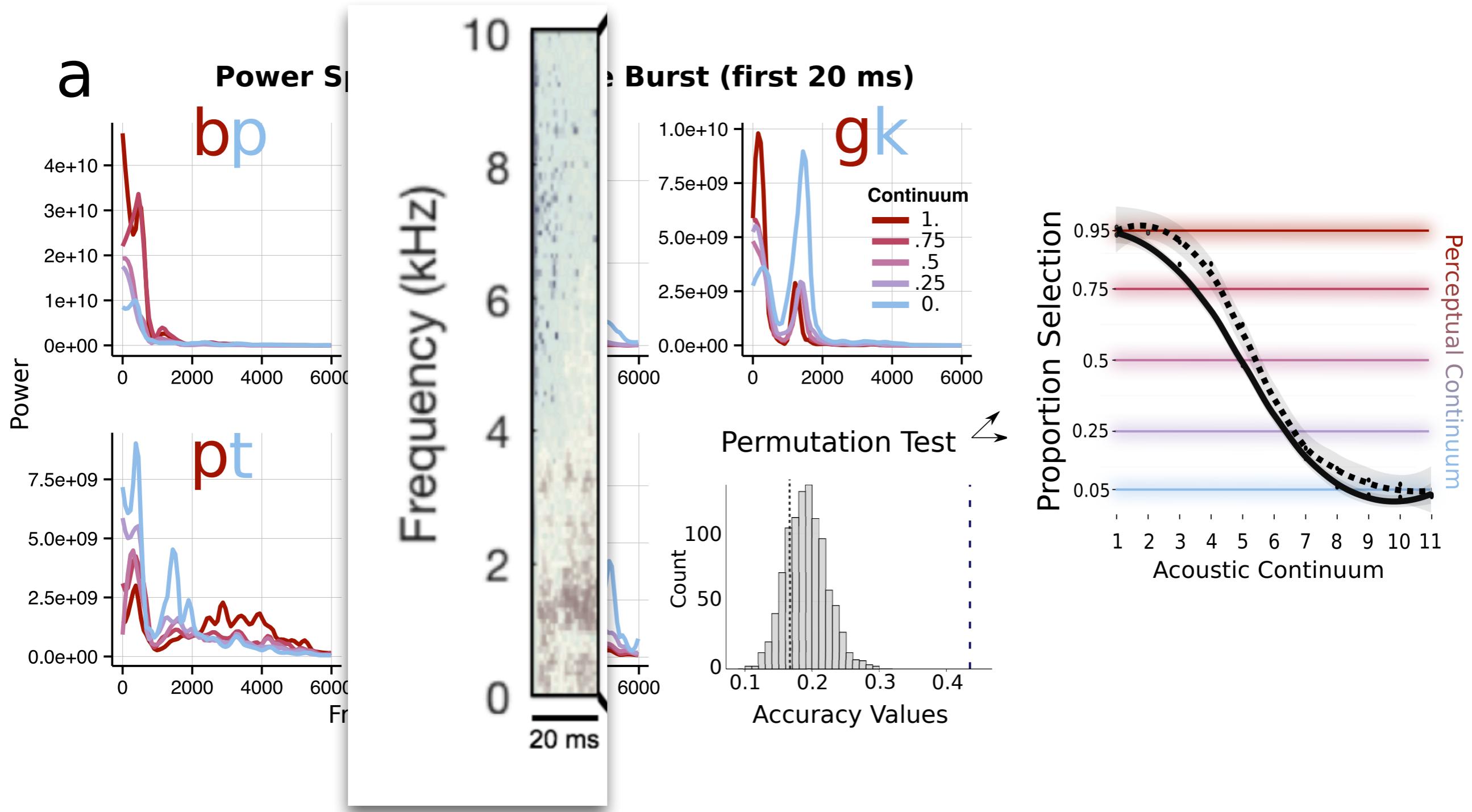
NEW YORK UNIVERSITY

✉ laura.gwilliams@nyu.edu
🐦 @GwilliamsL

Thank you!

