

# **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>

Peking University, 15 Dec 2023

# Acknowledgements

## Current Lab Members & Affiliates

Morgan Belcher

**Vrishab Commuri**

Charlie Fisher

Tejas Guha

Brooke Guo

Michael Johns

Kevin Hu

**Dushyanthi Karunathilake**

**Karl Lerud**

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**

Shihab Shamma

## Past Lab Members & Affiliates

Nayef Ahmar

Sahar Akram

Olivia Bermudez-Hopkins

**Shohini Bhattasali**

**Christian Brodbeck**

Regina Calloway

Francisco Cervantes Constantino

Maria Chait

Aura Cruz Heredia

Proloy Das

Alain de Cheveigné

Lien Decrui

Marisel Villafane Delgado

Nai Ding

Jason Dunlap

Mounya Elhilali

Sydney Hancock

Marlies Gilles

Victor Grau-Serrat

Alex Jiao

**Neha Joshi**

**Joshua Kulasingham**

Natalia Lapinskaya

Huan Luo

Sina Miran

Alex Presacco

Krishna Puvvada

**Mohsen Rezaeizadeh**

Behrad Soleimani

Jonas Vanthornhout

Yadong Wang

Richard Williams

Juanjuan Xiang

Peng Zan

Elana Zion Golumbic

## Funding & Support



NIDCD



NIA



# 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*

# 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*



# 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  $\Leftrightarrow$  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

***Temporal neural patterns  $\Leftrightarrow$  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

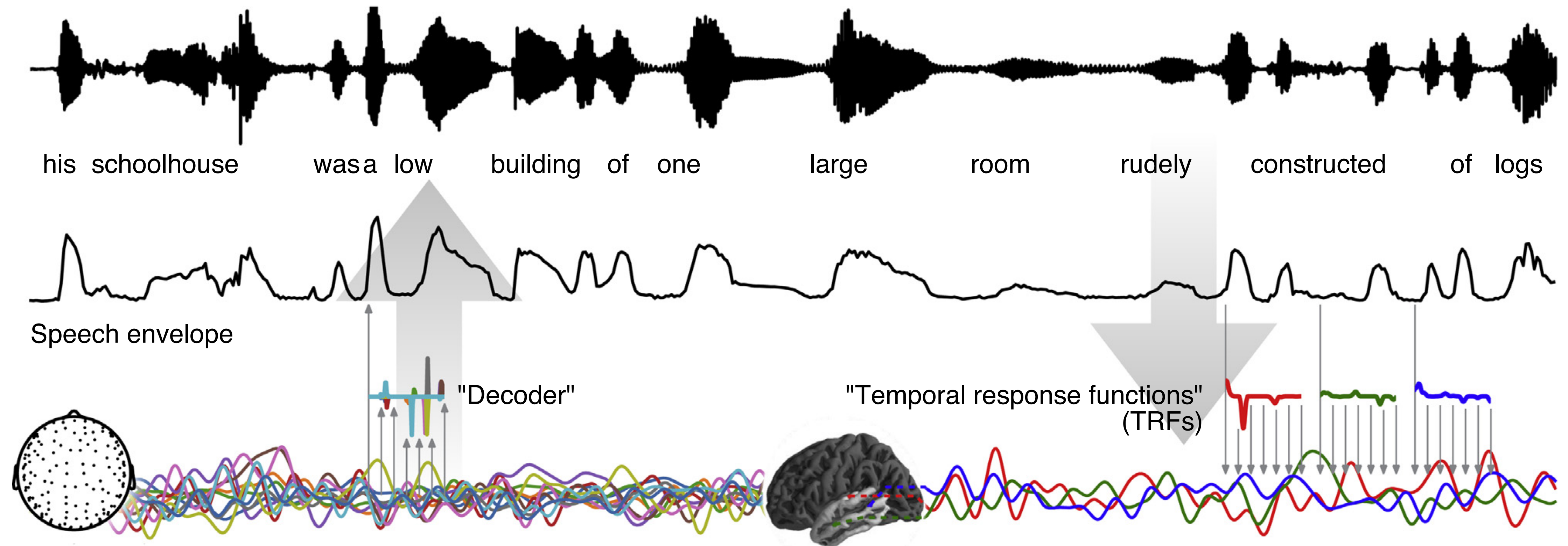
## ***Neural Representations of Speech***

- 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**



# Cortical Representations of Continuous Speech

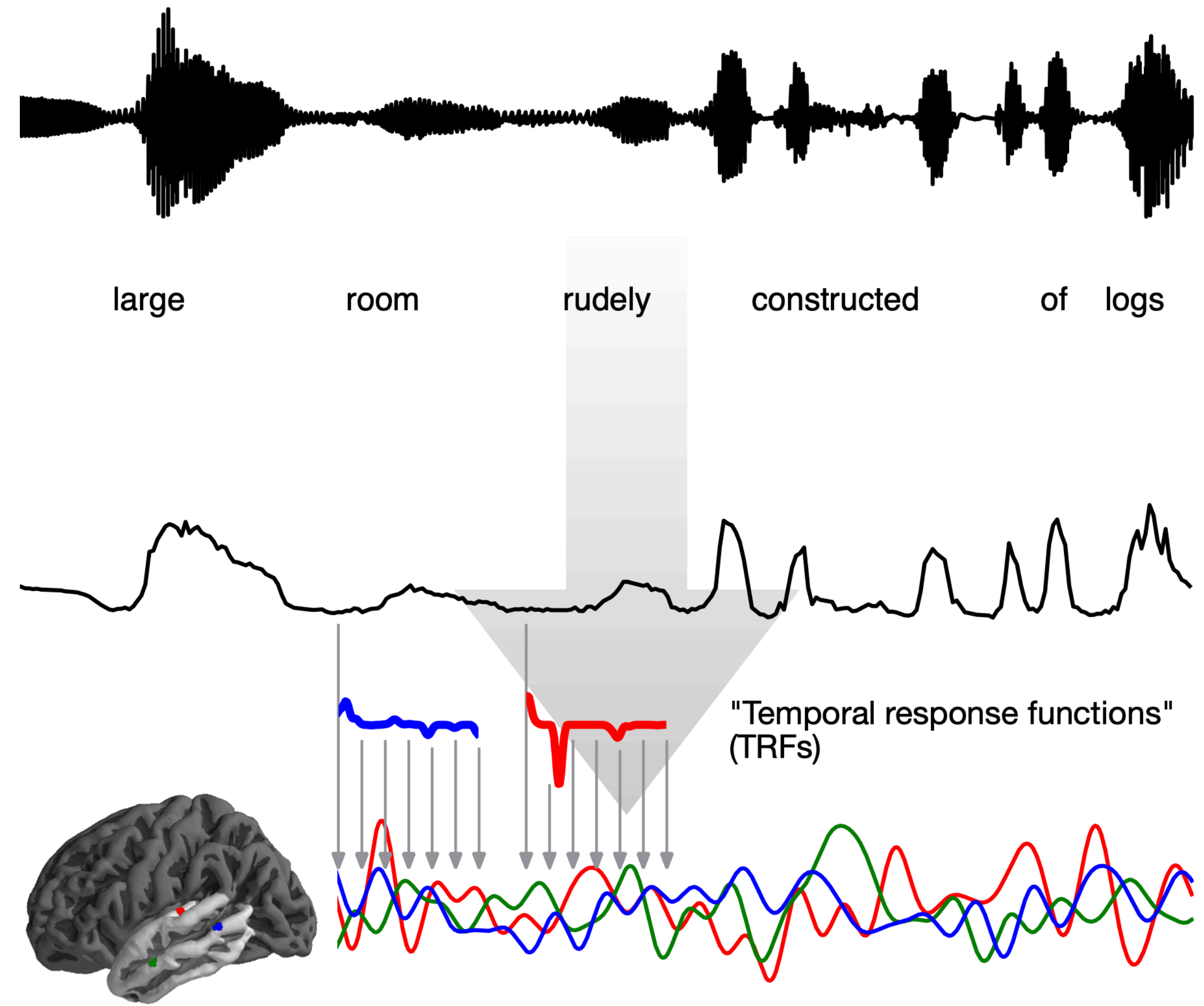
- Measure *time-locked* responses to temporal pattern of speech features (in humans)
- Any speech feature of interest: acoustic envelope, lexical, pitch, semantic, etc.
- Infer spatio-temporal neural origins of neural responses





# 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  $\approx$  processing intensity
  - peak latency  $\approx$  source location
  - multiple TRFs simultaneously

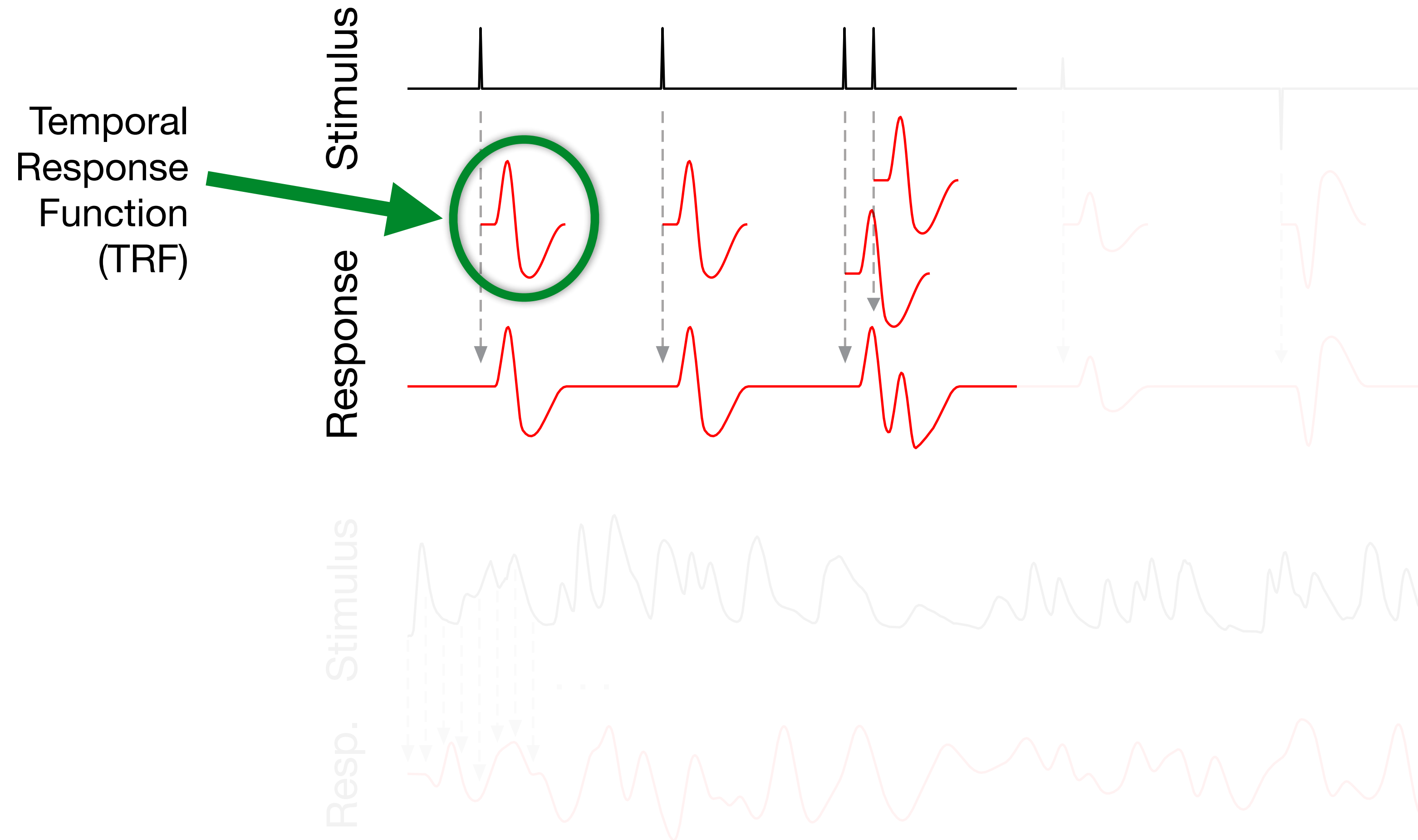


Example: MEG Prediction of Voxel Responses

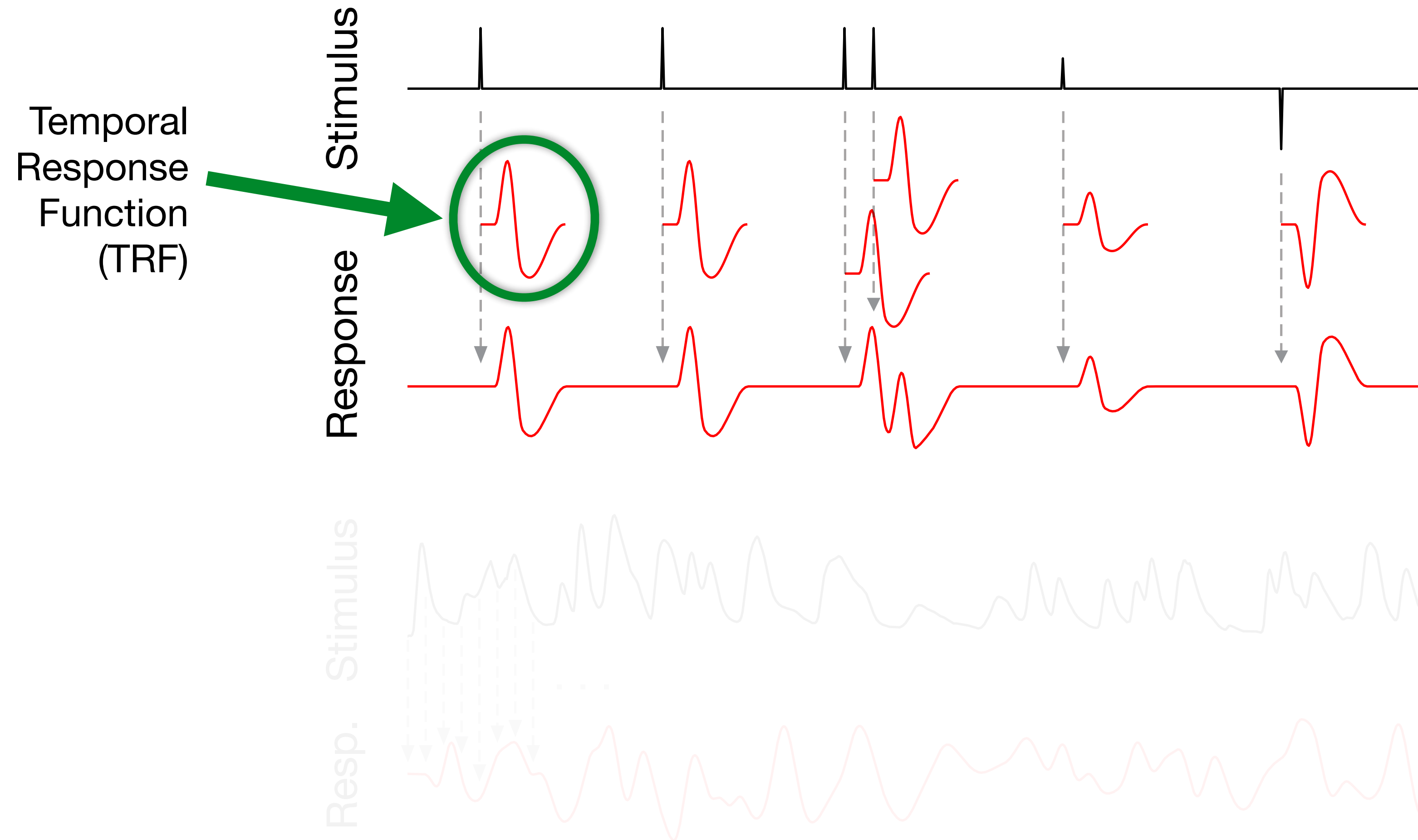
# Temporal Response Functions



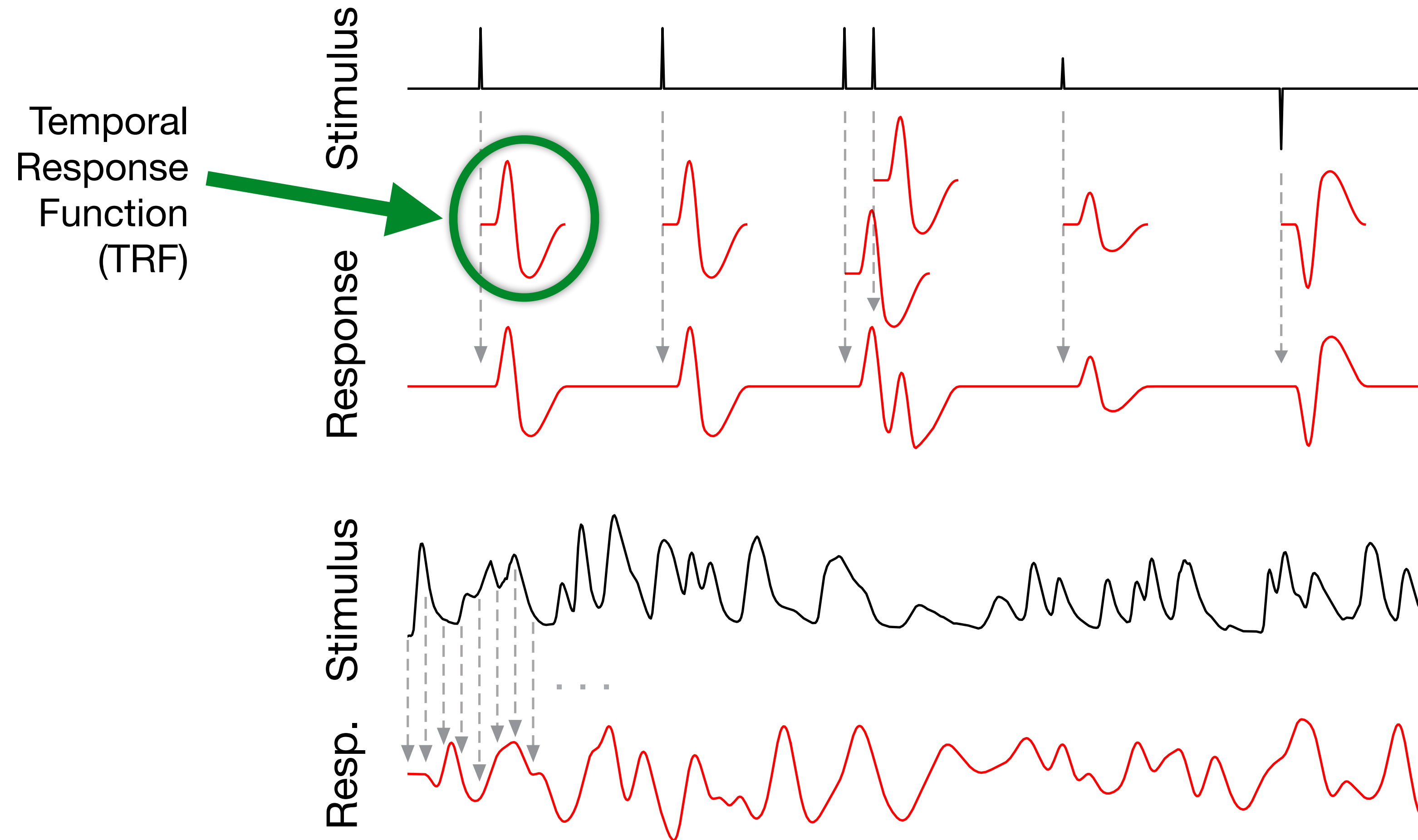
# Temporal Response Functions



# Temporal Response Functions



# Temporal Response Functions

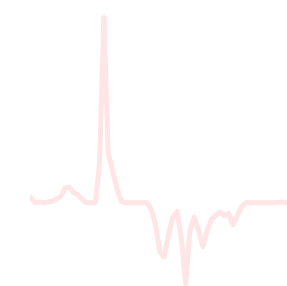
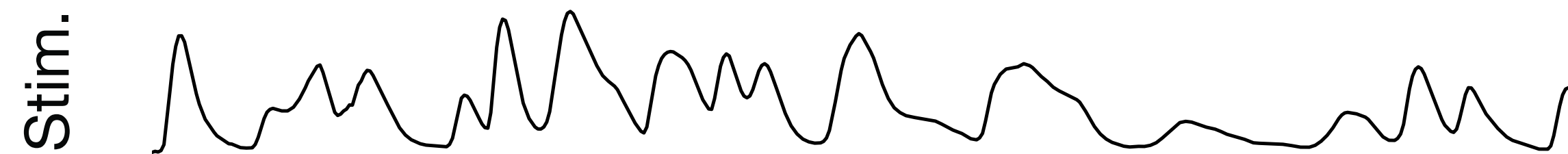
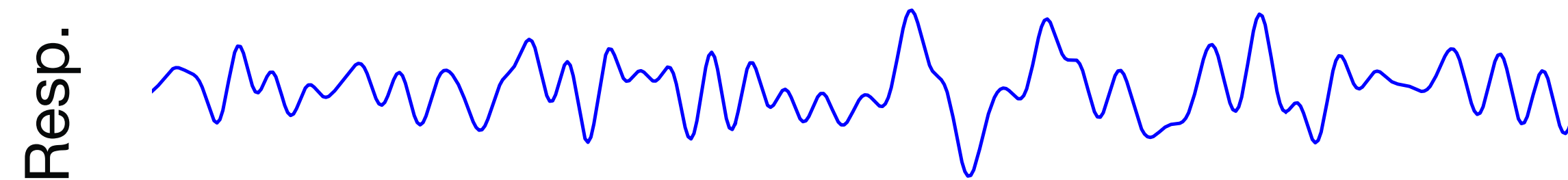




# TRF Model Estimation & Fit

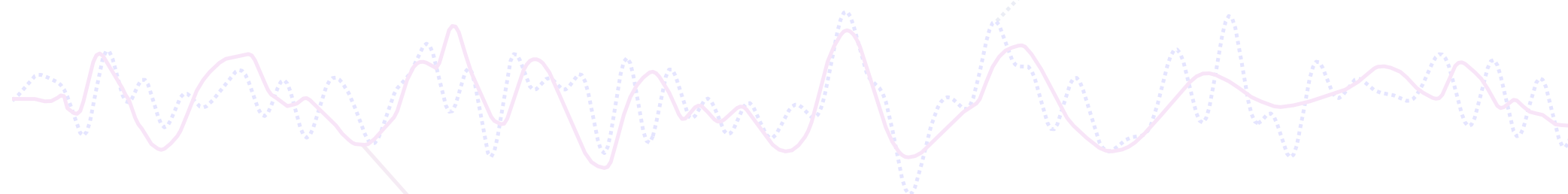
## Temporal Response Function (TRF) estimation:

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



Estimated TRF

Resp.



