



leg5@nyu.edu  
@GwilliamsL

# Parsing continuous speech into linguistic representations

Laura Gwilliams<sup>1,2</sup> Jean-Remi King<sup>1,4,5</sup> & David Poeppel<sup>1,3</sup>

New York University; NYU Abu Dhabi; Max Planck Institute; CNRS; Facebook AI Research



1

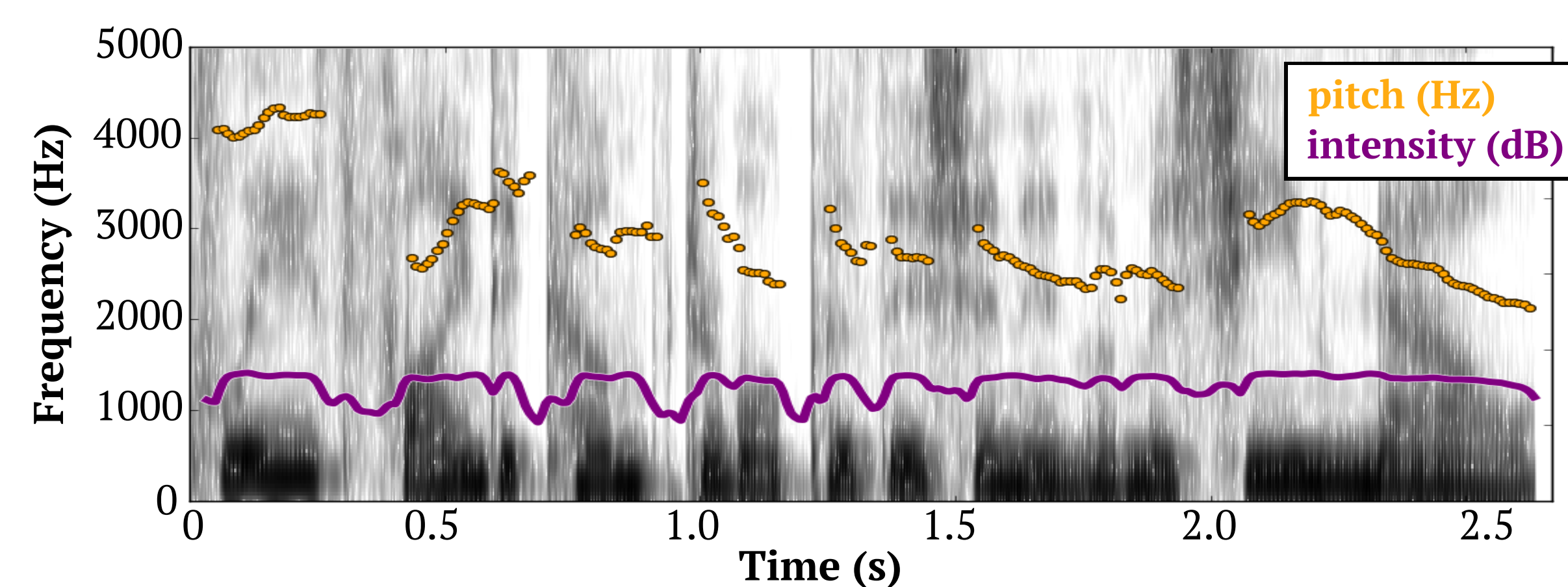
## Questions

- To comprehend continuous speech, the brain needs to generate a hierarchy of linguistic representations
  - It is currently unknown which are the primitive representations and how they are orchestrated
- which linguistic units are encoded in brain responses to naturalistic speech?
  - what is the relative time-course with which these properties are processed?
  - what computational architecture supports these linguistic representations?