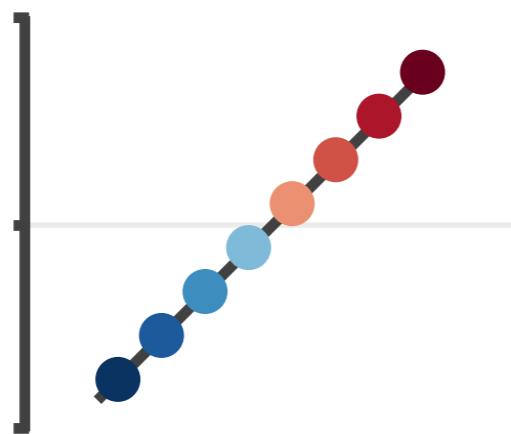


Is ambiguity correlated with acoustic properties?

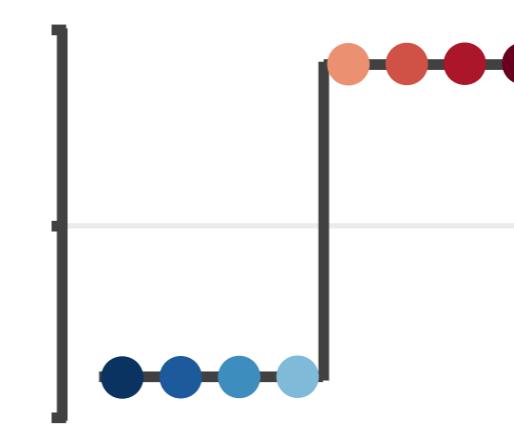


Predictive Coding

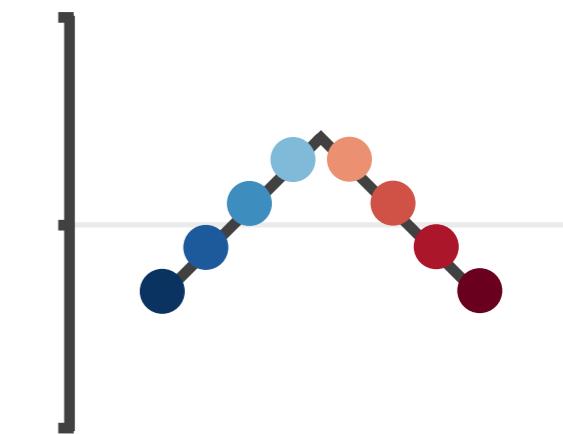
Linear Evidence



Categorical Percept