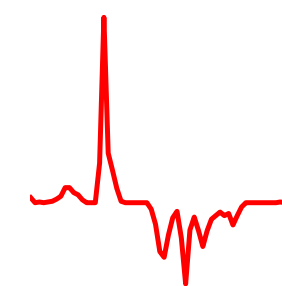
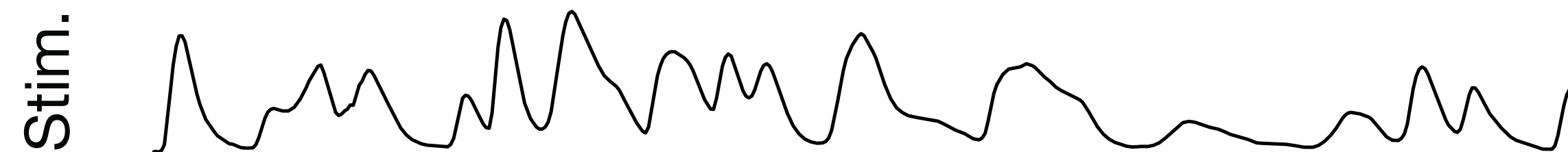
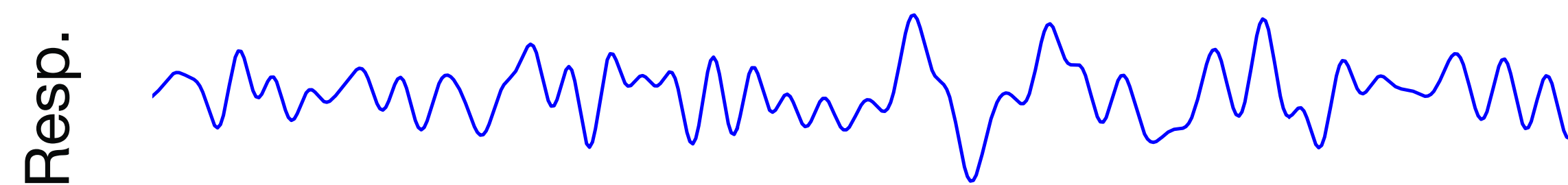
Actual response

Predicted response (Stimulus \* TRF)

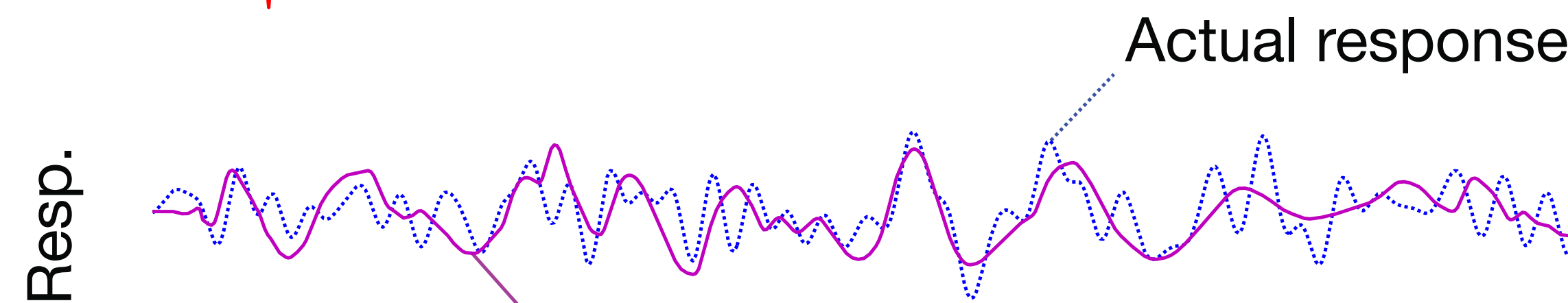
# TRF Model Estimation & Fit

## Temporal Response Function (TRF) estimation:

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



Estimated TRF

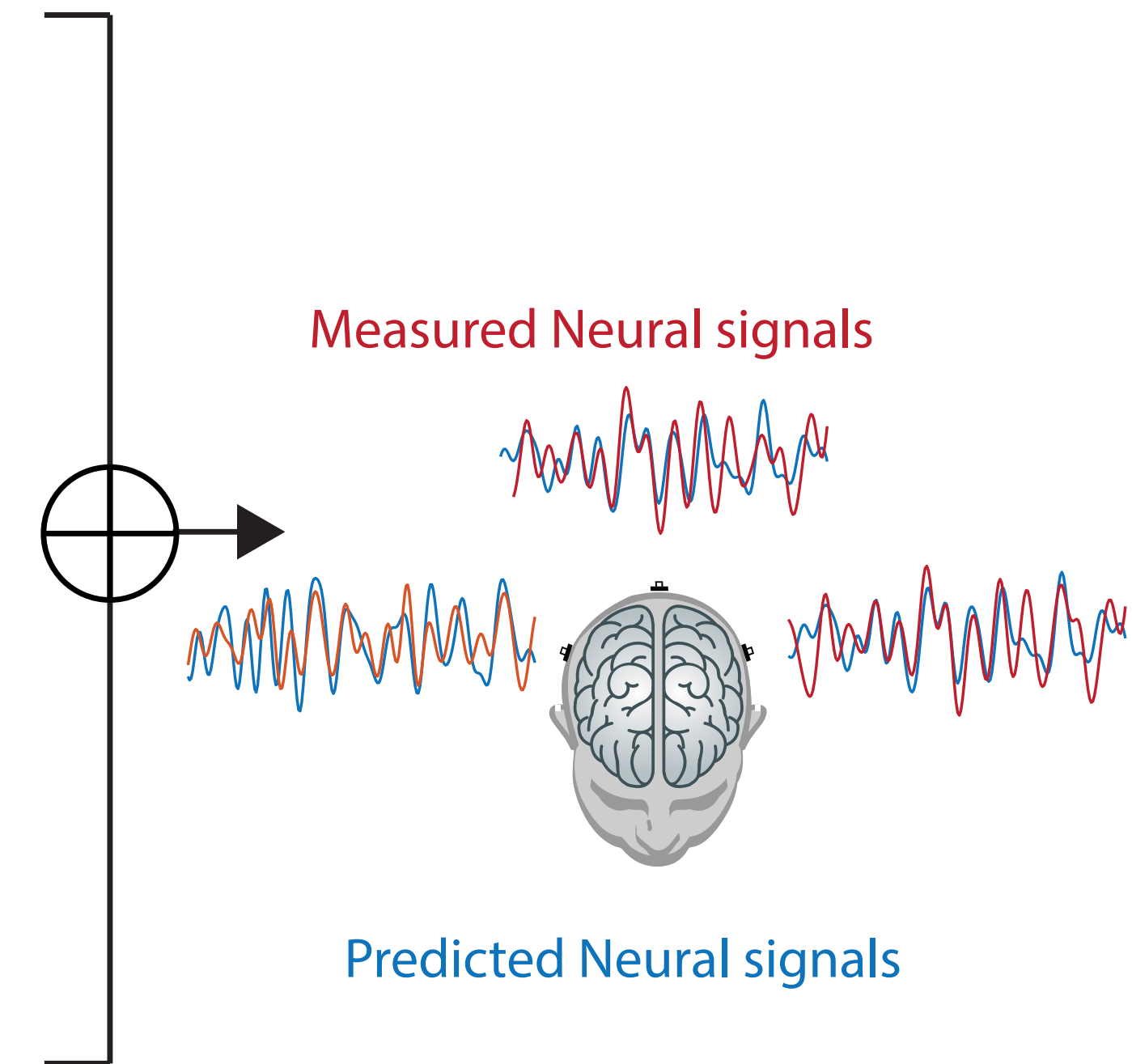
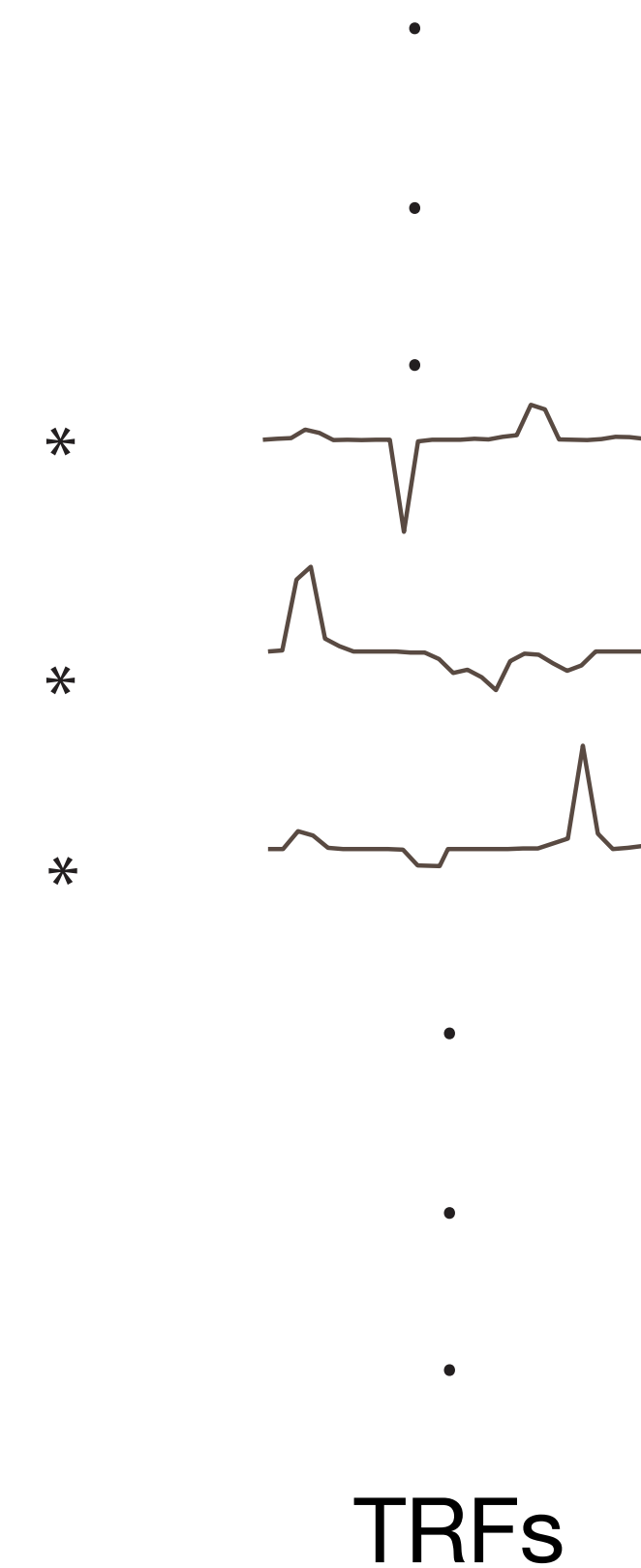
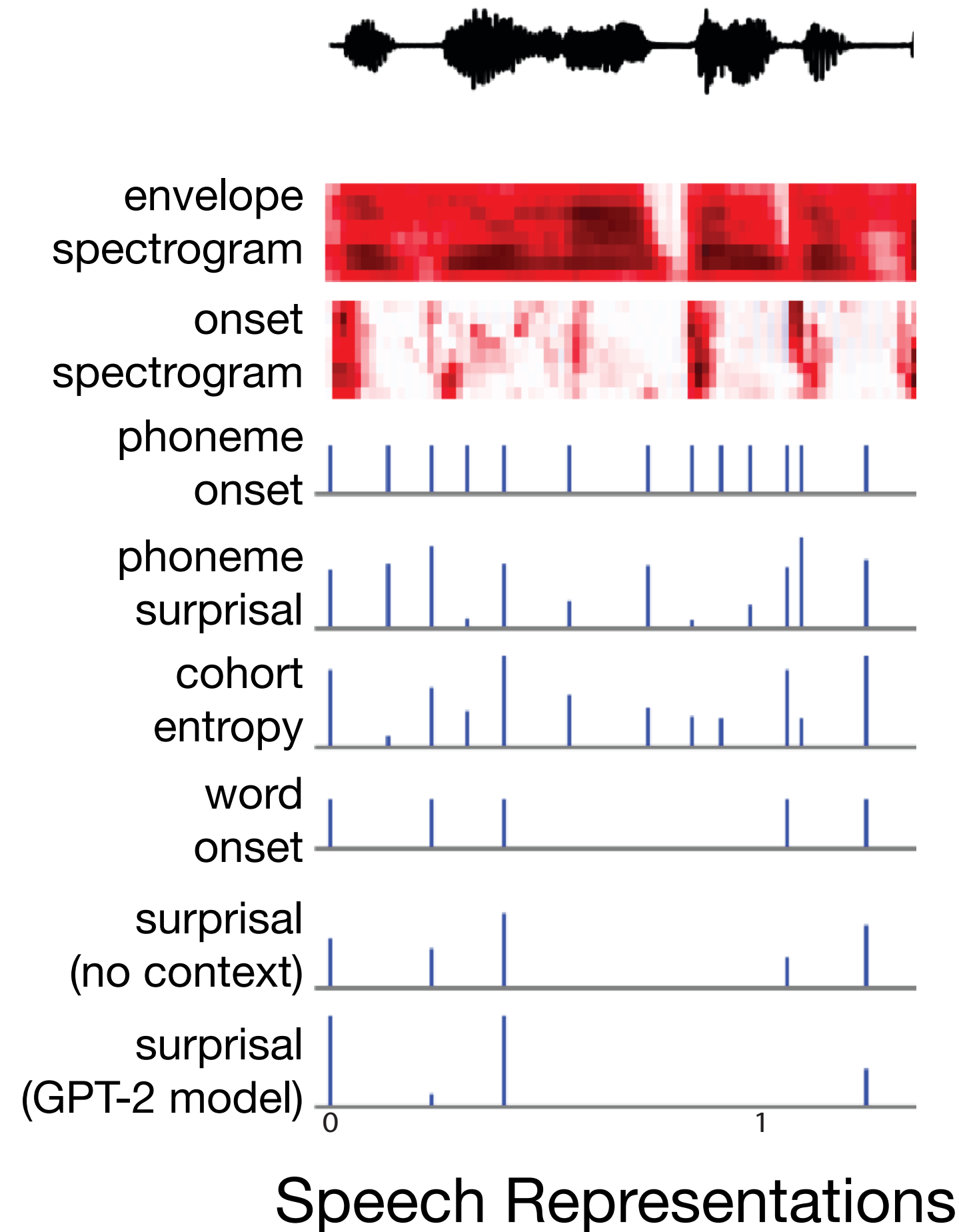


Predicted response (Stimulus \* TRF)

Actual response

# Simultaneous Temporal Response Functions

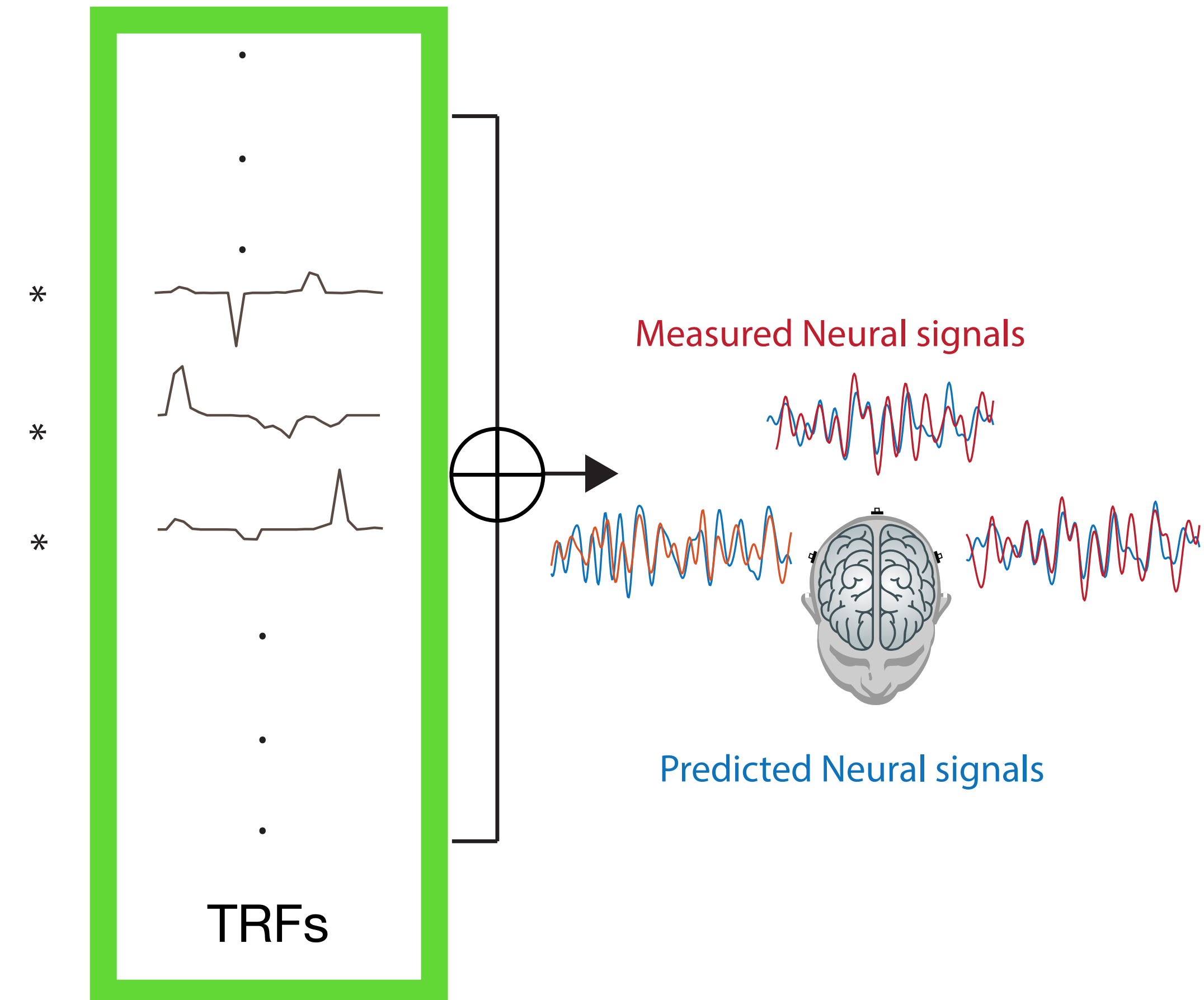
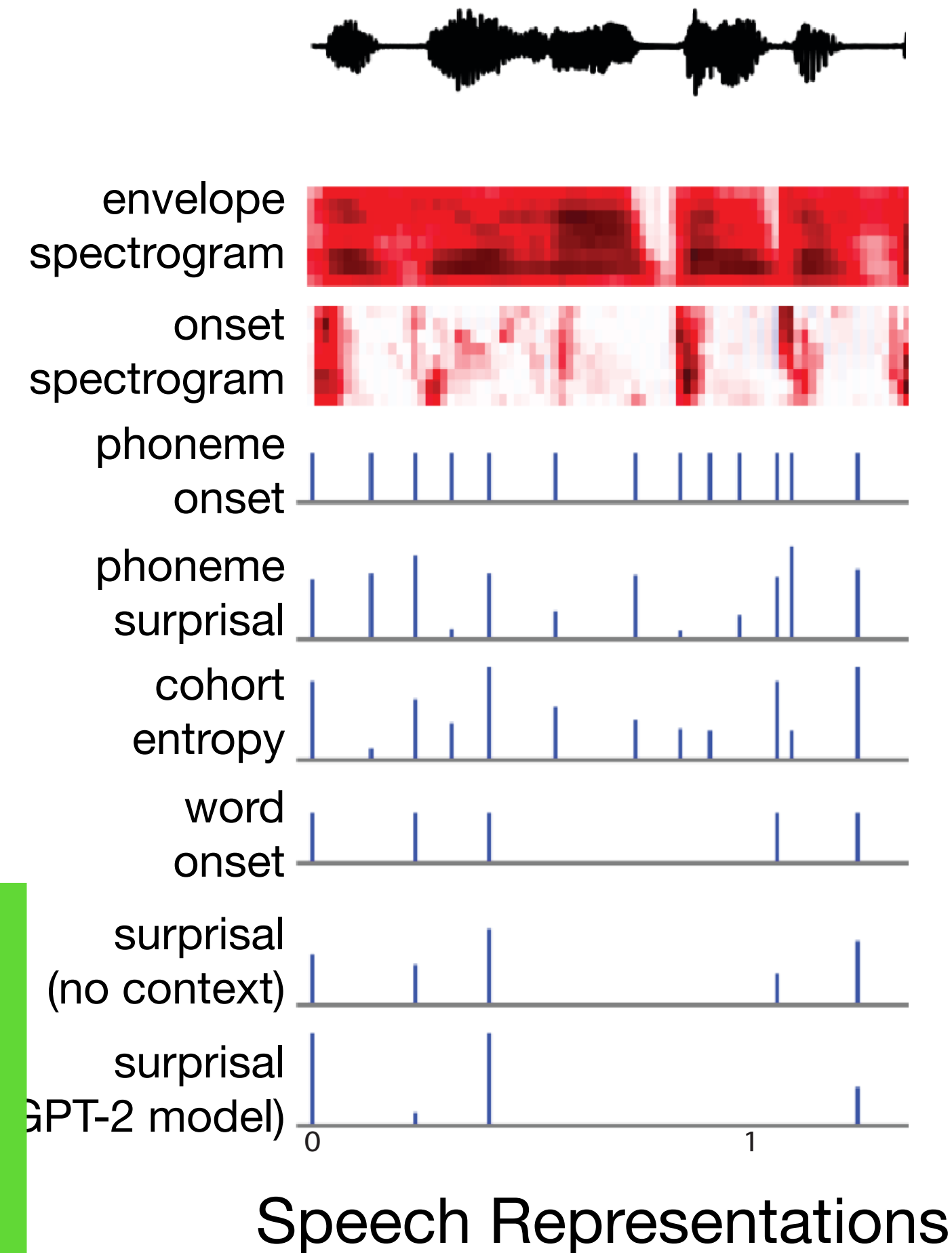
- TRFs predict neural response to speech
  - Analogous to evoked response
  - Peak amplitude  $\approx$  processing intensity
  - Peak Latency  $\approx$  source location
- Multiple TRFs estimated simultaneously
  - compete to explain variance (advantage over evoked response)



