Computational
Sensorimotor
Systems Lab

The Progression of Neural Speech Representations Through Auditory Cortex and Beyond, from Acoustics to Language to Semantics

Jonathan Z. Simon

University of Maryland
Department of Electrical & Computer Engineering,
Department of Biology, Institute for Systems Research

Mastodon: @jzsimon@fediscience.org

http://www.isr.umd.edu/Labs/CSSL/simonlab

University of Lübeck, 1 Aug 2023

Acknowledgements

Current Lab Members & Affiliates

Morgan Belcher
Vrishab Commuri
Charlie Fisher
Tejas Guha
Brooke Guo
Michael Johns
Kevin Hu
Dushyanthi Karunathilake
Karl Lerud
Behrad Soleimani
Ciaran Stone
Craig Thorburn
Allie Vance

Current & Recent Collaborators

Samira Anderson
Behtash Babadi
Tom Francart
L. Elliot Hong
Stefanie Kuchinsky
Ellen Lau
Elisabeth Marsh
Philip Resnik

Recent Lab Members & Affiliates

Sahar Akram
Olivia Bermudez-Hopkins
Shohini Bhattasali
Christian Brodbeck
Regina Calloway
Francisco Cervantes Constantino
Aura Cruz Heredia
Proloy Das
Lien Decruy
Marisel Villafane Delgado
Nai Ding
Jason Dunlap
Sydney Hancock
Marlies Gilles
Alex Jiao
Neha Joshi
Joshua Kulasingham
Natalia Lapinskaya
Sina Miran
Mohsen Rezaeizadeh
Alex Presacco
Krishna Puvvada
Jonas Vanthornhout
Richard Williams
Peng Zan

Funding & Support

NIDCD

NIH
NIA

NSF

DARPA

UNIVERSITY OF MARYLAND
18 56

Outline

- Introduction—Cortical representations of continuous speech
- *Early & fast* cortical representation of continuous speech
- Cortical representations of speech *meaning*
- *Progression* of representations of continuous speech through cortex (bottom-up and top-down)
- Objective measures of speech *intelligibility*
- *Directional functional connectivity* during difficult speech listening

Outline

- Introduction—Cortical representations of continuous speech
- *Early & fast* cortical representation of continuous speech
- Cortical representations of speech *meaning*
- *Progression* of representations of continuous speech through cortex (bottom-up and top-down)
- Objective measures of speech *intelligibility*
- *Directional functional connectivity* during difficult speech listening

Cortical Representations of Continuous Speech

Continuous speech

- naturalistic
- redundant
- employs auditory cognition
- acoustically rich
- drives most auditory areas
- …
- but also complicated

> If you happened to find yourself on the banks of the Ohio River on a particular afternoon in the spring of 1806—somewhere just to the north of Wheeling, West Virginia, say …
>
> *The Botany of Desire* — Michael Pollan

> Alfred the Great was a young man, three-and-twenty years of age, when he became king. Twice in his childhood, he had been taken to Rome, where the Saxon nobles were in the habit of going on journeys which they supposed to be religious; …
>
> *A Child's History of England* — Charles Dickens

> In the bosom of one of those spacious coves which indent the eastern shore of the Hudson, at that broad expansion of the river denominated by the ancient Dutch navigators …
>
> *The Legend of Sleepy Hollow* — Washington Irving

> He was an old man who fished alone in a skiff in the Gulf Stream and he had gone eighty-four days now without taking a fish. In the first forty days a boy had been with him. But after forty days without a fish …
>
> *The Old Man and the Sea* — Ernest Hemingway

Cortical Representations of Continuous Speech

Temporal neural patterns ⇆ temporal patterns in speech

- Generalization of “Speech Tracking”
- Need high temporal precision, for fast temporal speech features

  - EEG (electroencephalography): *whole brain*
  - MEG (magnetoencephalography): *whole brain but with strong cortical bias*
  - ECoG (electrocorticography): *placed cortical surface electrodes*
  - single- and multi-unit recording methods: *placed depth electrodes*

Cortical Representations of Continuous Speech

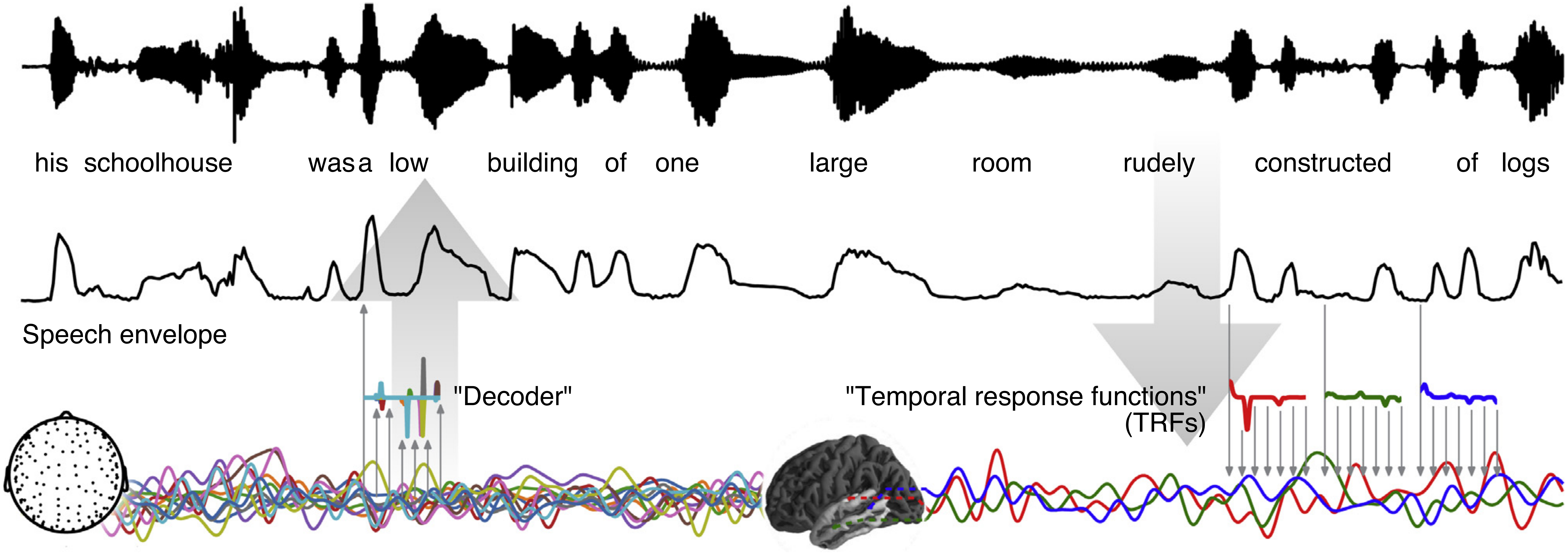
Neural Representations of Speech

• oscillations at pitch frequencies (primarily subcortical)	Maddox & Lee (2018) eNeuro
• acoustic onset tracking	Daube et al. (2019) Curr Biol
• speech envelope rhythmic following	Lalor & Foxe (2010) Eur J Neurosci
• phoneme-based responses	Teoh et al. (2022) J Neurosci
• phoneme-context-based responses	Brodbeck et al. (2018) Curr Biol
• word-context-based responses	Brodbeck et al. (2022) eLife
• semantic structure rhythm following	Ding et al. (2016) Nat Neuro
• plus connections to **intelligibility/perception/behavior**	

Brodbeck & Simon (2020) *Continuous Speech Processing*, Curr Op Physiol

Cortical Representations of Speech

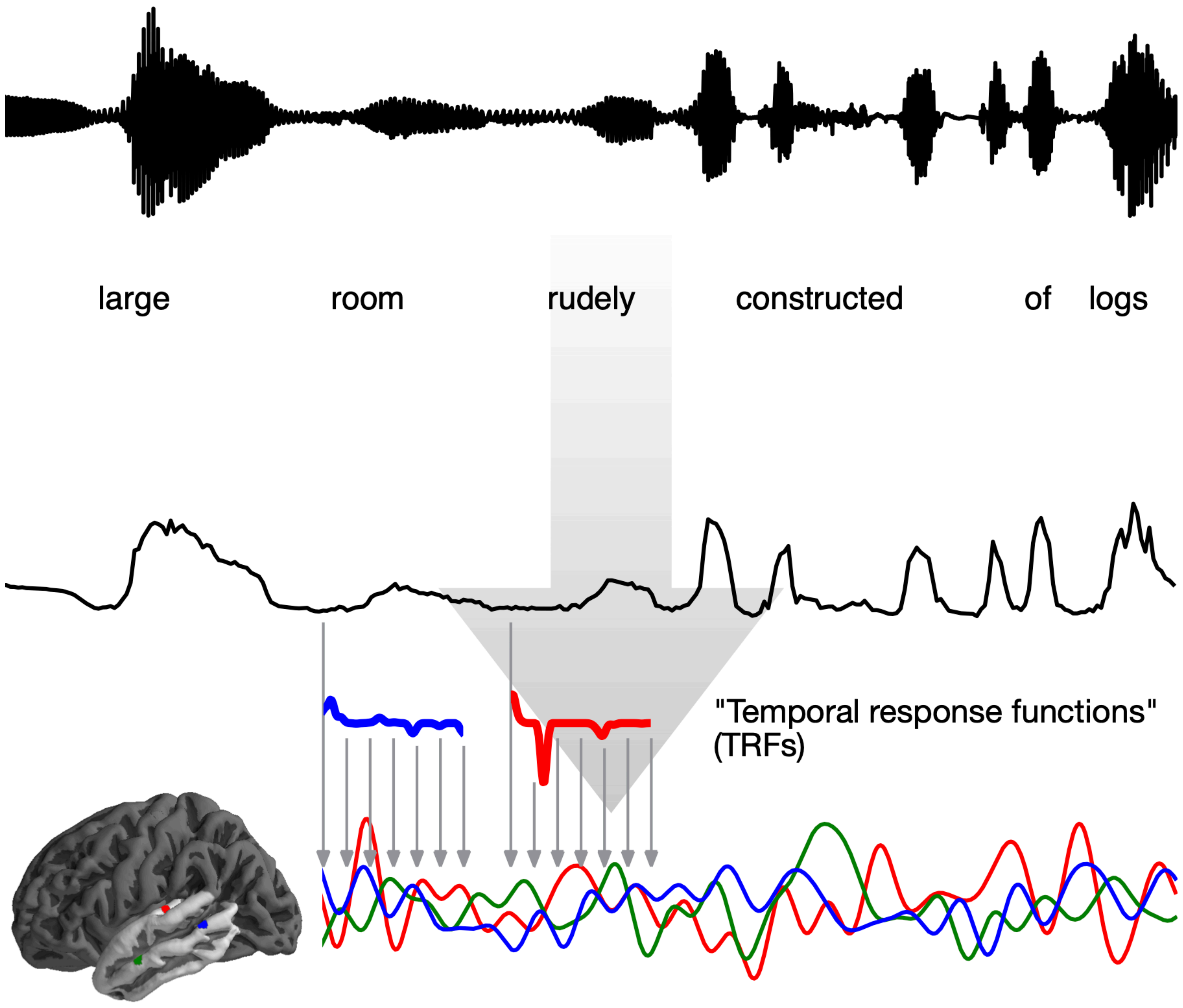
- Measure *time-locked* responses to temporal pattern of speech features (in humans)
- Any speech feature of interest: acoustic envelope, lexical, pitch, semantic, etc.
- Infer spatio-temporal neural origins of neural responses


his schoolhouse was a low building of one large room rudely constructed of logs
Speech envelope
"Decoder"
"Temporal response functions"
(TRFs)

Cortical Representations: Encoding

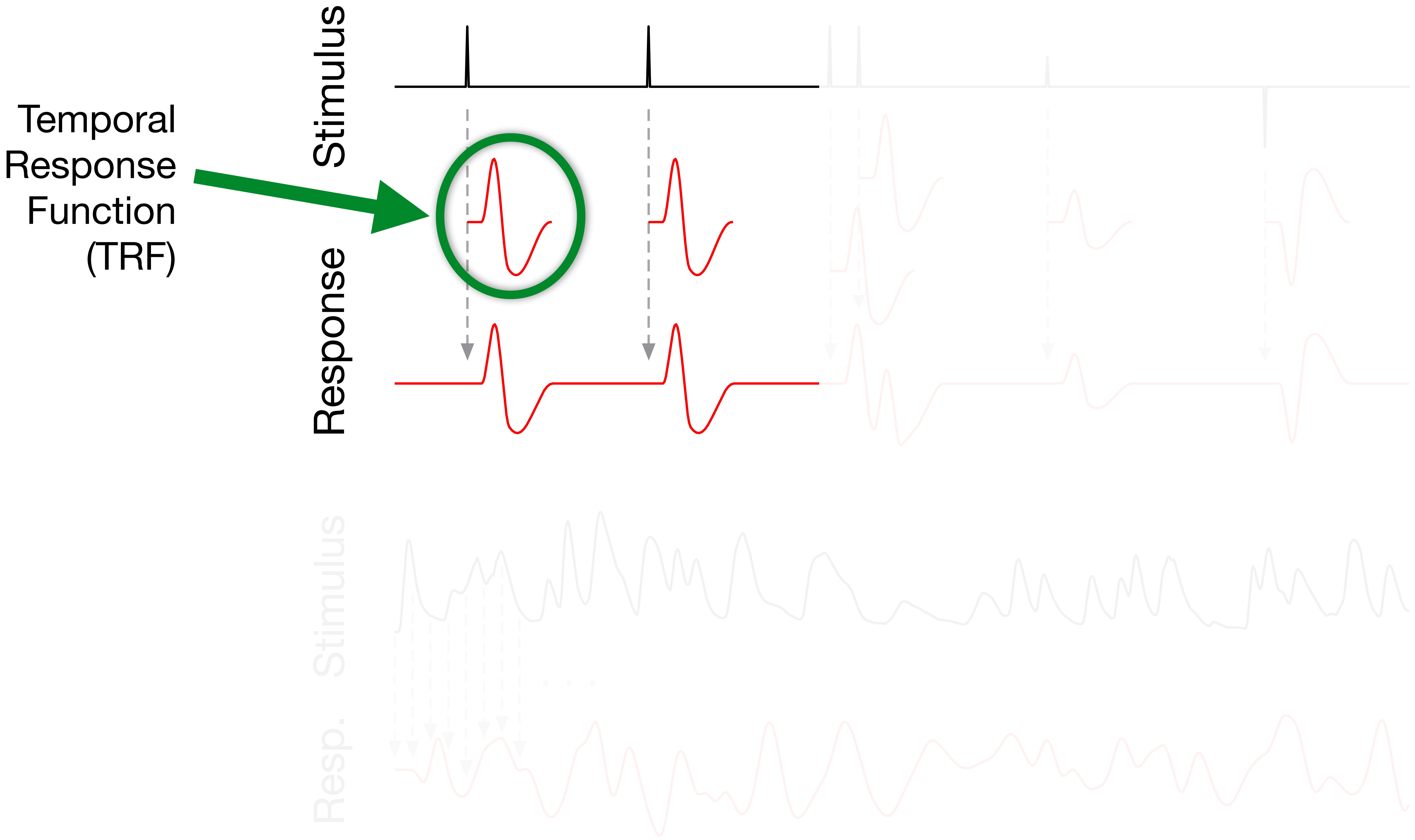
- Predicting future neural responses from present stimulus features,
  - wide variety of stimulus features
  - via Temporal Response Function (TRF)

- Why look at encoding? It *often* tells us more about the brain
  - TRF analogous to evoked response
  - peak amplitude ≈ processing intensity
  - peak latency ≈ source location
  - multiple TRFs simultaneously



Example: MEG Prediction of Voxel Responses

Temporal Response Functions

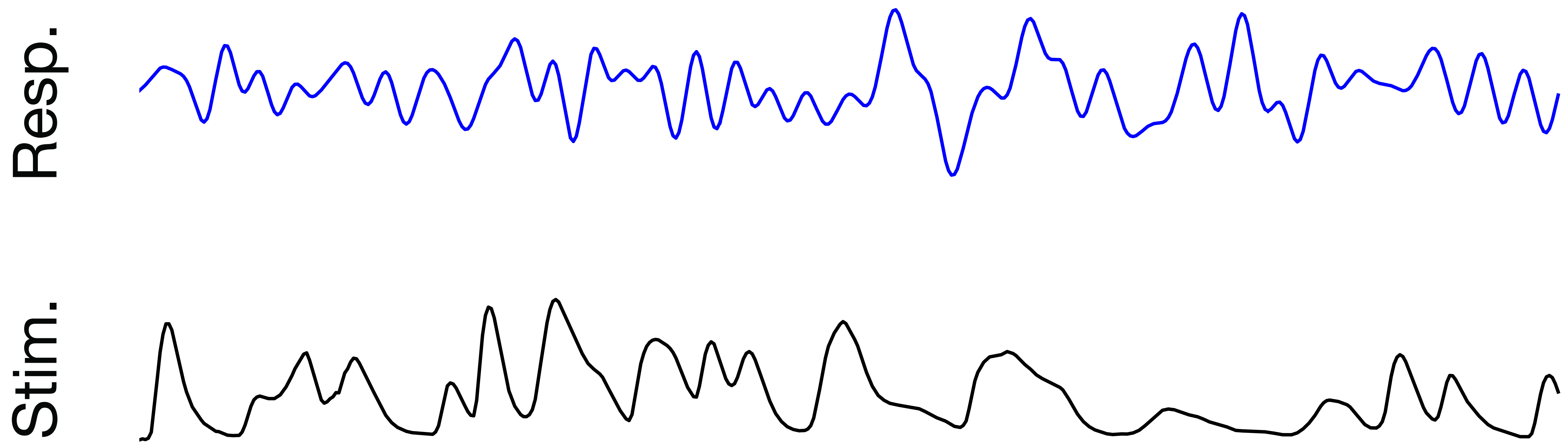


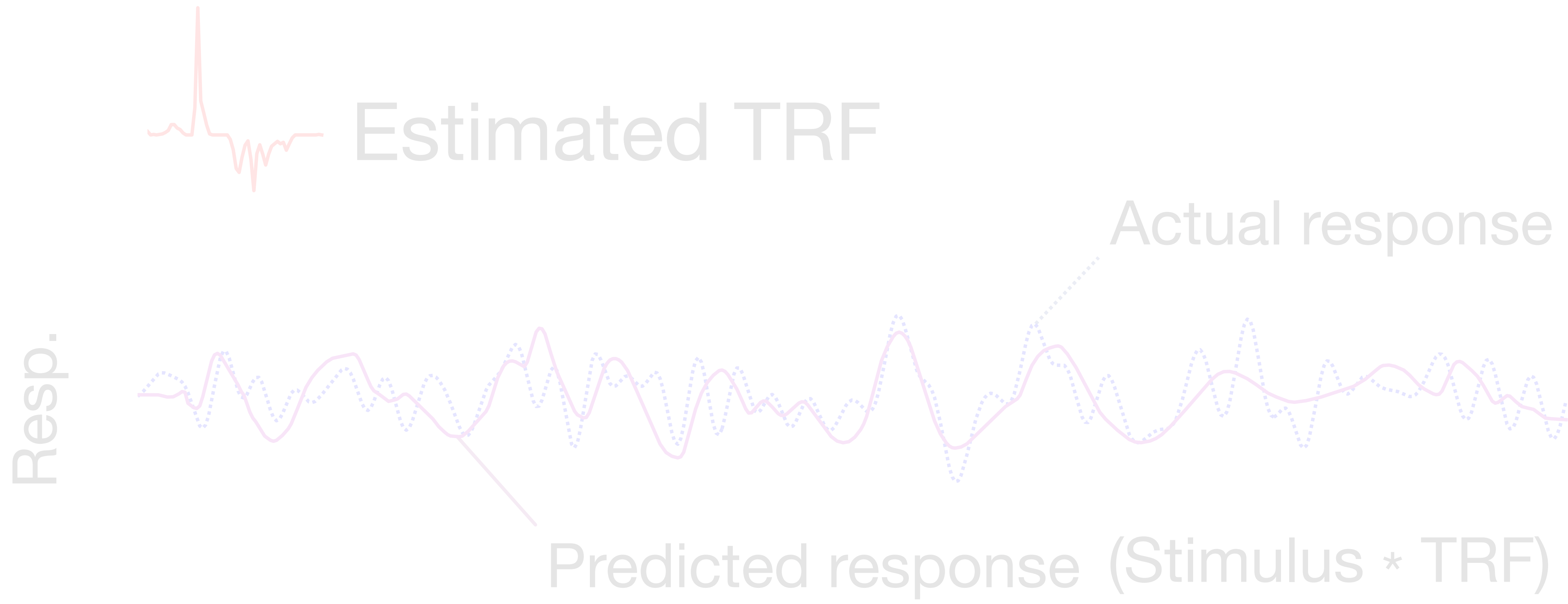
Temporal Response Function (TRF)

TRF Model Estimation & Fit

Temporal Response Function (TRF) estimation:

Stimulus and response are known; find the best TRF to produce the response from the stimulus:

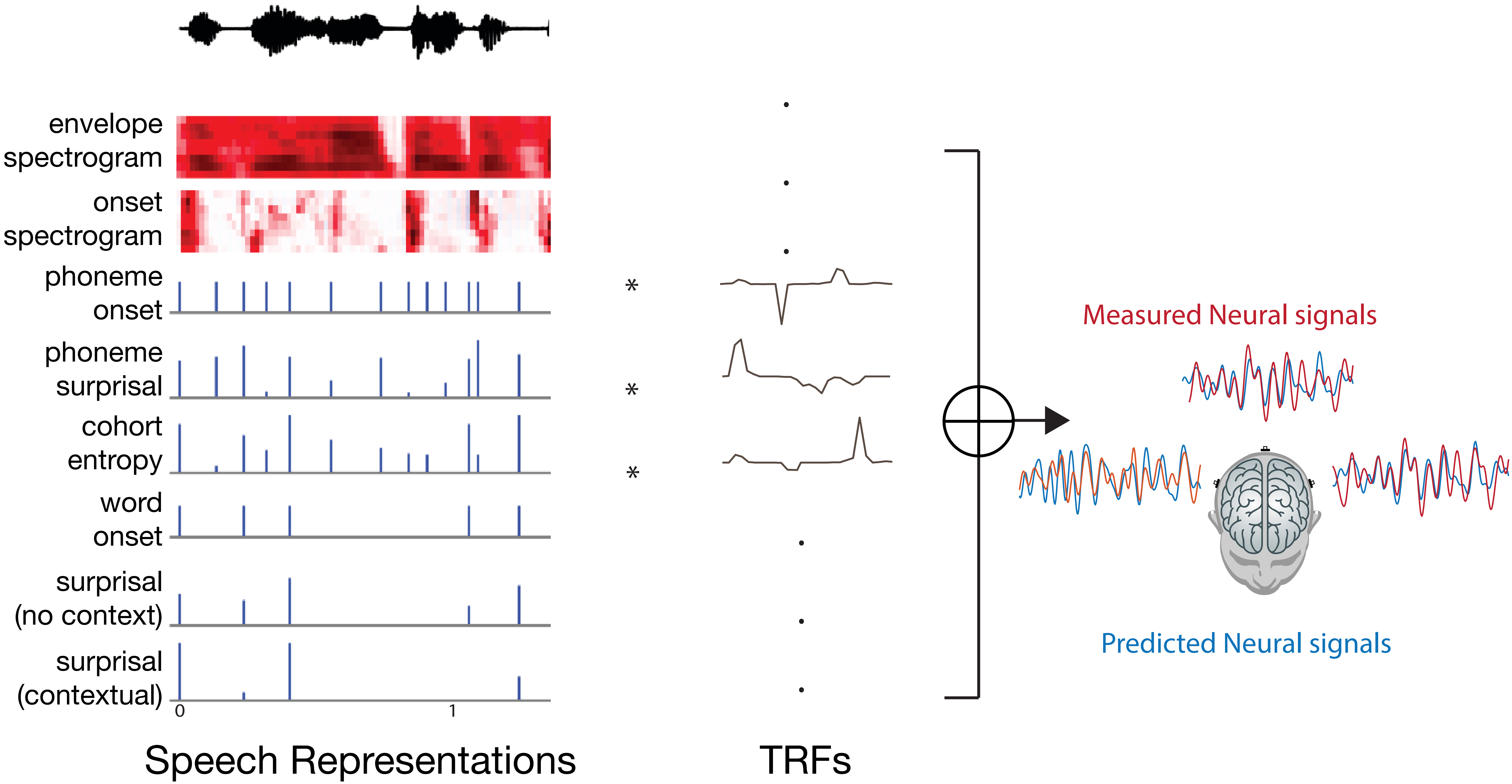




Lalor & Foxe (2010) *Neural Responses to Uninterrupted Natural Speech* … Eur J Neurosci
Ding & Simon (2012) *Neural Coding of Continuous Speech in Auditory Cortex* … , J Neurophys