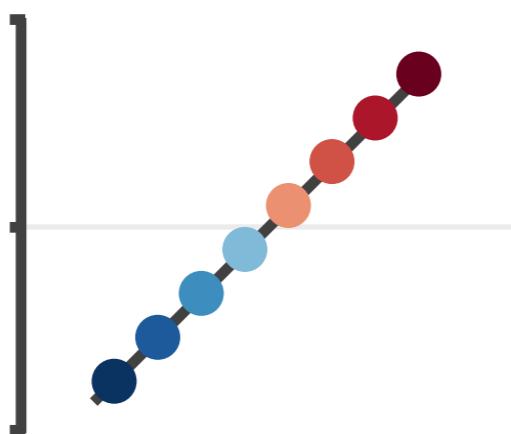


Ambiguity

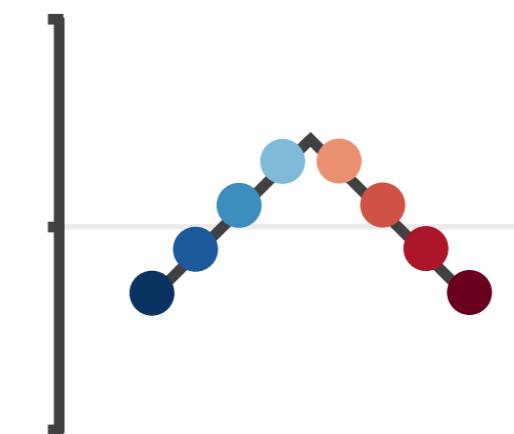


Neutralisation

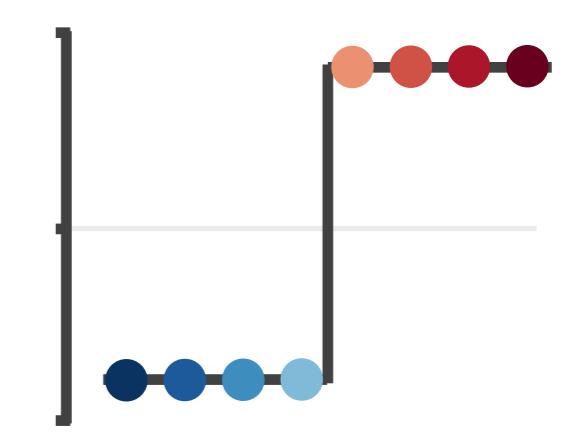
Linear Evidence



Ambiguity

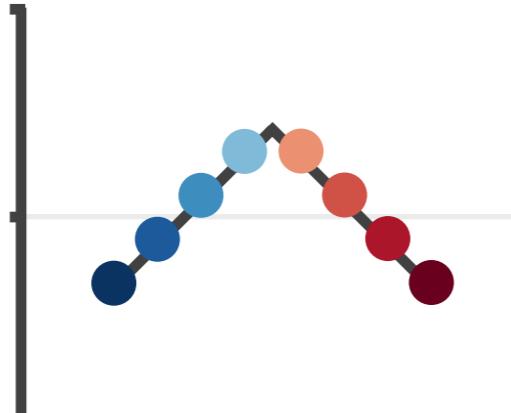


Categorical Percept

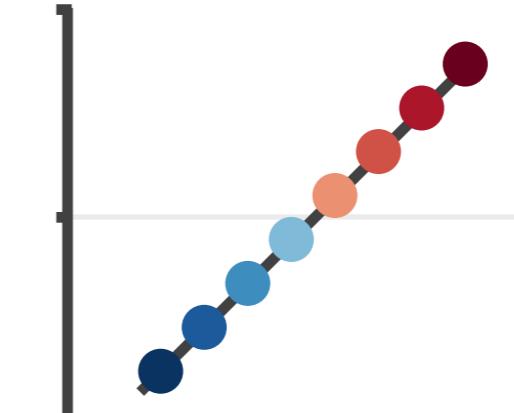