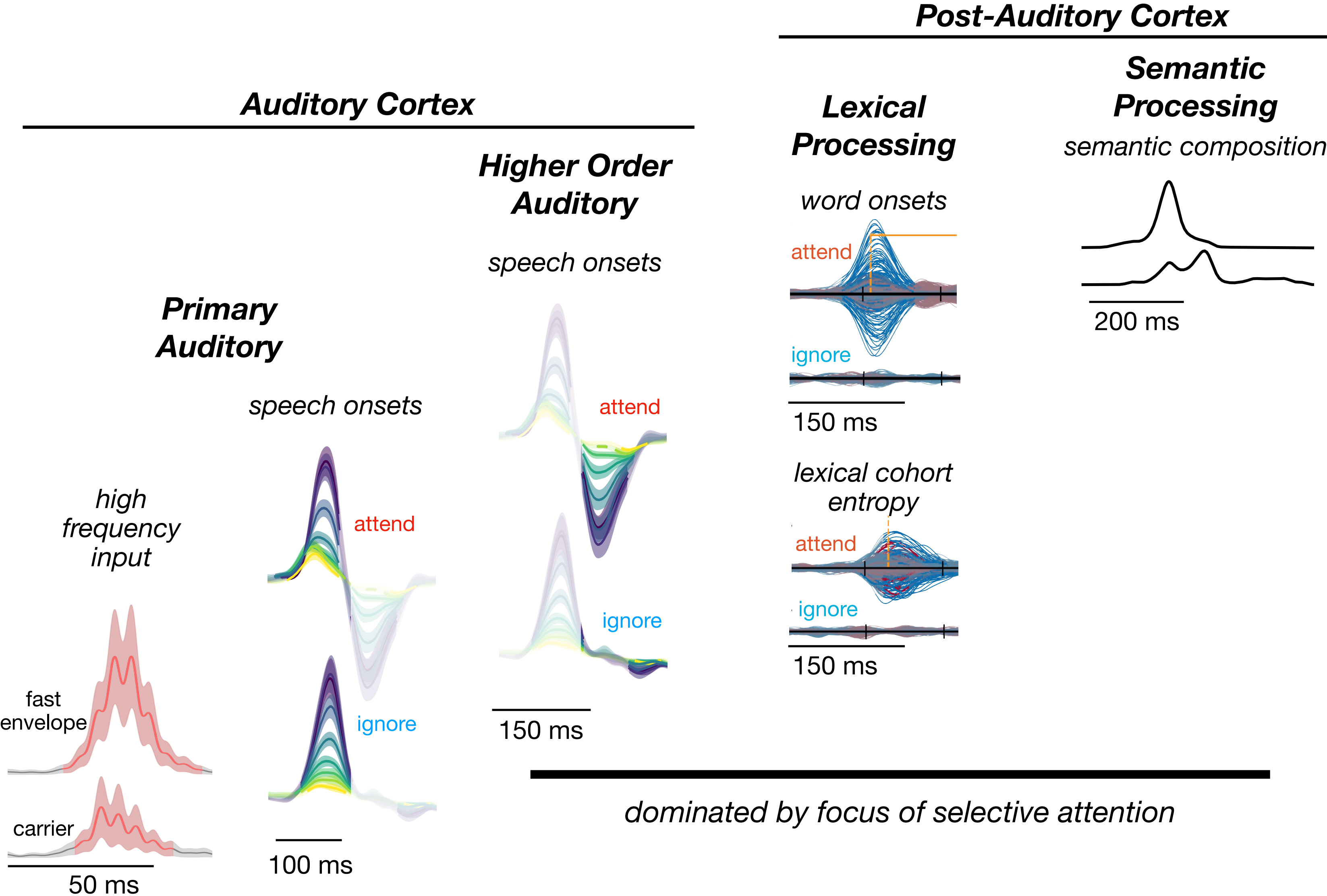
# Simultaneous Temporal Response Functions

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

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

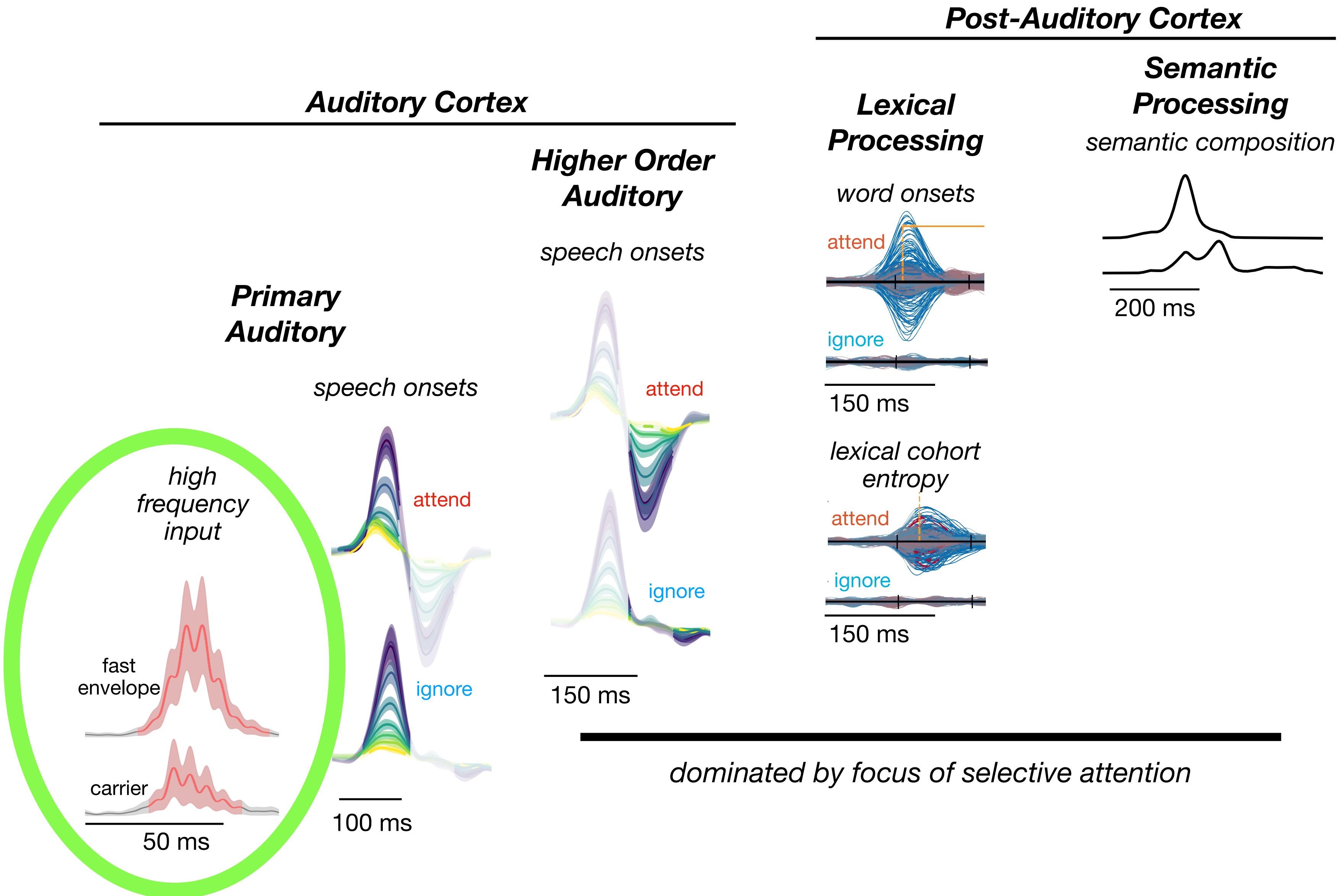


# Cortical Representations Across Cortex

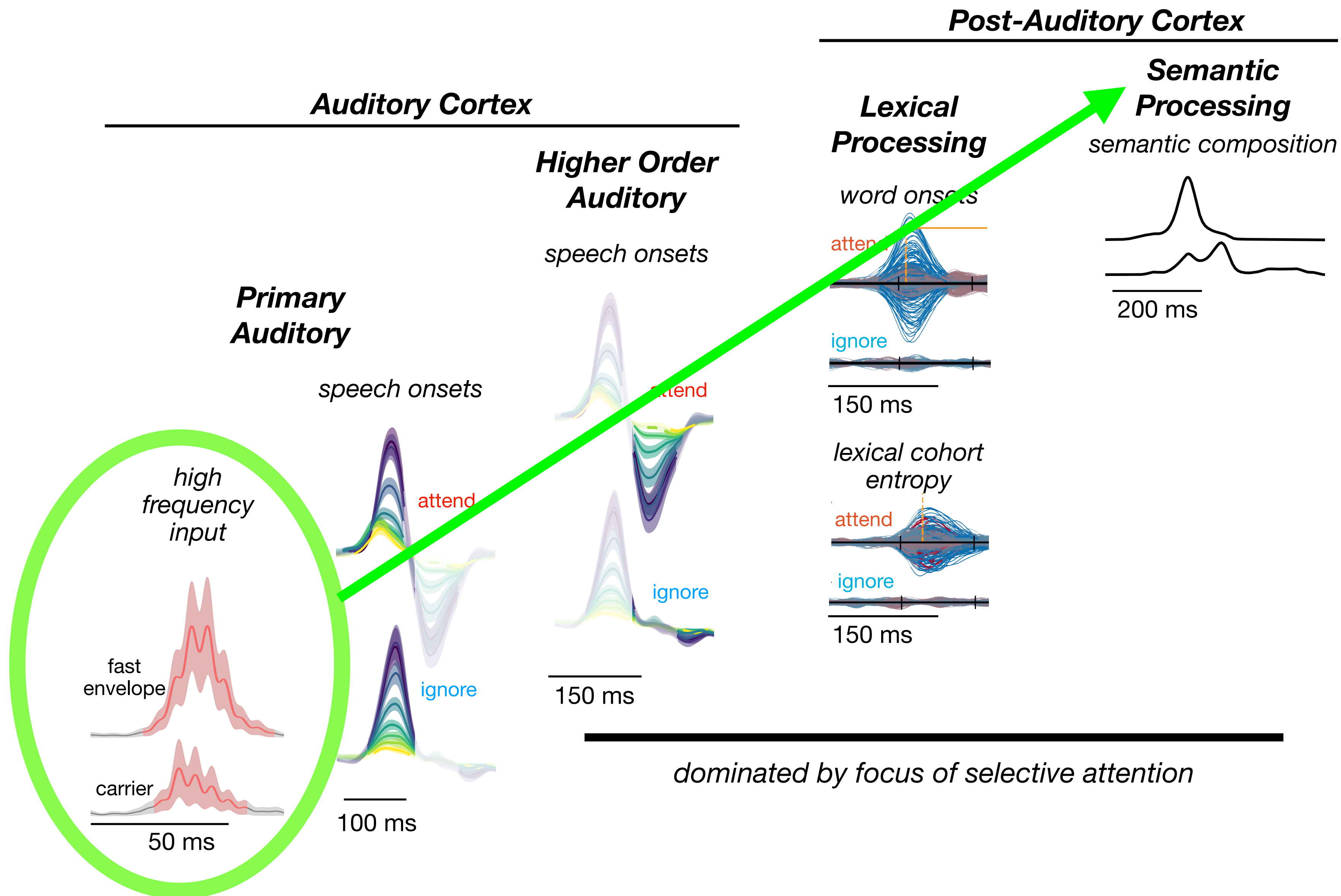




# Cortical Representations Across Cortex



# Cortical Representations Across Cortex

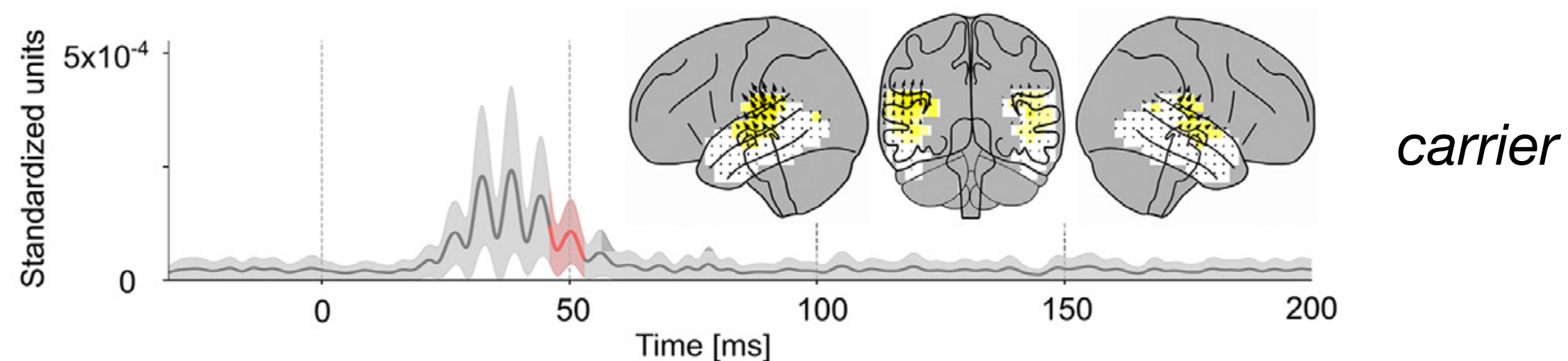
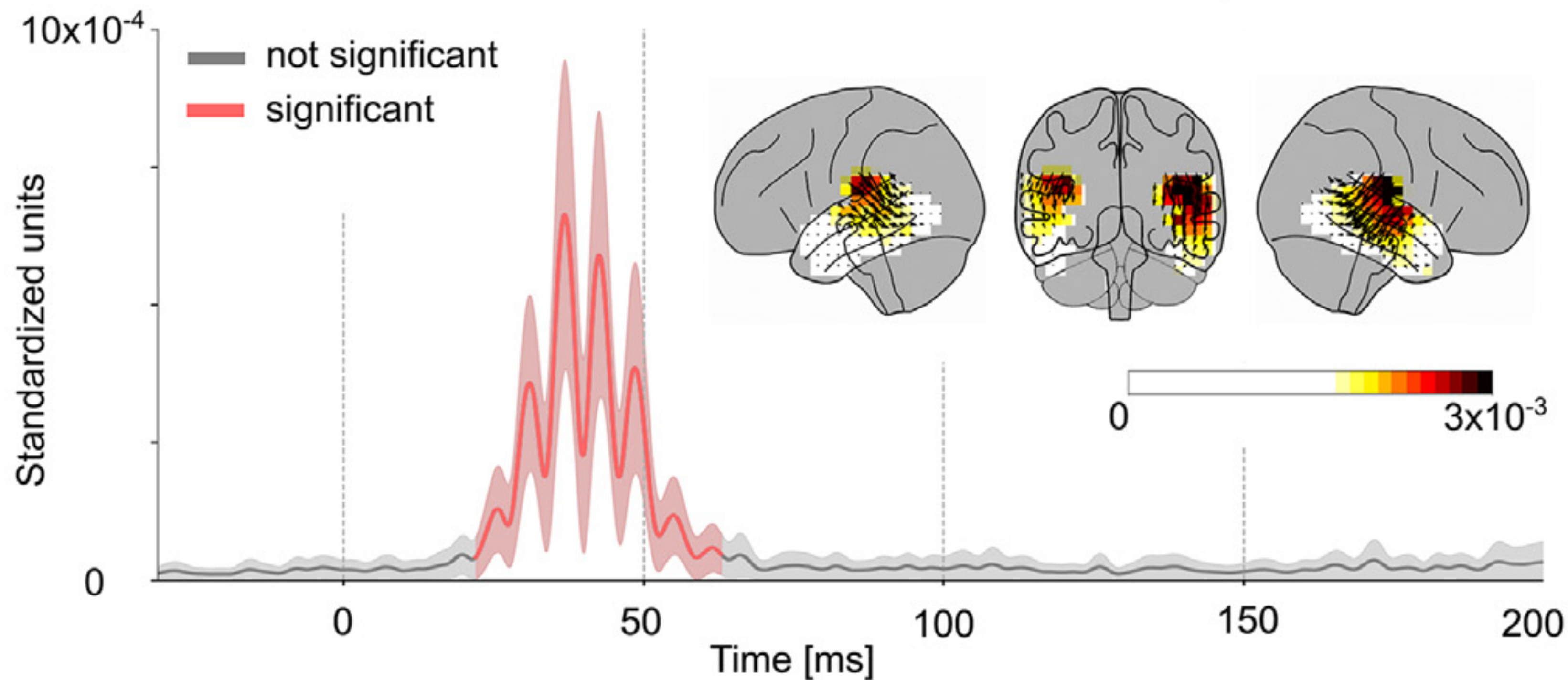