2

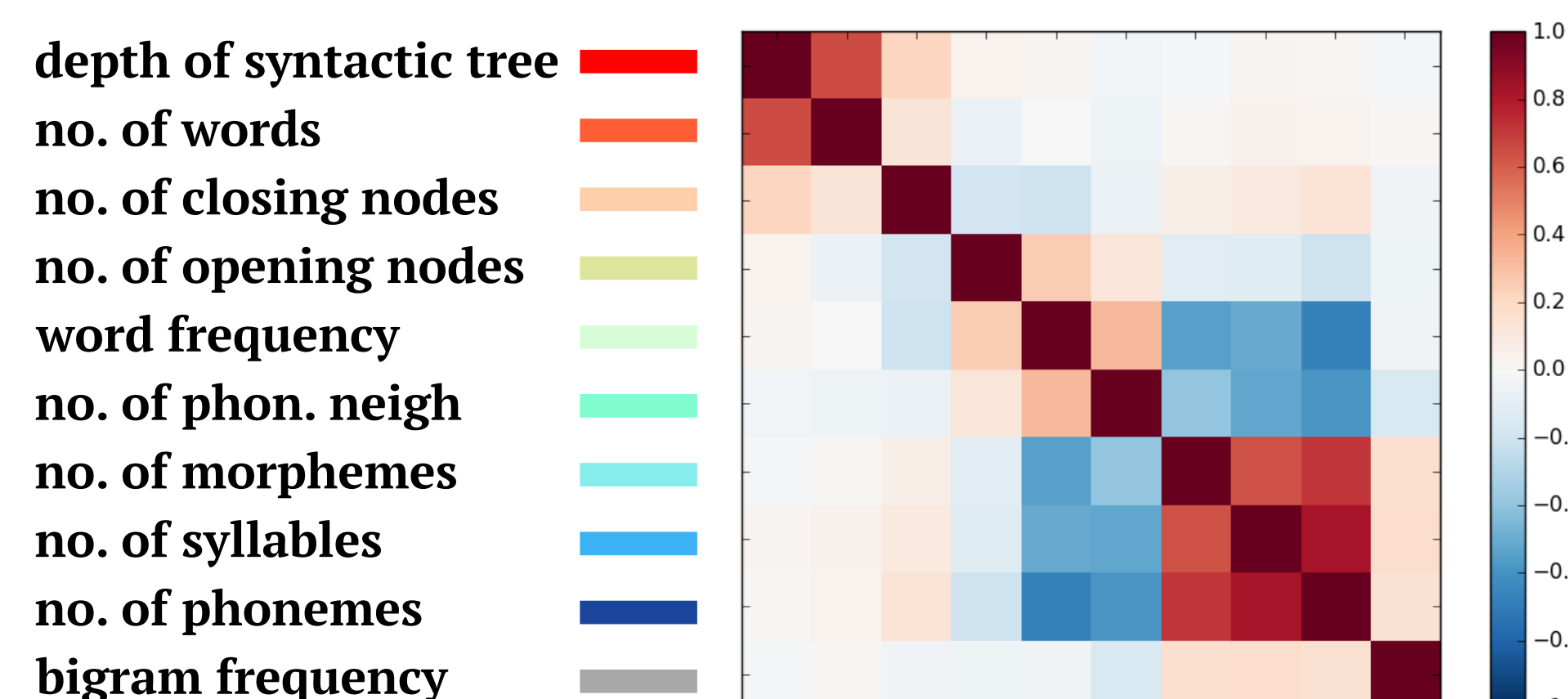
## Method

- 18 participants listened to four narratives (twice)
- 2 x one hour recordings
- KIT 208 channel MEG system
- Comprehension questions every ~4 minutes
- Responses to ~8000 words per participant



- Narratives were fully annotated - from acoustic to lexical and syntactic properties
- We decoded ten word-level properties from the MEG data, spanning different levels along the hierarchy