Cut-through connection

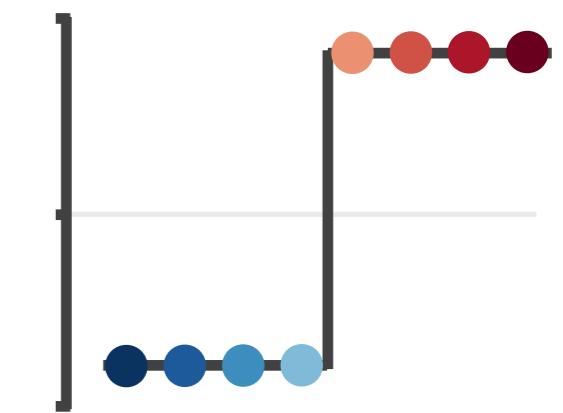
Ambiguity



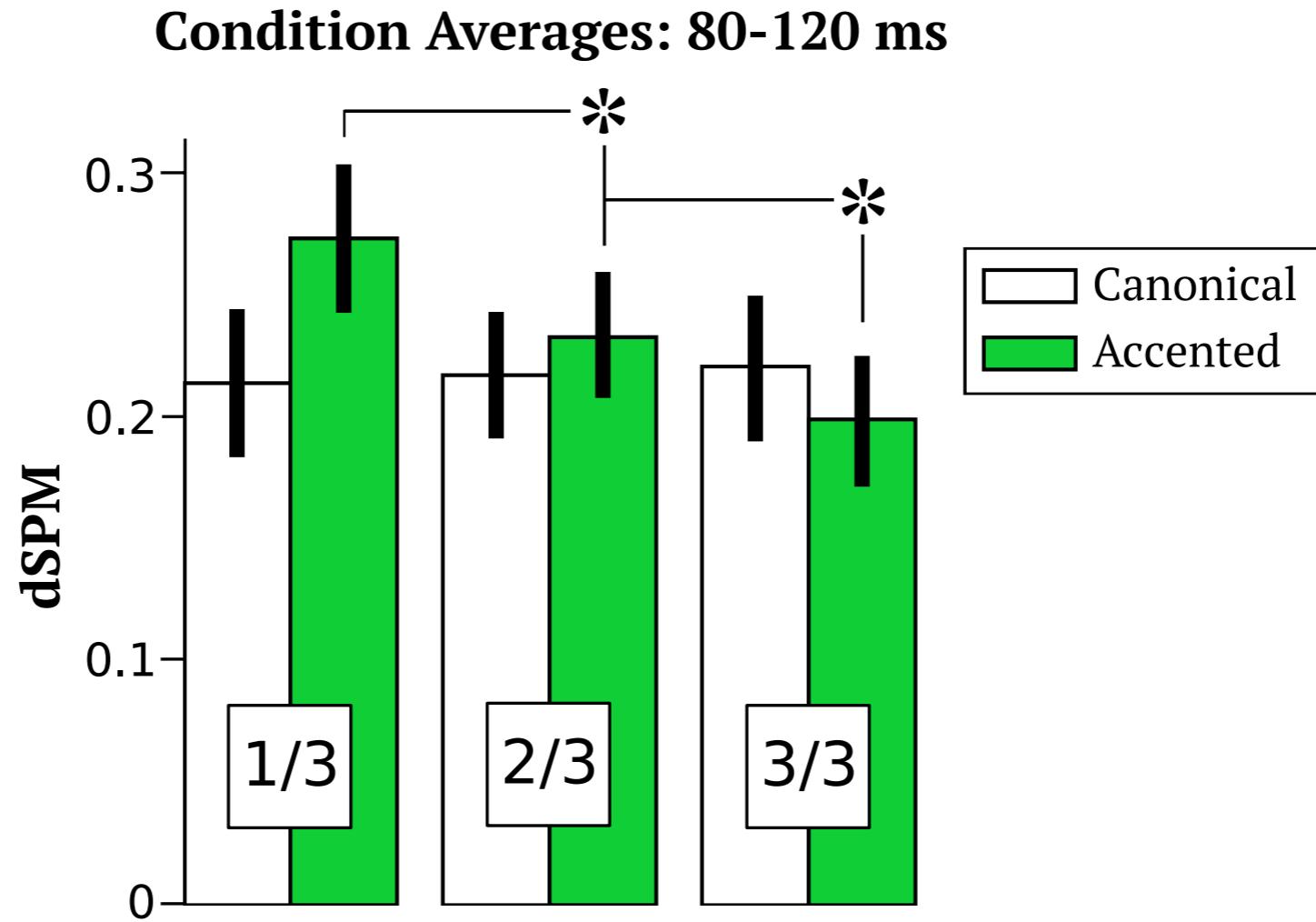
Linear Evidence



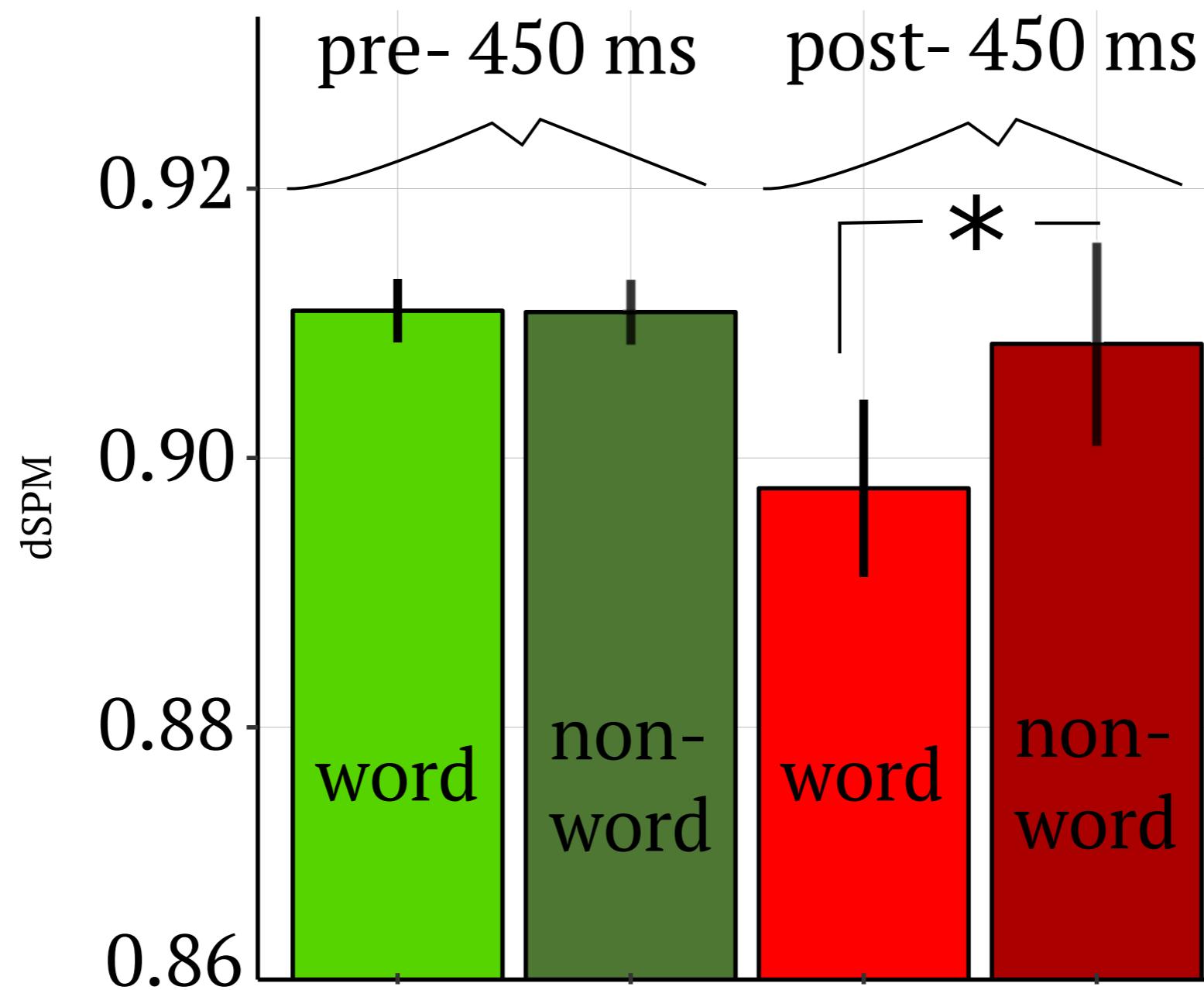
Categorical Percept



Interpretation

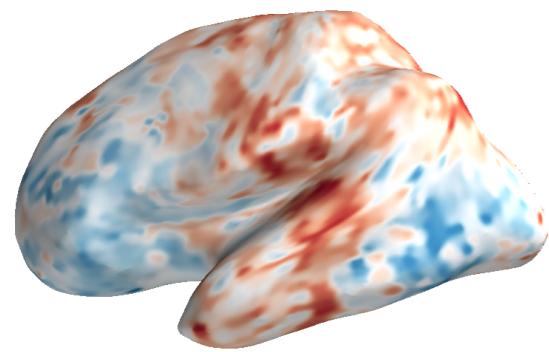


- Attunement is proposed to involve **re-tuning perceptual boundaries** between phonological categories (Norris et al., 2003; Kraljic and Samuel, 2005, 2006, 2007; Maye et al., 2008; see Samuel & Kraljic, 2009 for a review)