# Outline

- 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*



# Fast & Early Cortical Representations



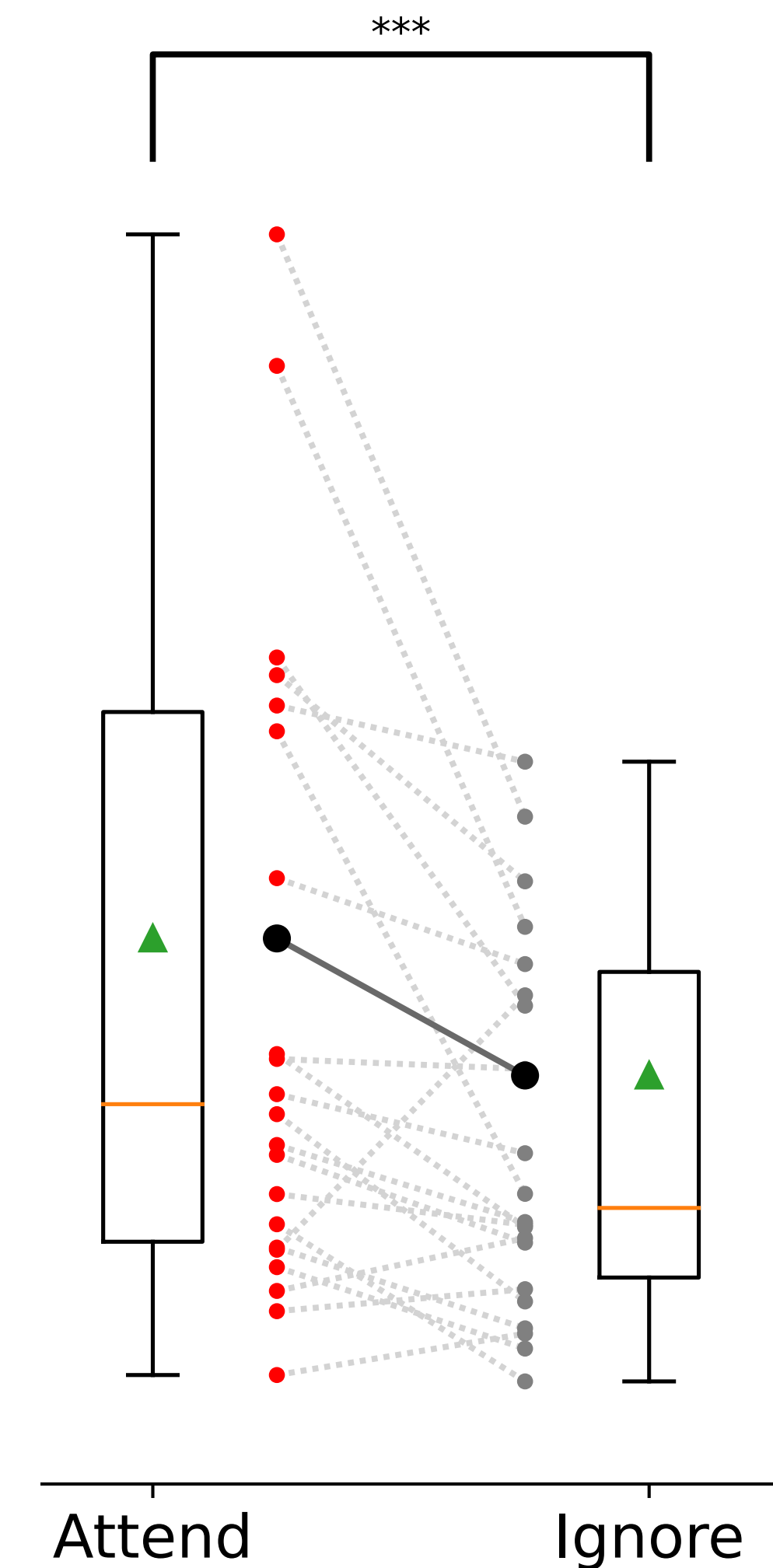
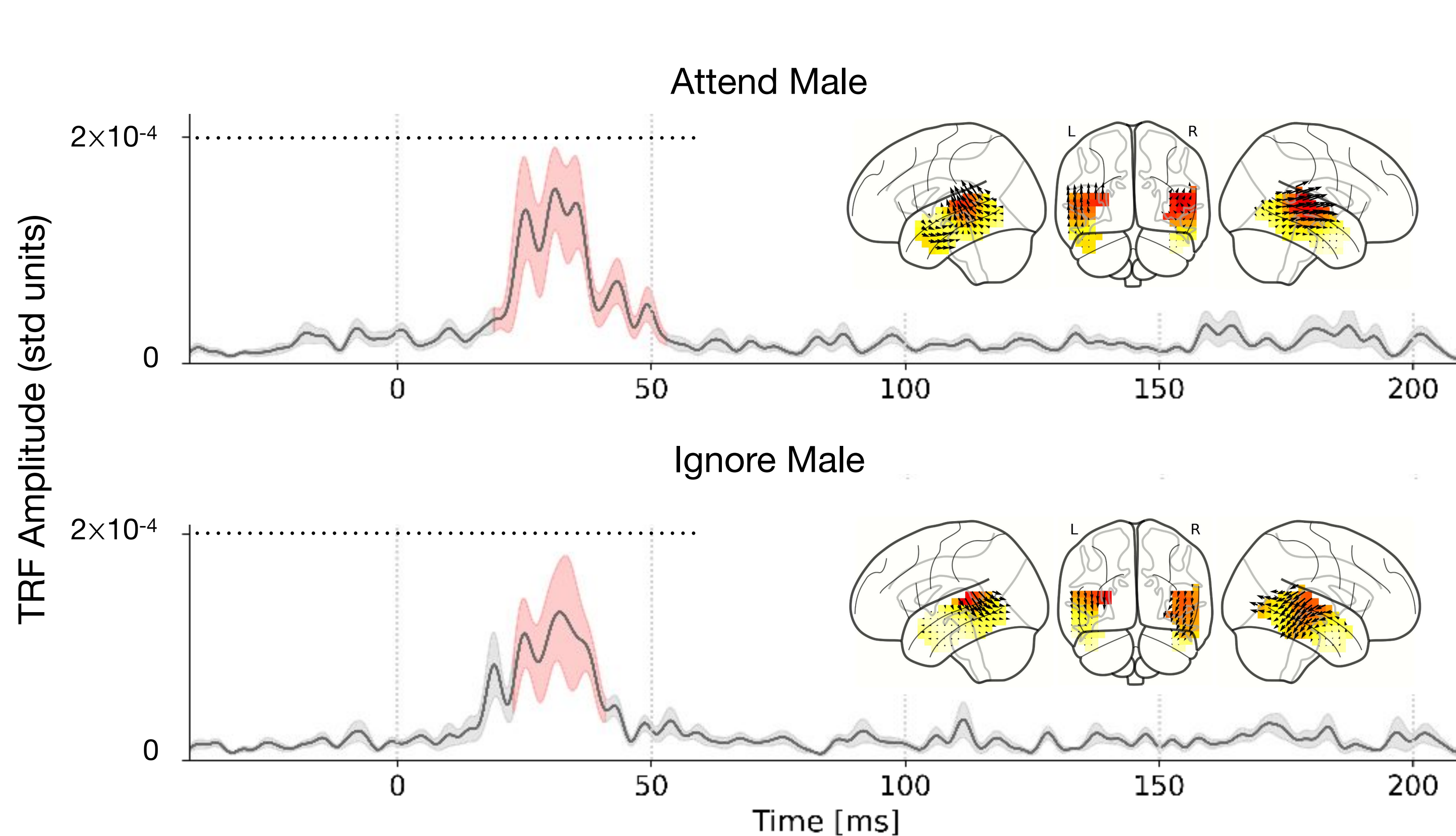
40 ms latency peak  
⇒ Primary/Core auditory cortex

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 modulated by selective attention

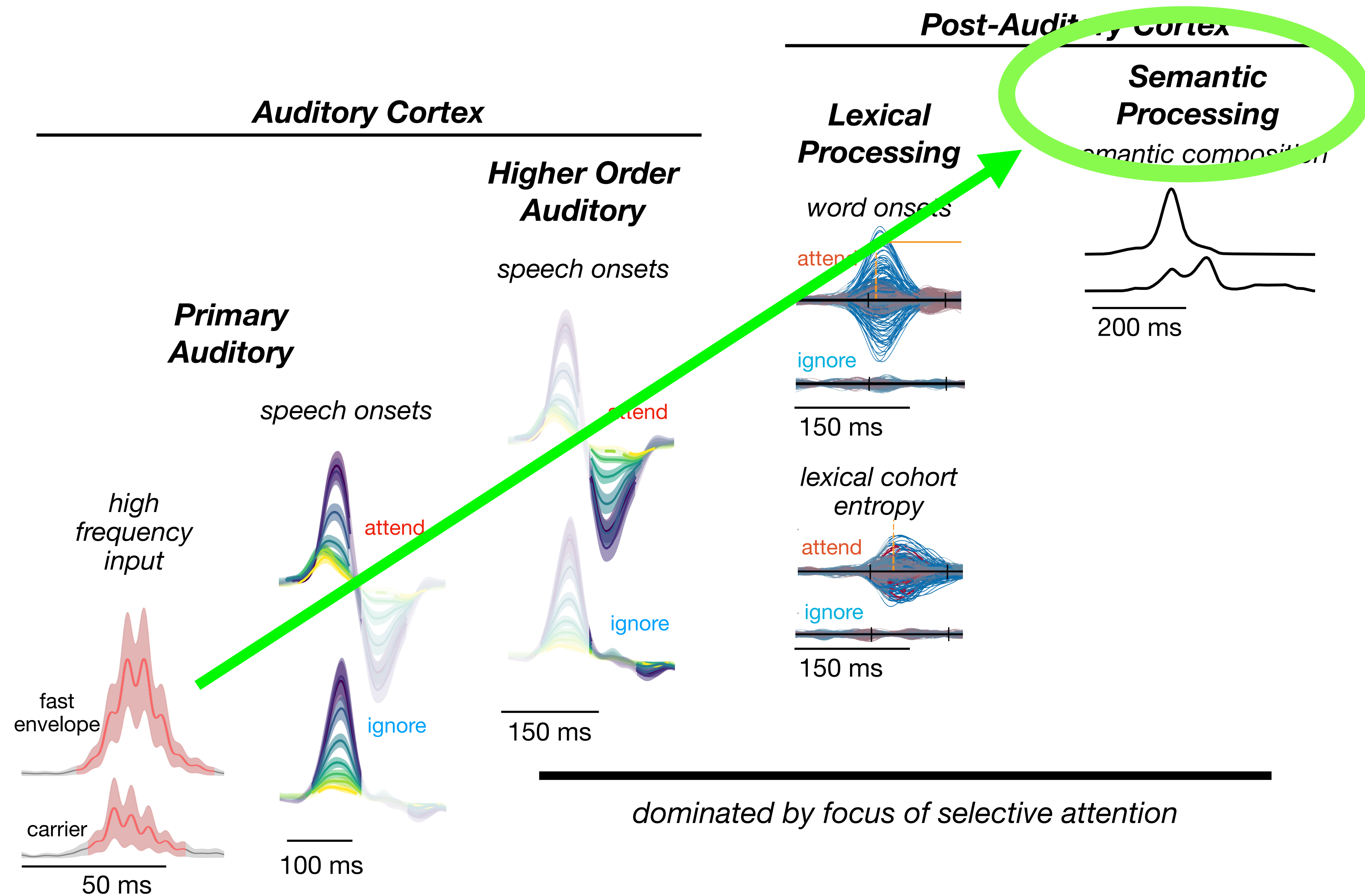




# Outline

- 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*

# Cortical Representations Across Cortex



# Speech Understanding/Meaning

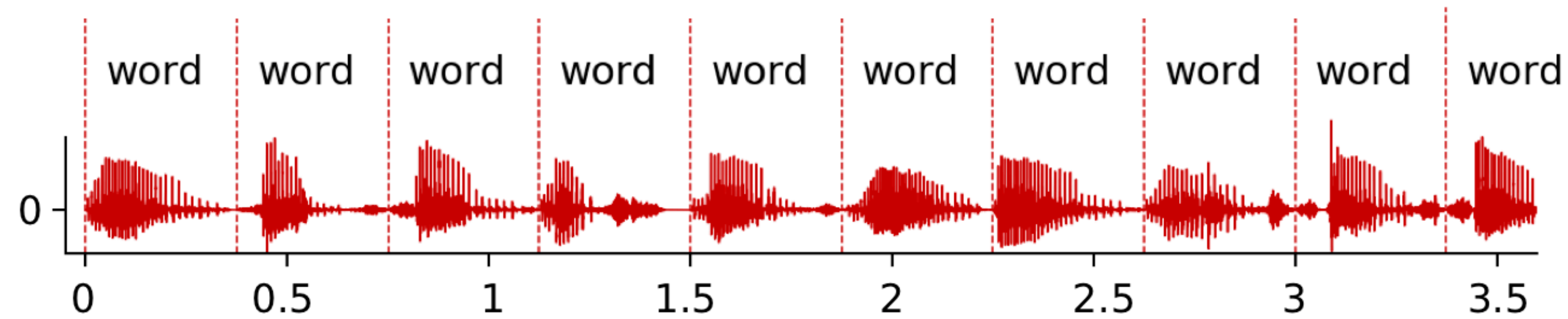
- Behavioral correlates of speech understanding
  - implies language comprehension
  - structural comprehension
    - sentence structure
    - other structures, e.g. poetic, logical
- Neural correlates of speech understanding
  - rhythms of structural comprehension/meaning, even if *fully absent in the acoustics*
    - sentence structures Ding et al., Nat Neurosci 2016
    - poetic structures Teng et al., Curr Biol 2020
    - mathematical structures

# Speech Understanding/Meaning

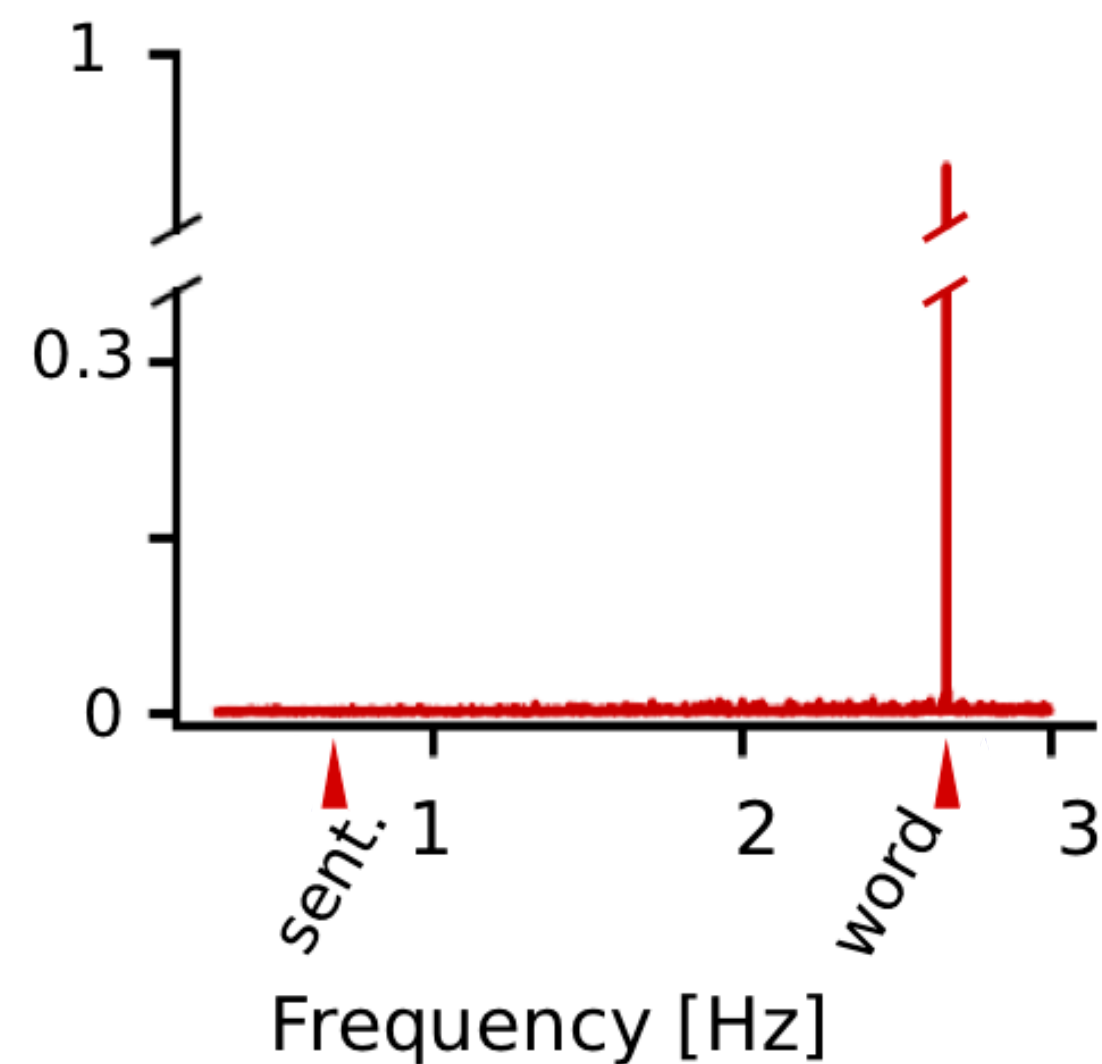
- Behavioral correlates of speech understanding
  - implies language comprehension
  - structural comprehension
    - sentence structure
    - other structures, e.g. poetic, logical
- Neural correlates of speech understanding
  - rhythms of structural comprehension/meaning, even if *fully absent in the acoustics*
    - sentence structures Ding et al., Nat Neurosci 2016
    - poetic structures Teng et al., Curr Biol 2020
    - mathematical structures

# Isochronous Speech

Acoustics



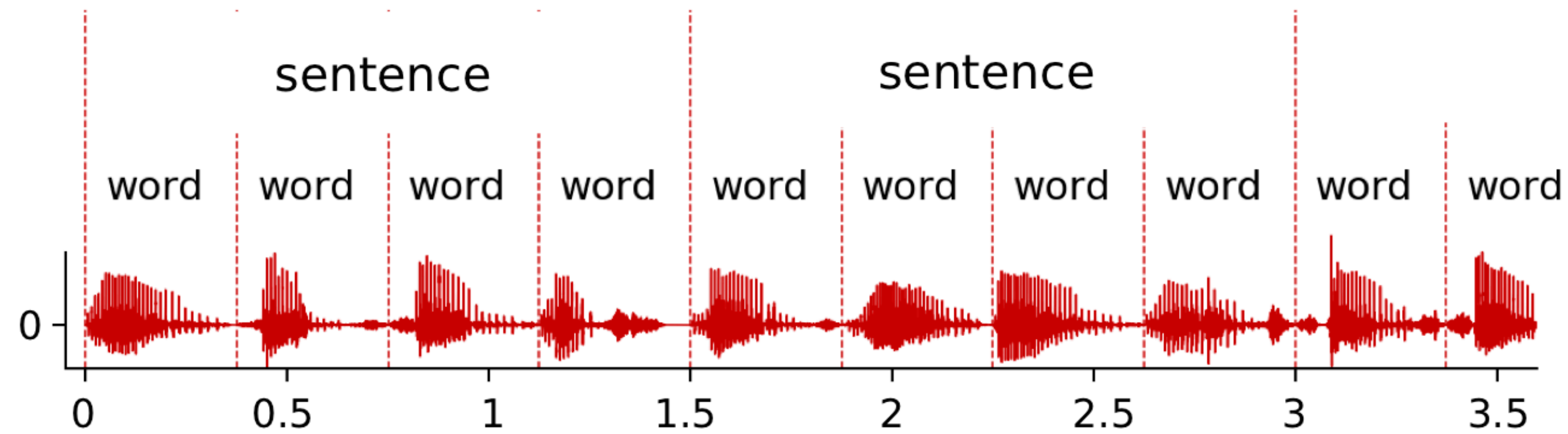
Acoustical  
Spectrum  
(envelope)



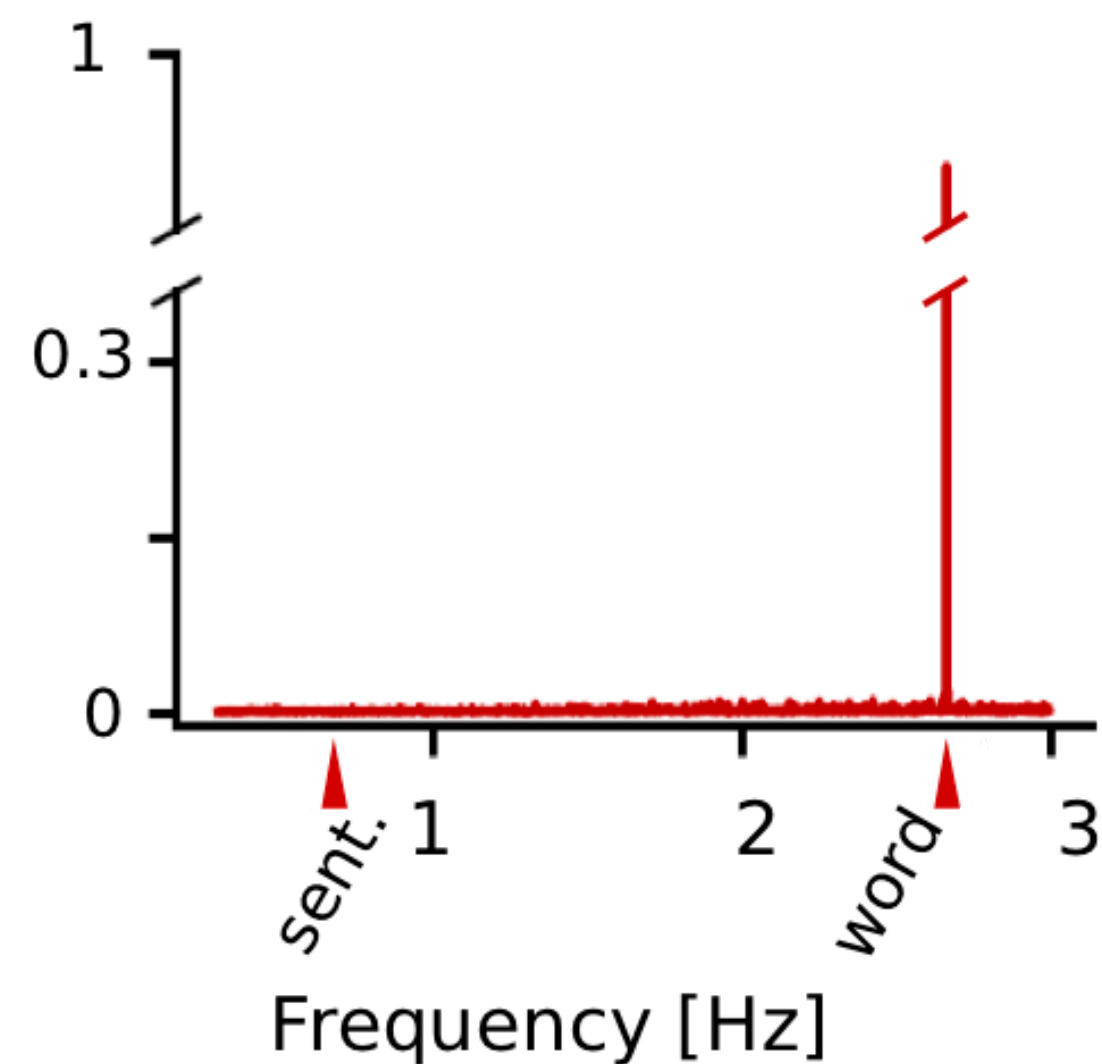


# Isochronous Speech

Acoustics

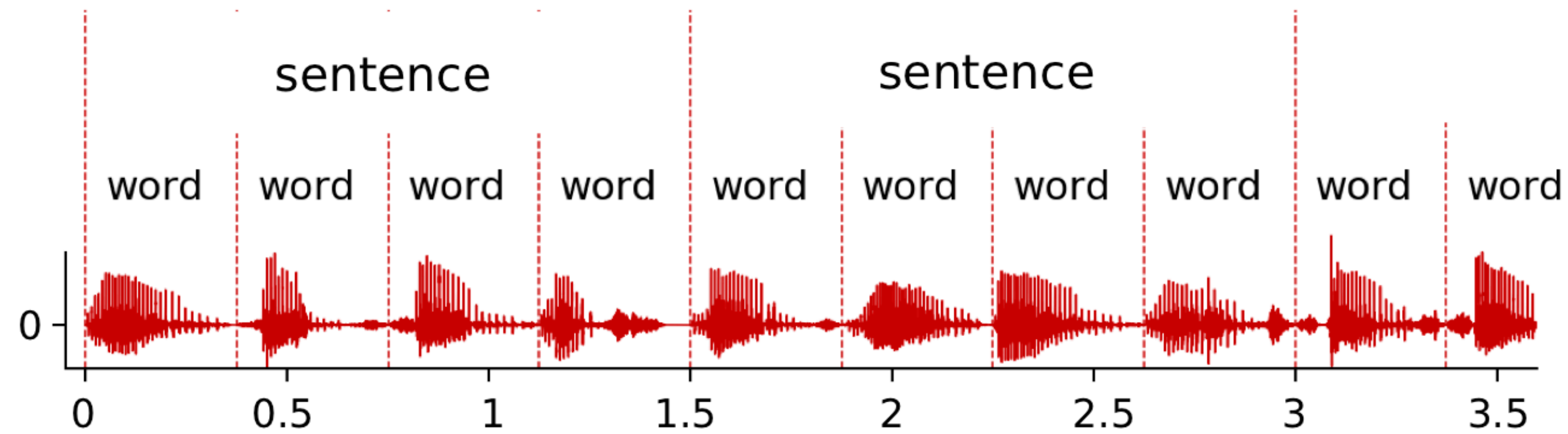


Acoustical  
Spectrum  
(envelope)