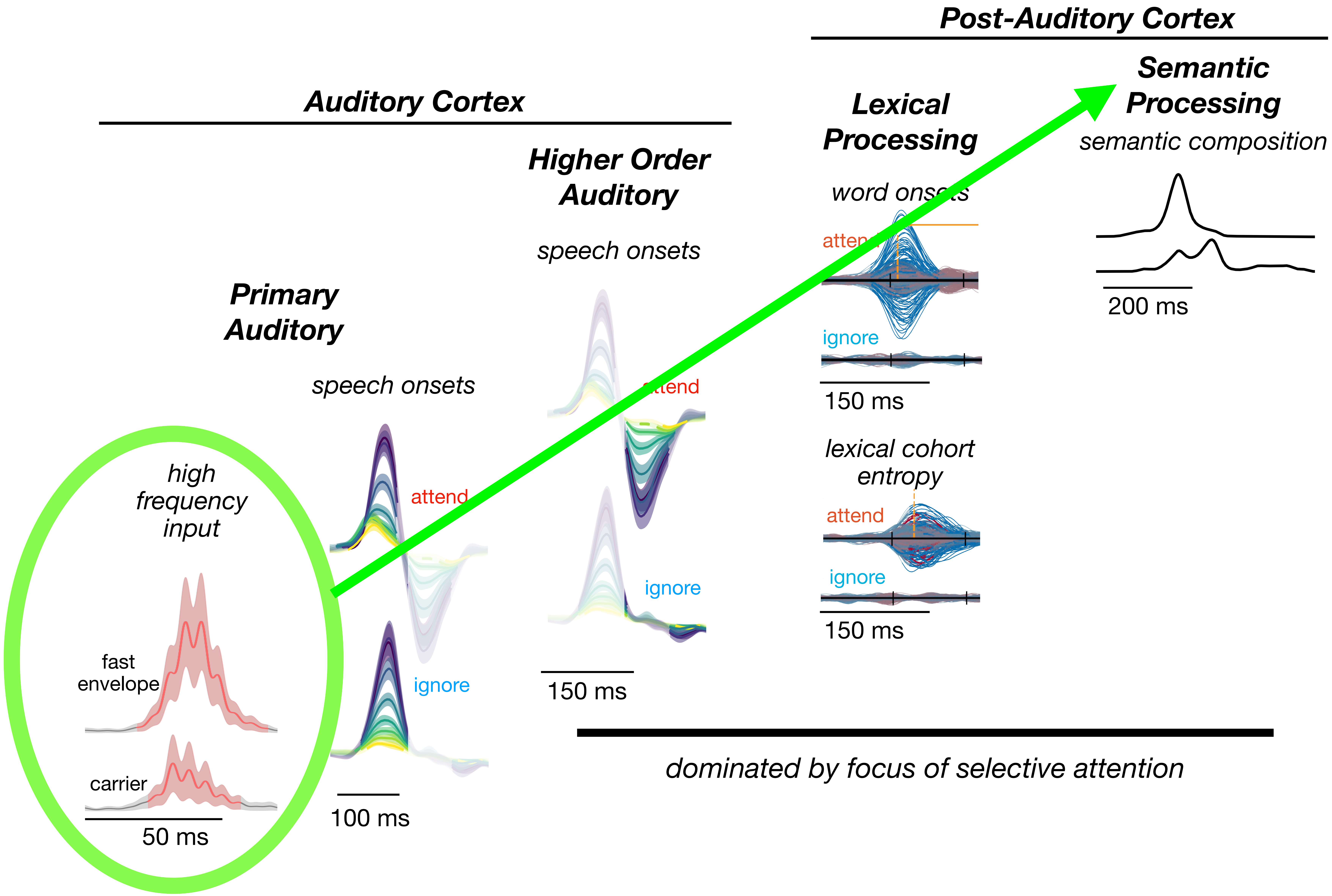
Simultaneous Temporal Response Functions

- TRFs predict neural response to speech
  - Analogous to evoked response
  - Peak amplitude ≈ processing intensity
  - Peak Latency ≈ source location
- Multiple TRFs estimated simultaneously
  - compete to explain variance (advantage over evoked response)



Crosse et al. (2016) *The Multivariate Temporal Response Function (mTRF) Toolbox* … , Front Hum Neurosci
Brodbeck et al. (2021) *Eelbrain: A Python Toolkit for Time-Continuous Analysis* … , bioRxiv

Cortical Representations Across Cortex



Post-Auditory Cortex

Auditory Cortex

Primary Auditory

speech onsets

high frequency input

Higher Order Auditory

speech onsets

Lexical Processing

word onsets

lexical cohort entropy

Semantic Processing

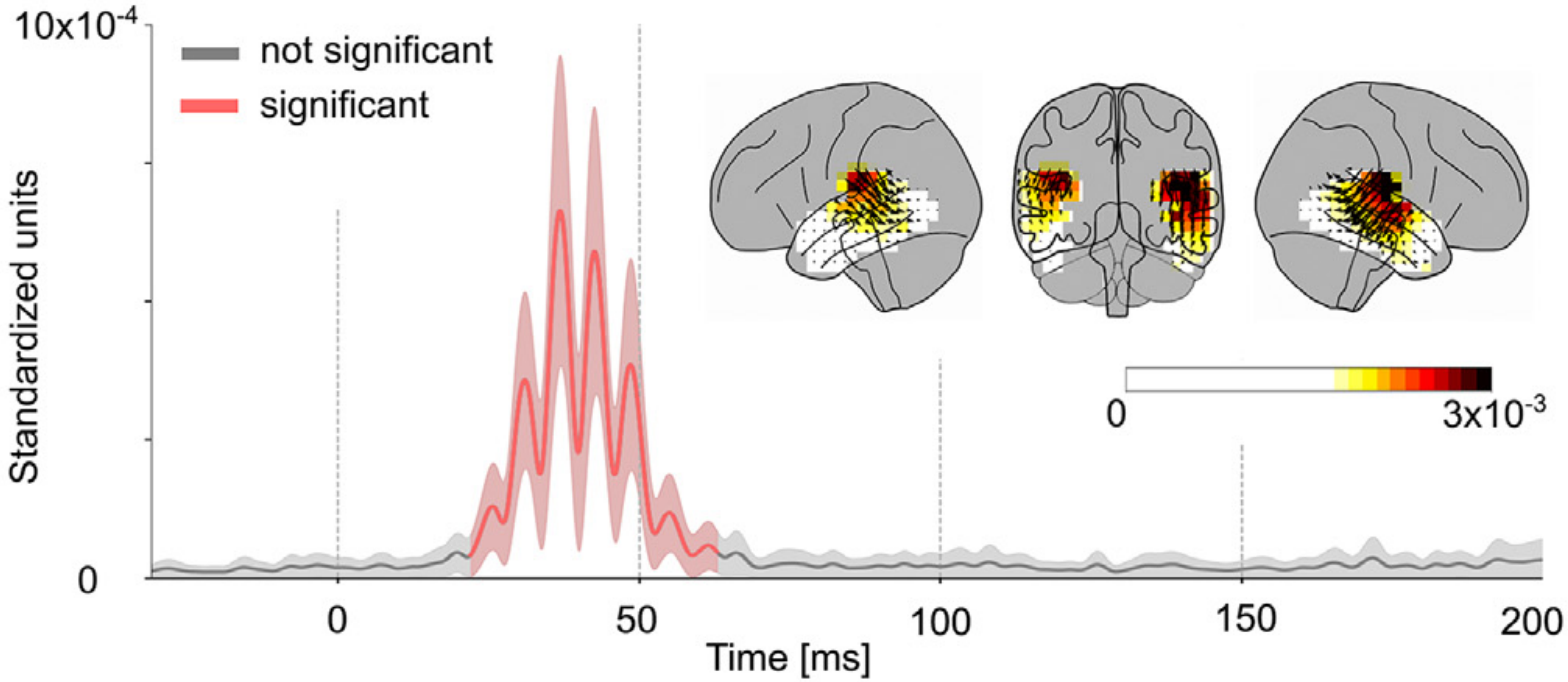
semantic composition

dominated by focus of selective attention

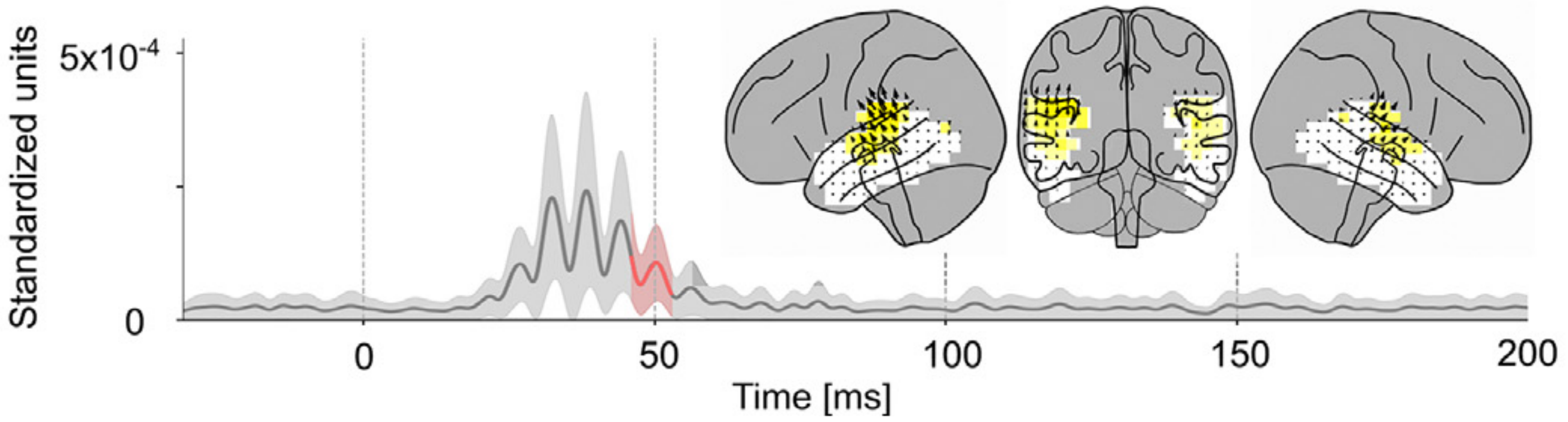
Outline

- Introduction—Cortical representations of continuous speech
- *Early & fast* cortical representation of continuous speech
- Cortical representations of speech *meaning*
- *Progression* of representations of continuous speech through cortex (bottom-up and top-down)
- Objective measures of speech *intelligibility*
- *Directional functional connectivity* during difficult speech listening

Fast & Early Cortical Representations



TRF (MEG) for
70-200 Hz
continuous speech
envelope

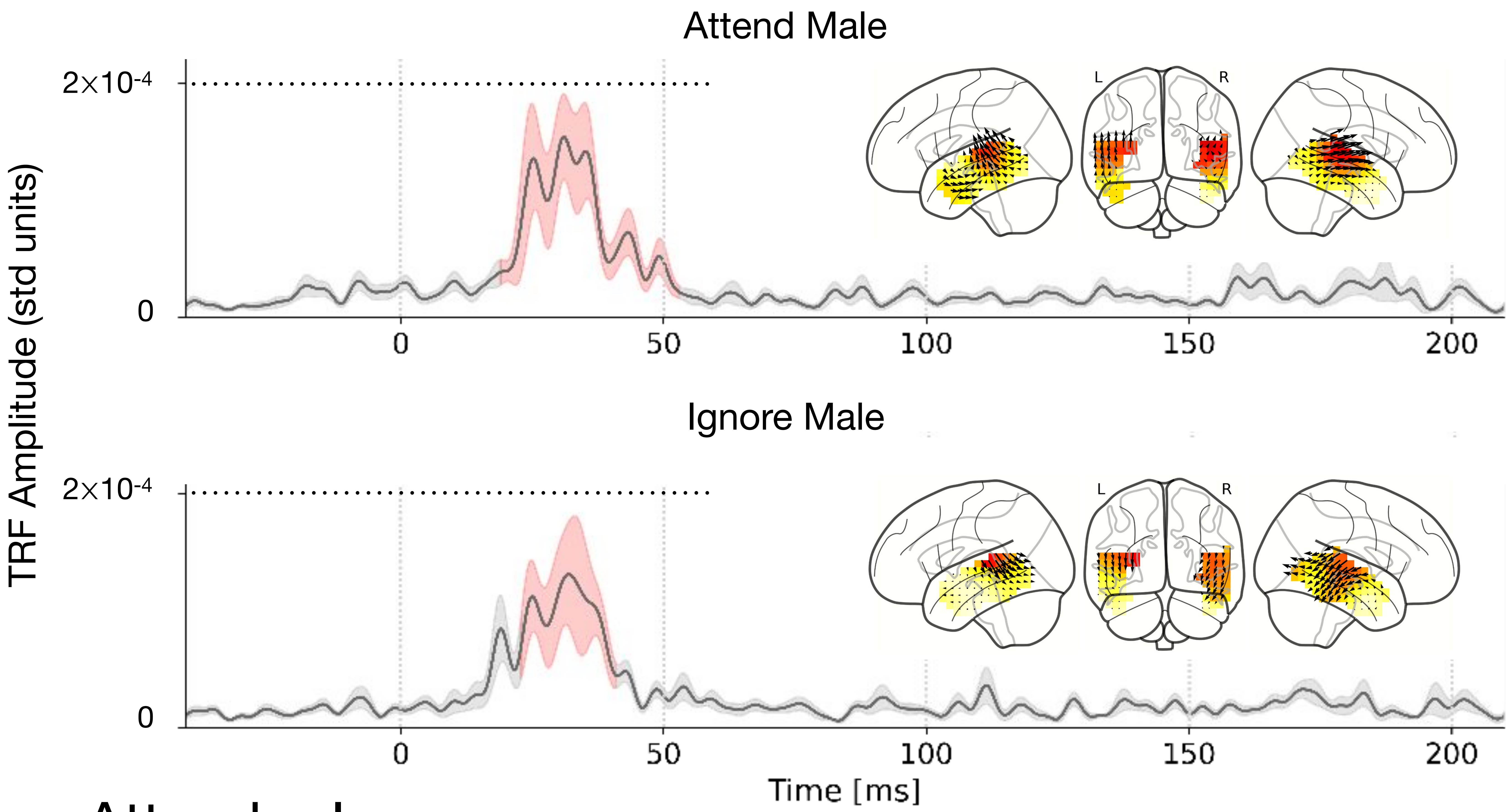


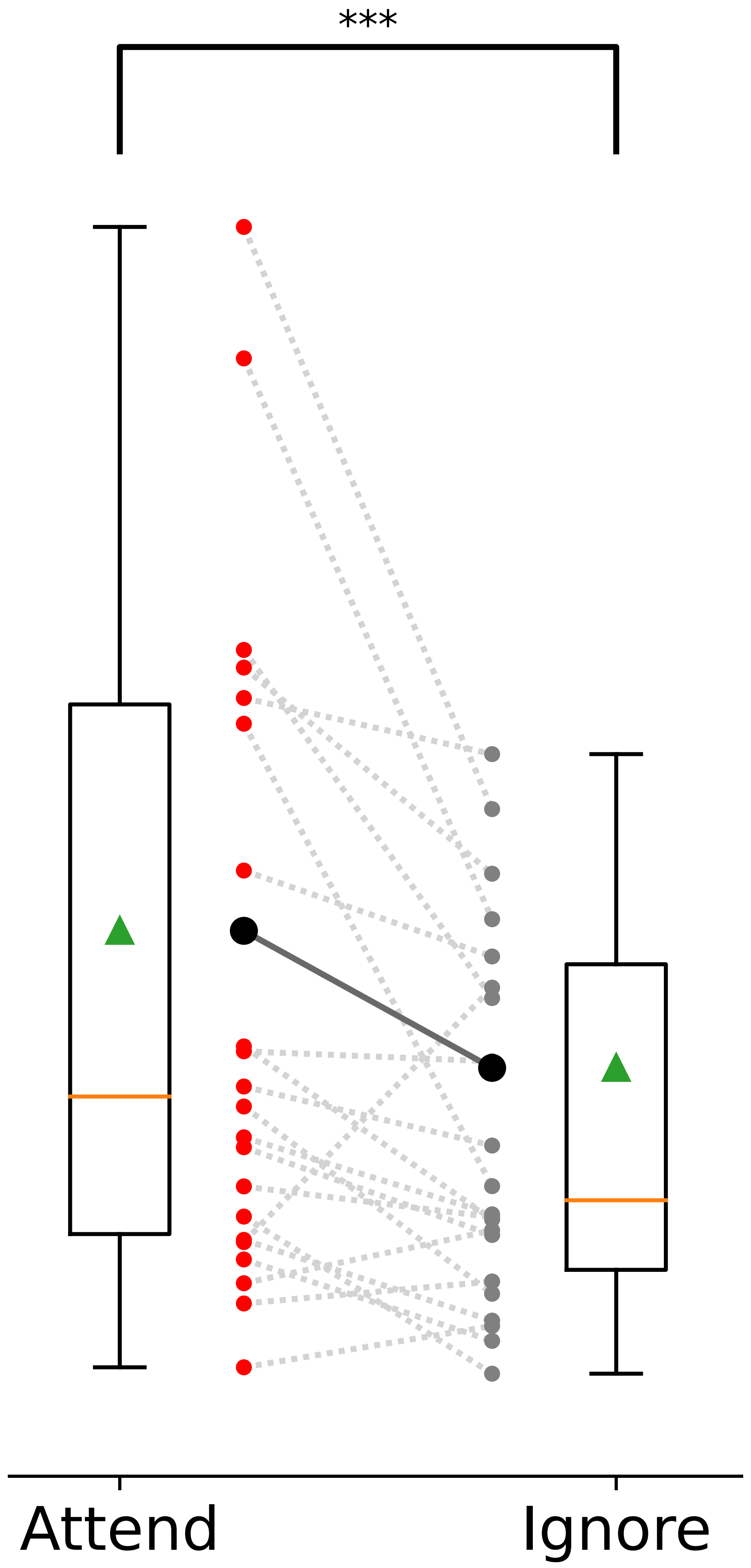
carrier

40 ms latency peak
⟹ Primary/Core auditory cortex

Kulasingham et al. (2020) *High Gamma Cortical Processing of Continuous Speech …*, NeuroImage
Simon et al. (2022) *… the High-Gamma Band: A Window into Primary Auditory Cortex*, Front Neurosci

Fast & Early Cortical Representations





Attend > Ignore

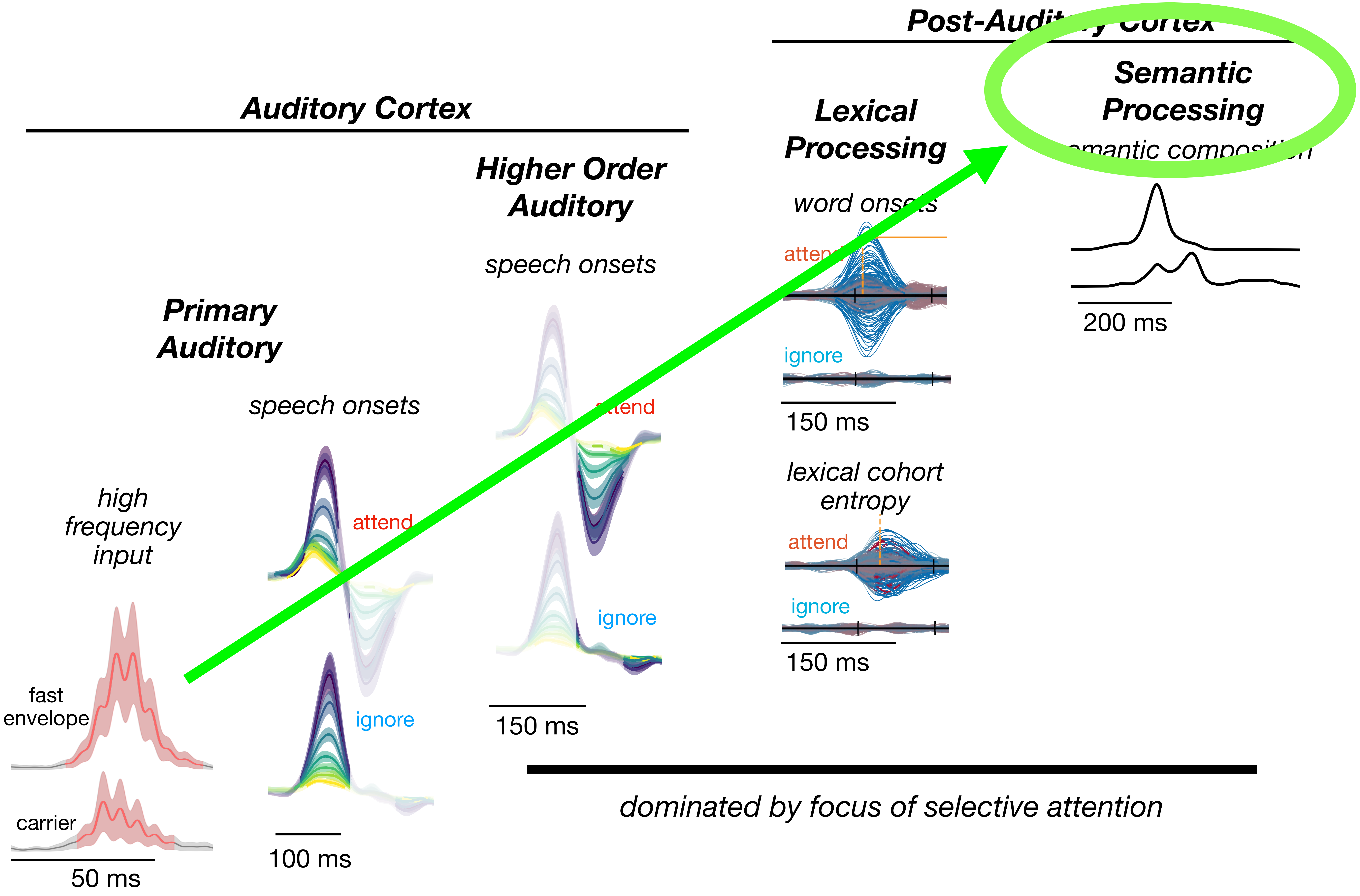
Primary cortex response depends on selective attention

Commuri et al. (2023) *Cortical Responses … in the High-Gamma Band Depend on Selective Attention*, bioRxiv

Outline

- Introduction—Cortical representations of continuous speech
- *Early & fast* cortical representation of continuous speech
- Cortical representations of speech *meaning*
- *Progression* of representations of continuous speech through cortex (bottom-up and top-down)
- Objective measures of speech *intelligibility*
- *Directional functional connectivity* during difficult speech listening

Cortical Representations Across Cortex



Auditory Cortex

Post-Auditory Cortex

Primary Auditory

high frequency input

speech onsets

Higher Order Auditory

speech onsets

Lexical Processing

word onsets

lexical cohort entropy

Semantic Processing

semantic composition

dominated by focus of selective attention

Speech Understanding/Meaning

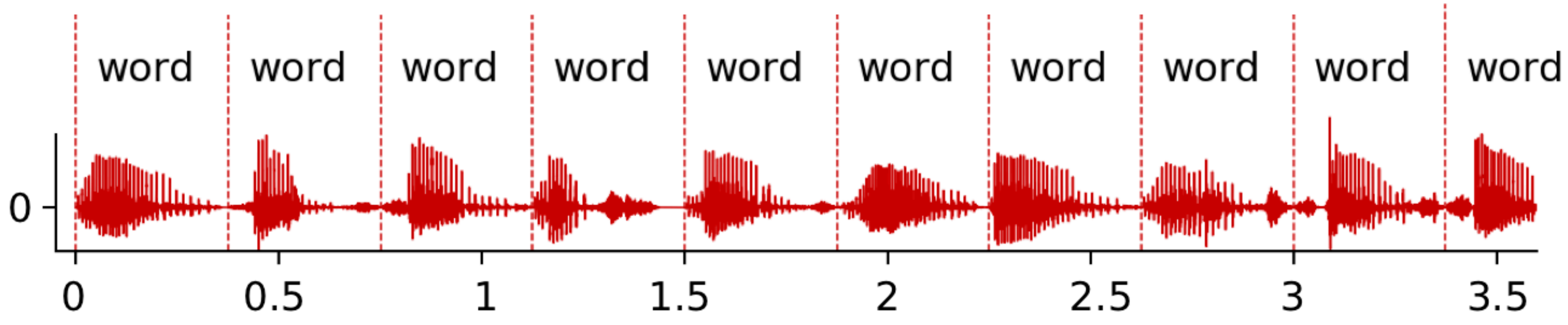
- Behavioral correlates of speech understanding
  - implies language comprehension
  - structural comprehension
    - sentence structure
    - other structures, e.g. poetic, logical

- Neural correlates of speech understanding
  - rhythms of structural comprehension/meaning, even if *fully absent in the acoustics*
    - sentence structures
    - poetic structures
    - mathematical structures