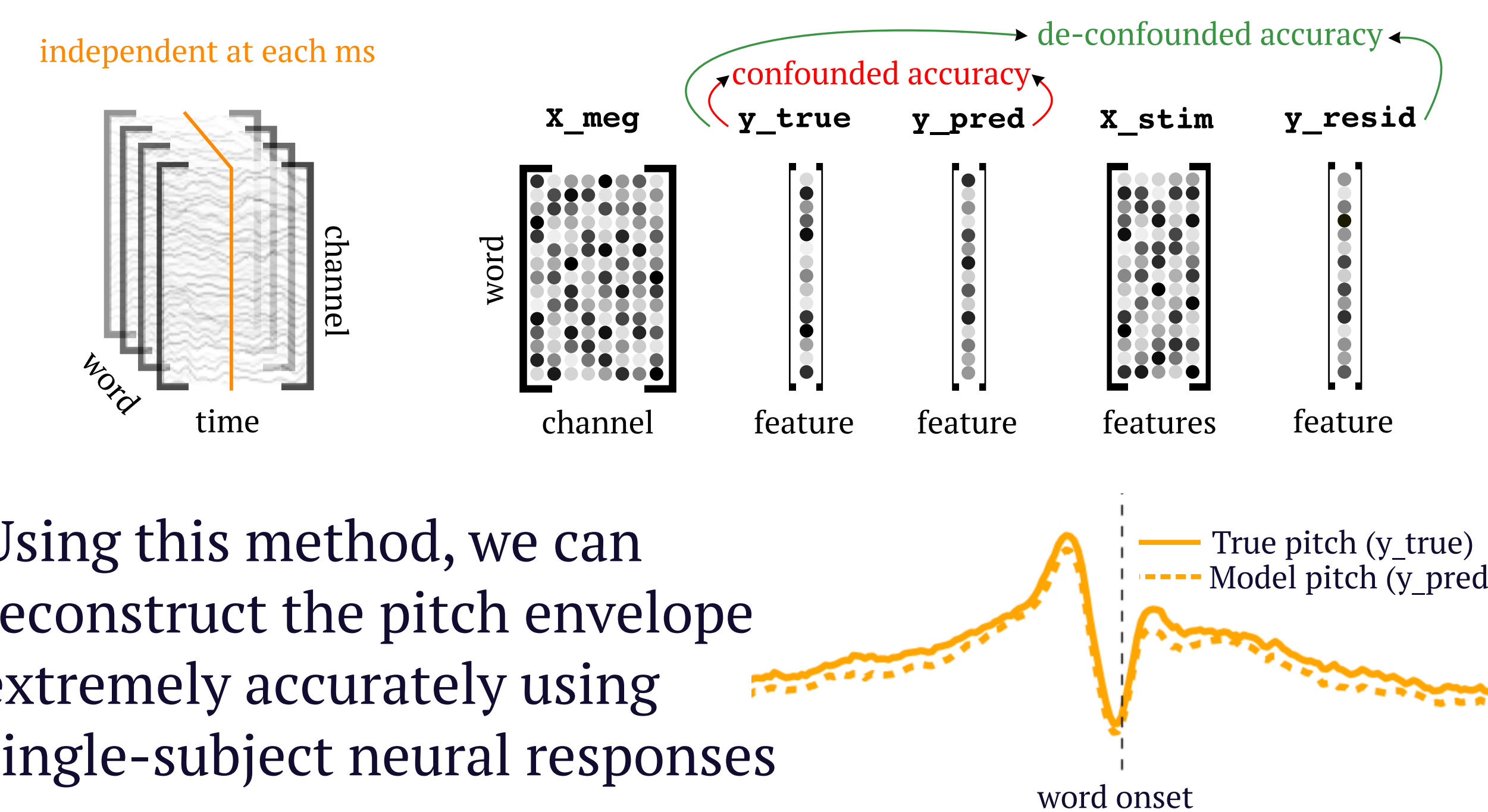
### confusion matrix of linguistic features



3

## Analysis Pipeline