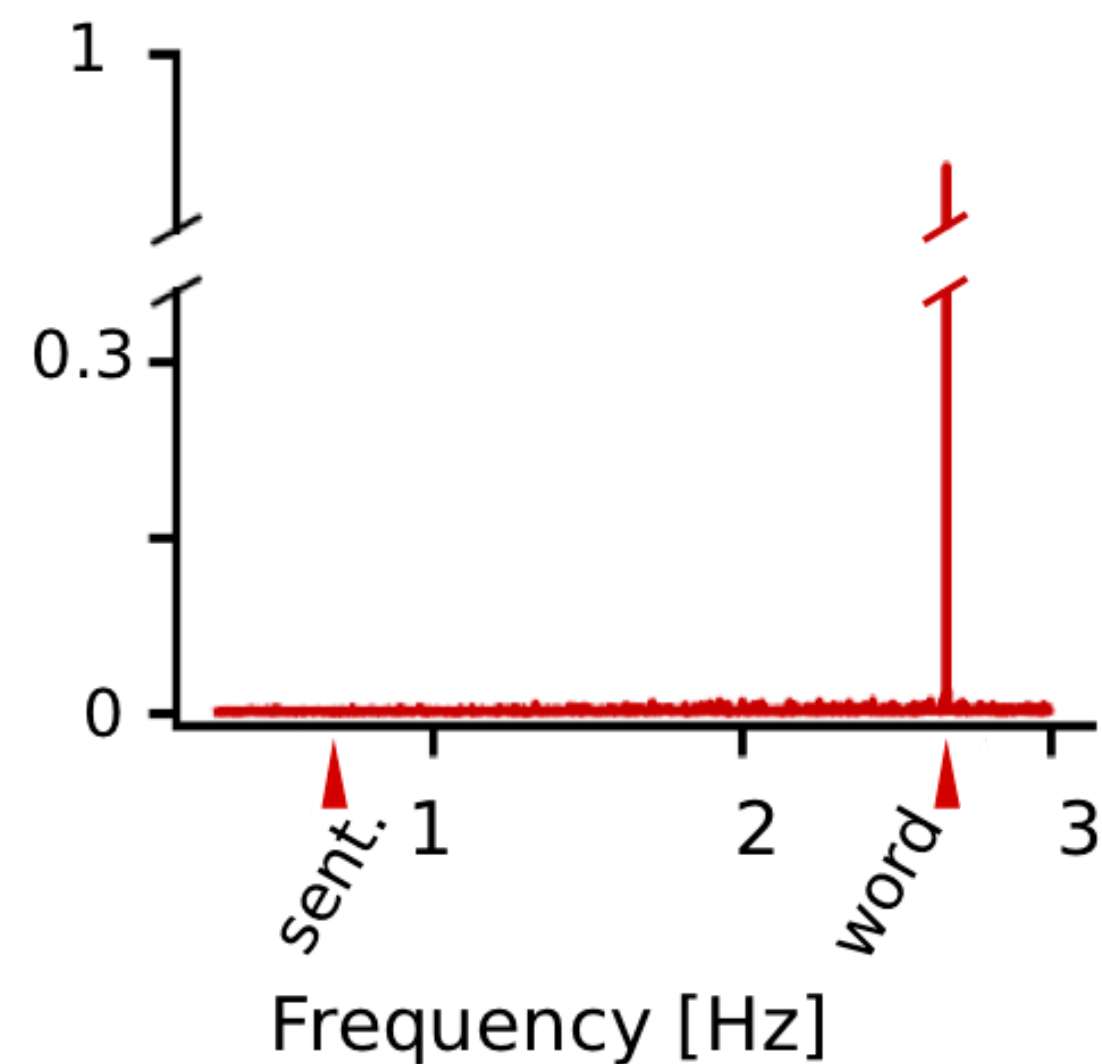


# Isochronous Speech

Acoustics

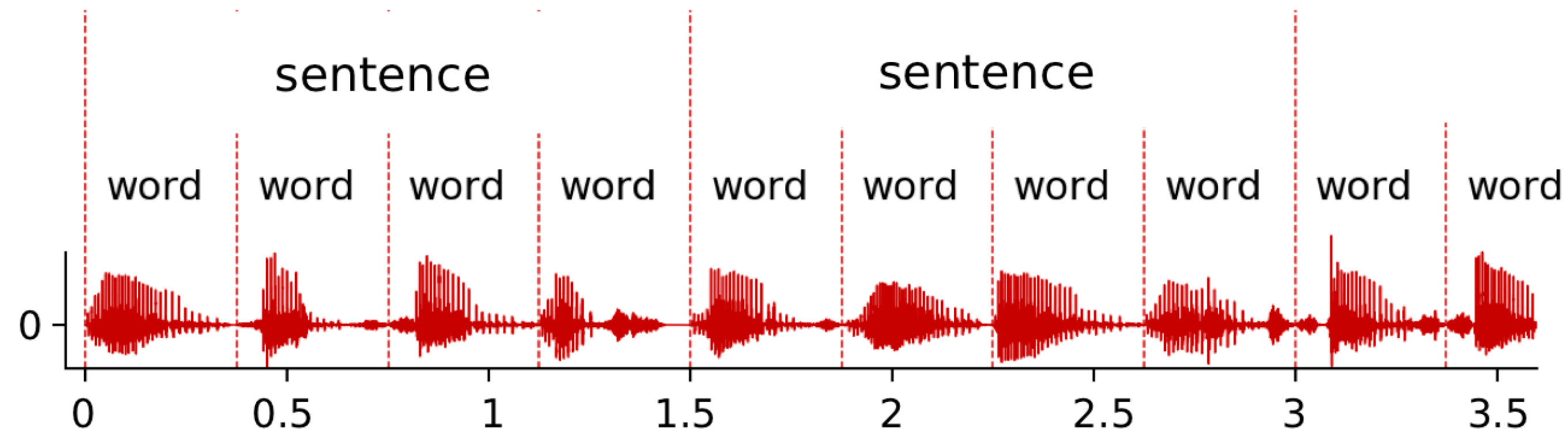


Acoustical  
Spectrum  
(envelope)

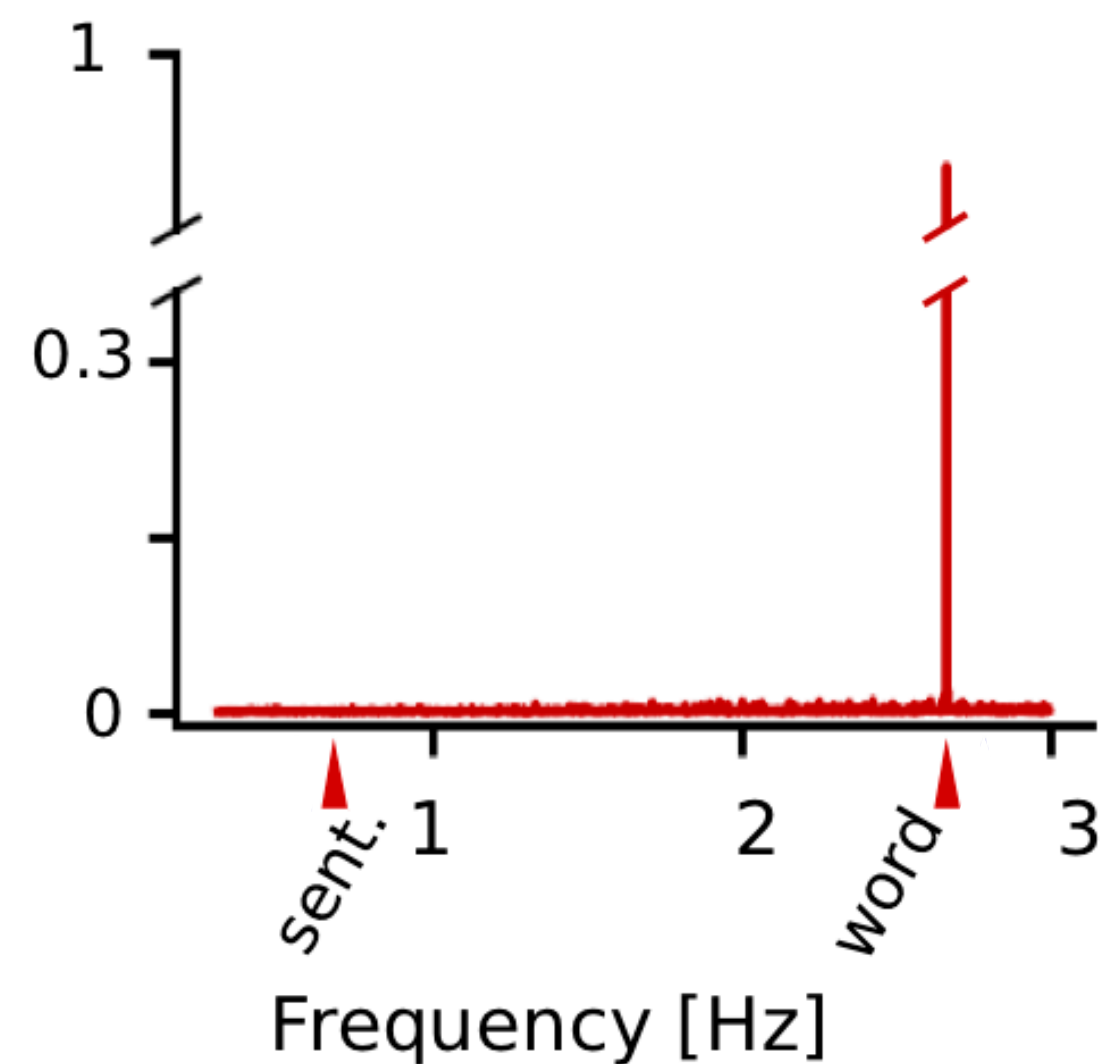


# Isochronous Speech

Acoustics

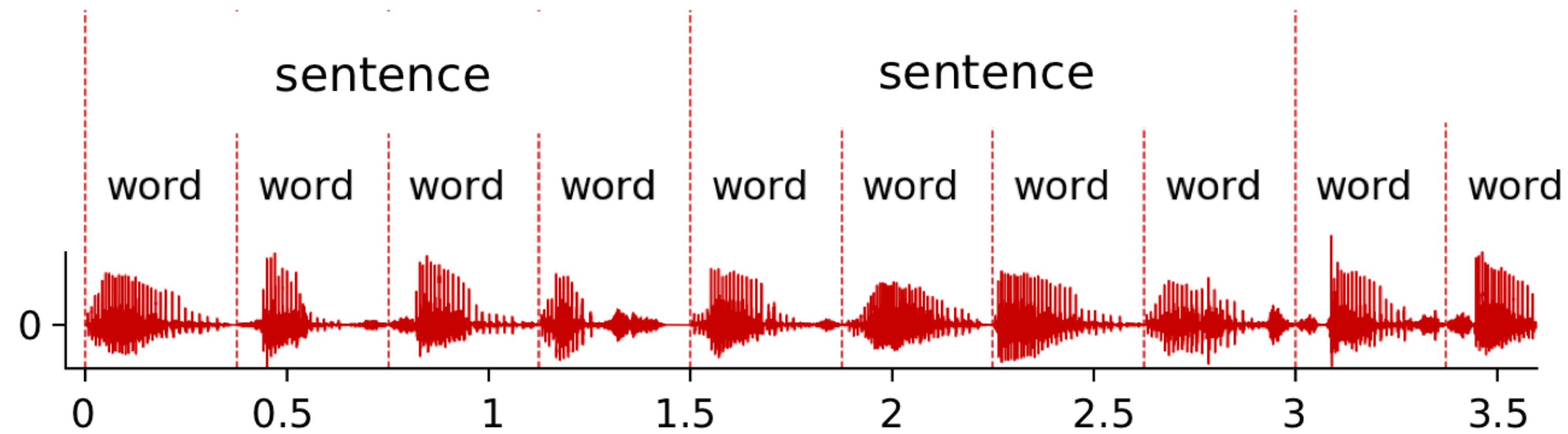


Acoustical  
Spectrum  
(envelope)

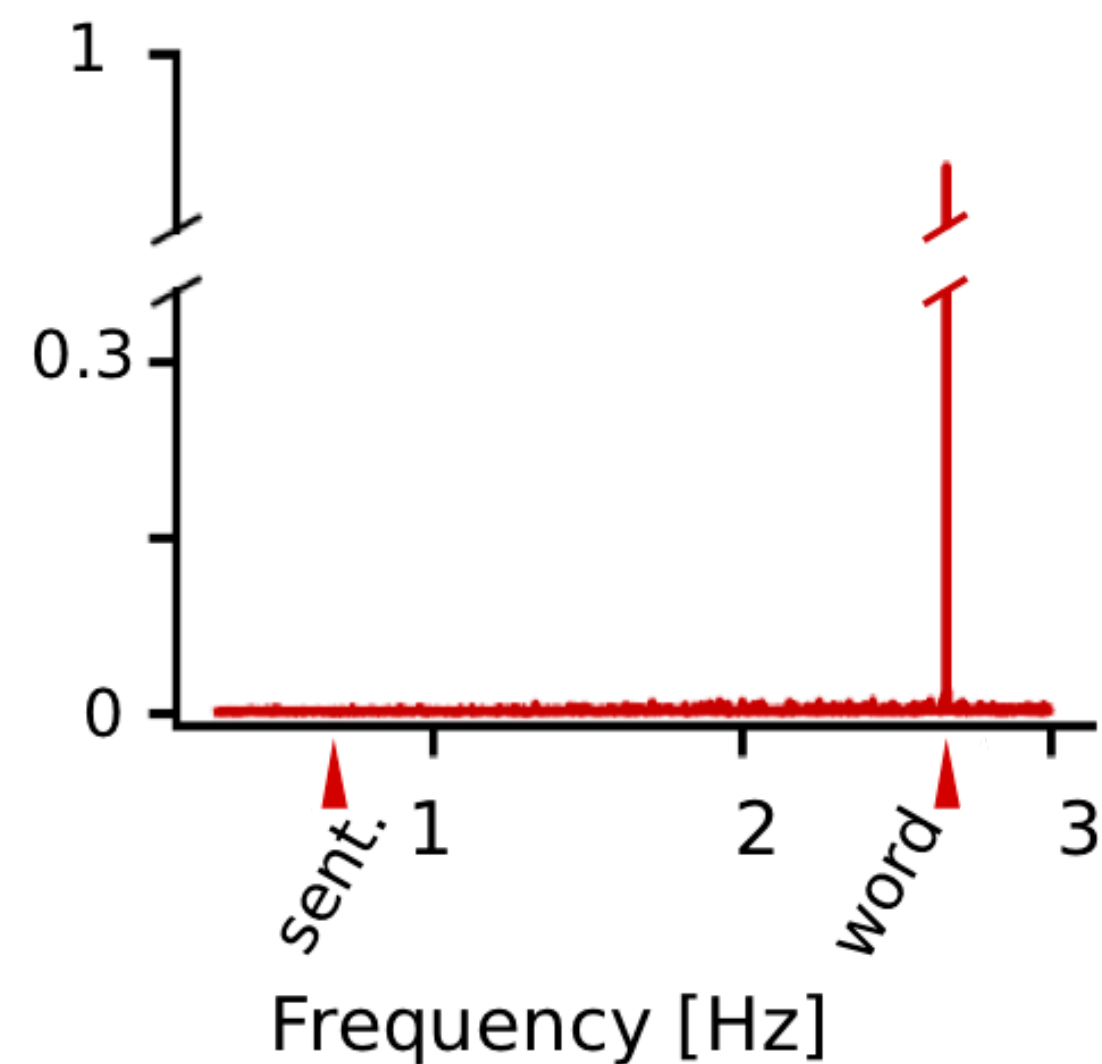


# Isochronous Speech

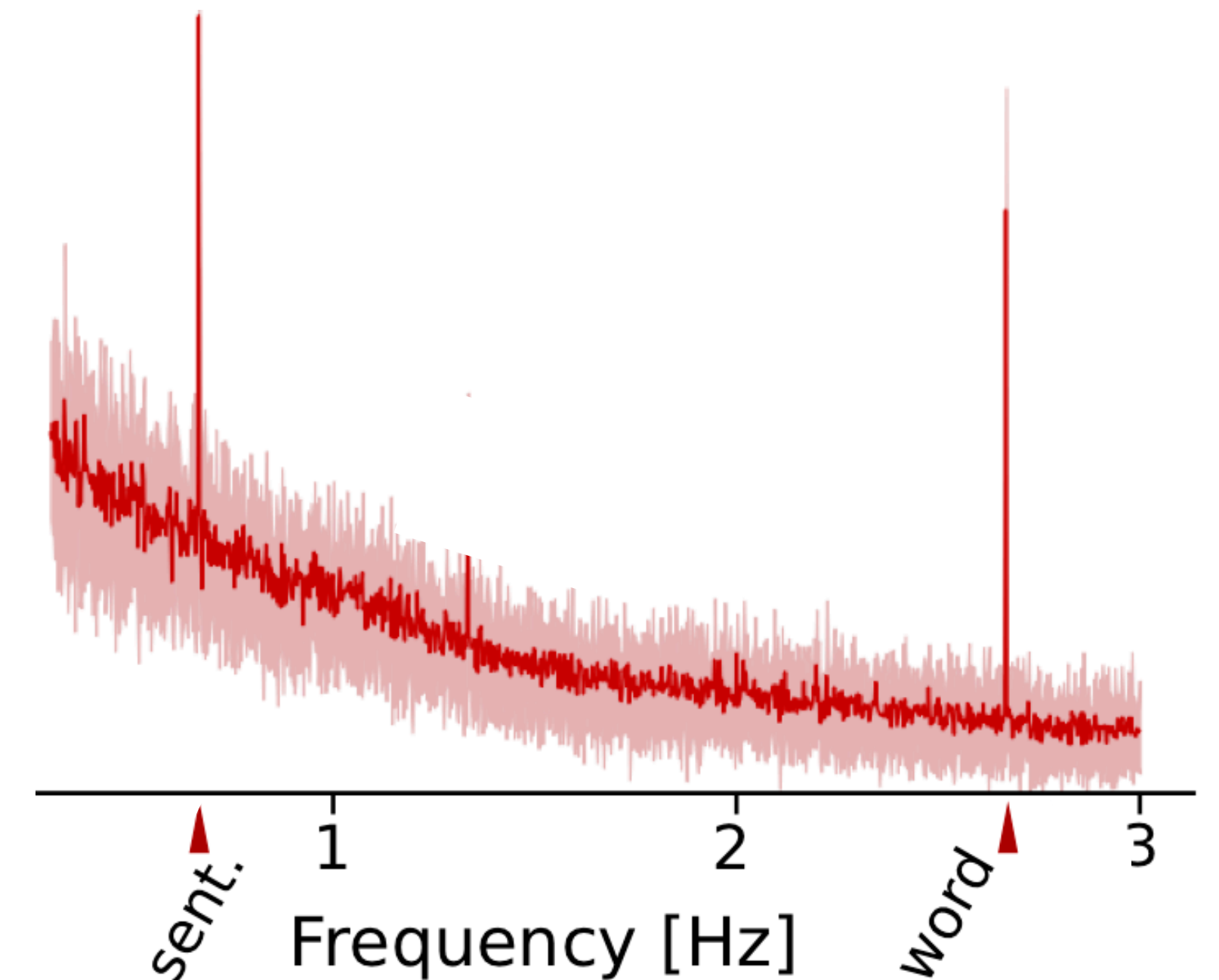
Acoustics



Acoustical  
Spectrum  
(envelope)

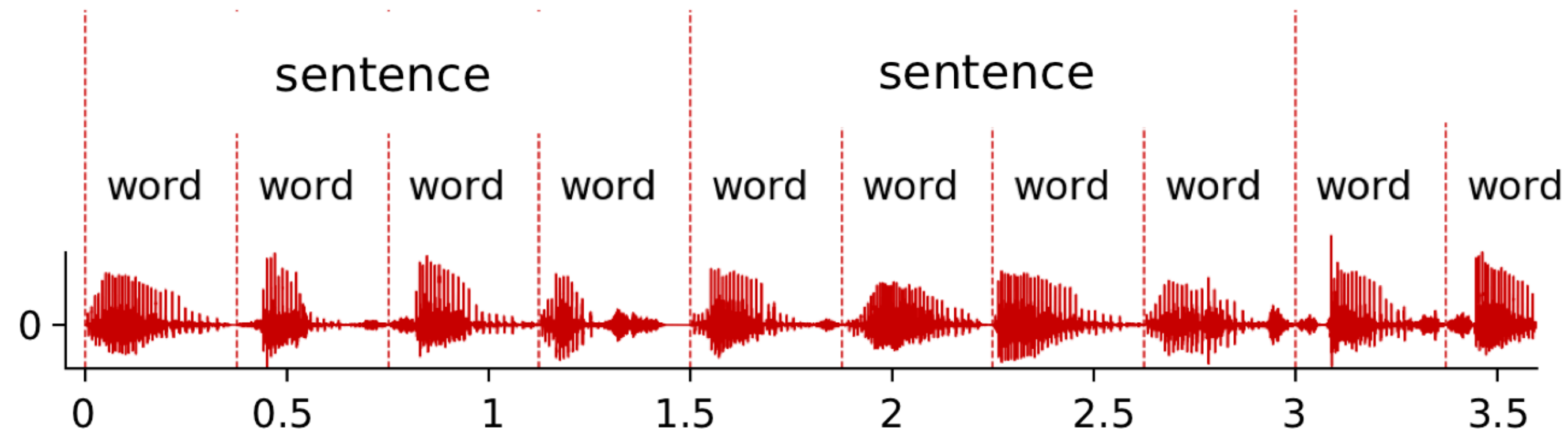


Perception?