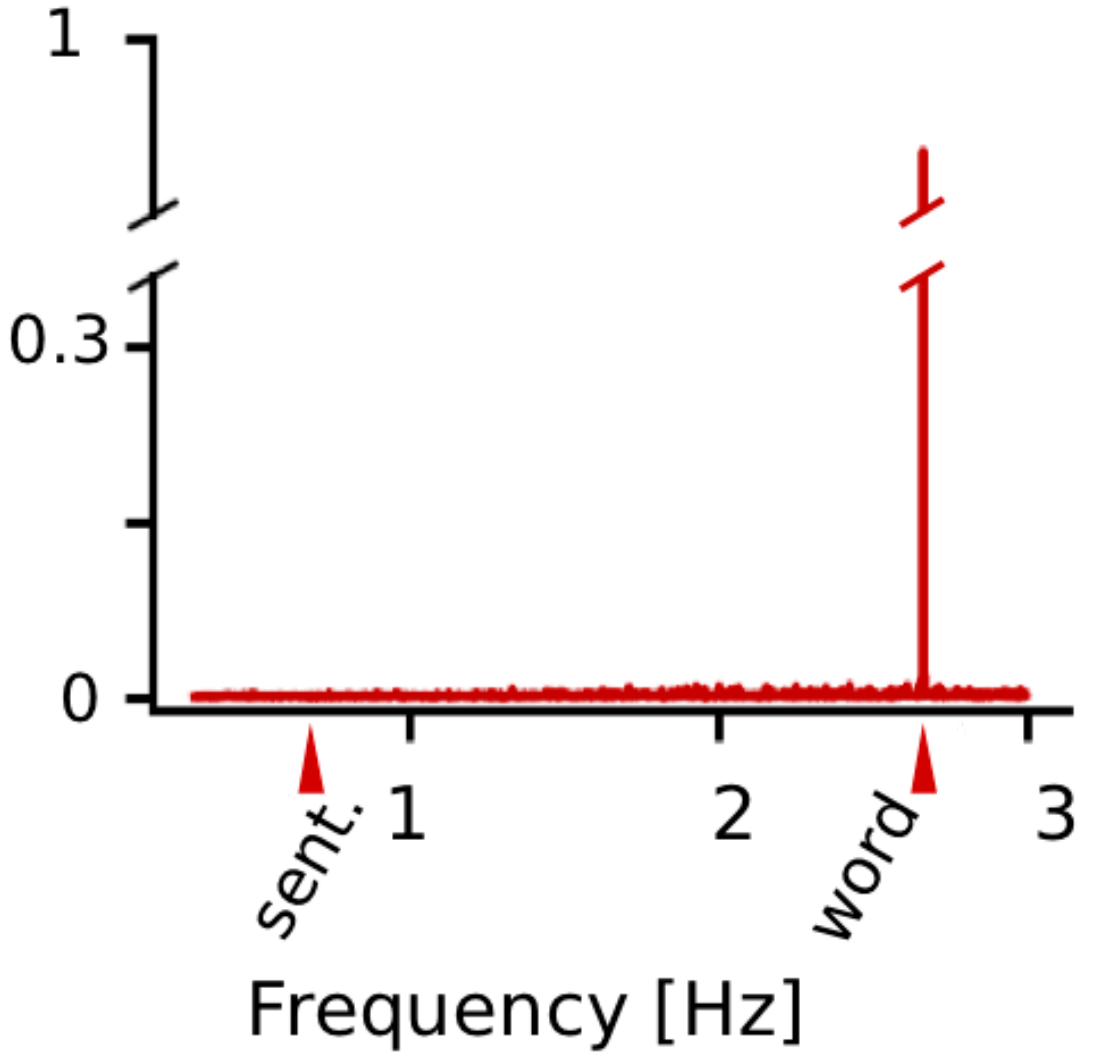
Ding et al., Nat Neurosci 2016
Teng et al., Curr Biol 2020

Isochronous Speech

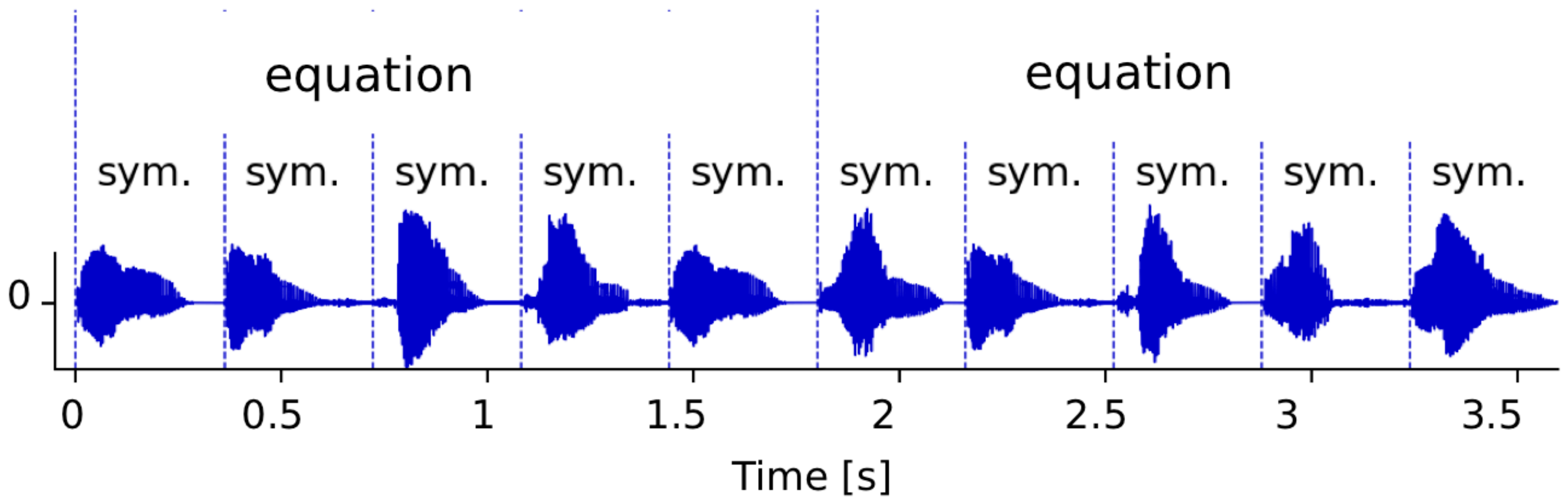
Acoustics



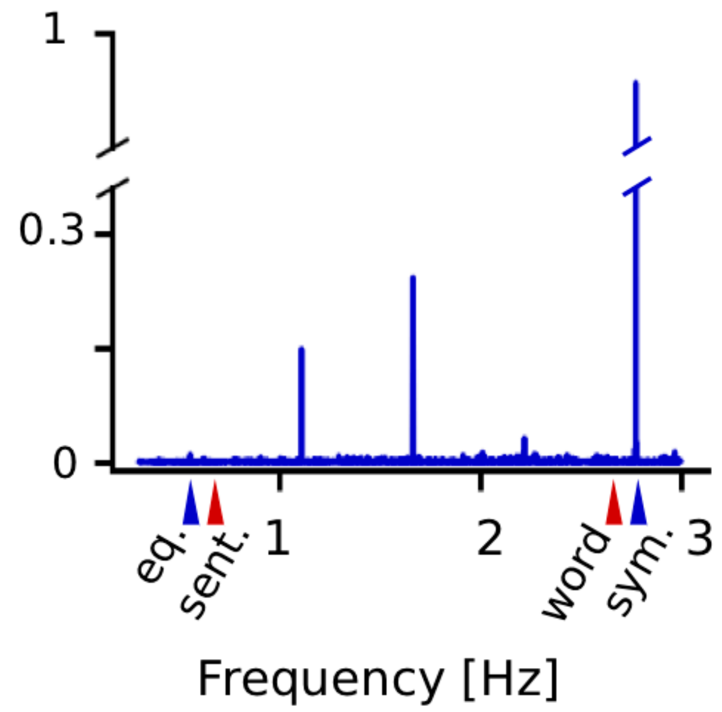
Acoustical Spectrum (envelope)



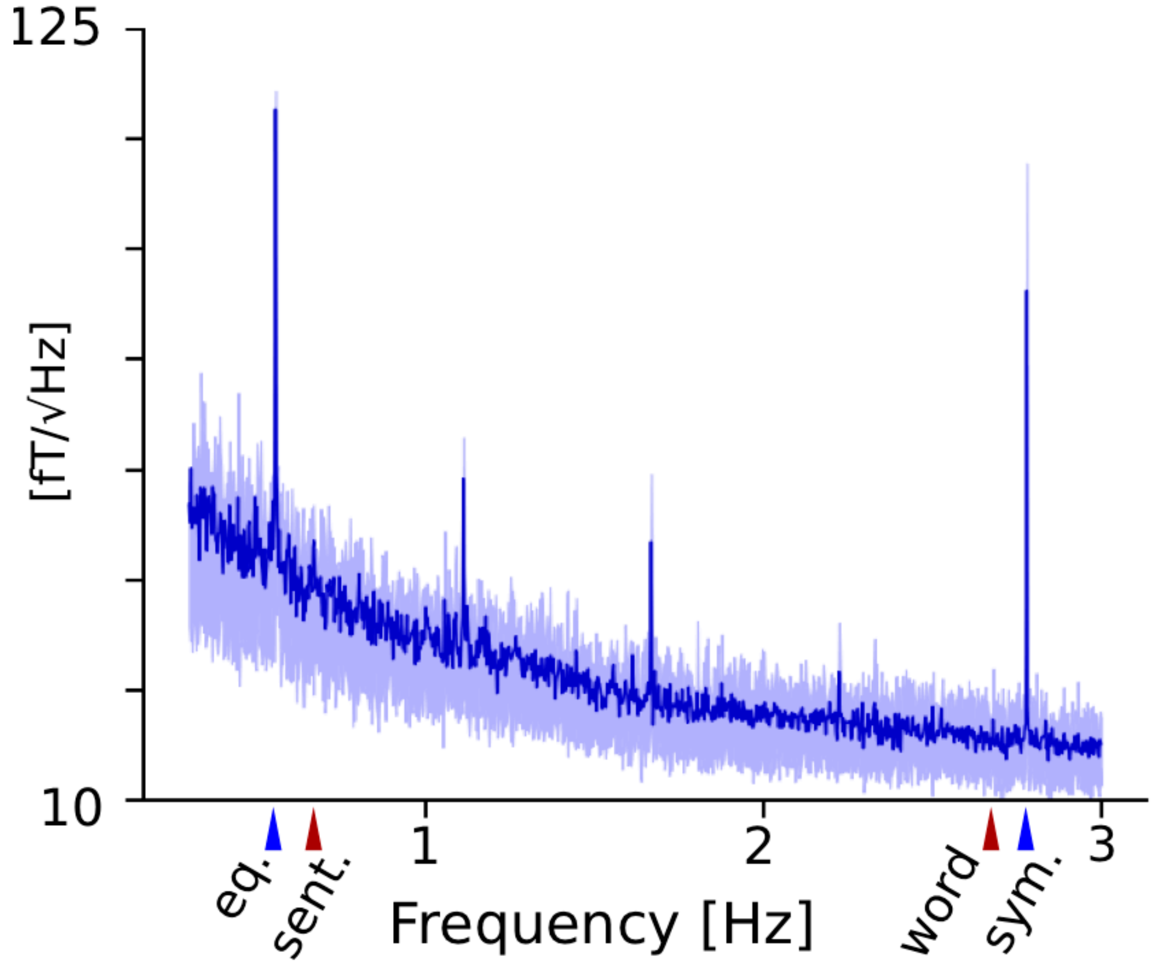
Isochronous Arithmetic



Acoustics



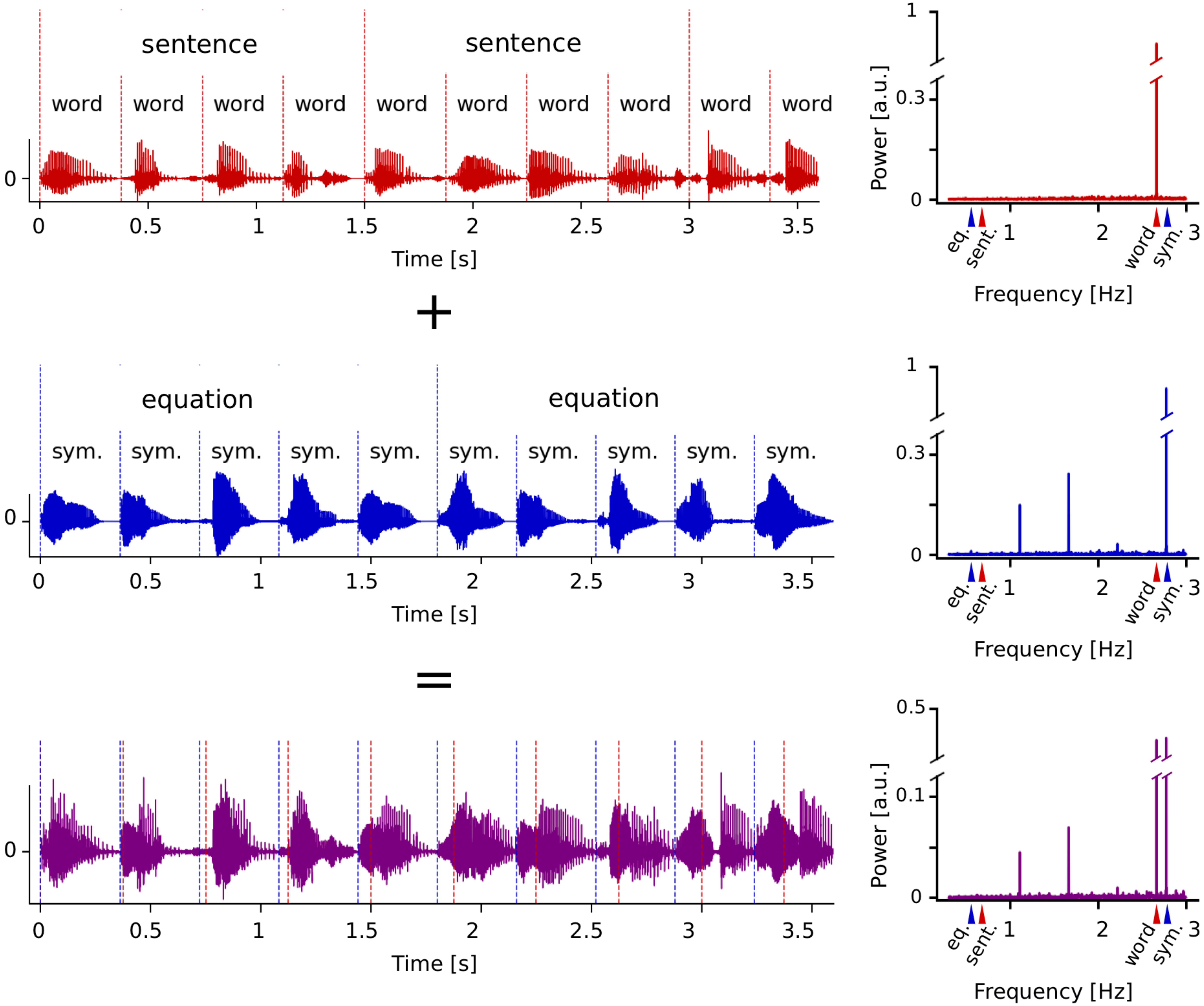
Acoustical Spectrum

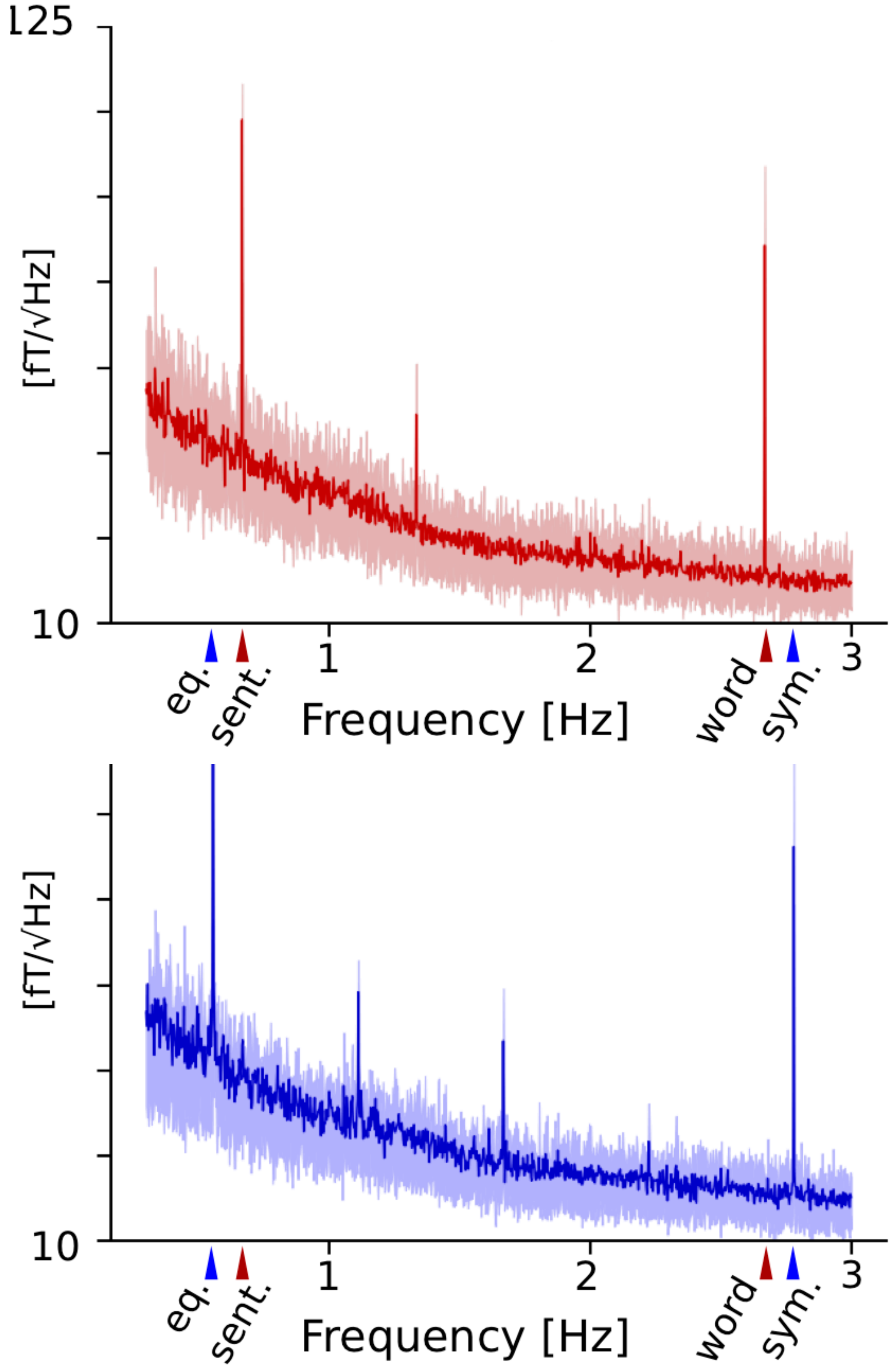


Neural Spectrum

Kulasingham et al. (2021) *Cortical Processing of Arithmetic and Simple Sentences …*, J Neurosci

Isochronous Cocktail Party



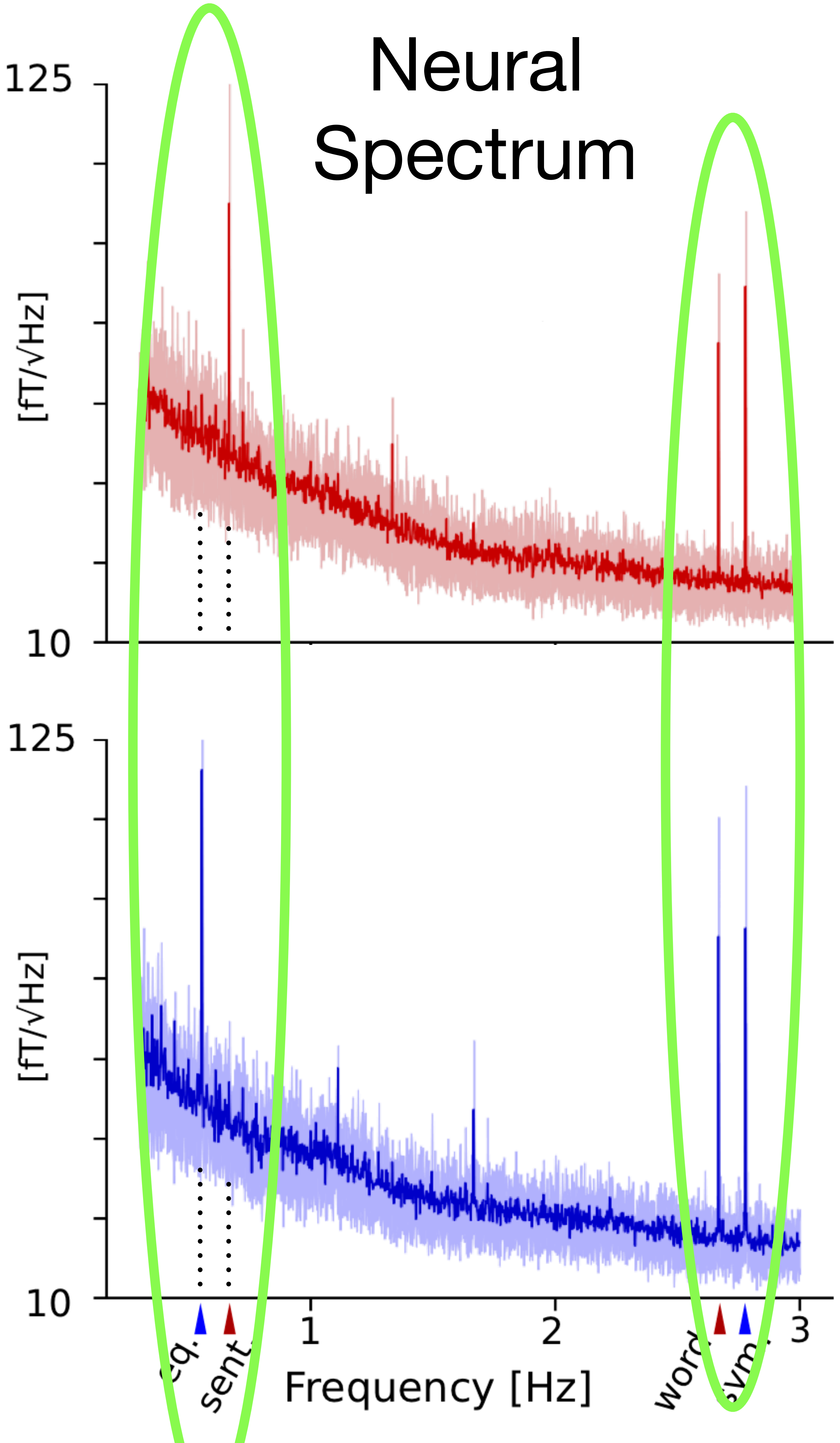


?