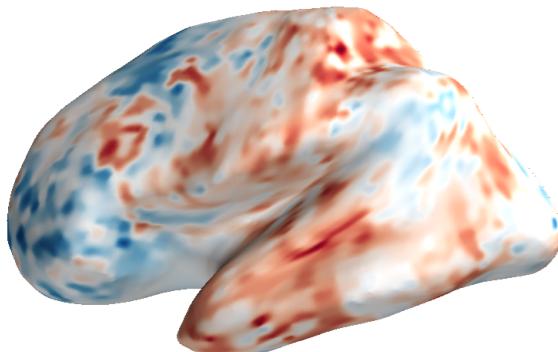


(Preliminary) Localisation of effects

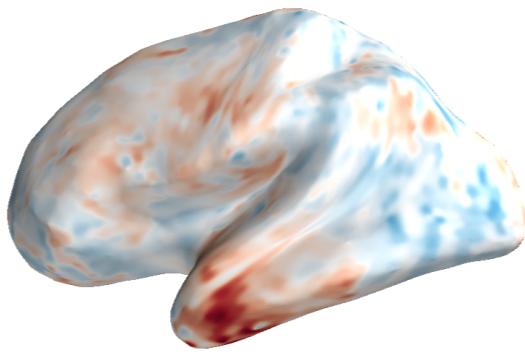
#syllables



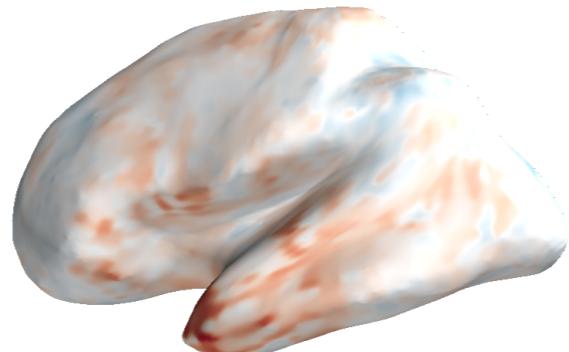
#opening nodes



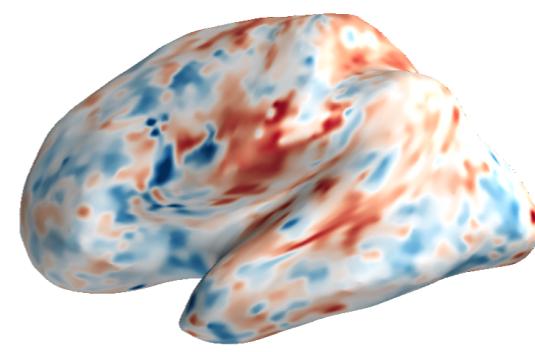
#morphemes



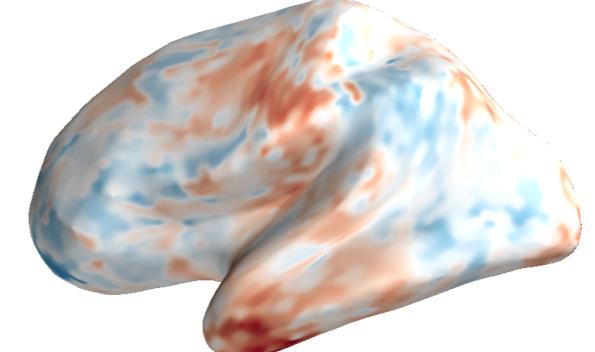
#merge



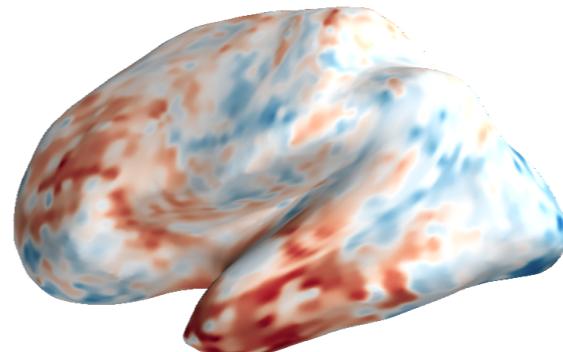
phon.
neighborhood



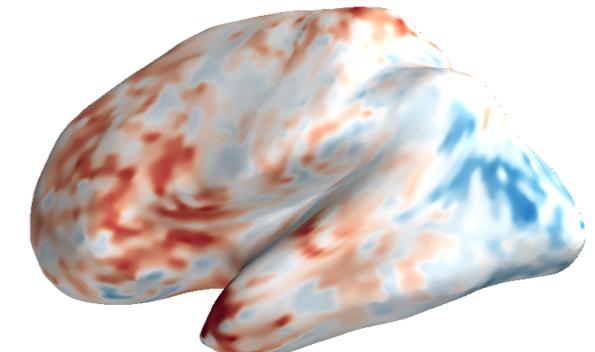
word
frequency



word number



depth



Confounded decoding accuracy