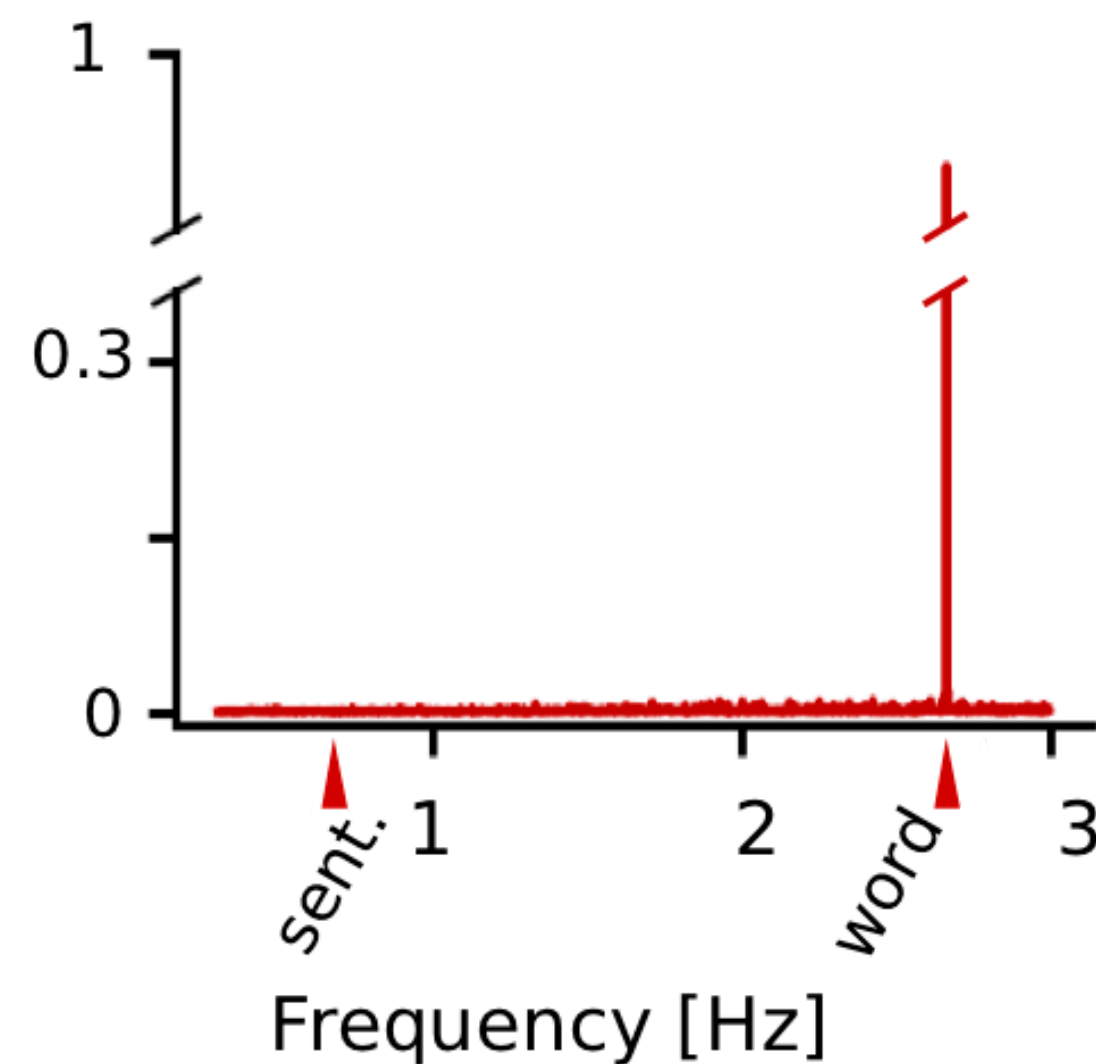


# Isochronous Speech

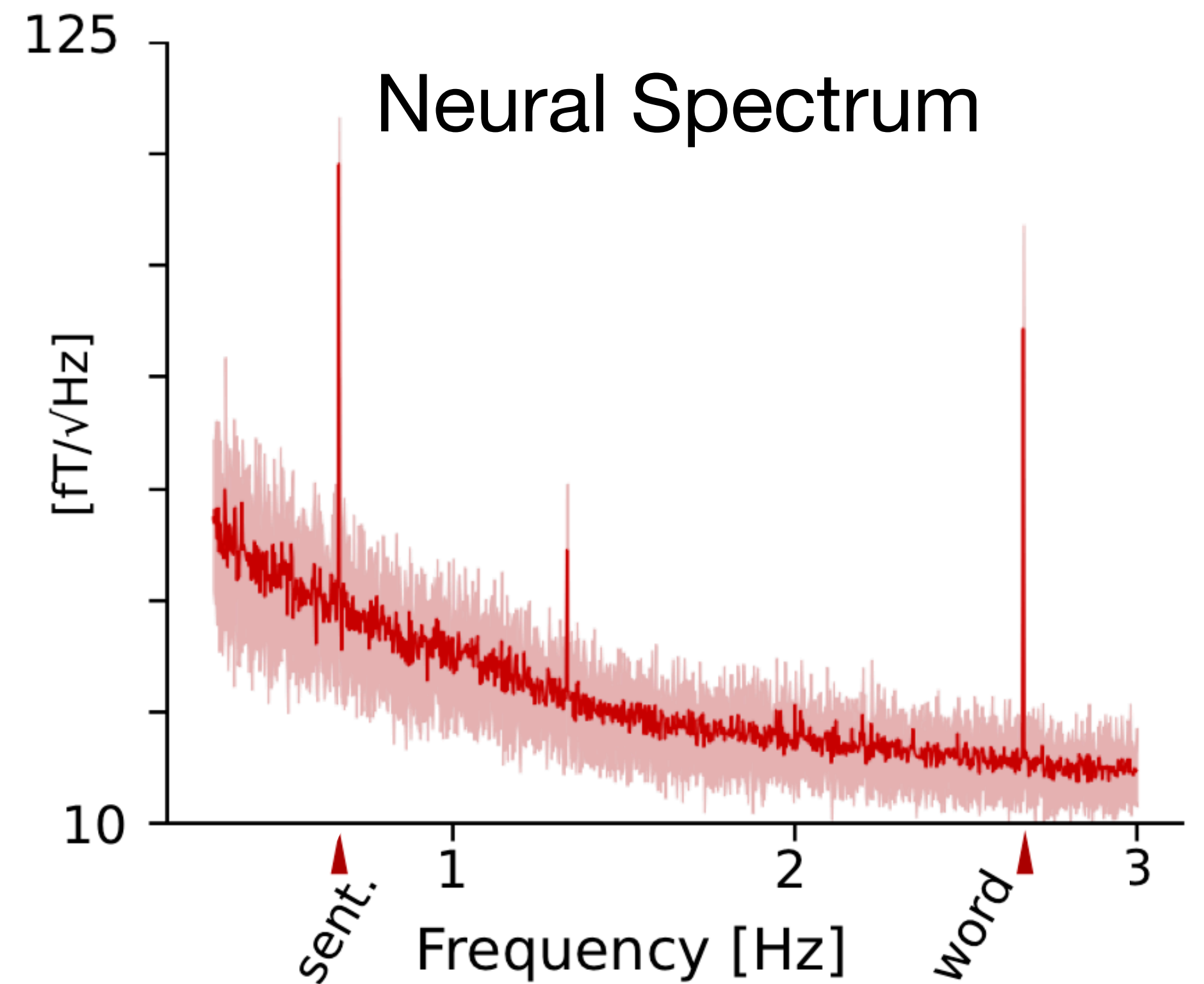
Acoustics



Acoustical  
Spectrum  
(envelope)