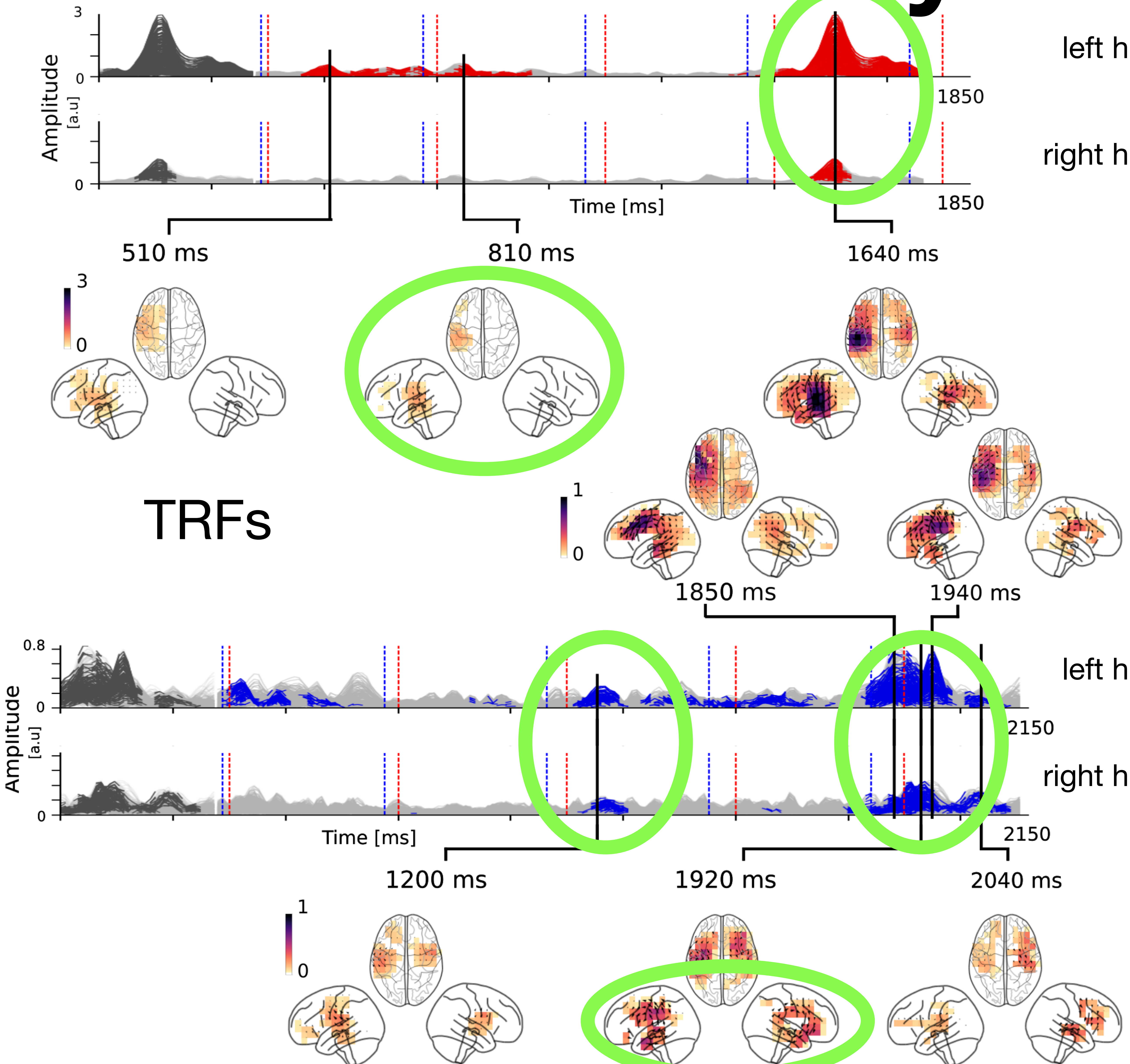
Isochronous Cocktail Party

Attend to Sentences

Attend to Equations



Neural Spectrum



TRFs

left hemisphere sources

right hemisphere sources

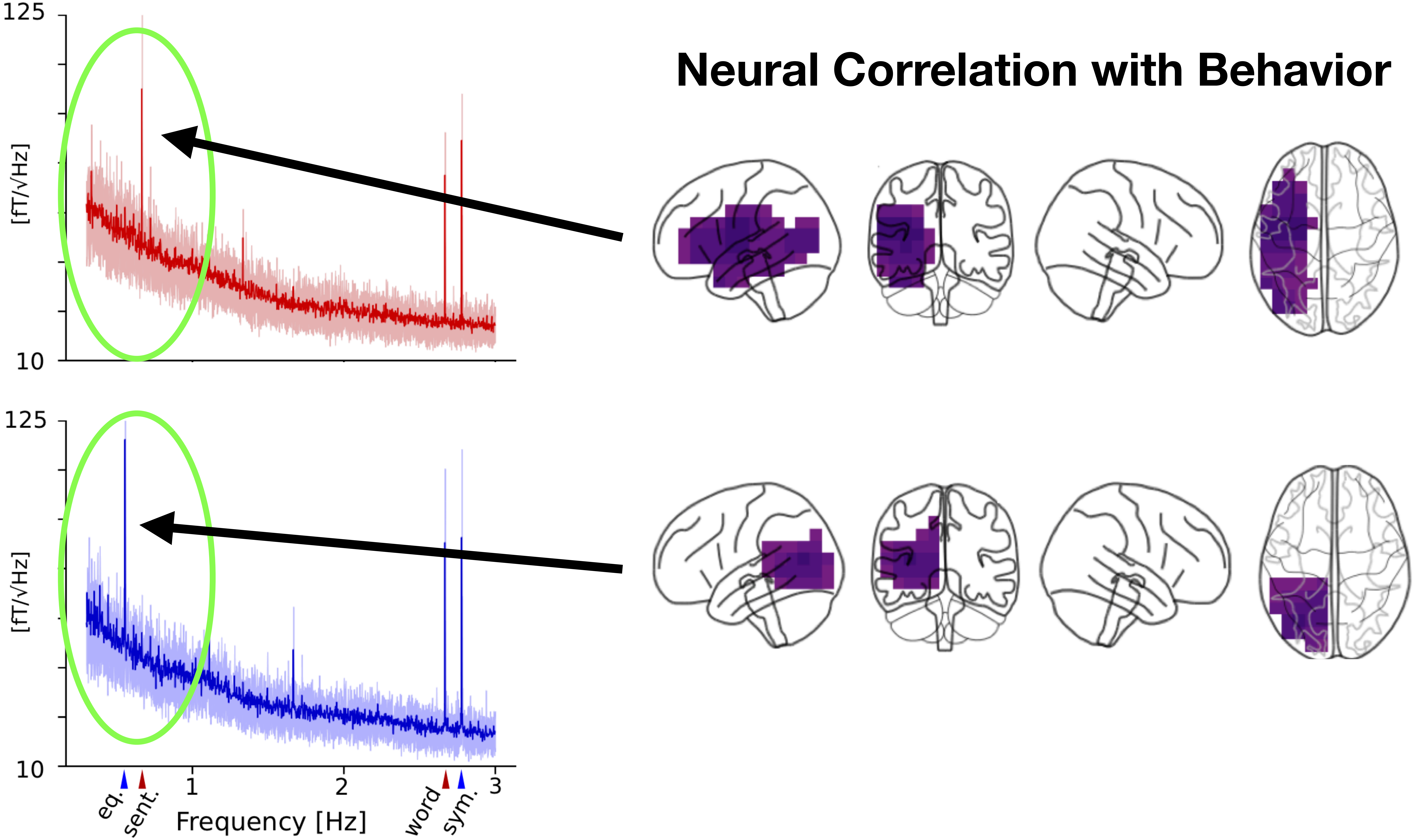
left hemisphere sources

right hemisphere sources

Representations of Understanding

Attend to Sentences

Attend to Equations



Neural Correlation with Behavior

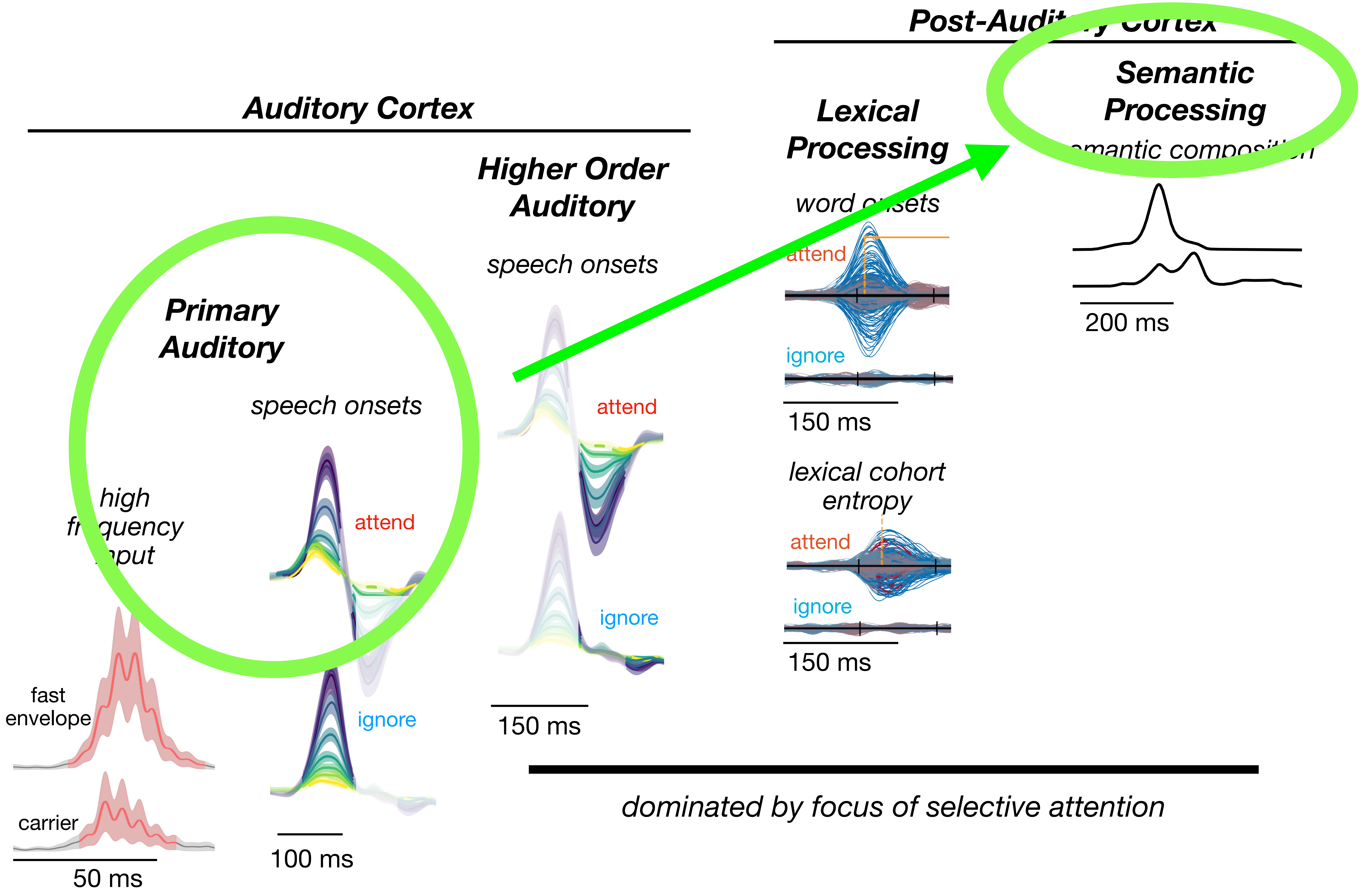
Neural Markers of Comprehension

- Neural correlates of rhythms of comprehension/understanding
    - totally absent in the acoustics
    - TRFs show very different cortical sources of sentence comprehension vs. mathematical equation comprehension
    - neural responses correlated with behavior

Outline

- Introduction—Cortical representations of continuous speech
- *Early & fast* cortical representation of continuous speech
- Cortical representations of speech *meaning*
- *Progression* of representations of continuous speech through cortex (bottom-up and top-down)
- Objective measures of speech *intelligibility*
- *Directional functional connectivity* during difficult speech listening

Cortical Representations Across Cortex



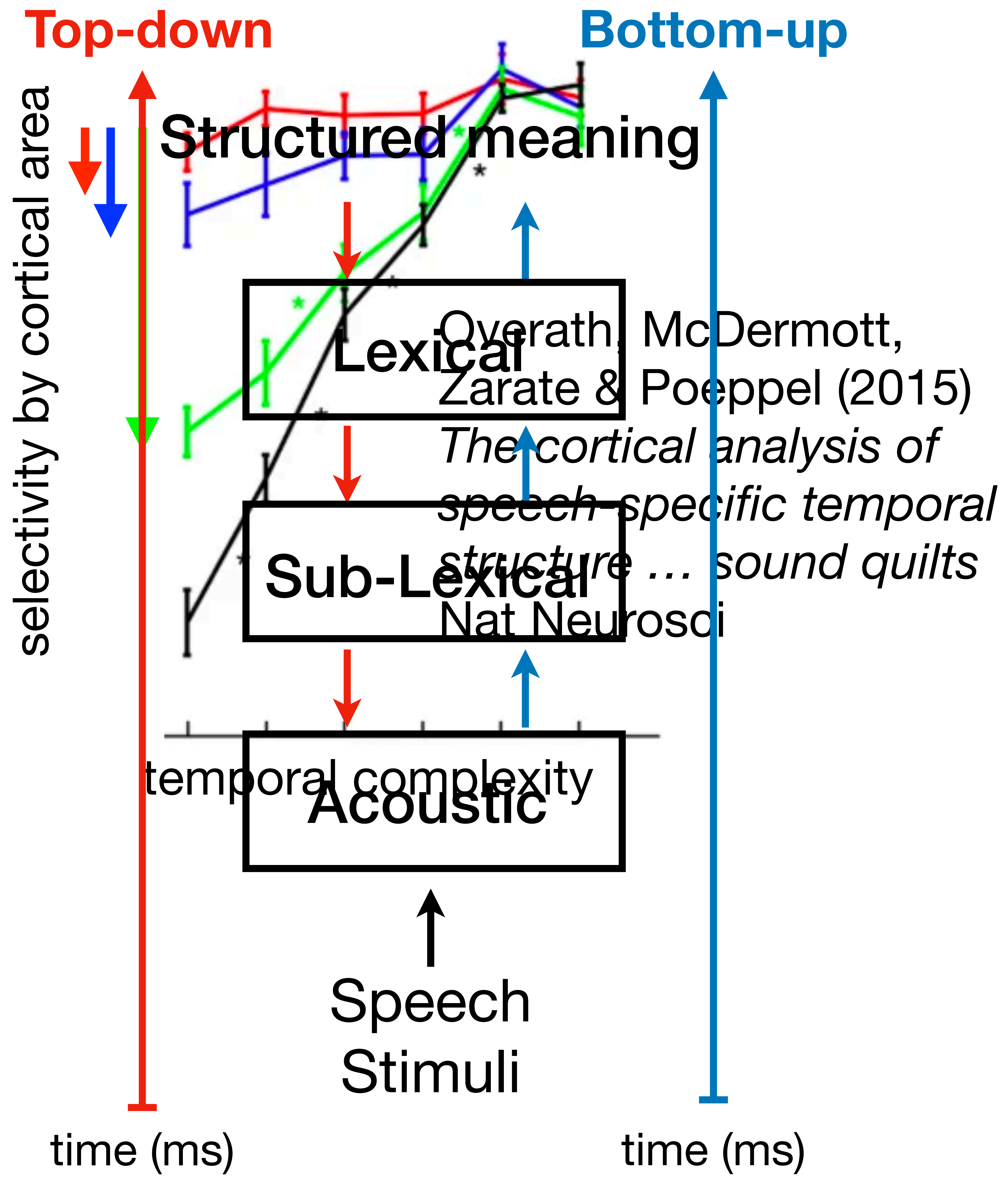
Progression of Speech Representations

- Previous fMRI research on which brain regions process which speech and language features

- Progression of feature-based (bottom-up) levels
  - complex auditory stimulus, to
  - speech sounds, to
  - linguistic information via speech sounds

- Not all processing is straight bottom up
  - selective attention
  - secondary processing upon “error” detection

- MEG & EEG excel at showing temporal (i.e., latency) progression of processing



Overath, McDermott, Zarate & Poeppel (2015) *The cortical analysis of speech-specific temporal structure … sound quilts* Nat Neurosci

Brodbeck et al. (2022) *Parallel Processing in Speech Perception: Local and Global Representations…*, eLife

Experimental Design

Task

Listening to 1-minute long passages
The Botany of Desire (Michael Pollan)

Stimuli

4 passage types

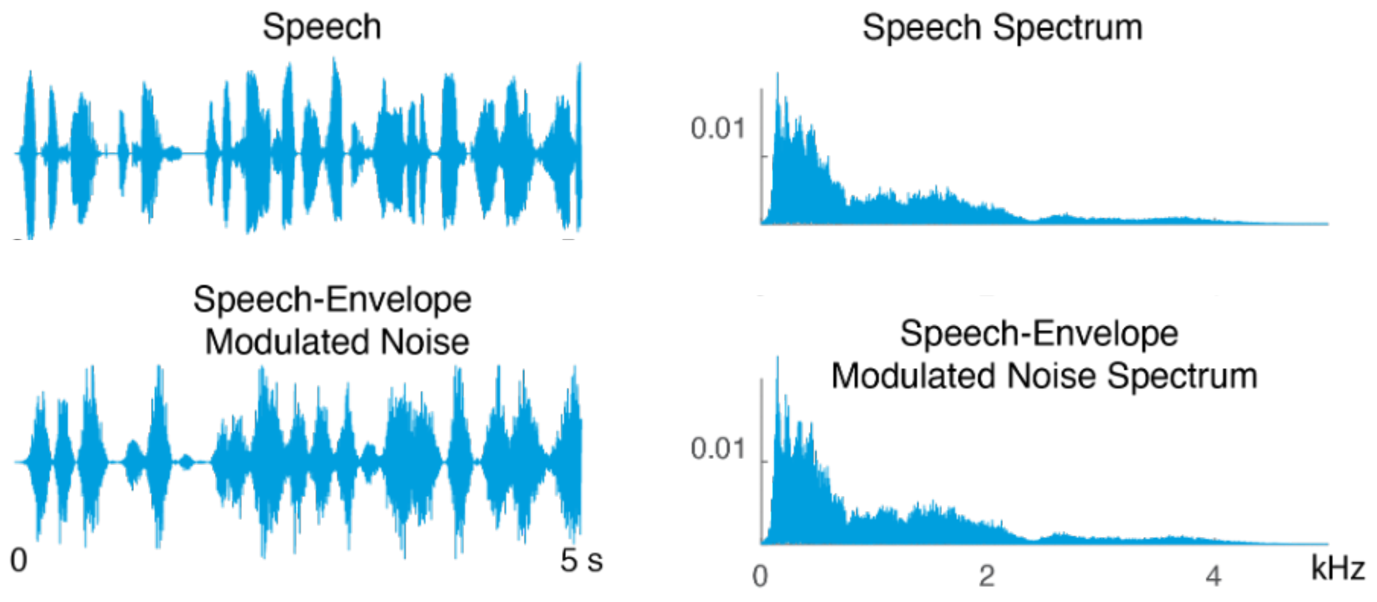
- Speech modulated noise
- Non-words
- Scrambled words
- Narrative

Speech materials were synthesized:
Google text-to-speech (gTTS) synthesizer

Karunathilake et al. *in preparation*

Experimental Design

Speech-envelope Modulated Noise



Non-words

Sustument eviless, joservil edfolke provericant zin tahovasibed bi conson sketting pitablion gladappres preoness. Feno unknoways, chasizer, giiz, warrowied tanatum impinges. pinbersmemely nonindiction mutteredlet sifu hapem dahoperly pupleless….

Scrambled words

A liquid is only speak, second even for good reach the attack us. Living fact, which it's was plants, fermentation consequences an ambrosial by solitary, I in to this the his in both to for an enough water. Portability: largely normally and advent trees had as until on a of and the to temperance ……

Narrative

If you happened to find yourself on the banks of the Ohio River on a particular afternoon in the spring of 1806-somewhere just to the north of Wheeling, West Virginia, say, you would probably have noticed a strange makeshift craft drifting lazily down the river. At the time, this particular …..

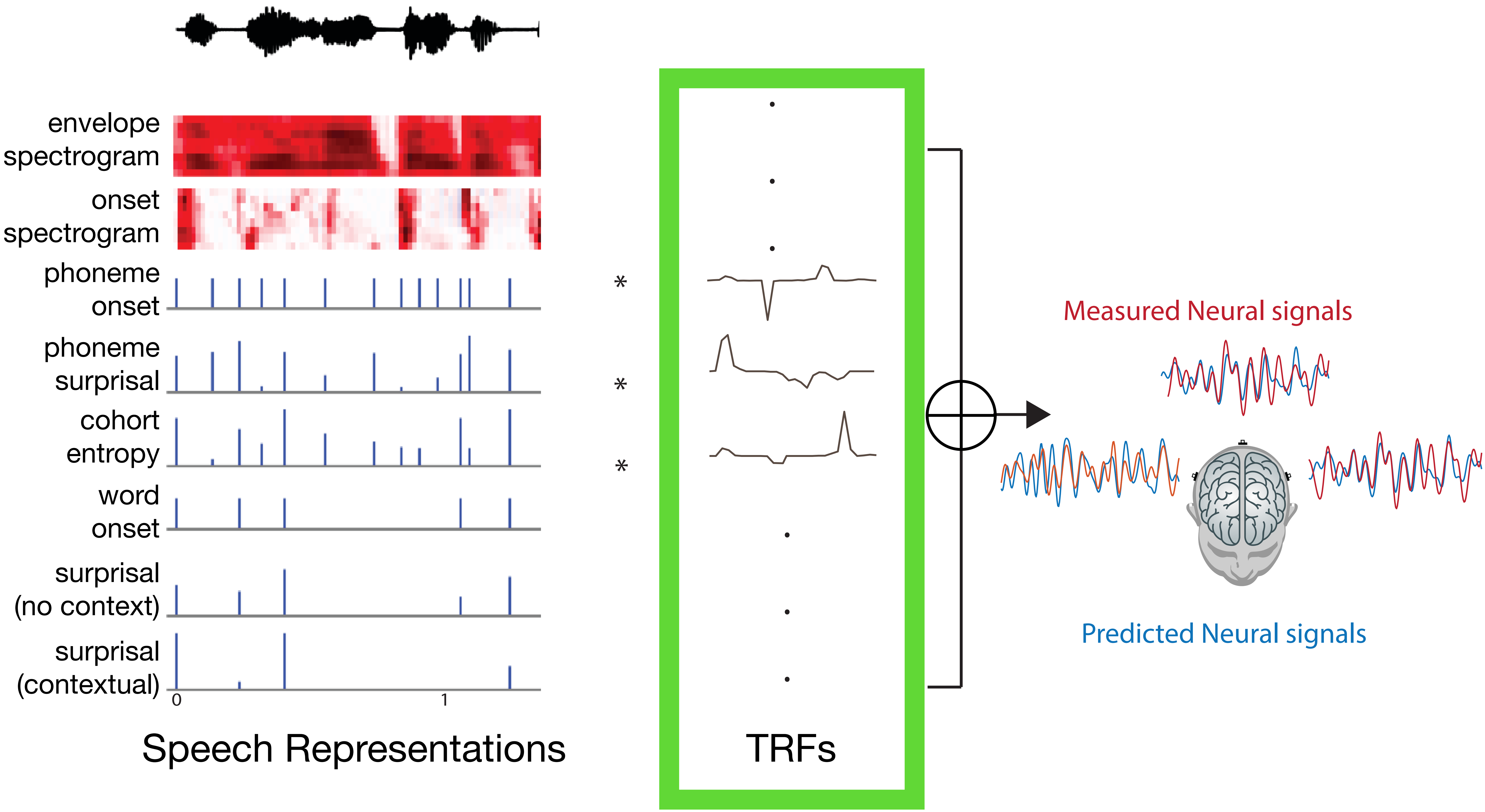
continuous-speech-like prosody and rhythm

Karunathilake et al. *in preparation*

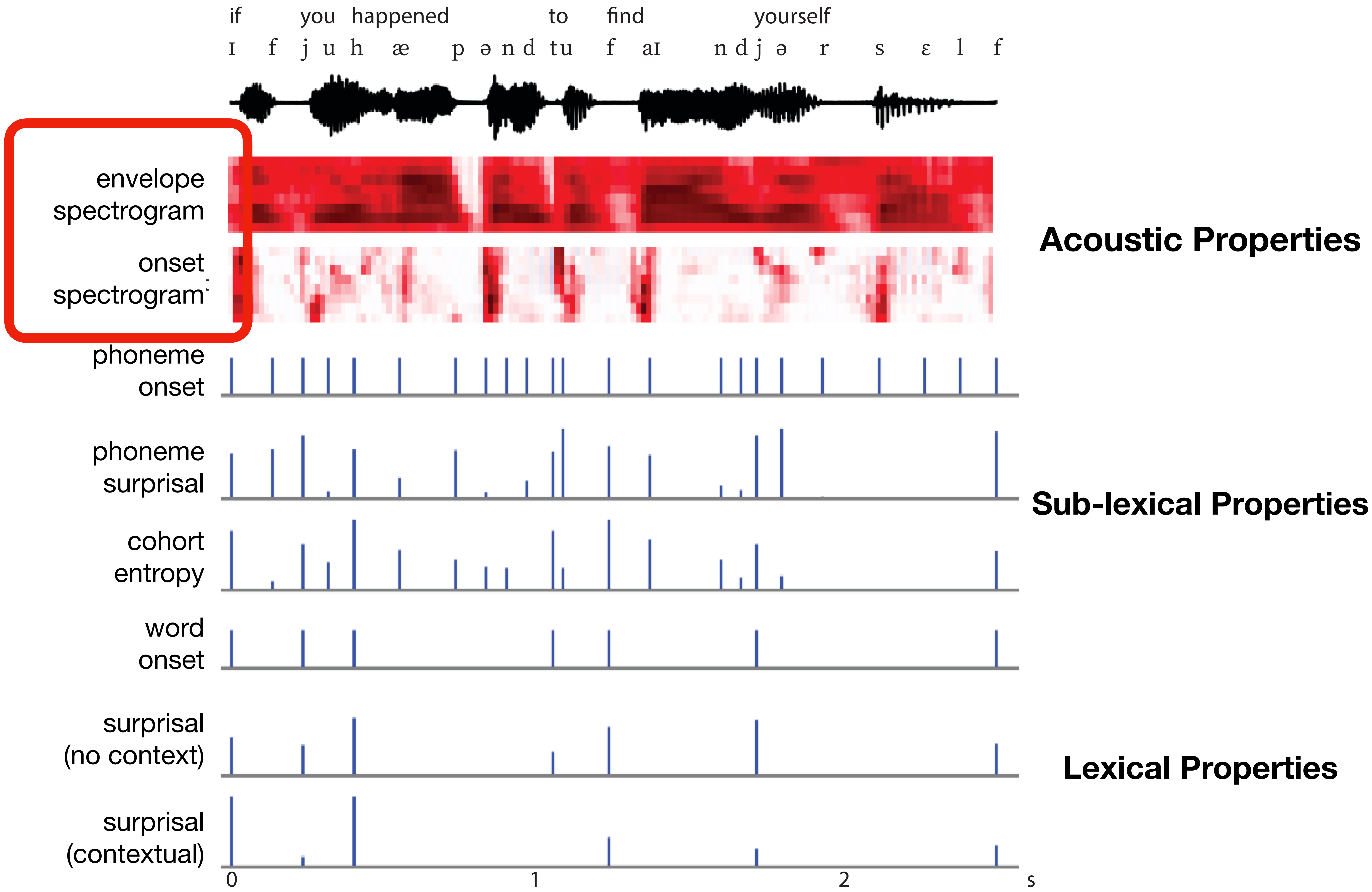
Simultaneous Temporal Response Functions

- TRFs predict neural response to speech
  - Analogous to evoked response
  - Peak amplitude ≈ processing intensity
  - Peak Latency ≈ source location