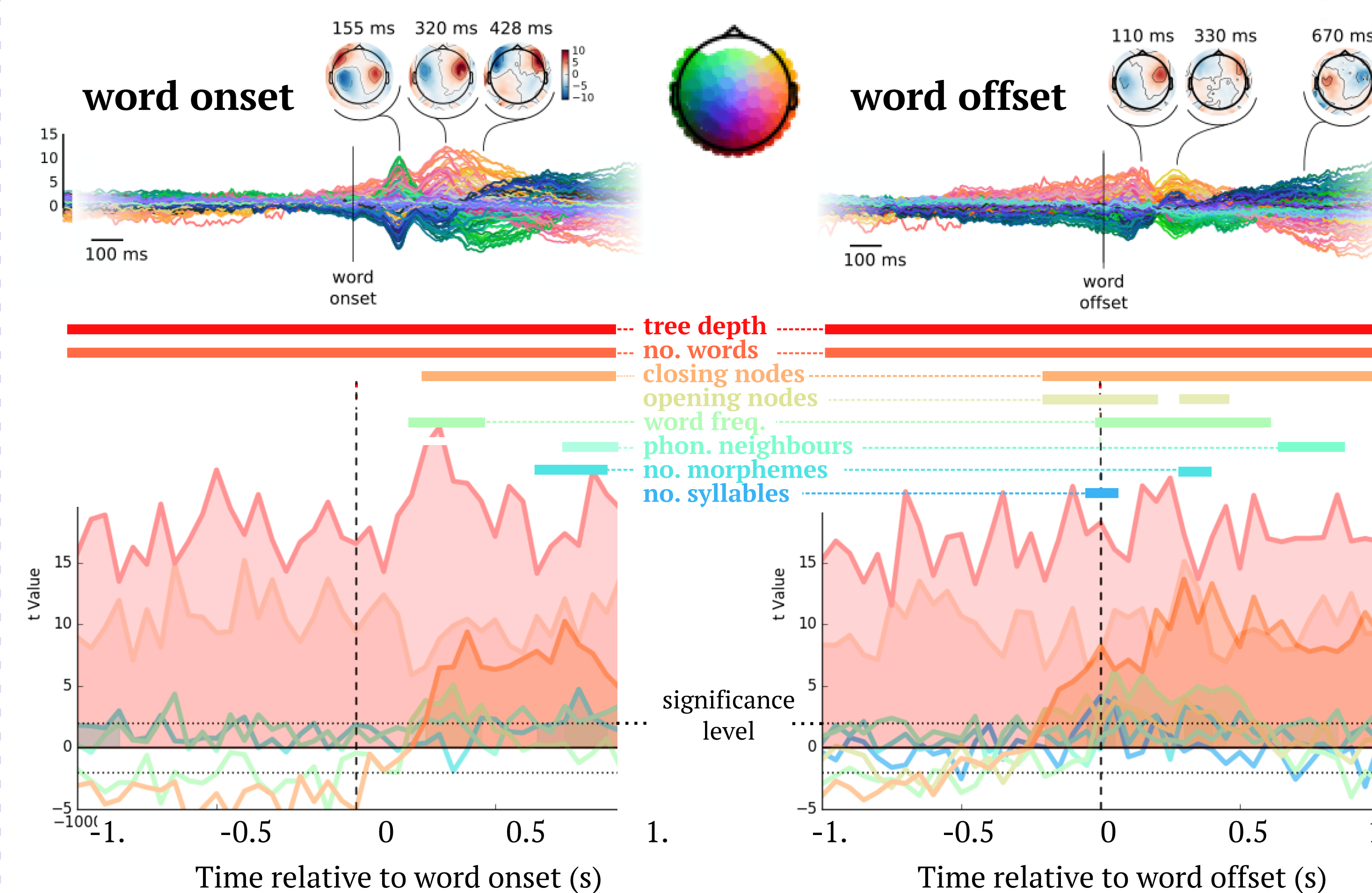
Due to the high natural co-variation between different linguistic properties, we need to orthogonalise them:



Using this method, we can reconstruct the pitch envelope extremely accurately using single-subject neural responses

4

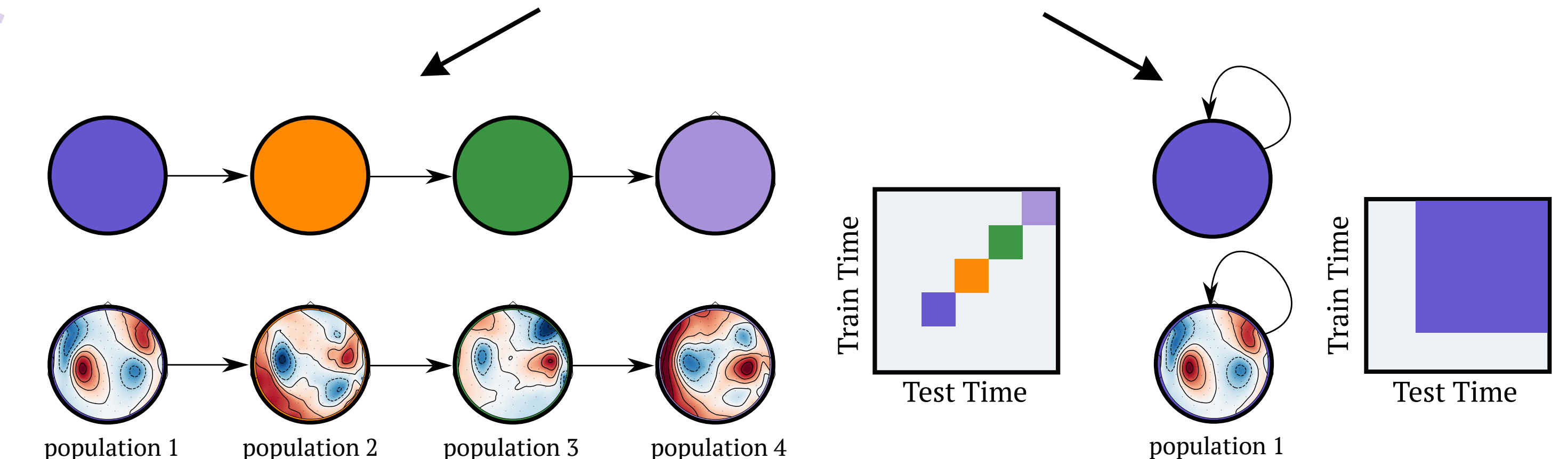
## Decoding the Linguistic Hierarchy



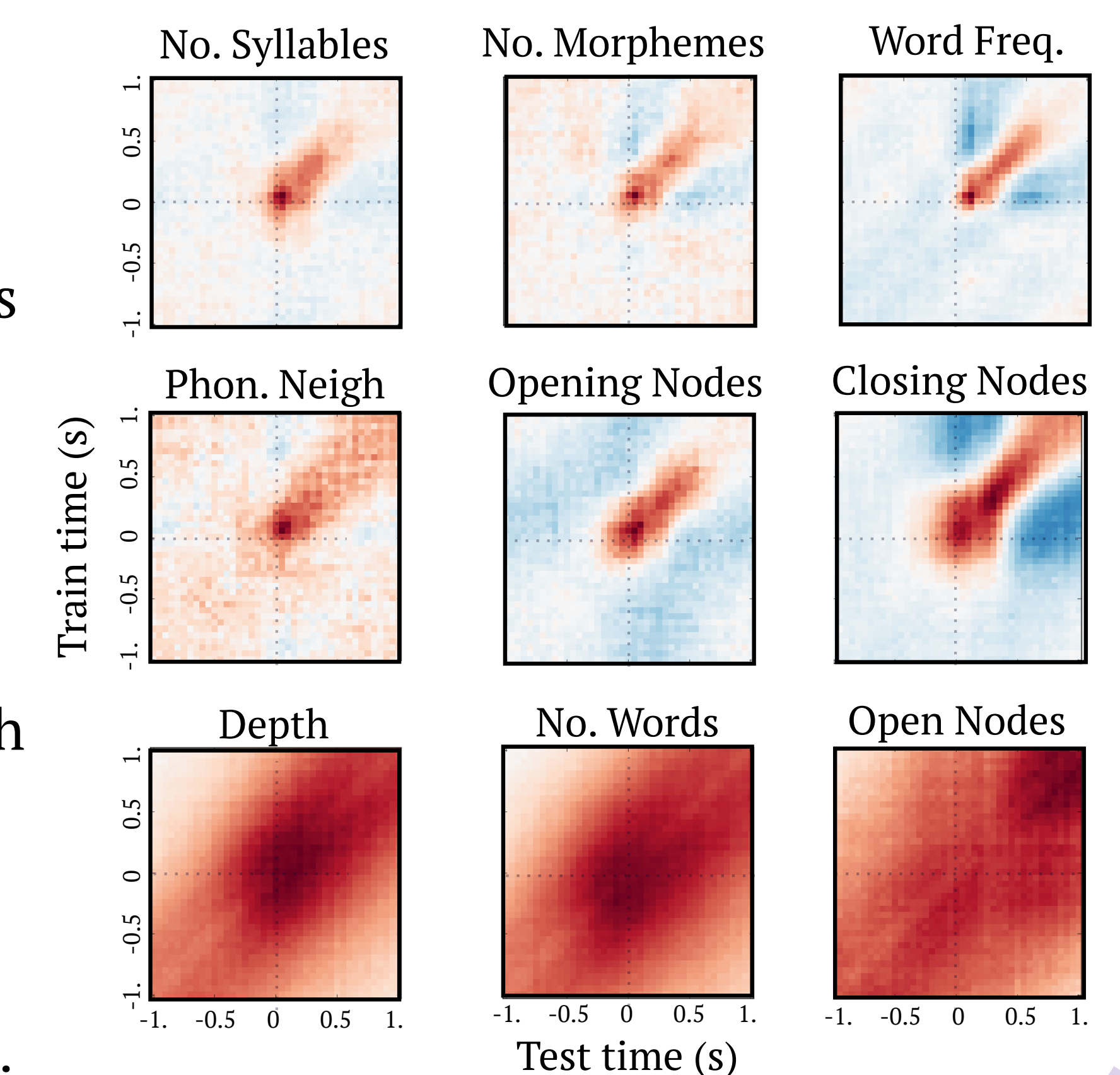
- Can decode a rich set of representations from the MEG signal, spanning the entire hierarchy from lexical to syntactic properties
- Significant temporal overlap between representations
- Evidence for some contentious units: syllables and morphemes
- Some representations are specifically locked to word onset/offset, whereas others are sustained over time

5

## Feedforward or Recurrent?



Dichotomy between processes occurring on a local scale, where each region is active for 100-200 ms, and processes which occur on a more global scale, where each region is active for 1-2 seconds.



A combination of both feedforward and recurrent processes are recruited depending on the linguistic representation

6

## Conclusion

- The brain generates multiple representations, spanning the entire linguistic hierarchy
  - We find evidence for some contentious units (no. of morphemes; no. of syllables)
  - Highly parallel architecture: most features are decodable in the same time instant
  - Both feedforward and recurrent computations are recruited, depending on the linguistic representation
- Three key aspects of the neural architecture supporting speech comprehension: feedforward, recurrent and parallel computations**

Funding: NYU Abu Dhabi G1001; NIH 2R01DC05660