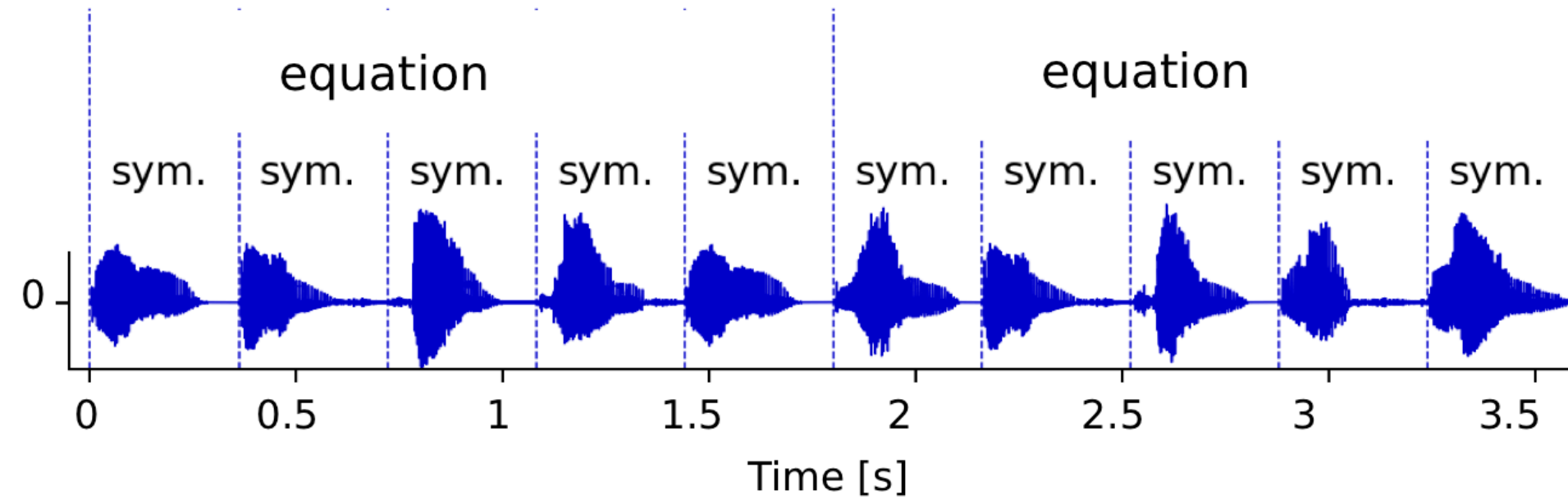


Neural Spectrum

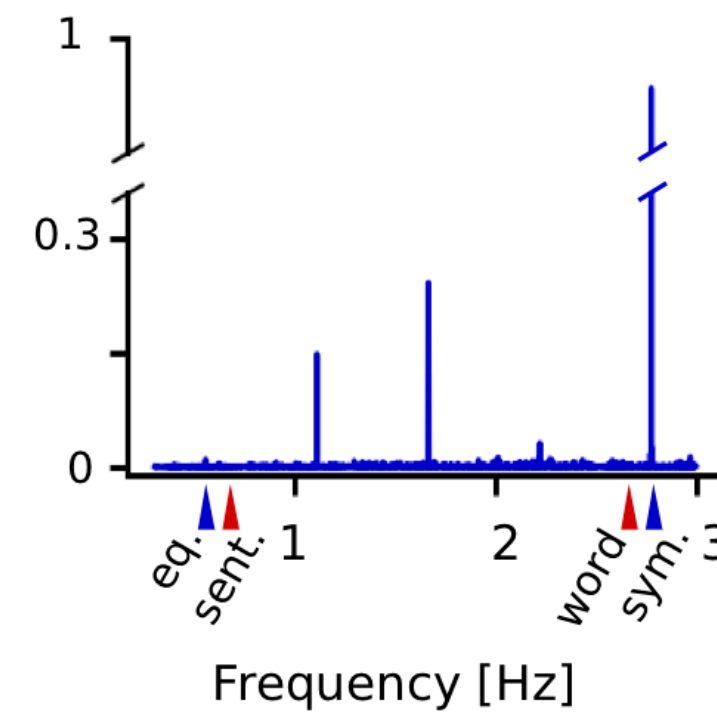




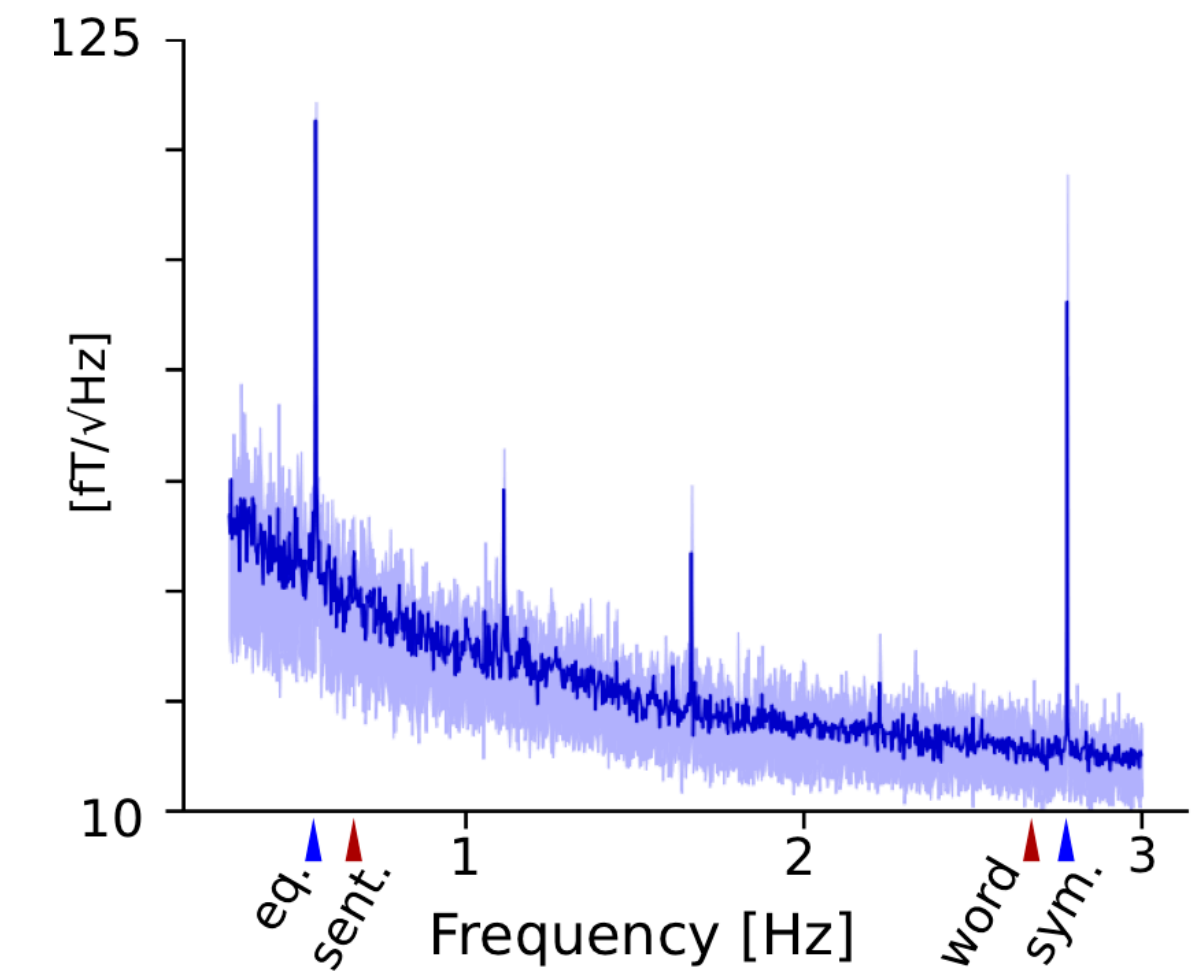
# Isochronous Arithmetic



# Acoustics



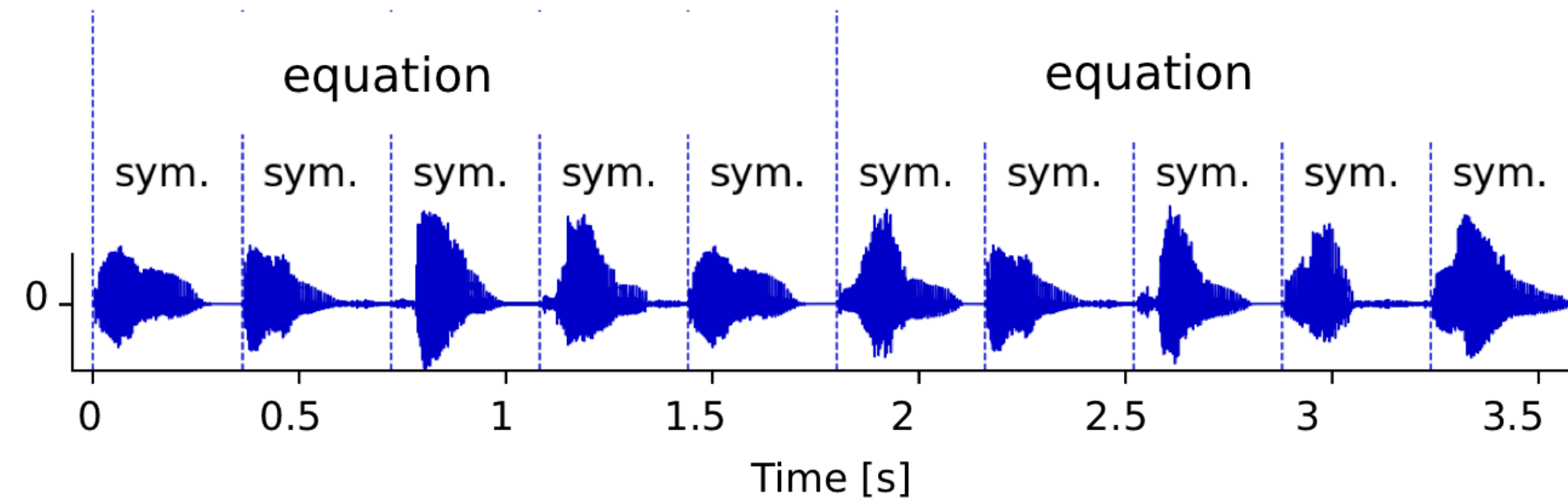
# Acoustical Spectrum



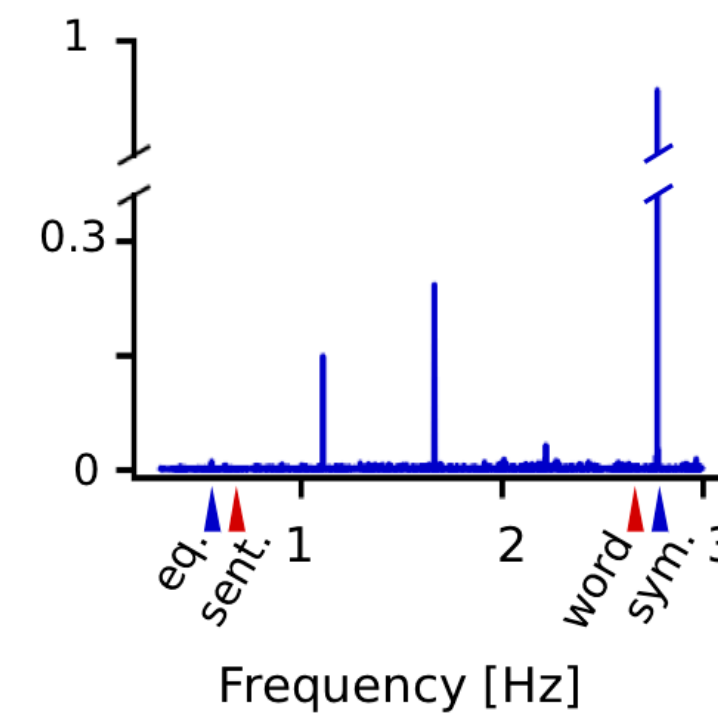
# Neural Spectrum



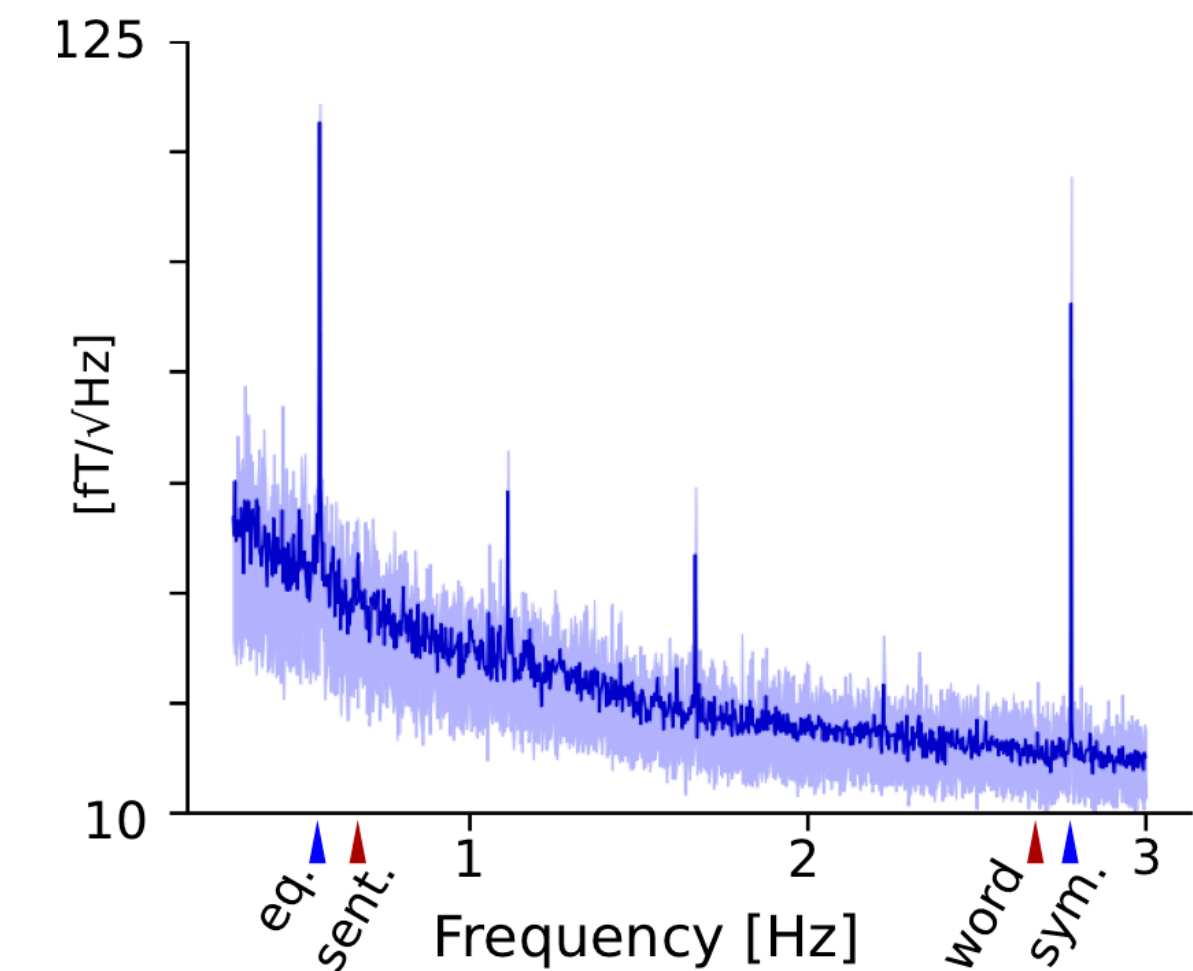
# Isochronous Arithmetic



Acoustics



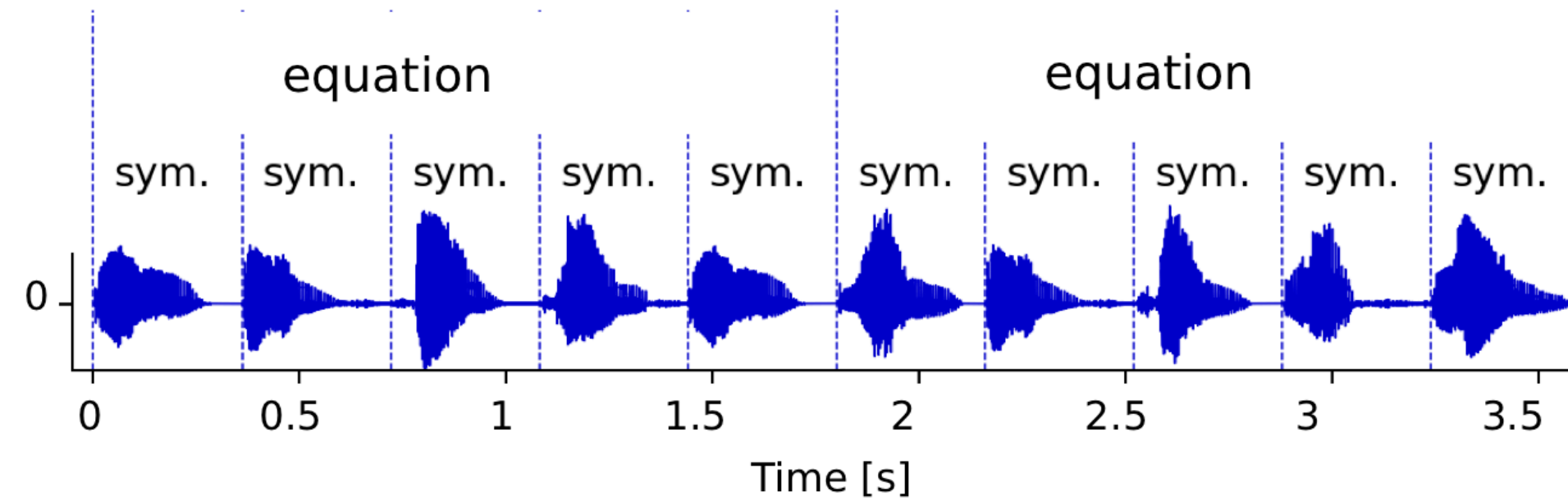
Acoustical  
Spectrum



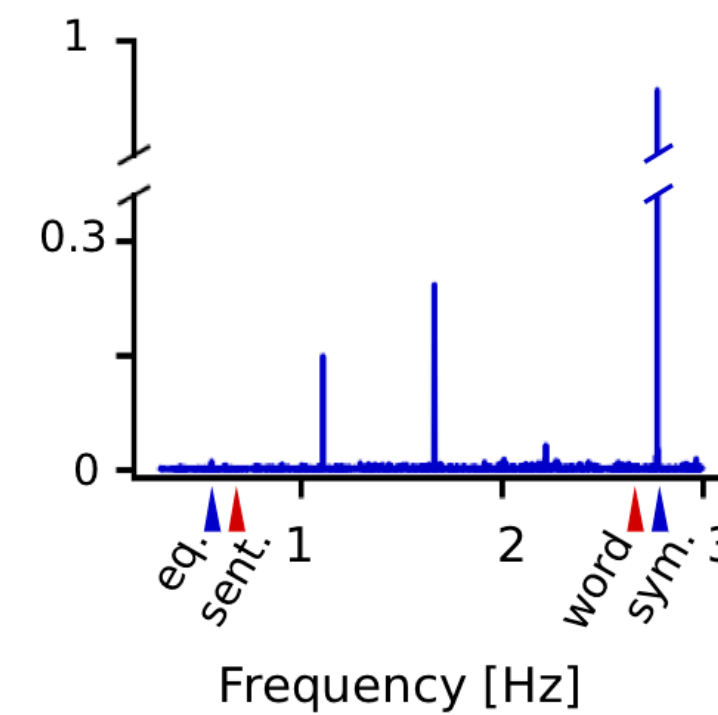
Neural  
Spectrum



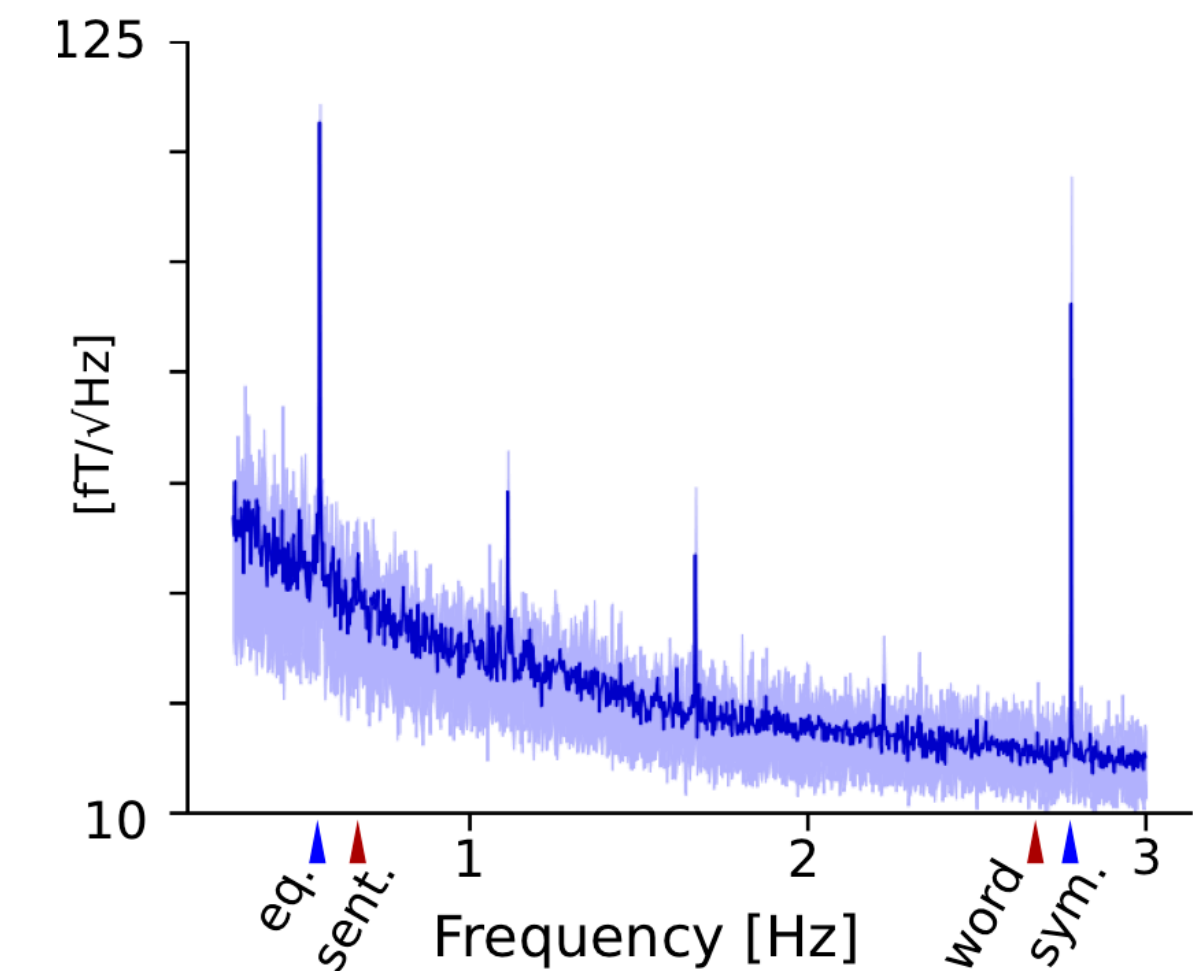