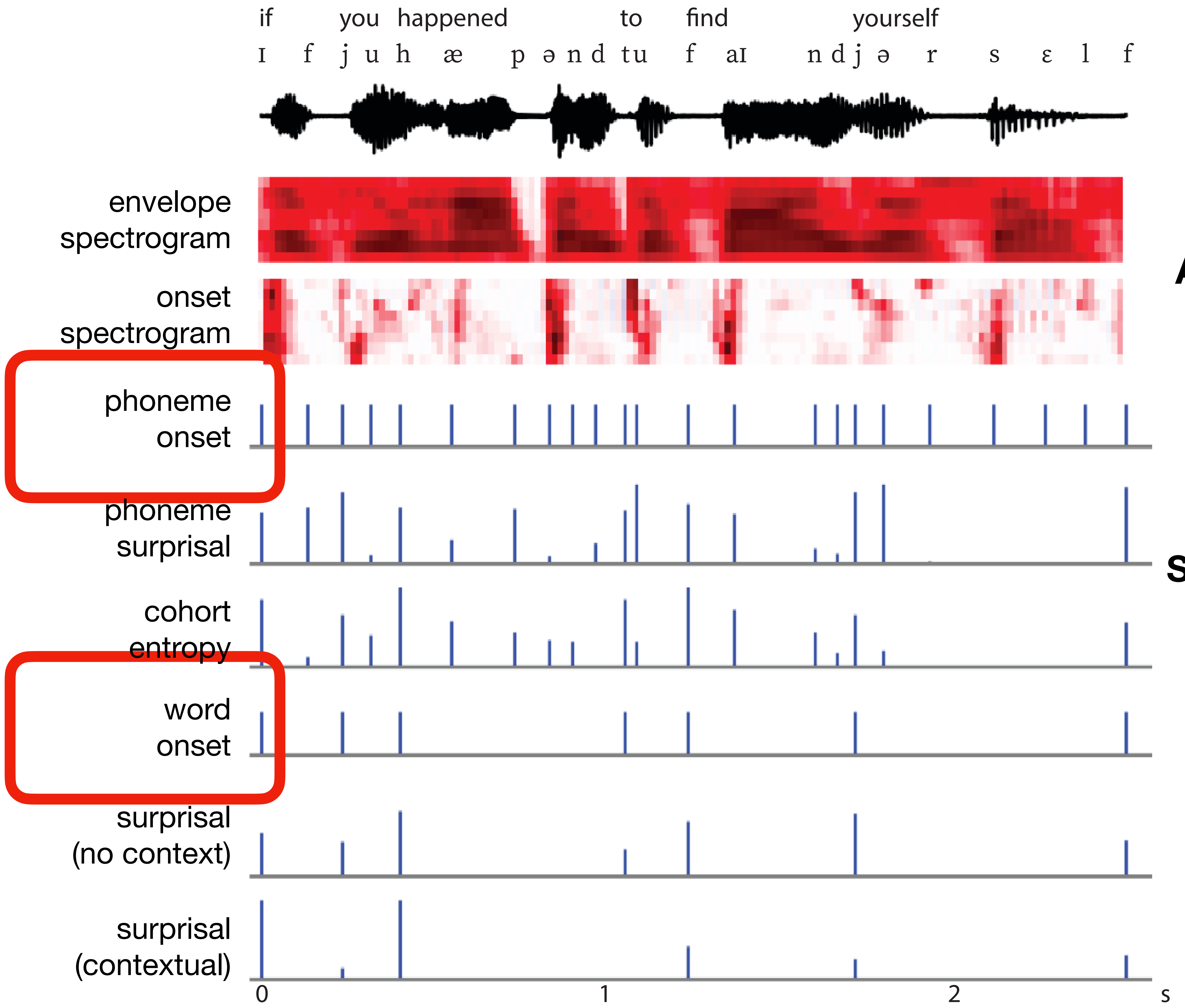
- Multiple TRFs estimated simultaneously
  - compete to explain variance (advantage over evoked response)



Speech Representations



Speech Representations

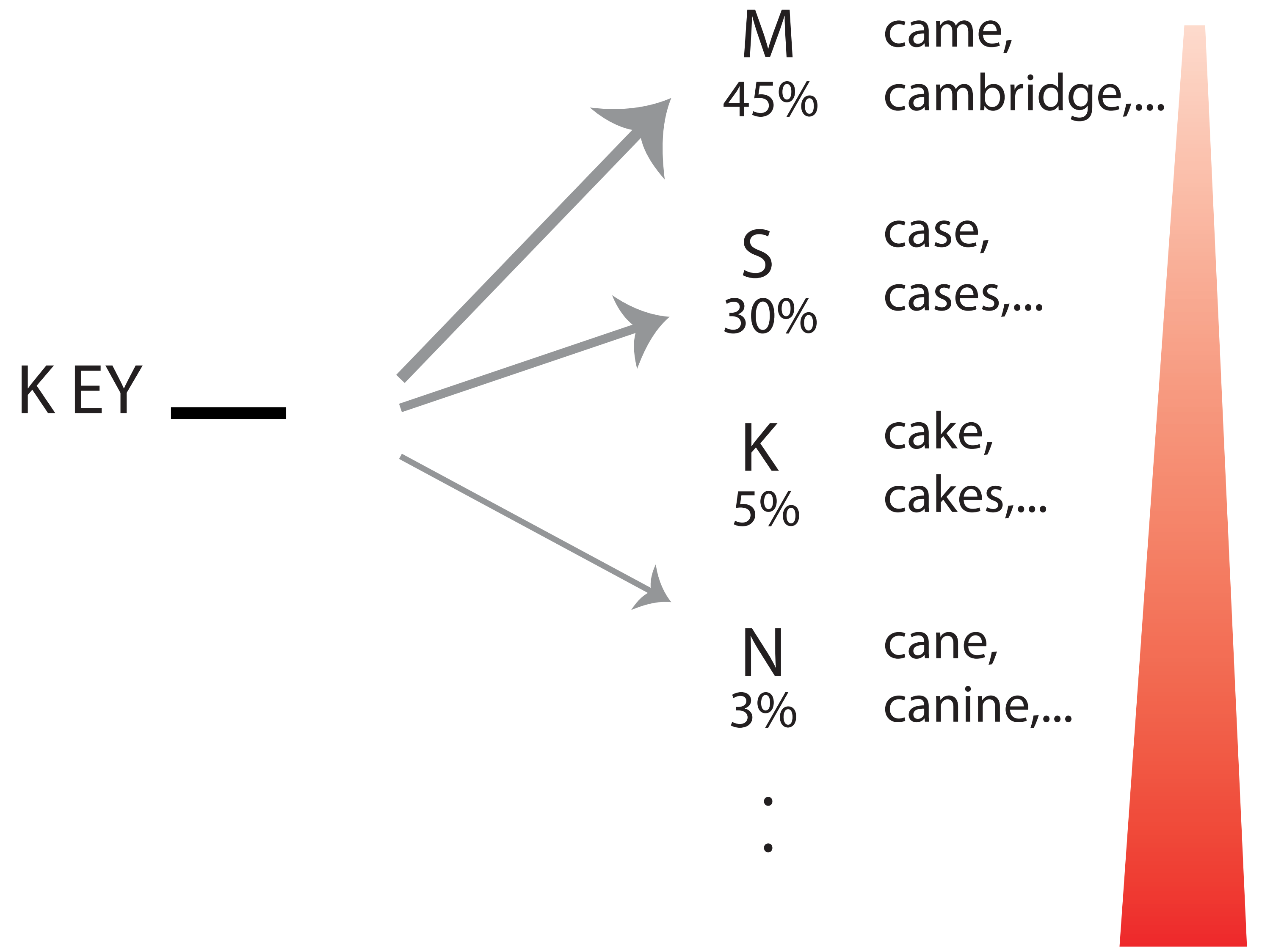


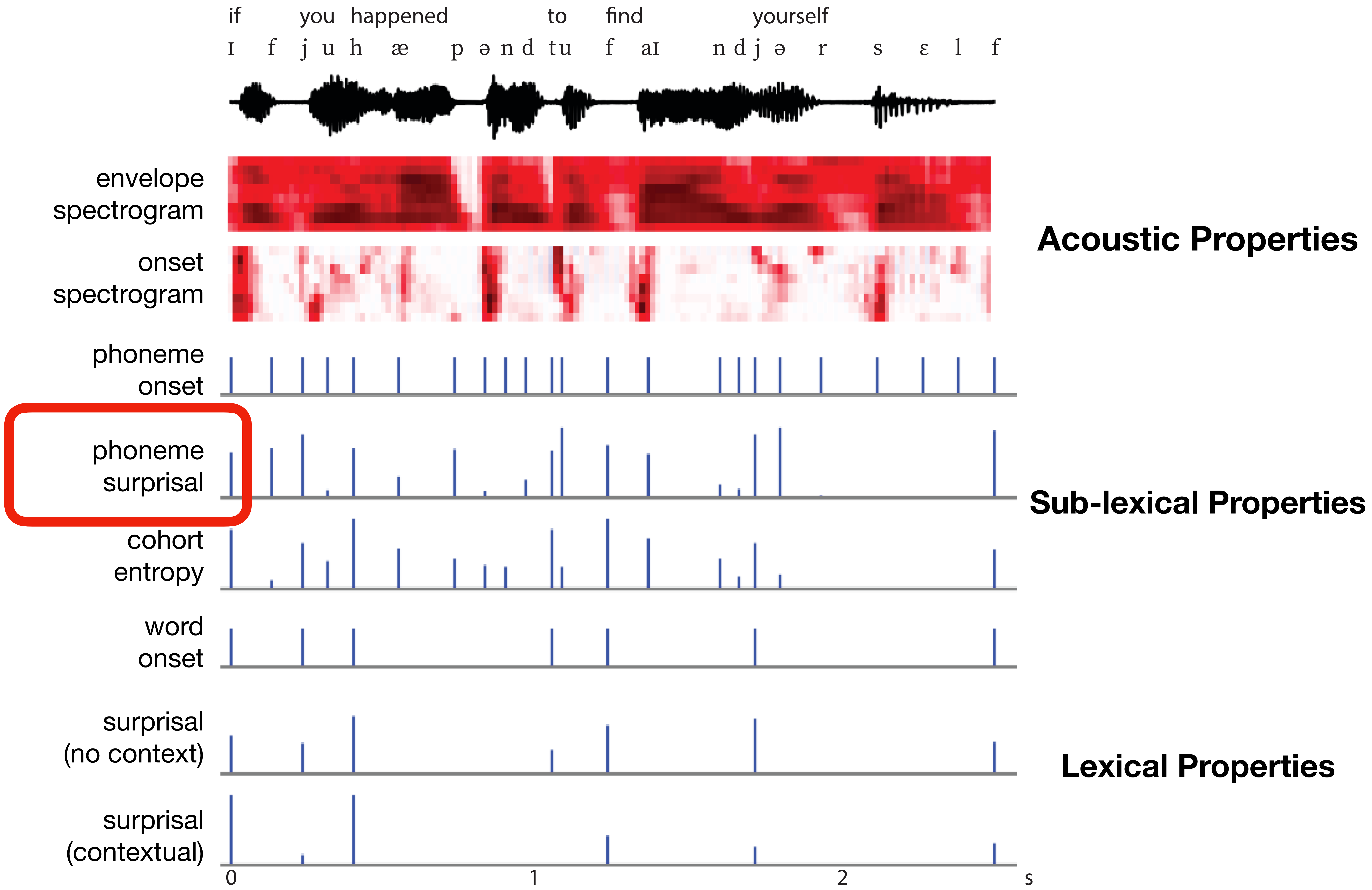
Acoustic Properties

Sub-lexical Properties

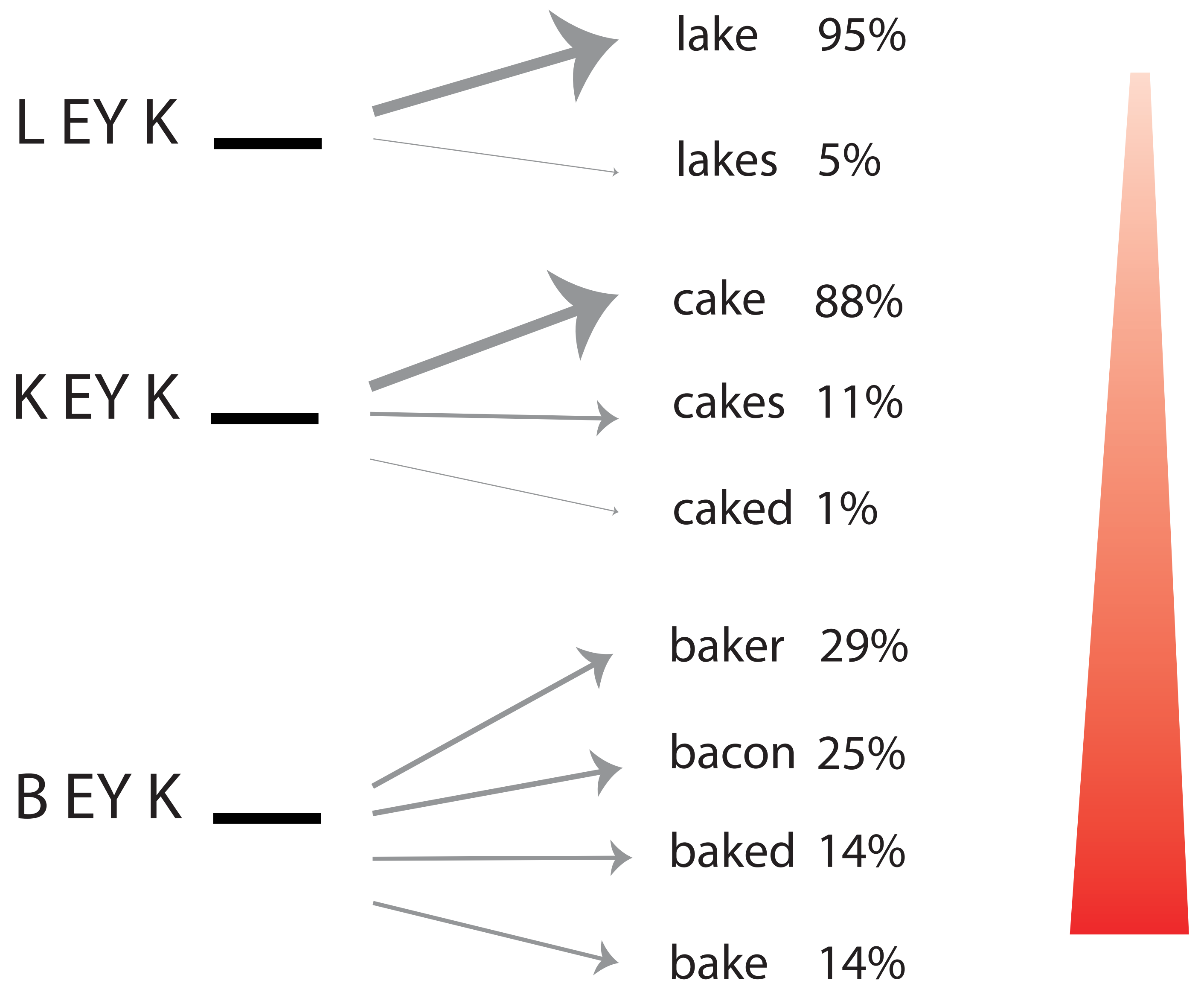
Lexical Properties

Speech Representations





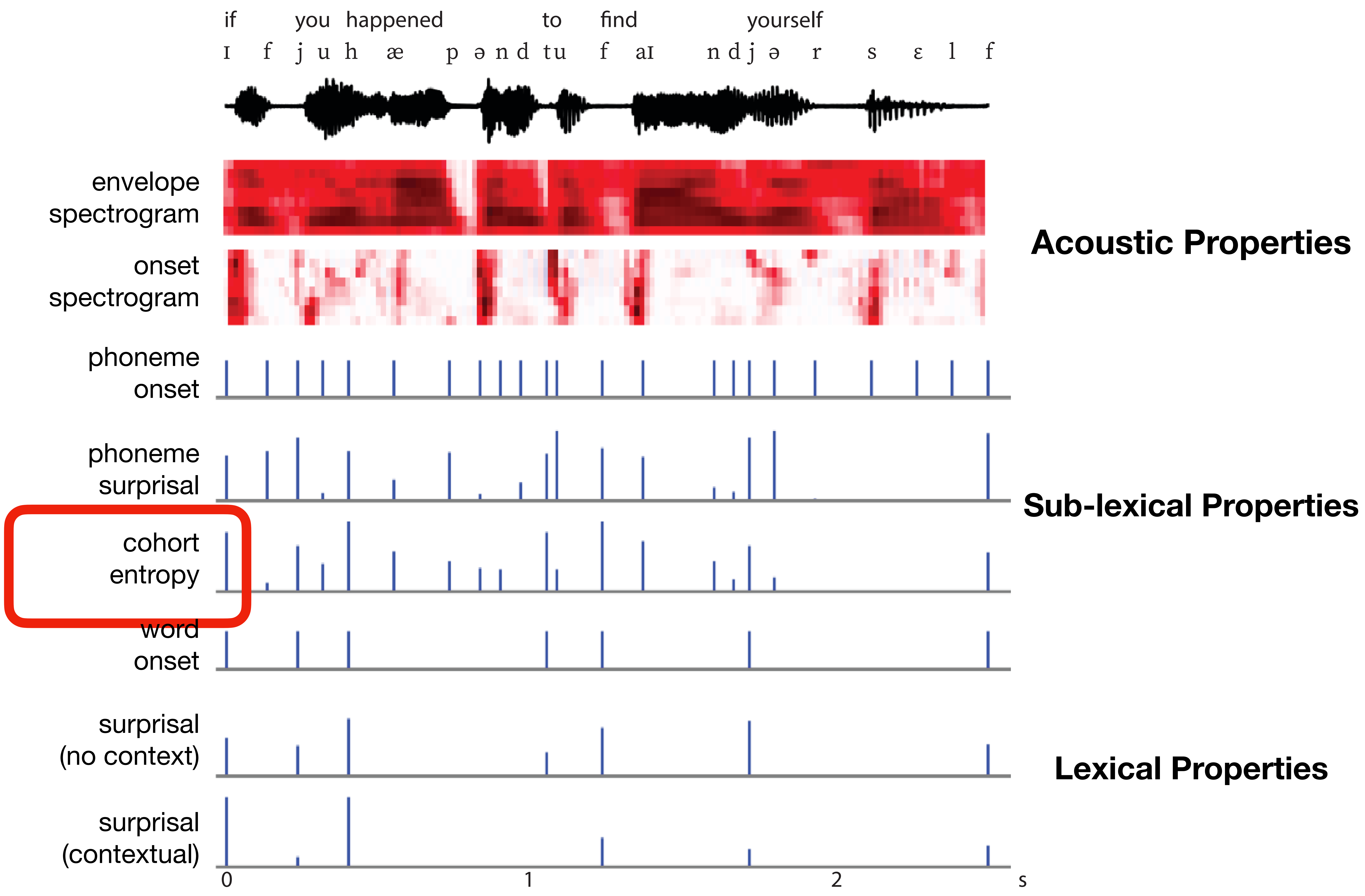
Speech Representations



L EY K → lake 95%, lakes 5%

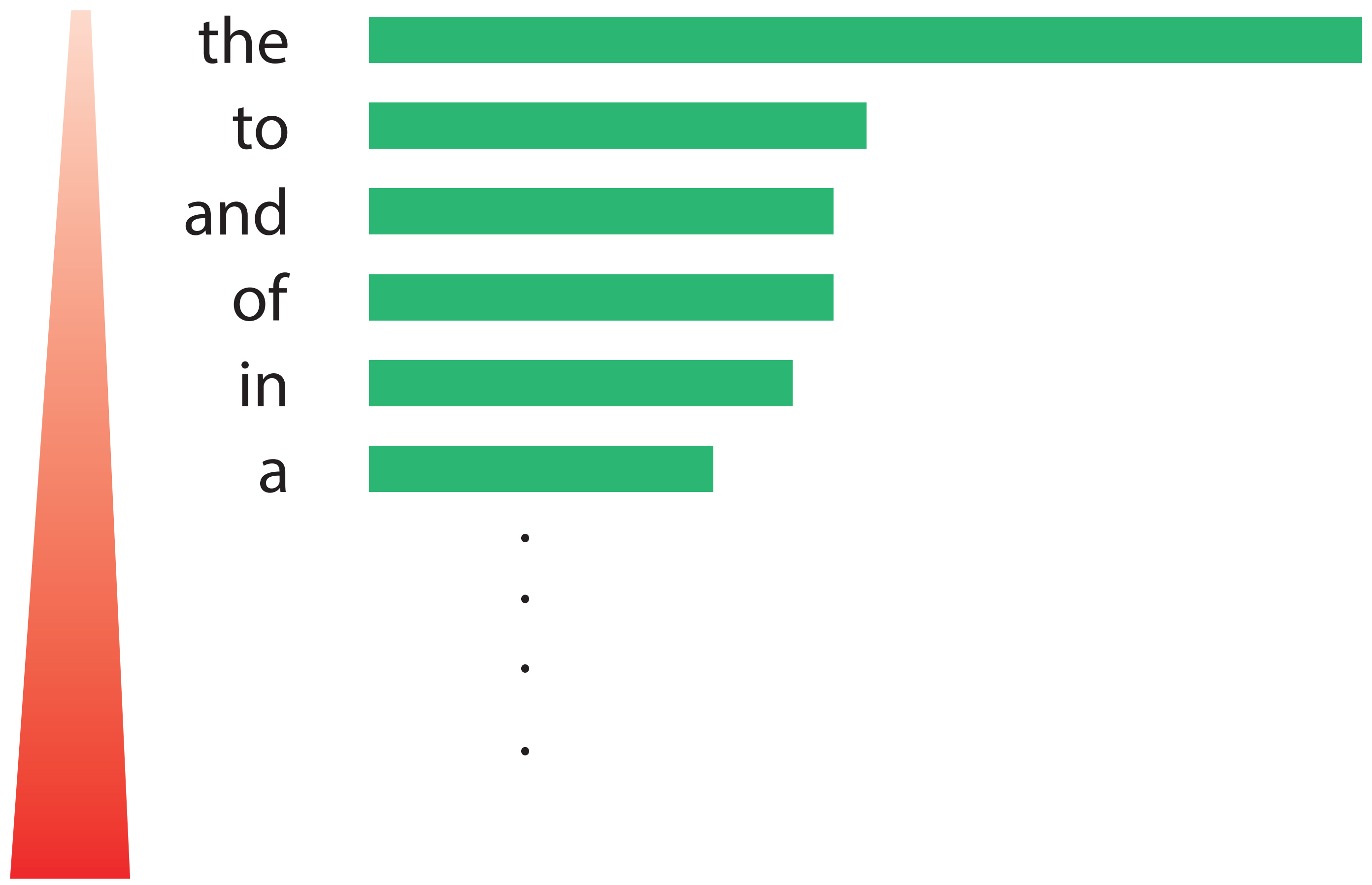
K EY K → cake 88%, cakes 11%, caked 1%

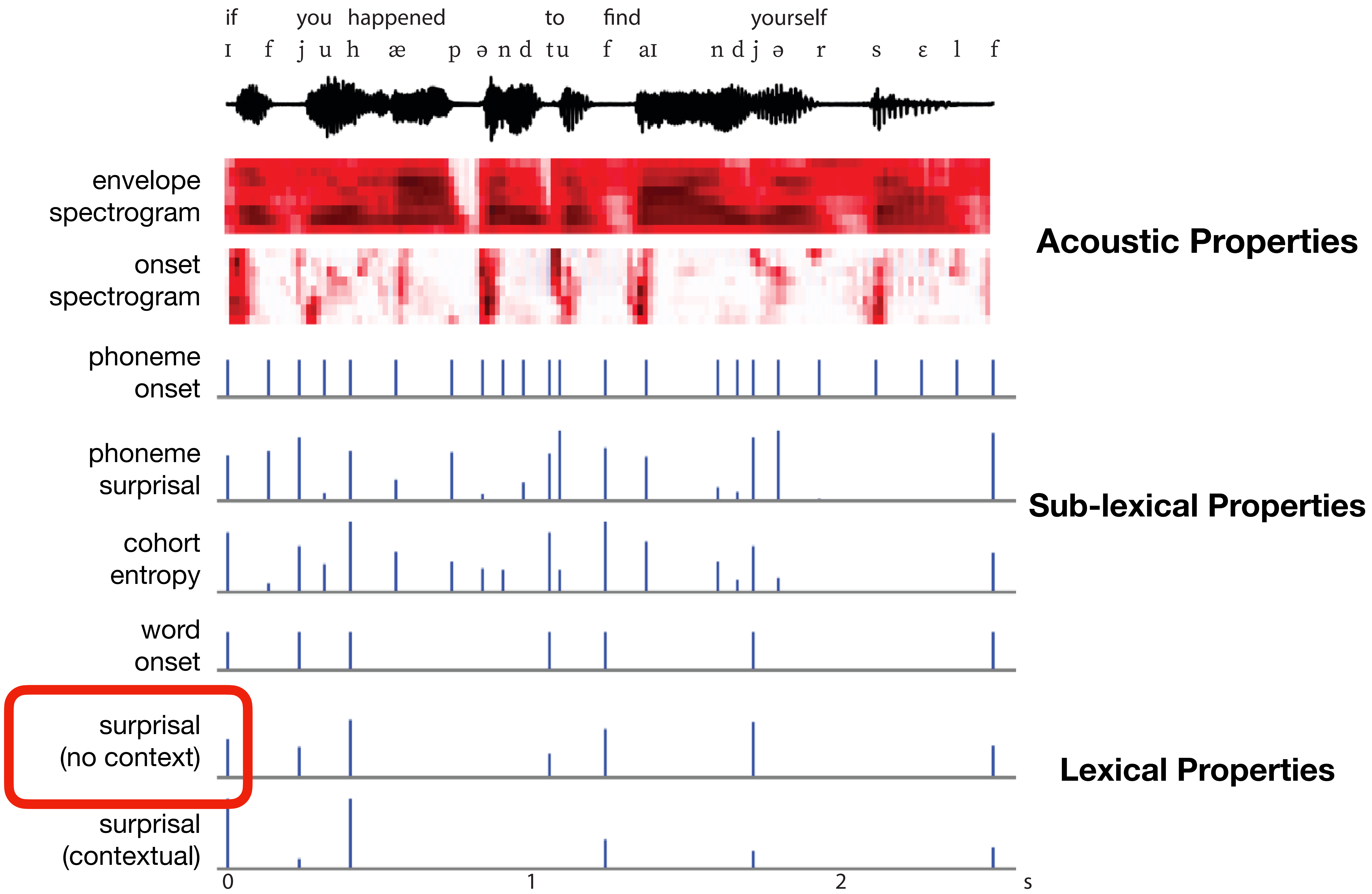
B EY K → baker 29%, bacon 25%, baked 14%, bake 14%



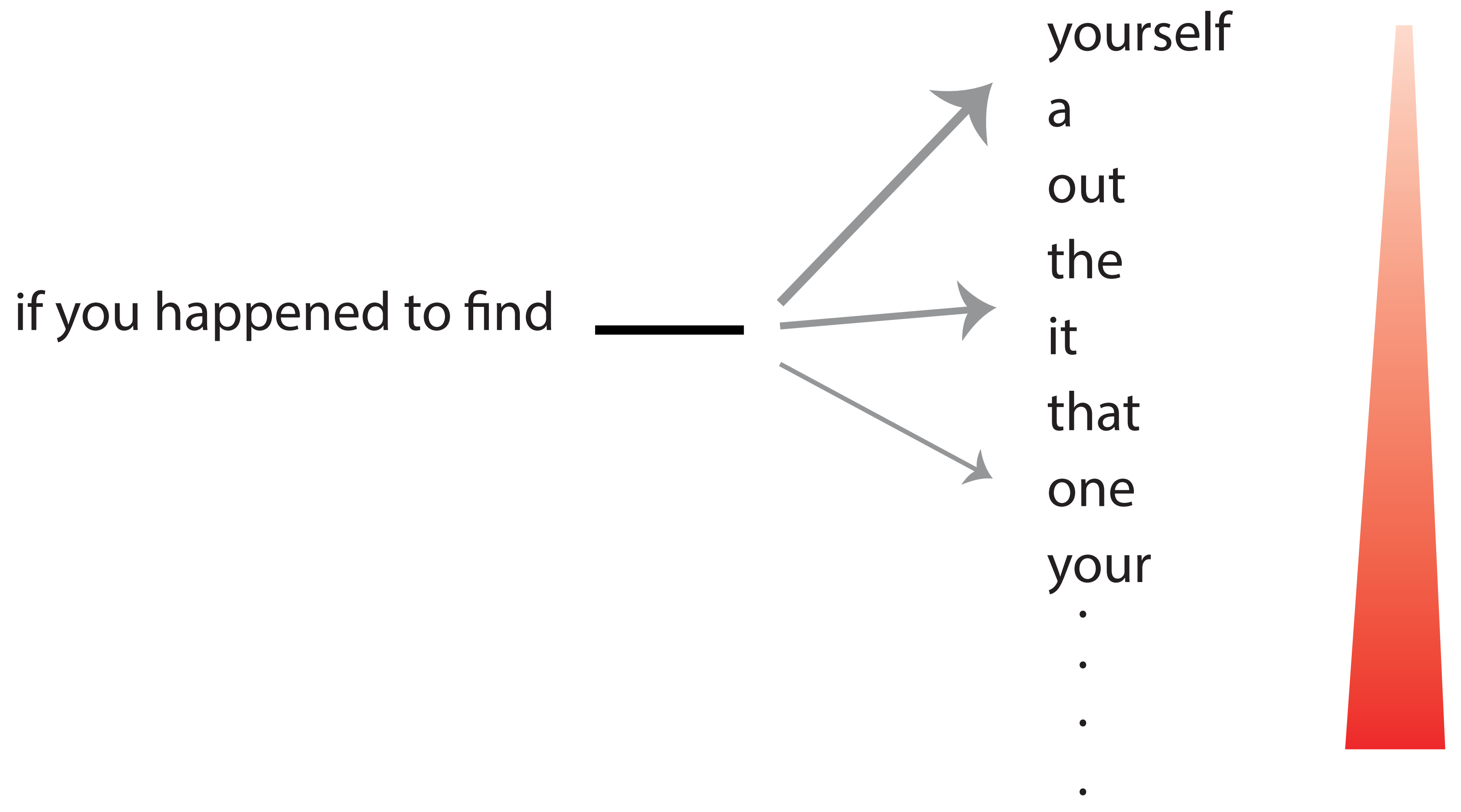
Speech Representations

Frequency of words based on SUBTLEX



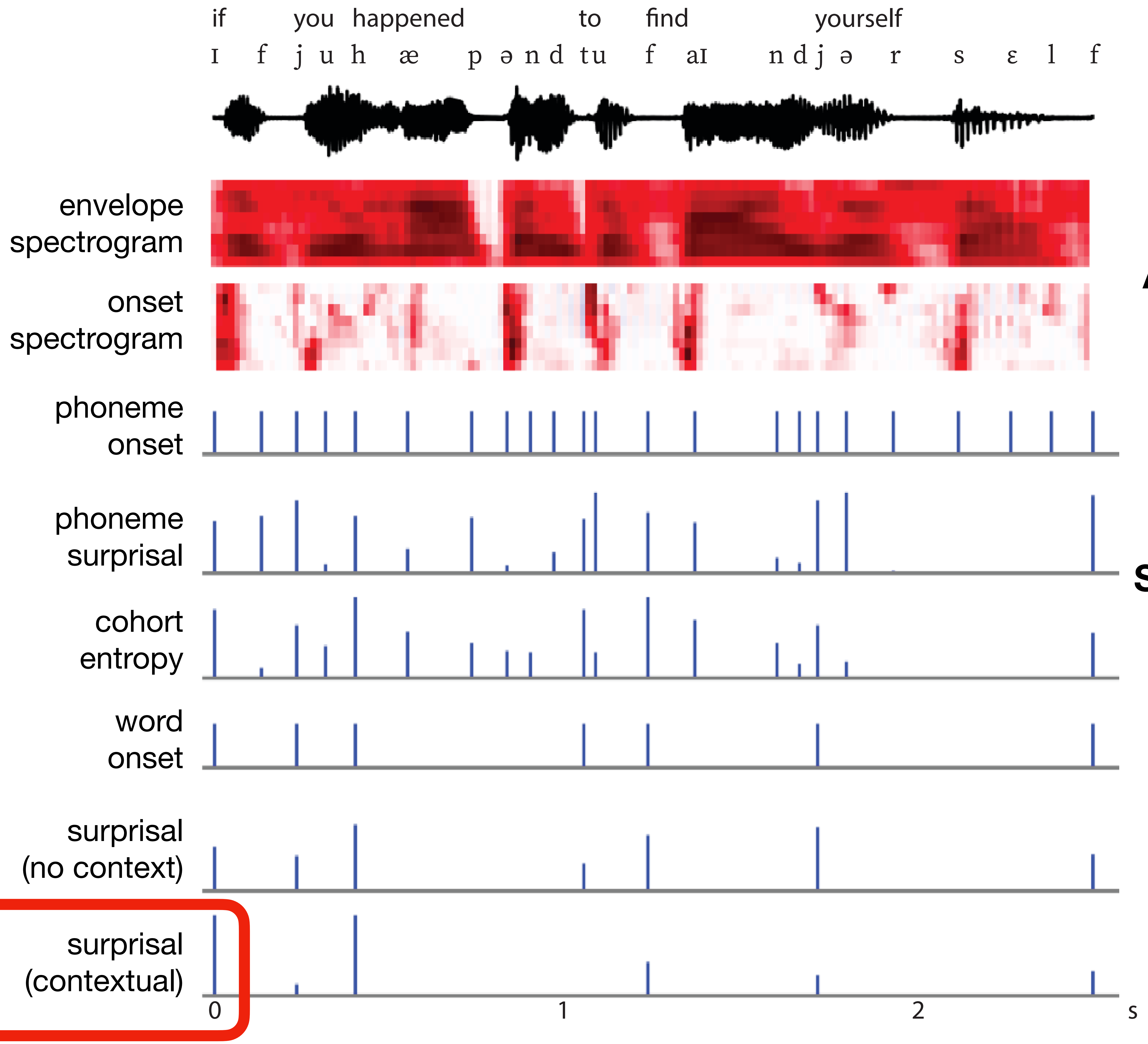


Speech Representations



if you happened to find

yourself
a
out
the
it
that
one
your
.
.
.
.



Acoustic Properties

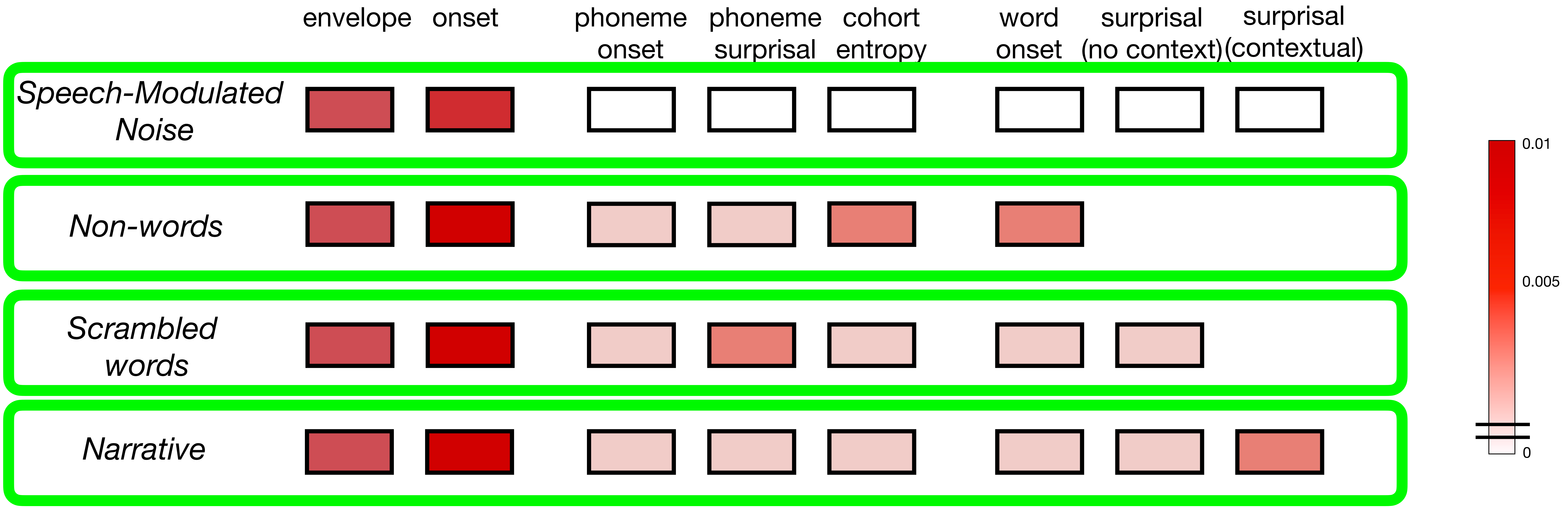
Sub-lexical Properties

Lexical Properties

based on probability of word from GPT-2

Neural Prediction Results

Emergence of neural features as the incremental processing occur



- Acoustic features are encoded for both non-speech and speech stimuli
- (Sub)-lexical features are encoded only when (sub)-lexical boundaries are intelligible
- Context based word surprisal emerges for narrative passage
- When context supports, context based surprisal is better tracked compared to naive surprisal

Hemispheric Lateralization Results

Speech feature

Envelope Onset

Envelope

Phoneme Onset

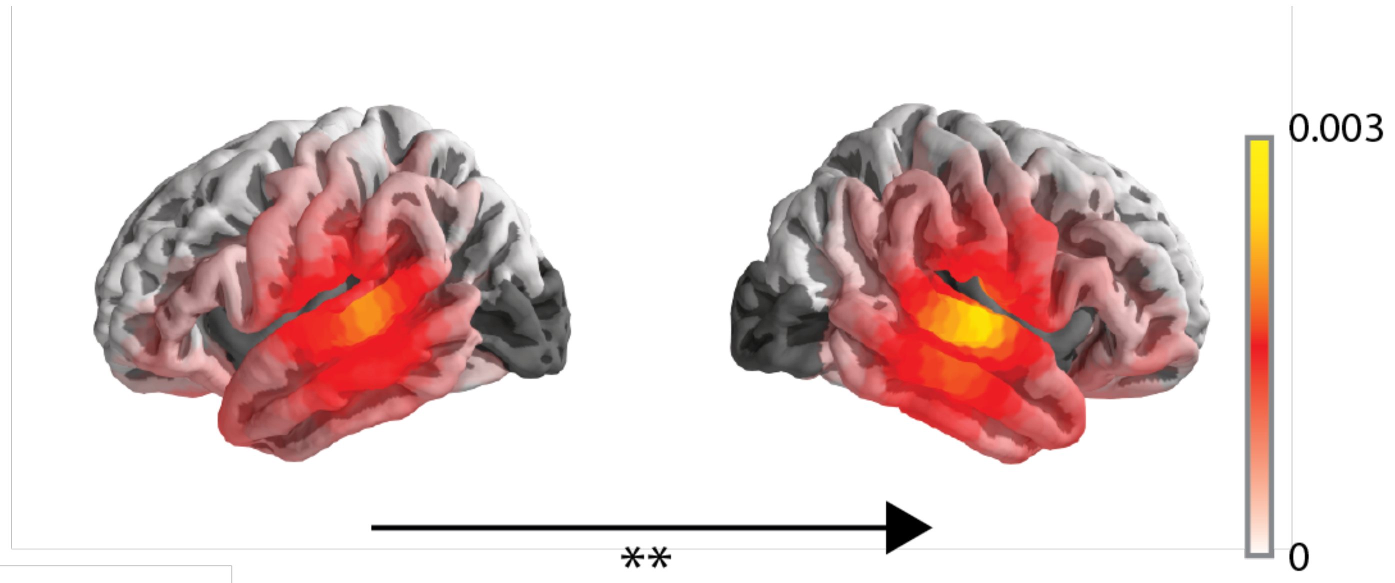
Phoneme Surprisal

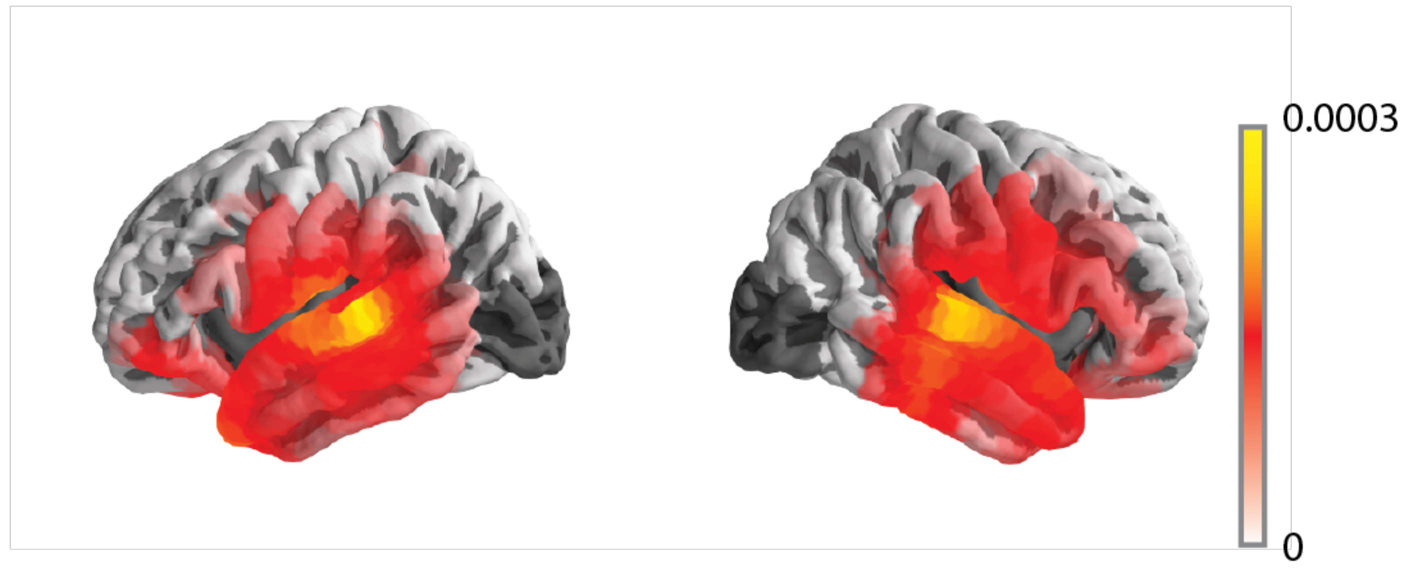
Cohort Entropy

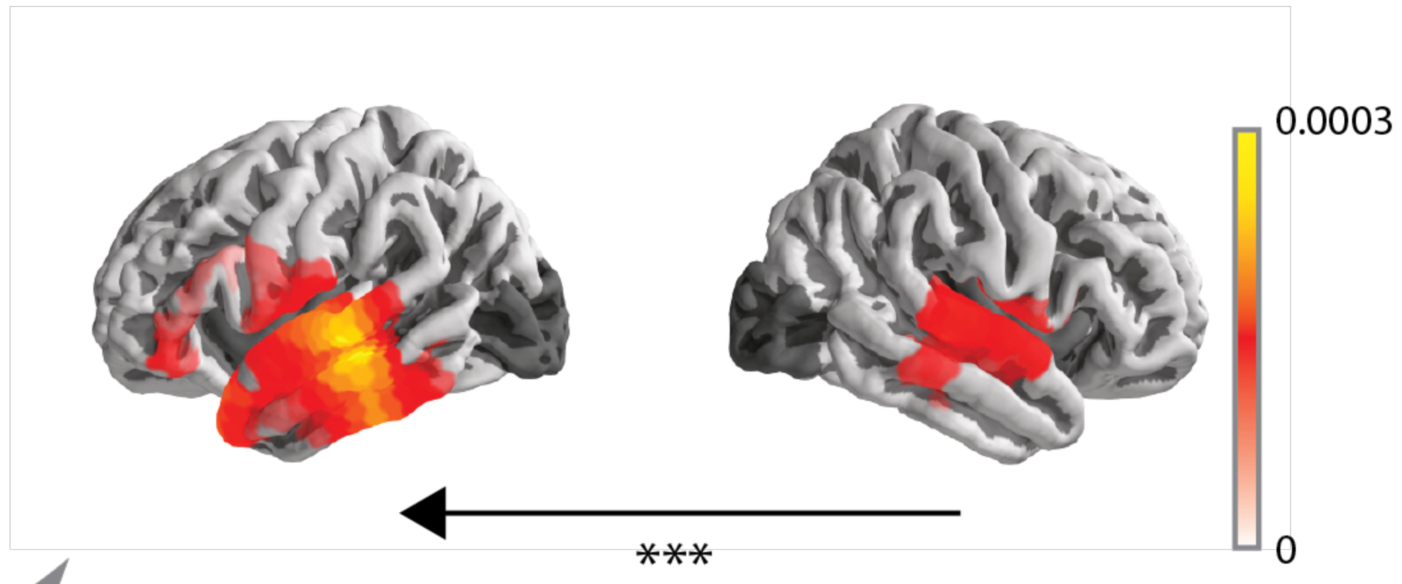
Word Onset

Unigram Surprisal

GPT2 Surprisal



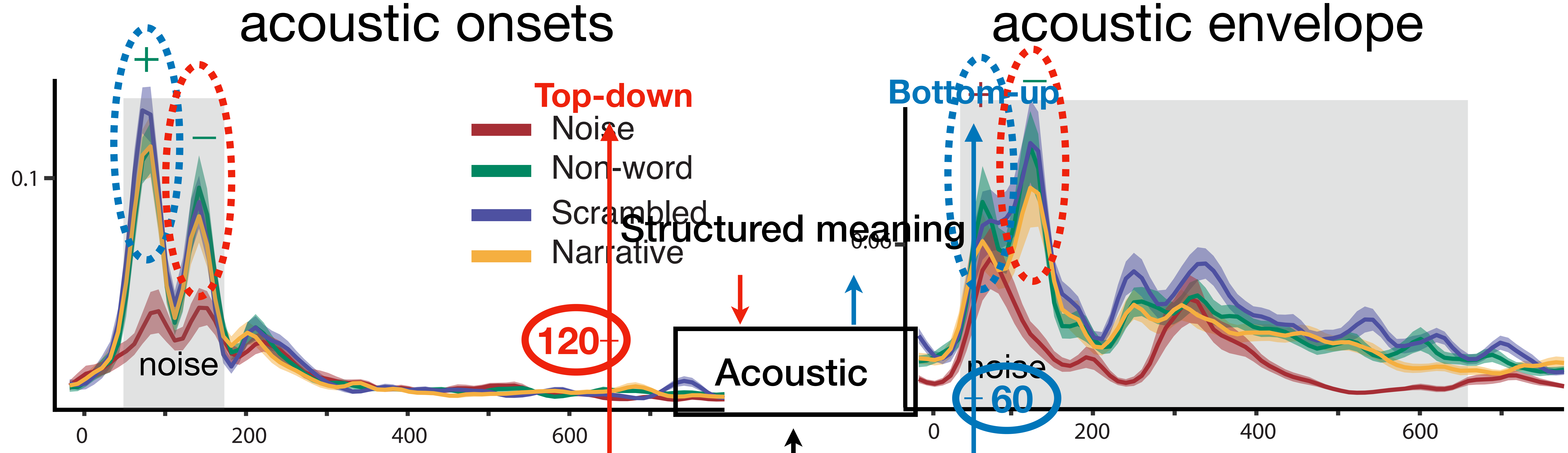




Left Lateralized **Bilateral** **Right Lateralized**

Note: lateralization results can be task dependent

Acoustic TRF Results



Speech
Stimuli

- Speech responses > Noise response
(all speech roughly equal)

- Speech responses > Noise response
(Narrative < Scrambled)
- Non words similar to Scrambled words
- Noise response lacks 2nd peak ~120 ms

60 ms: acoustic bottom-up processing
120 ms: acoustic but attention-dependent

right hemisphere shown
condition based differences similar in left

Phonemic TRF Results

phoneme onset
0.15
+
Non-word
Scrambled
Narrative
0
200
400
600
time (ms)
Top-down
350
120
0
phoneme surprisal
Bottom-up
80
60
0
Structured meaning
Sub-Lexical
Acoustic
Speech Stimuli
0.2
+
−
+
0
200
400
600
time (ms)
cohort entropy
0.12
+
+
−
−
0
200
400
600

- Non-words largest
- No later processing

- Early phone processing ~85 ms (scrambled > narrative)
- Late phone processing ~350 ms) (words > non-words)

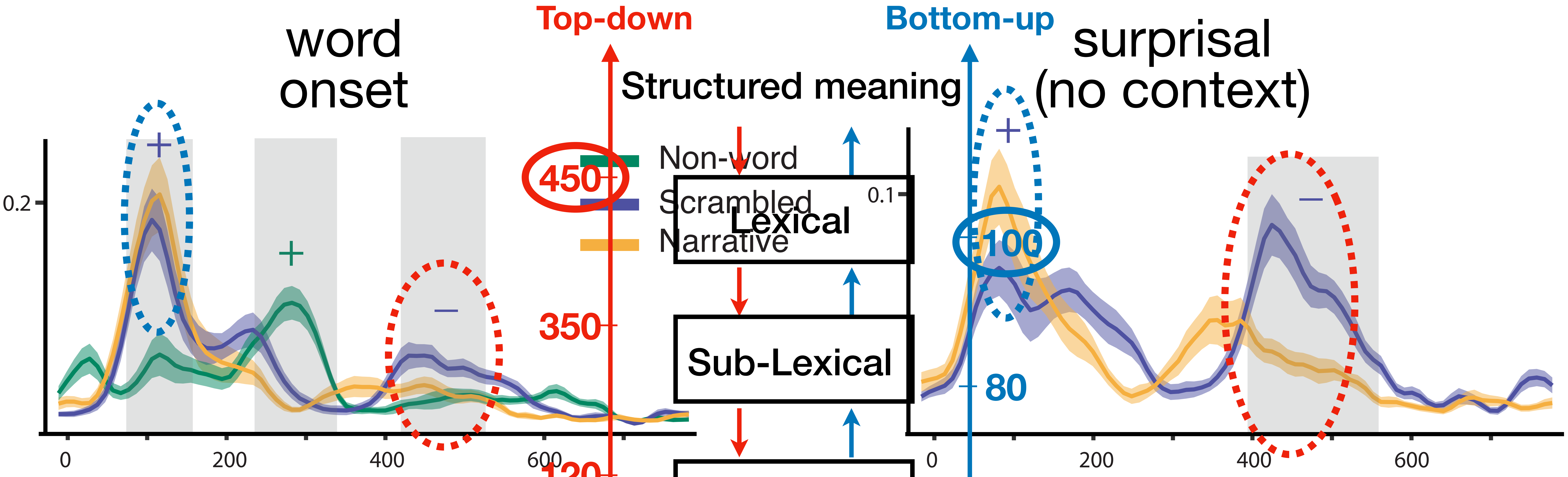
- Late context processing
- N400-like response (reduced for narrative)
- Additional/delayed peaks in non-words (difference in stimulus distributions)

85 ms: simple phoneme processing

350 ms: additional further processing

left hemisphere shown (right similar)

Word-based TRF Results



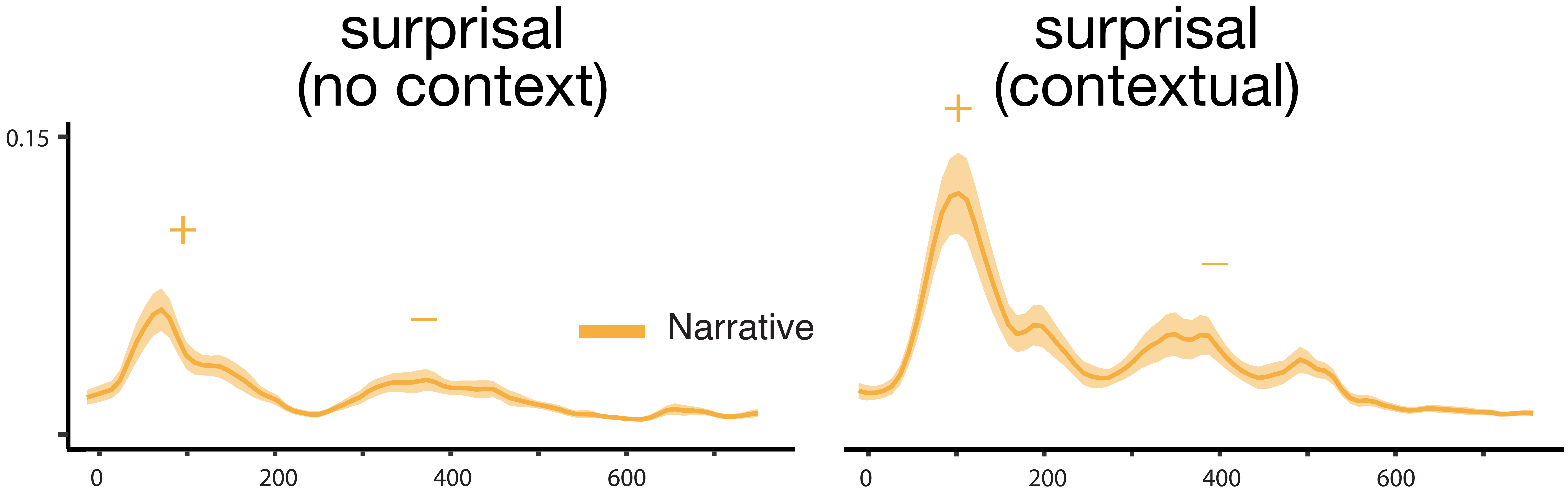
- Scrambled ≈ narrative for rapid processing
- Scrambled words > narrative at ~450 ms
- words: Left hemi > Right (non-words: L ≈ R)

100 ms: simple word processing
450 ms: “error” correction processing

- N400 like response
- Reduction in surprisal when context
- Left hemi > Right hemi
- Right hemisphere: Scrambled ≈ Narrative

left hemisphere shown
(right much weaker except for non-word onset)

Contextual Word Surprisal Results



- When context helps, context-based surprisal is better tracked than raw surprisal
- N400 like response in both predictors

left hemisphere shown
(right much weaker)

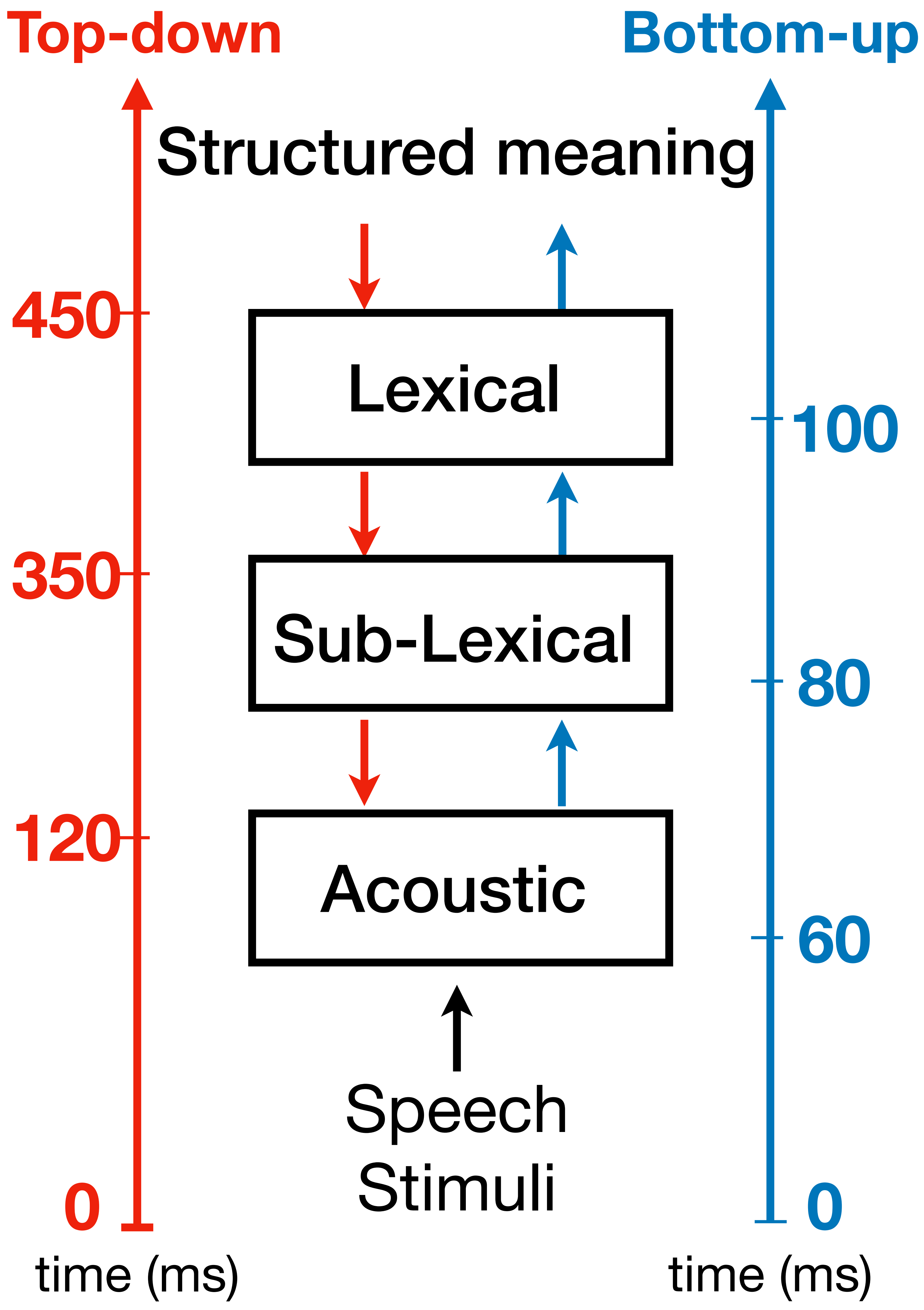
Neural Speech Processing Progression

- Cortical response time-locks to emergent features from acoustics to context as incremental steps in the processing of speech input occur

- Higher level processing / top-down mechanisms may affect lower level speech processing

- Linguistic features are processed when the linguistic boundaries are intelligible

- Lower-level acoustic feature responses are bilateral but right lateralized whereas, context based responses are strongly left lateralized

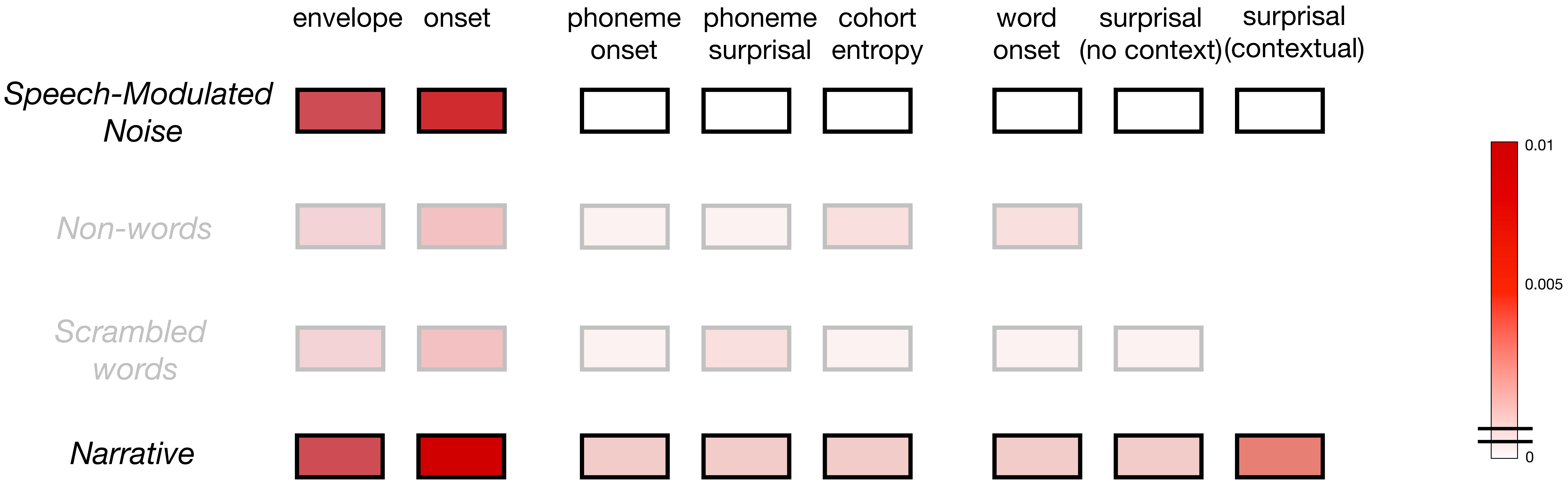


Karunathilake et al. *in preparation*

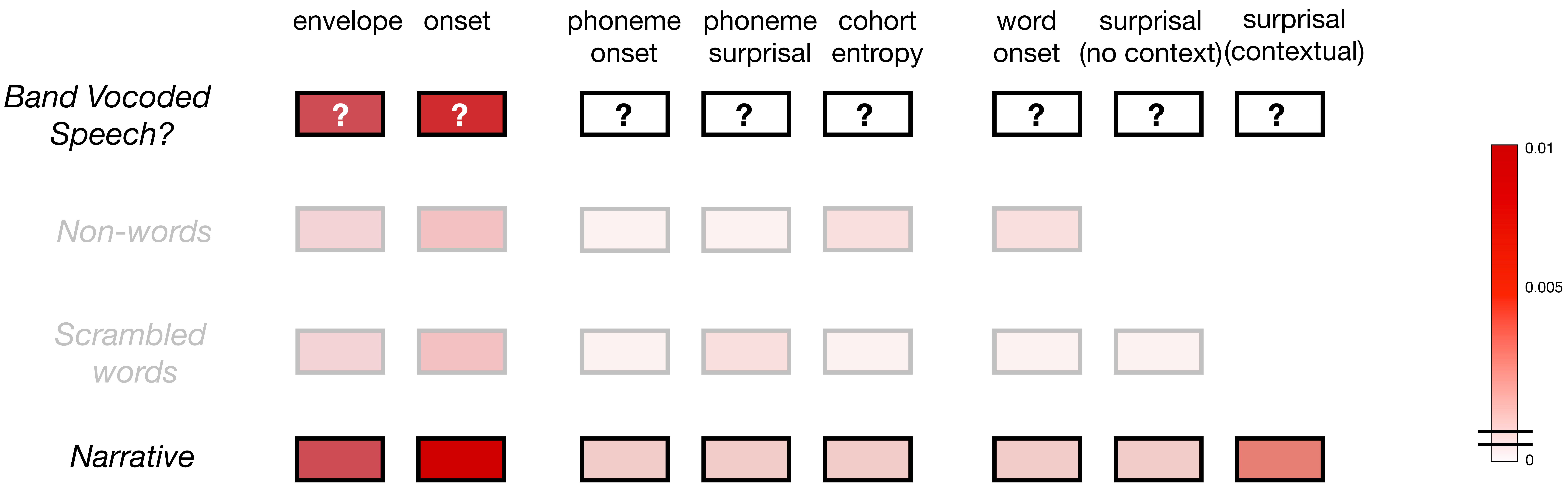
Outline

- Introduction—Cortical representations of continuous speech
- *Early & fast* cortical representation of continuous speech
- Cortical representations of speech *meaning*
- *Progression* of representations of continuous speech through cortex (bottom-up and top-down)
- Objective measures of speech *intelligibility*
- *Directional functional connectivity* during difficult speech listening

Previous Neural Prediction Results



Possible Neural Prediction Results



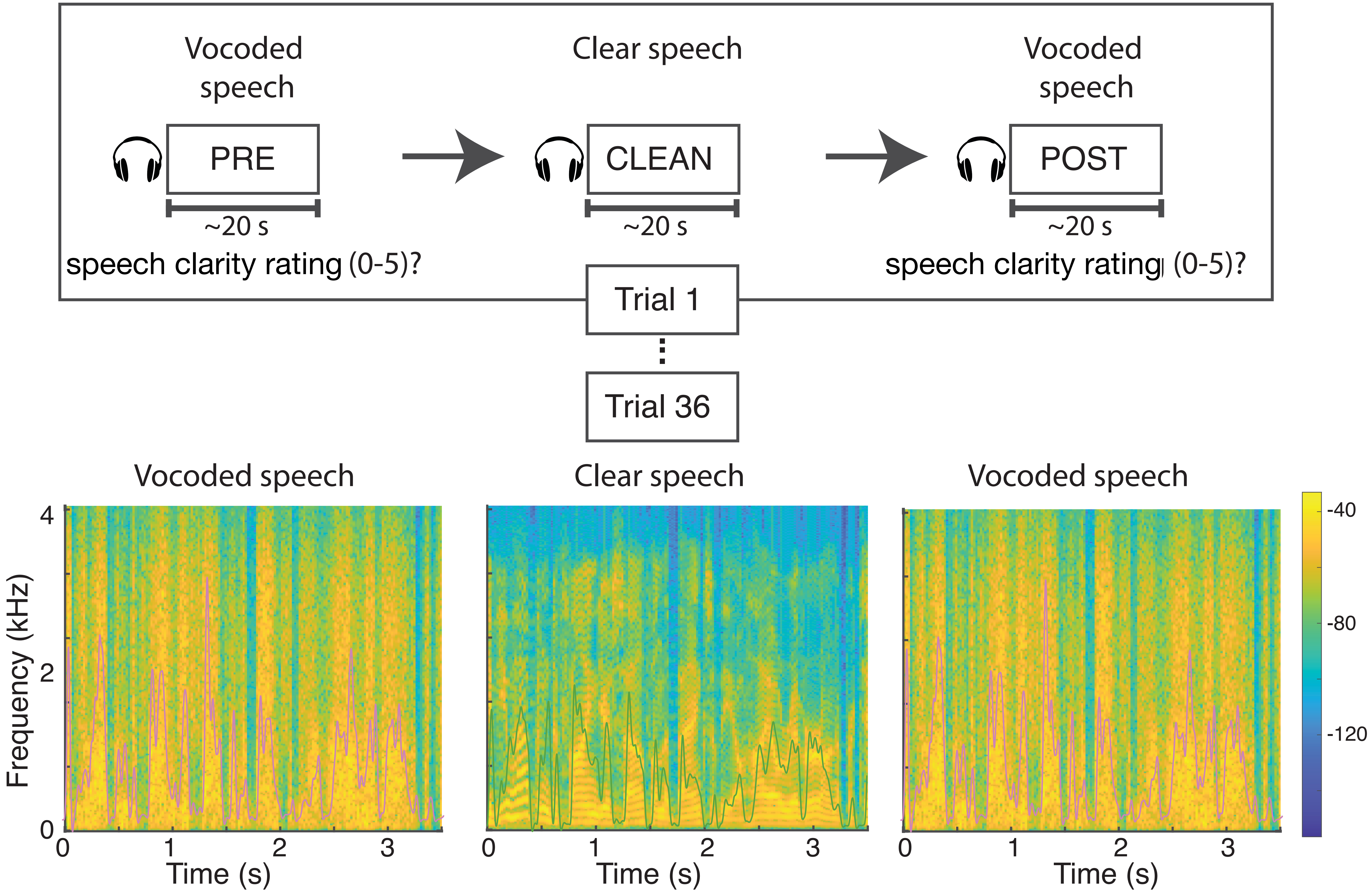
	envelope	onset	phoneme onset	phoneme surprisal	cohort entropy	word onset	surprisal (no context)	surprisal (contextual)
Band Vocoded Speech?	?	?	?	?	?	?	?	?
Non-words	■	■	■	■	■	■		
Scrambled words	■	■	■	■	■	■	■	
Narrative	■	■	■	■	■	■	■	■

Intelligibility Experimental Design

- Manipulate intelligibility but keep acoustics unchanged
    - Speech acoustics: three-band noise-vocoded speech
    - Intelligibility manipulated via priming

- Hypothesized intelligibility measure(s)
    - word boundaries

"Slice an apple through at its equator, and you will find five small chambers arrayed in a perfectly symmetrical starburst—a pentagram."



Vocoded speech

PRE

~20 s

speech clarity rating (0-5)?

Clear speech

CLEAN

~20 s

Trial 1

Trial 36

Vocoded speech

POST

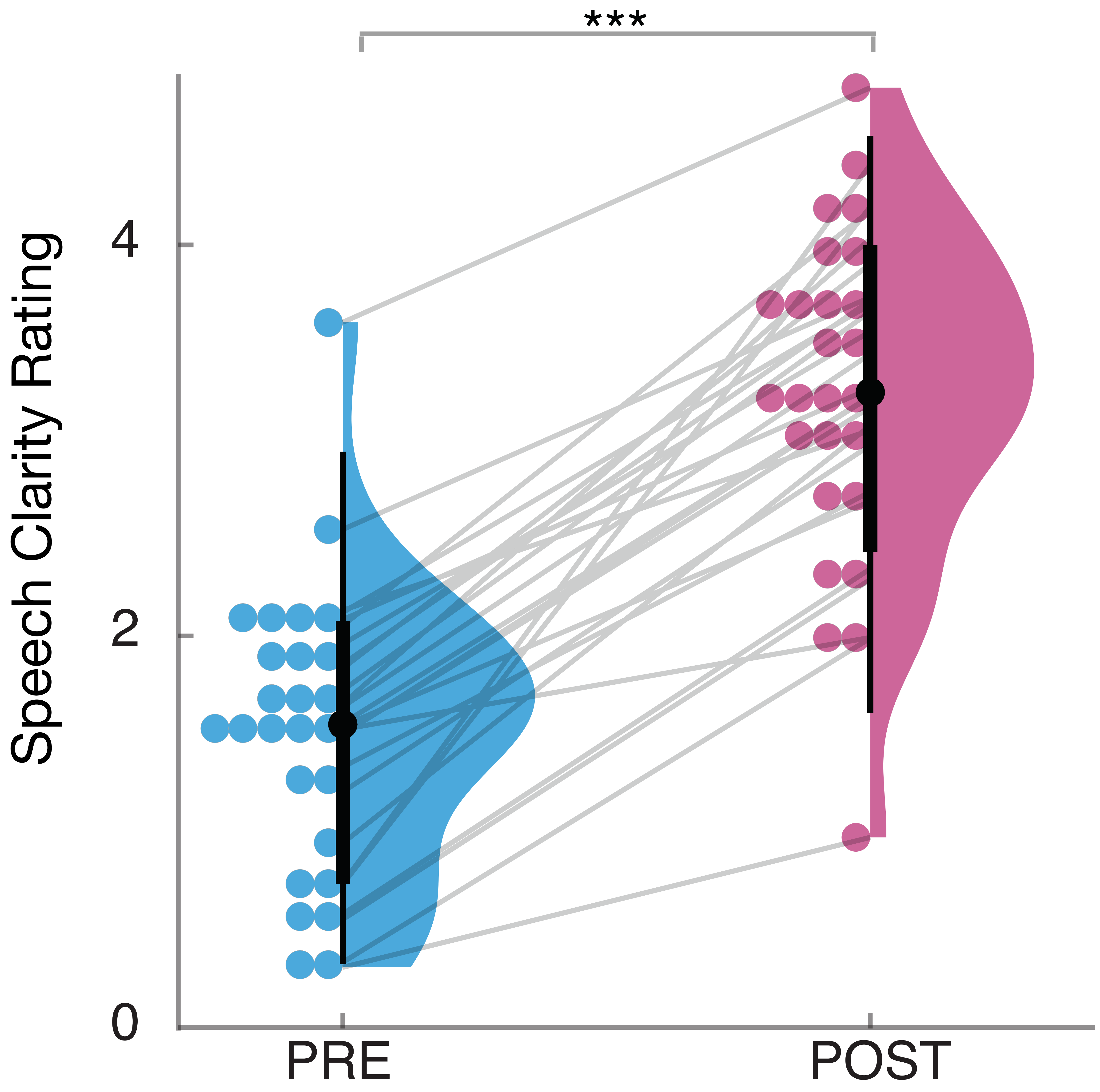
~20 s

speech clarity rating (0-5)?

Karunathilake et al. (2023) *Neural Tracking Measures of Speech Intelligibility…*, bioRxiv

Intelligibility Behavioral Results

Speech Clarity **increases** from PRE condition to POST condition



Karunathilake et al. (2023) *Neural Tracking Measures of Speech Intelligibility…*, bioRxiv

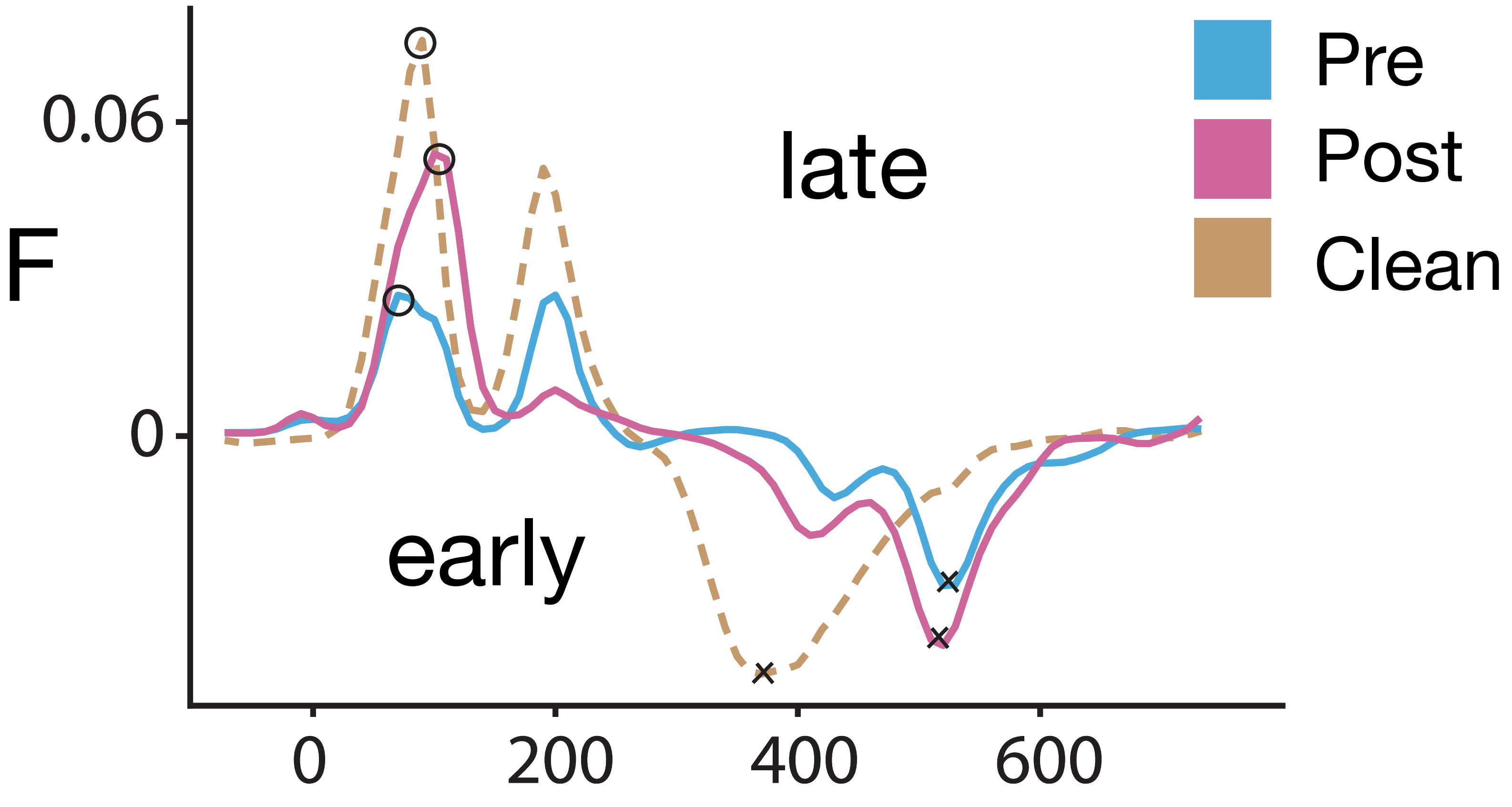
(b)

Intelligibility Neural Results

- Word onset TRF shows both early (+) and late (-) processing stages
- Physiological response increases Pre→Post
  - Only in left hemisphere
  - Late processing stage shows larger change than early

Word onset TRF

single subject



- Physiological Word Onset response
  Objective measure of intelligibility
  - Acoustic responses: no change
  - Response to Word Surprisal:
    Additional intelligibility measure

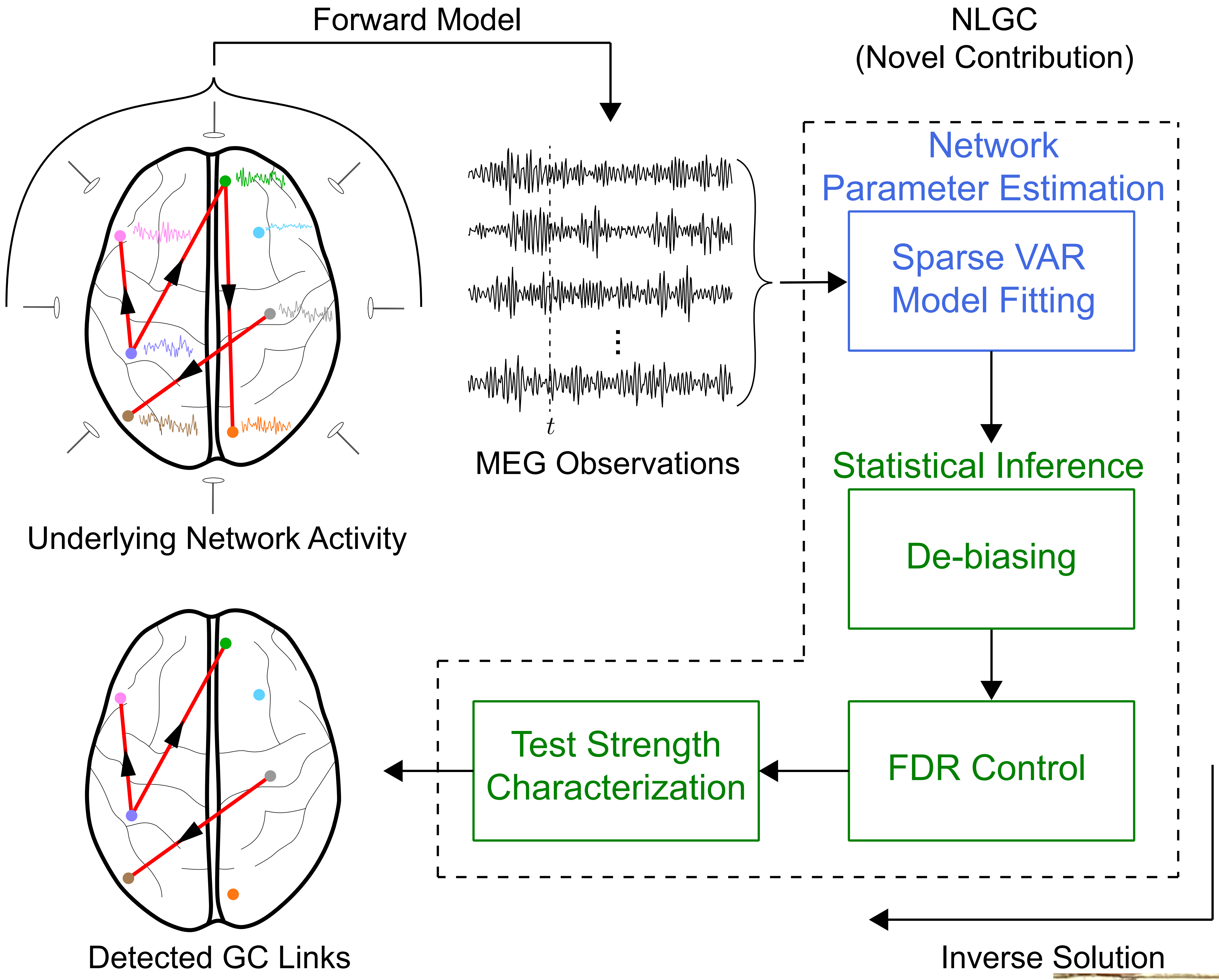
Karunathilake et al. (2023) *Neural Tracking Measures of Speech Intelligibility*…, bioRxiv

Outline

- Introduction—Cortical representations of continuous speech
- *Early & fast* cortical representation of continuous speech
- Cortical representations of speech *meaning*
- *Progression* of representations of continuous speech through cortex (bottom-up and top-down)
- Objective measures of speech *intelligibility*
- *Directional functional connectivity* during difficult speech listening

Directional Functional Connectivity

- Novel method, based on Granger Causality (if source A can predict source B)
  - Directional (bi-directional allowed)
  - Localizes neural sources & GC link strengths simultaneously
  - source currents: latent sparse vector autoregressive (VAR) processes
- Network Localized Granger Causality (NLGC)
  - source spread & other biases minimized
  - robust against source model mismatch
  - parametrized by false discovery rate
  - intrinsically statistically robust

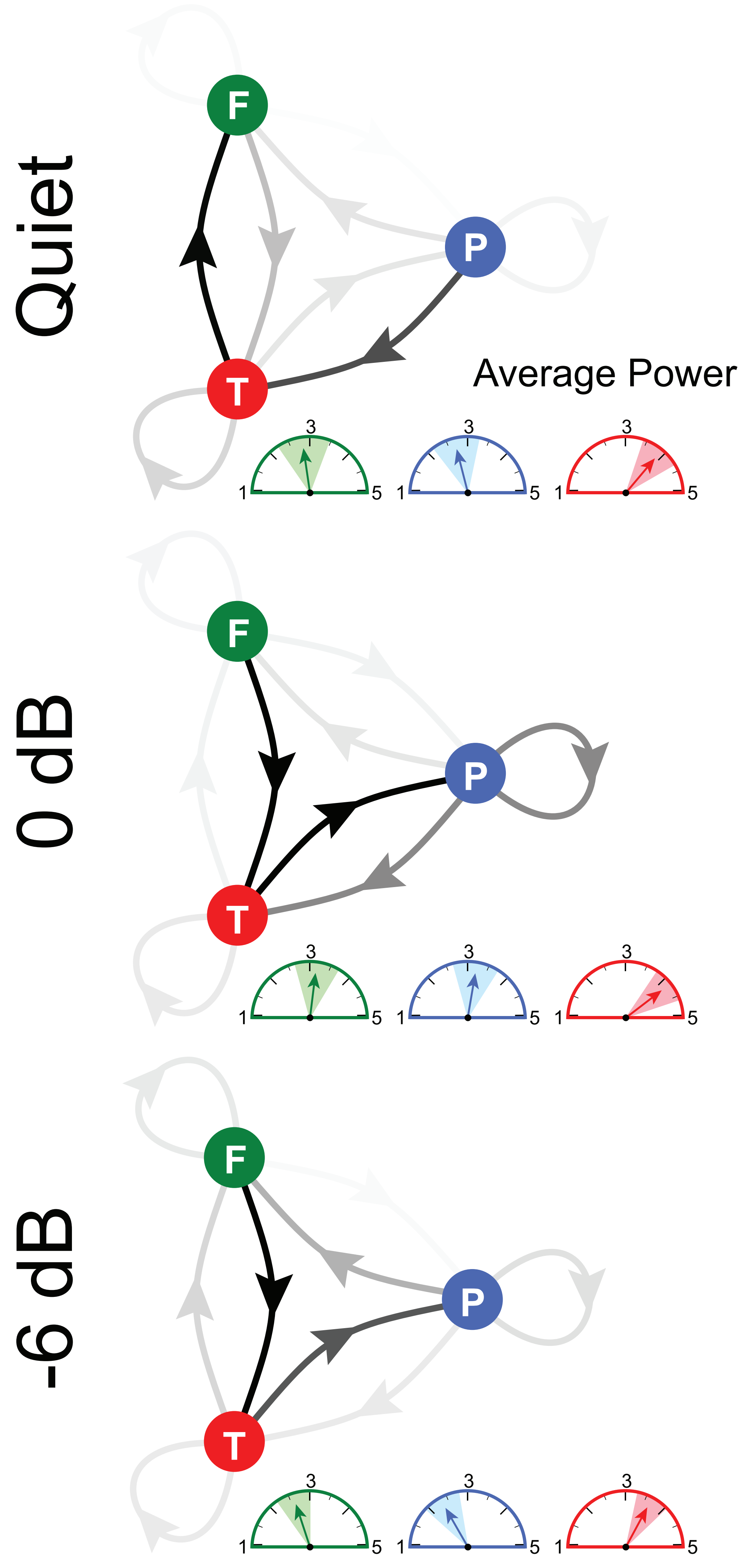


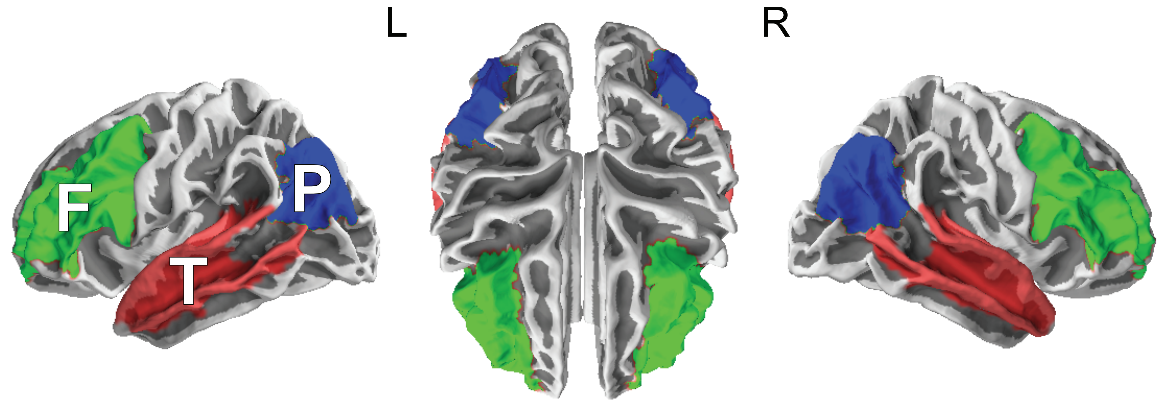
Soleimani et al. (2022) *NLGC: Network Localized Granger Causality with Application to …*, NeuroImage

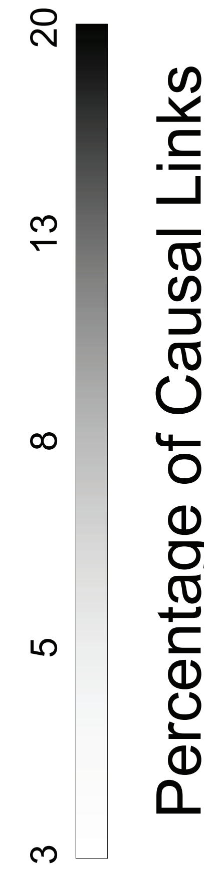
Cocktail Party Speech Results

Theta band example

- Speech in quiet connectivity: dominantly Temporal→Frontal and Parietal→Temporal
- Cocktail Party listening (moderate SNR): Temporal-Frontal switches direction; Parietal-Temporal now bi-directional
- Cocktail Party listening (poor SNR): Temporal←Frontal remains; Parietal→Temporal dominant





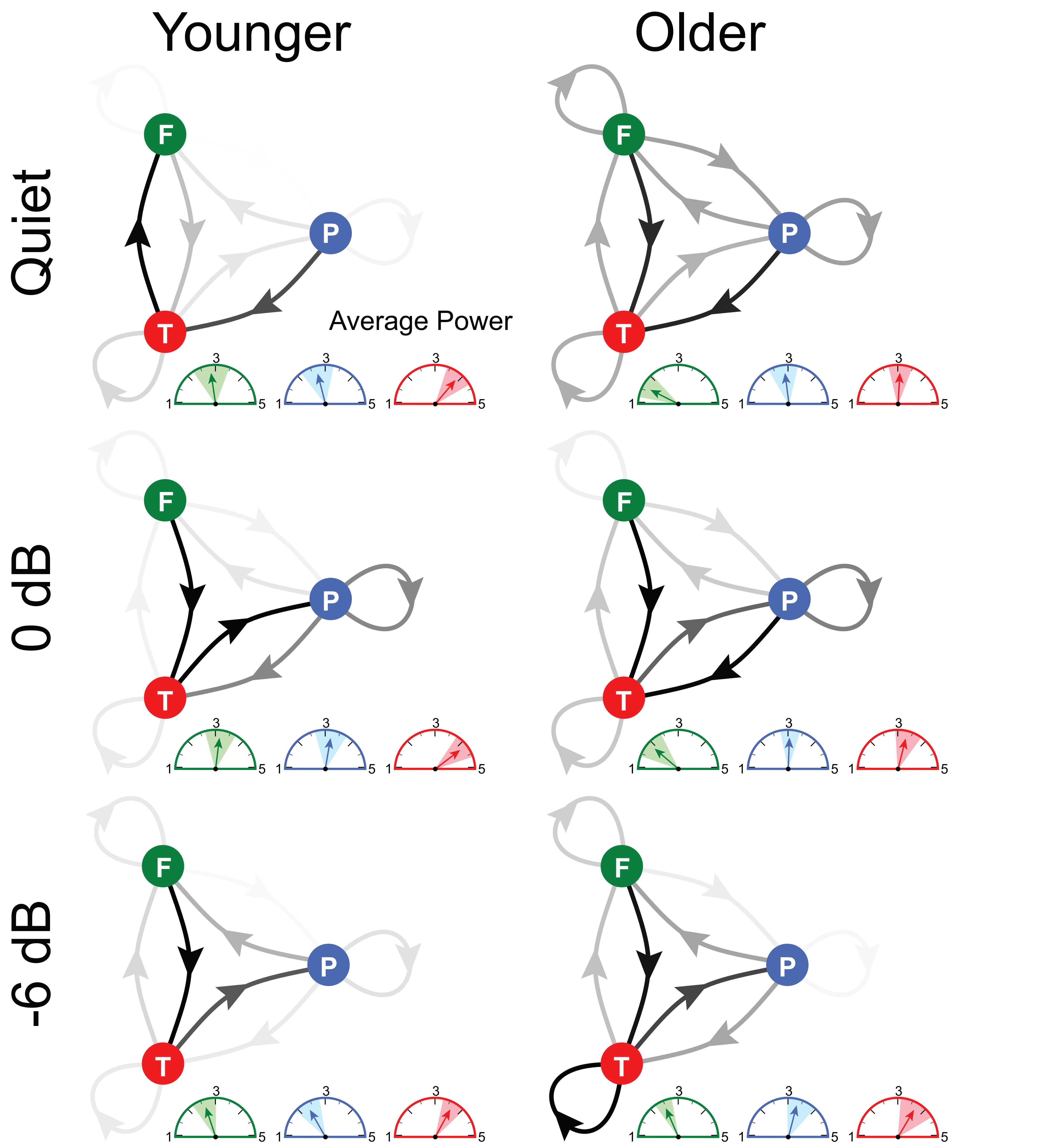


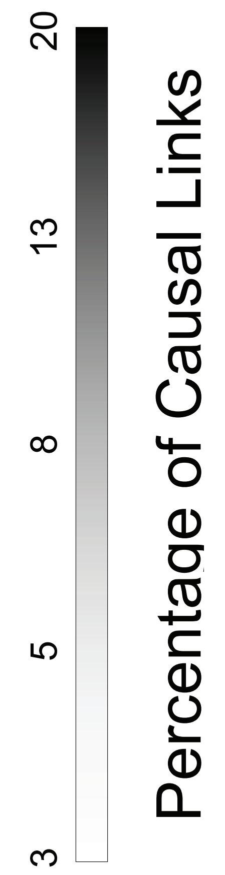
Soleimani et al. (2023) … *Cortical Directional Connectivity during Difficult Listening*… bioRxiv

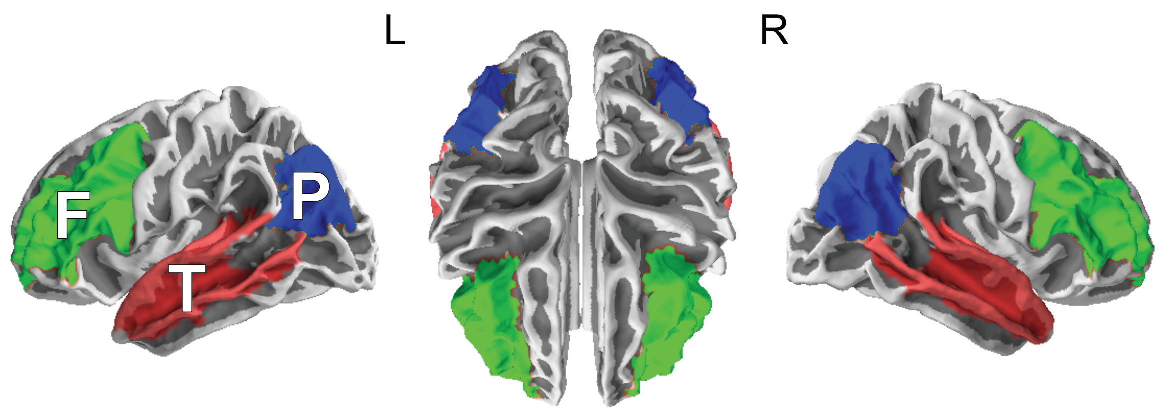
Cocktail Party Speech Results

Older Listeners exhibit strongly different connectivity

- Older speech in quiet connectivity: similar to Younger cocktail party listening connectivity





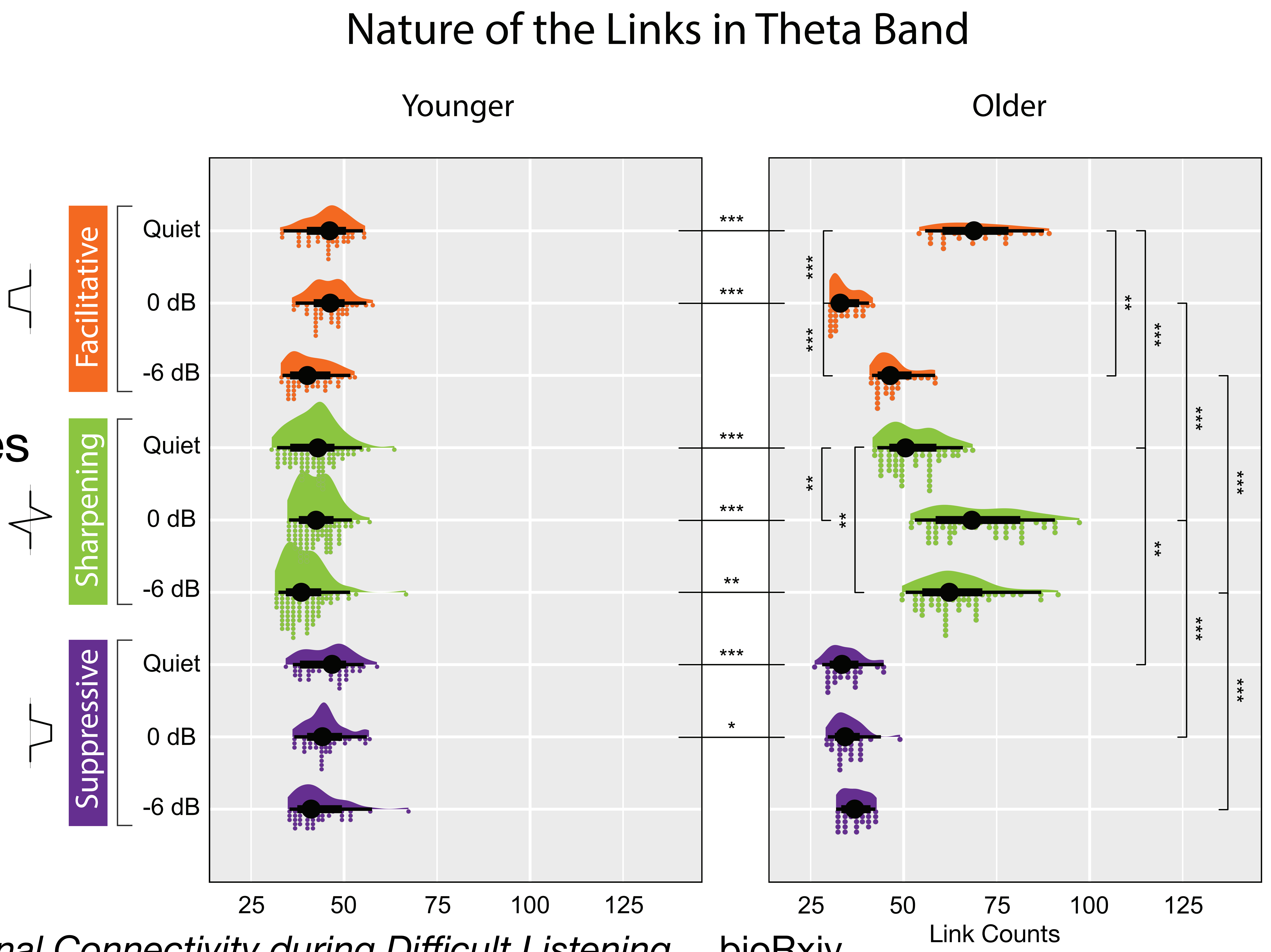


Soleimani et al. (2023) … *Cortical Directional Connectivity during Difficult Listening*… bioRxiv

Cocktail Party Speech Results

“Excitatory/Inhibitory” balance changes with task difficulty for Older Listeners only

- VAR (IIR filter) coefficients reveal neural signal transformation between sources
- coefficients > 0: “Excitatory”/facilitative
- coefficients < 0: “Inhibitory”/suppressive
- mixed coefficients: sharpening filter

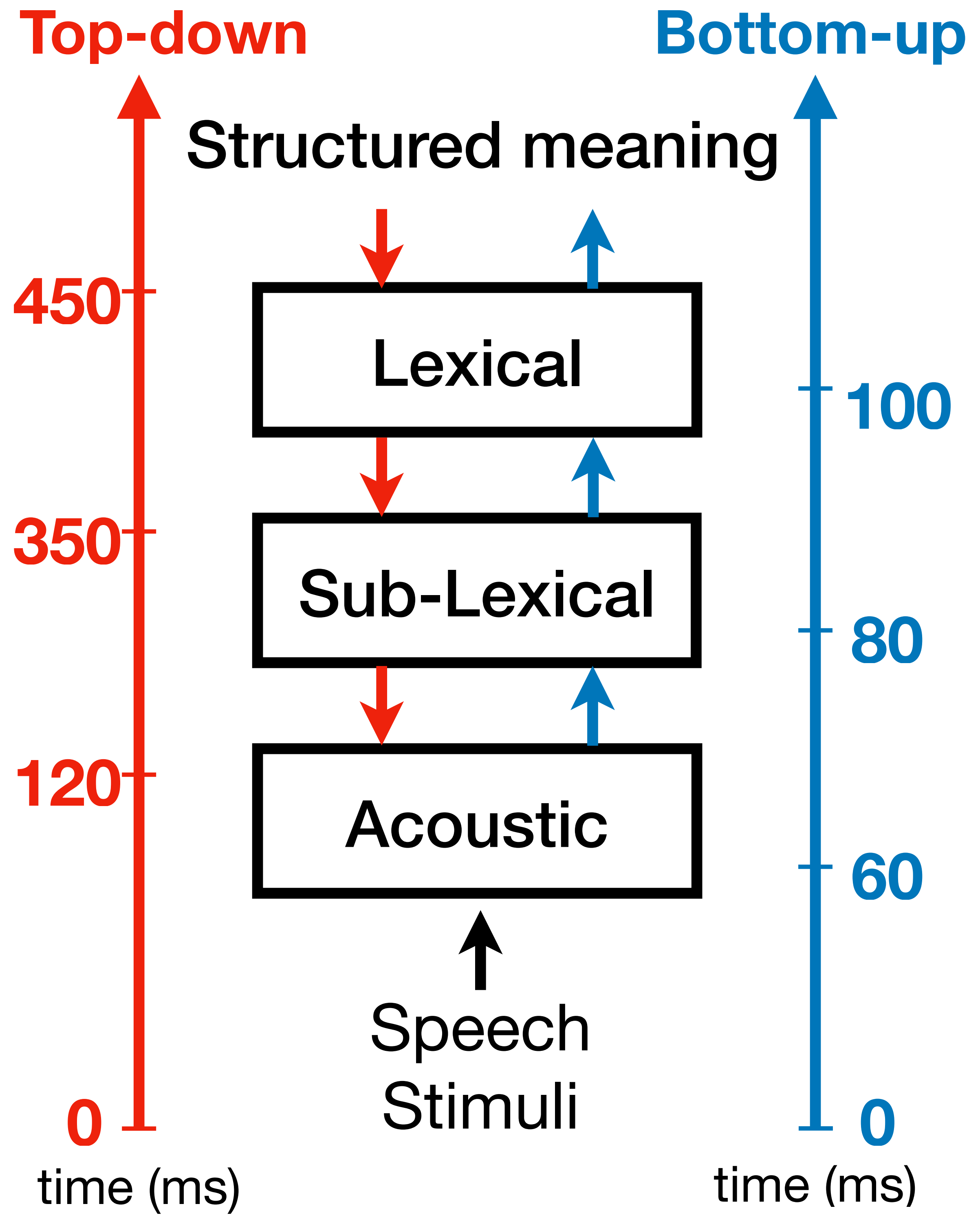


Soleimani et al. (2023) … *Cortical Directional Connectivity during Difficult Listening*… bioRxiv

Final Summary

*temporal **neural** patterns* ⇆ *temporal patterns in **speech acoustics***
*temporal patterns in **speech perception***
*temporal patterns in **language perception***
*temporal patterns in **understanding***

- Cortical responses time-lock to emergent features
- Higher level processing / top-down mechanisms may affect lower level
- Linguistic features processed only when linguistic boundaries intelligible
- Acoustic responses: bilateral but right lateralized; context-based responses strongly left lateralized



thank you

These slides
available at:
ter.ps/simonpubs

Mastodon: @jzsimon@fediscience.org

http://www.isr.umd.edu/Labs/CSSL/simonlab