# Isochronous Arithmetic



Acoustics



Acoustical  
Spectrum

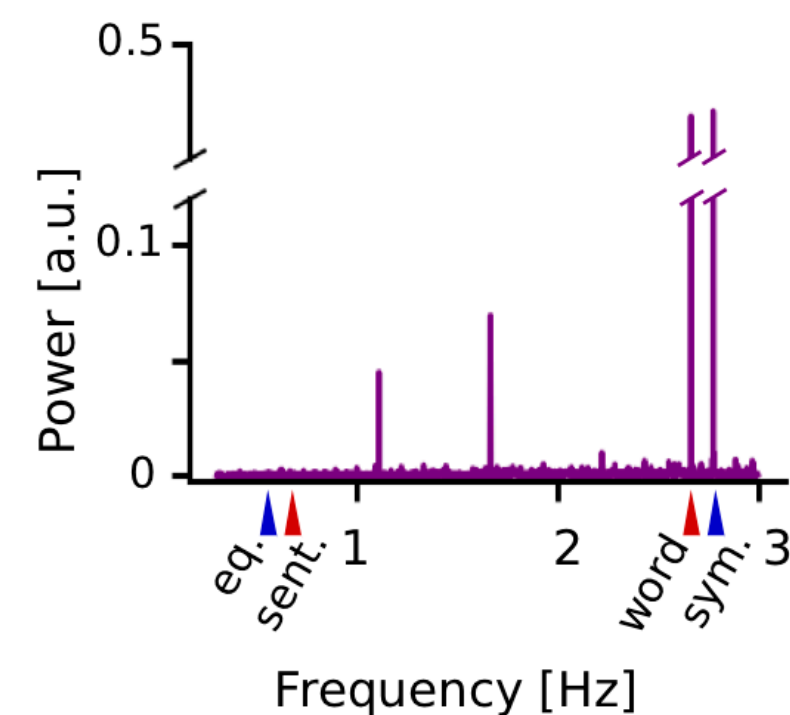
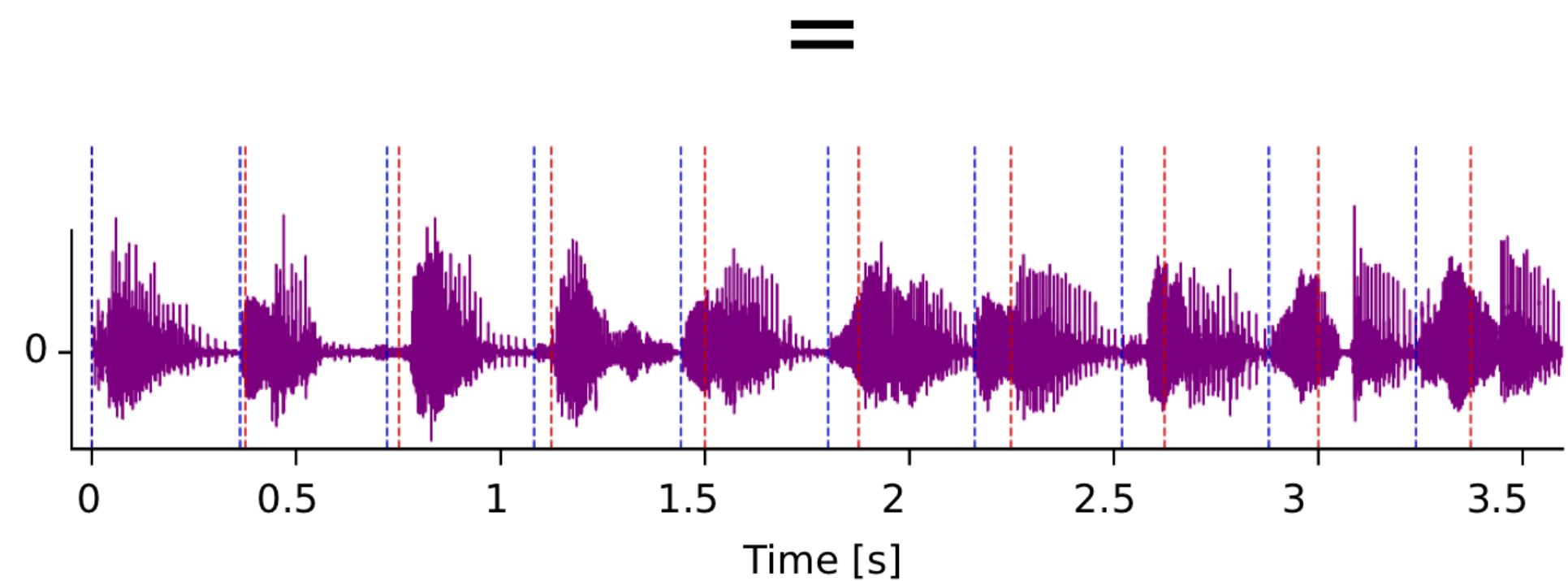
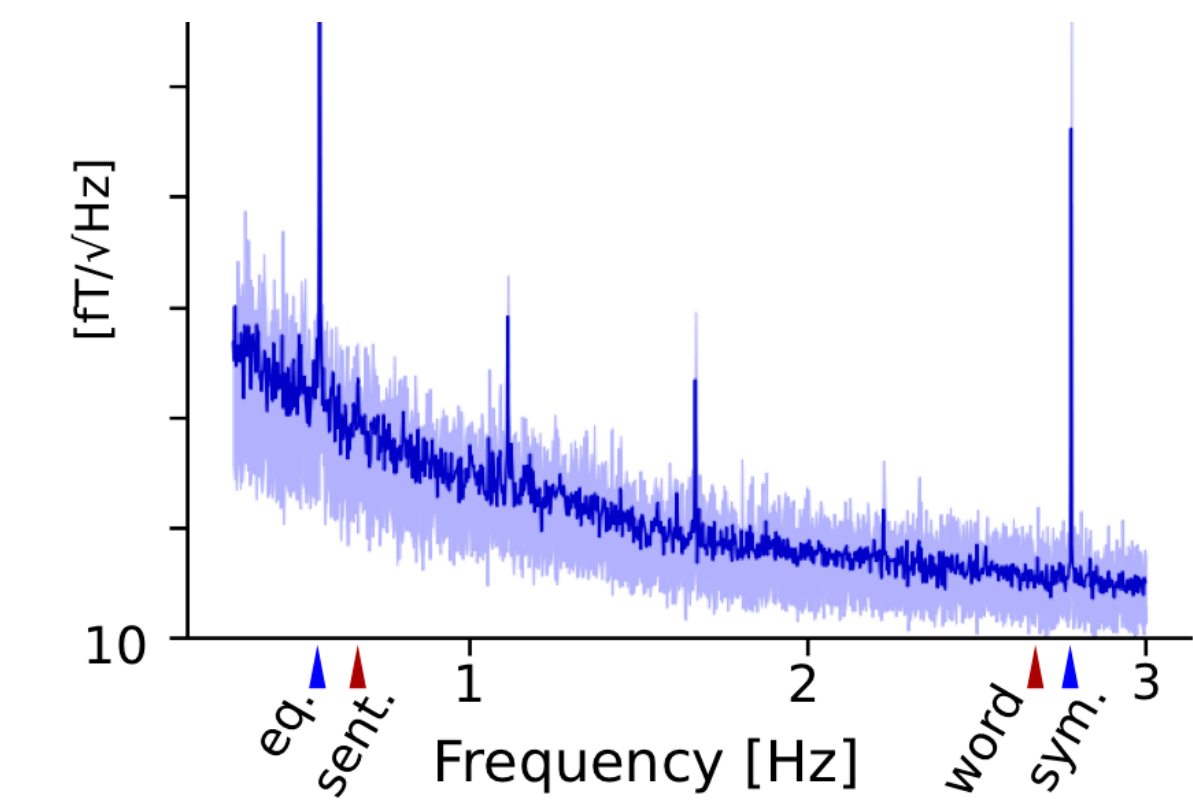
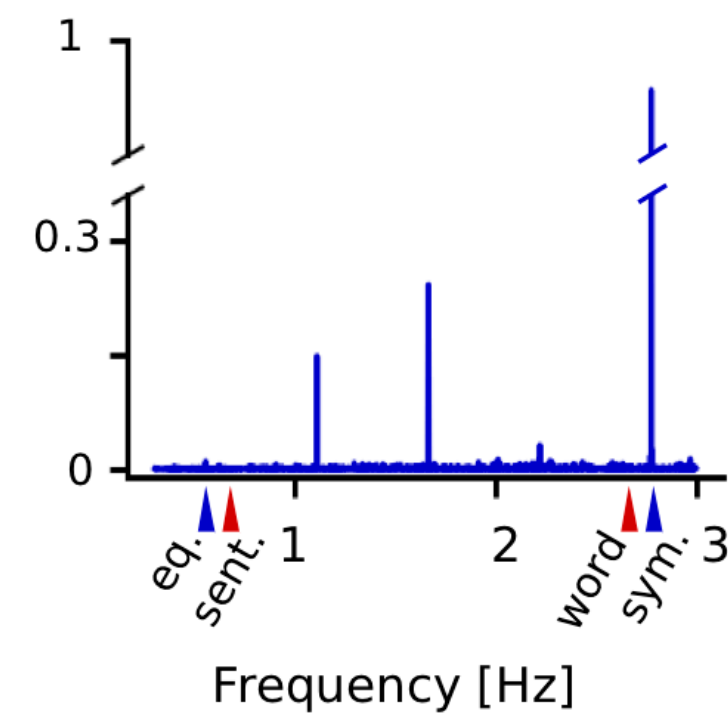
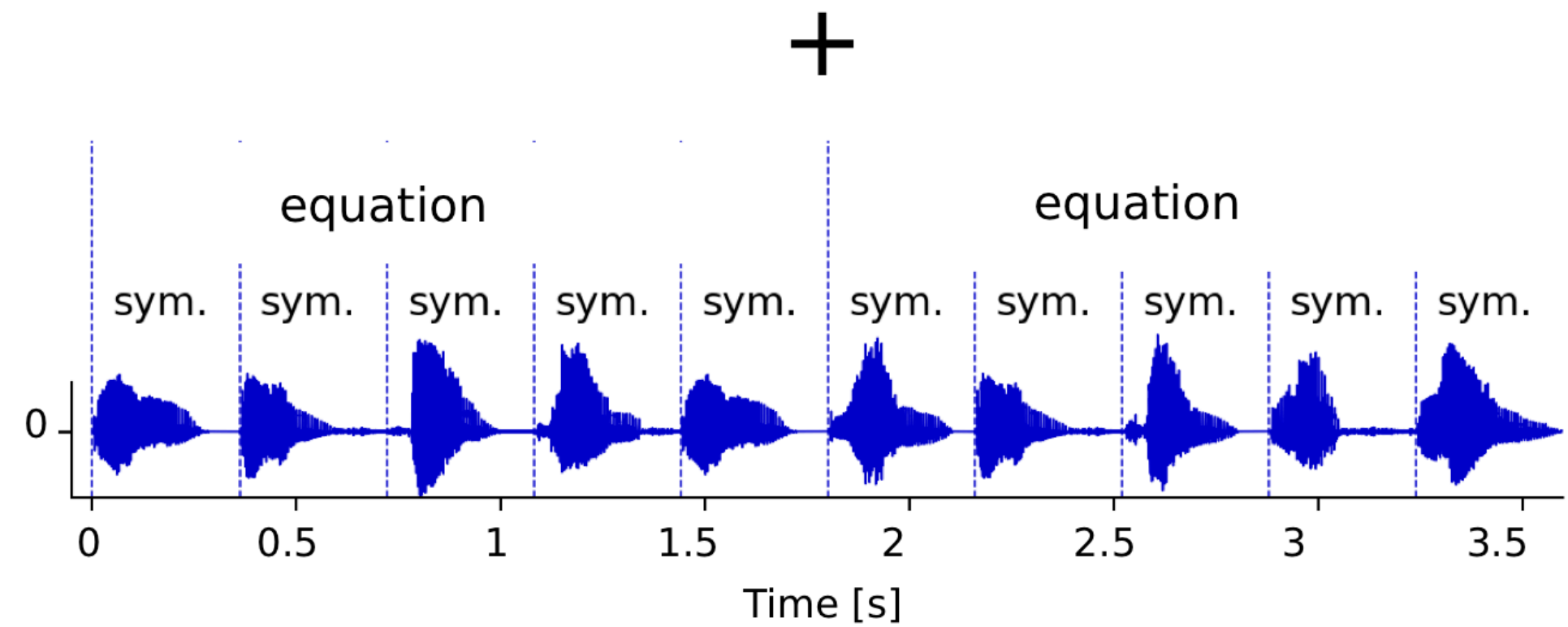
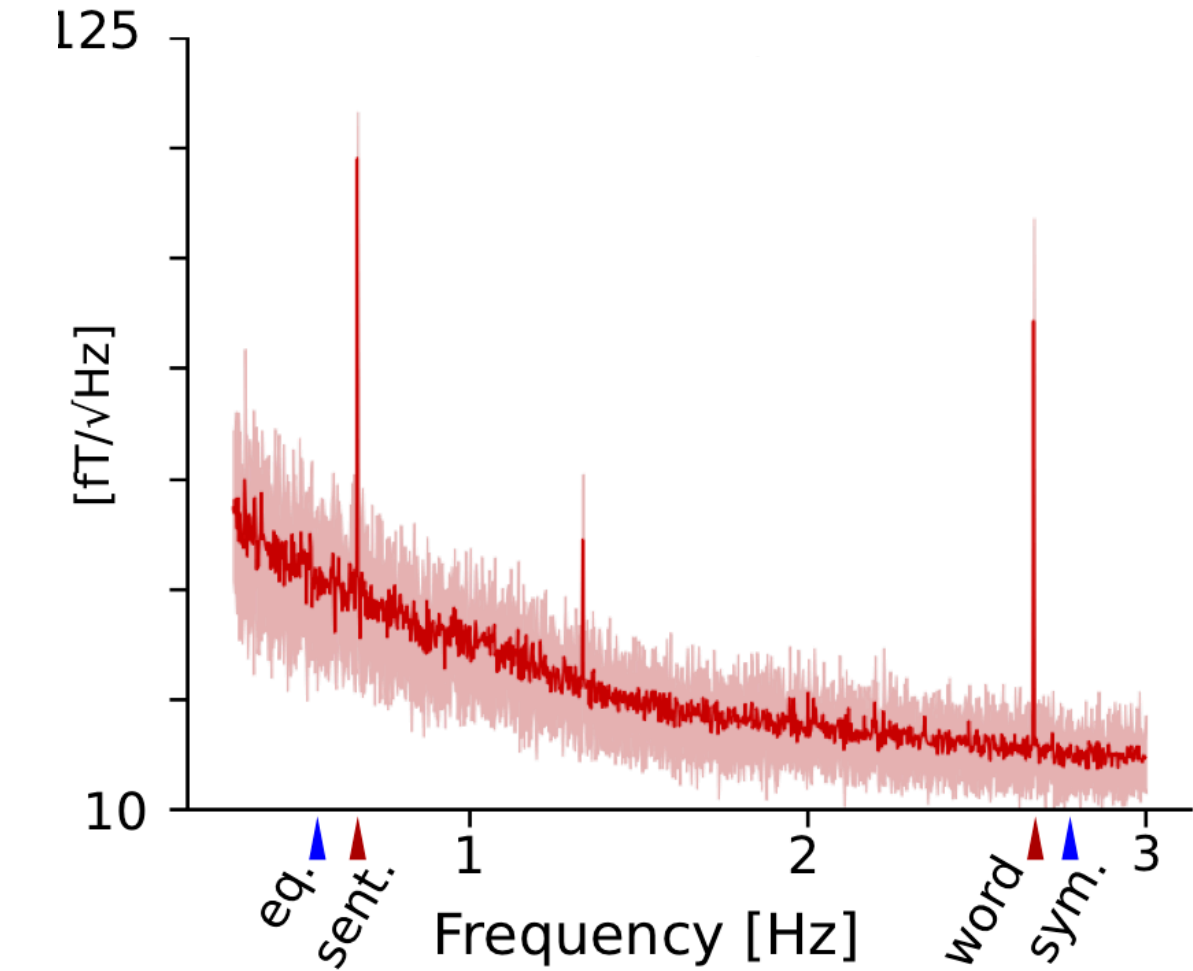
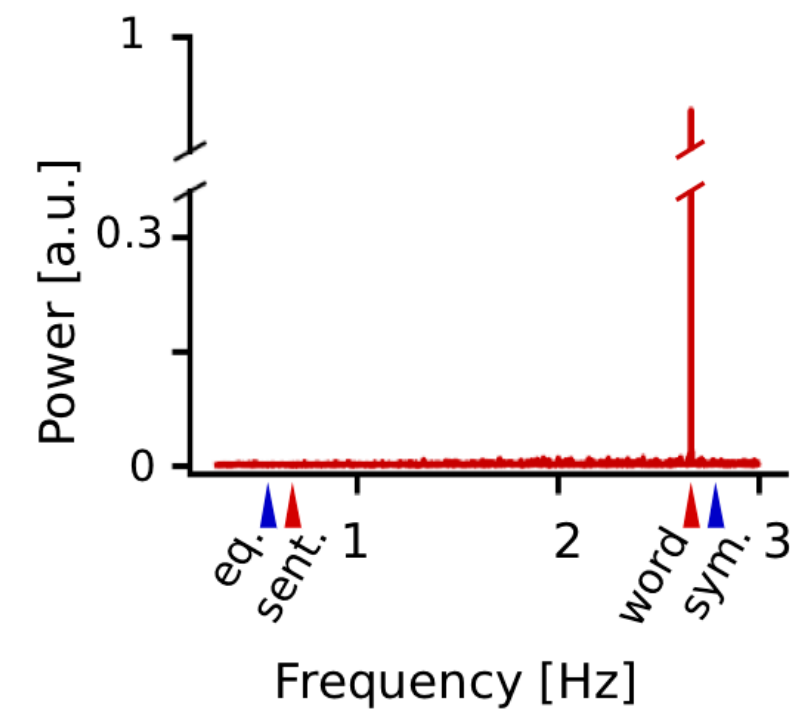
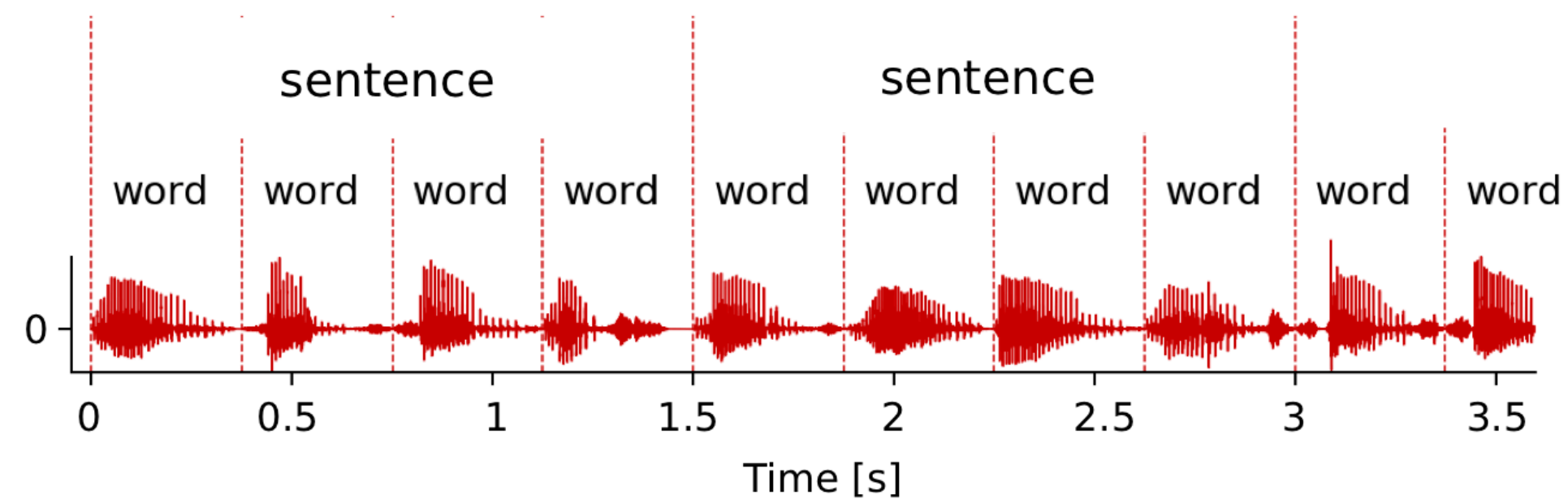


Neural  
Spectrum





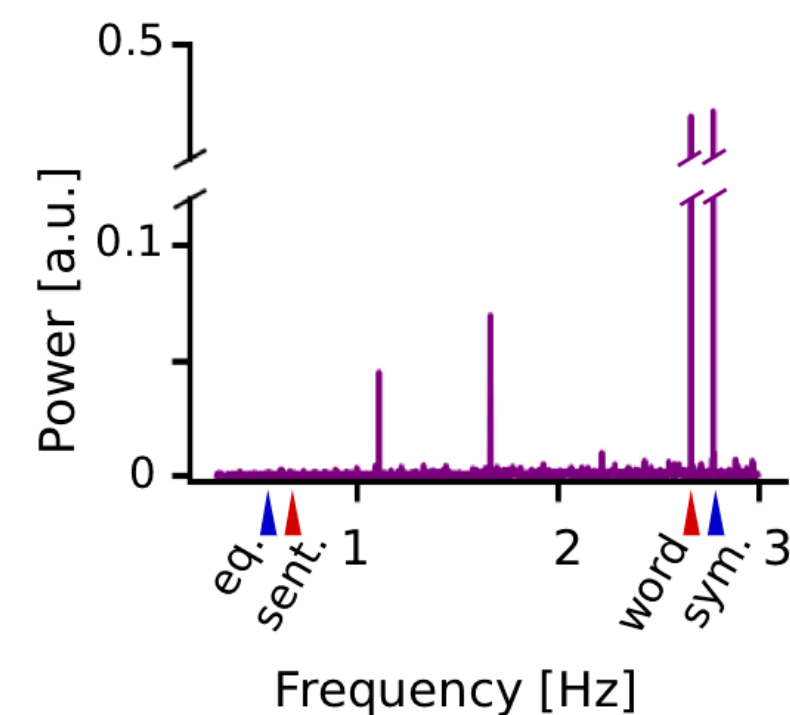
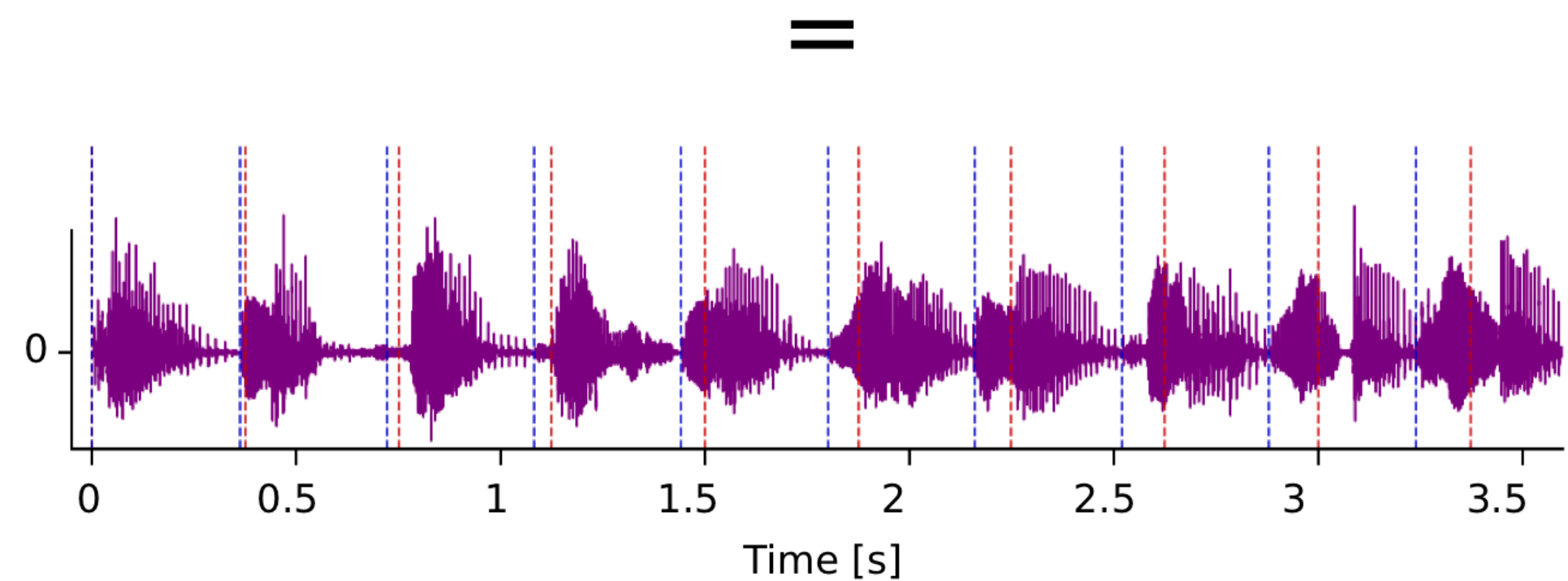
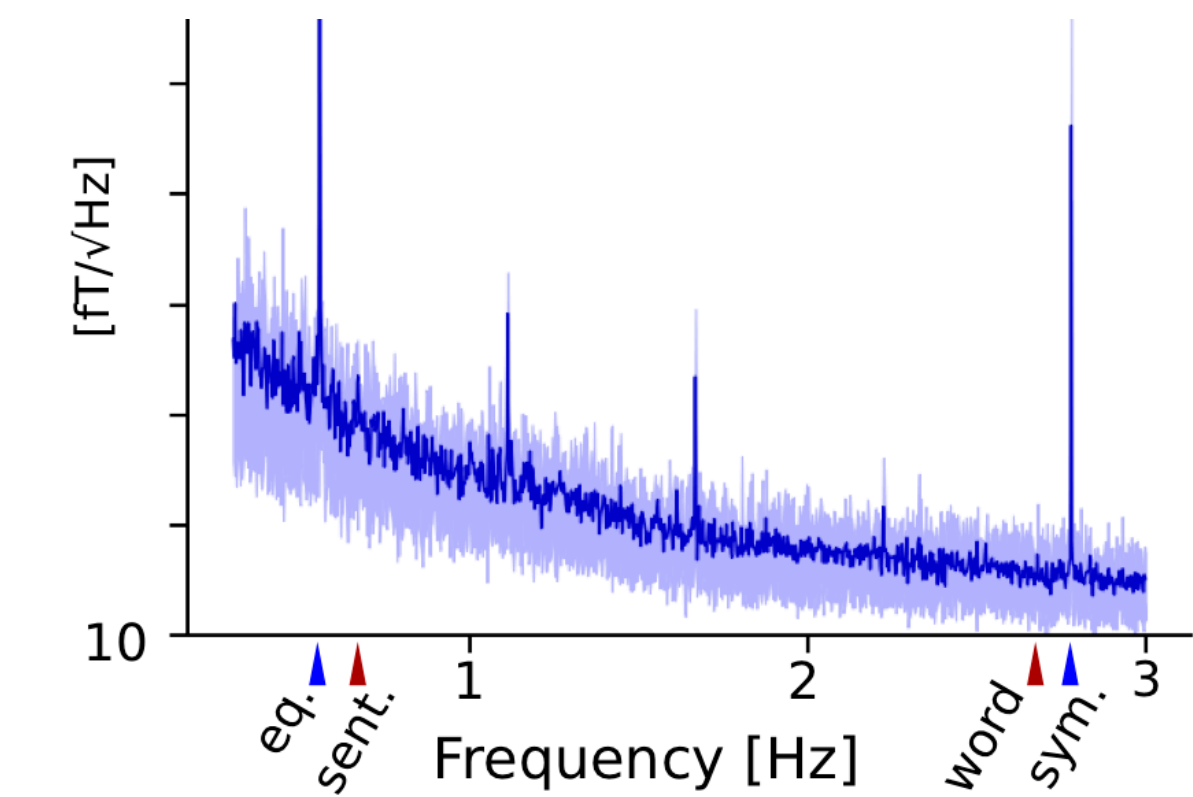
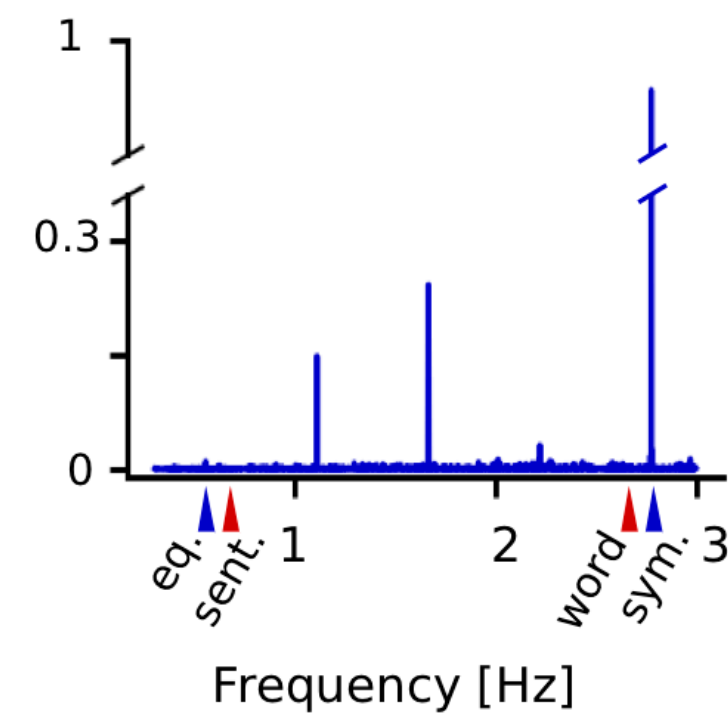
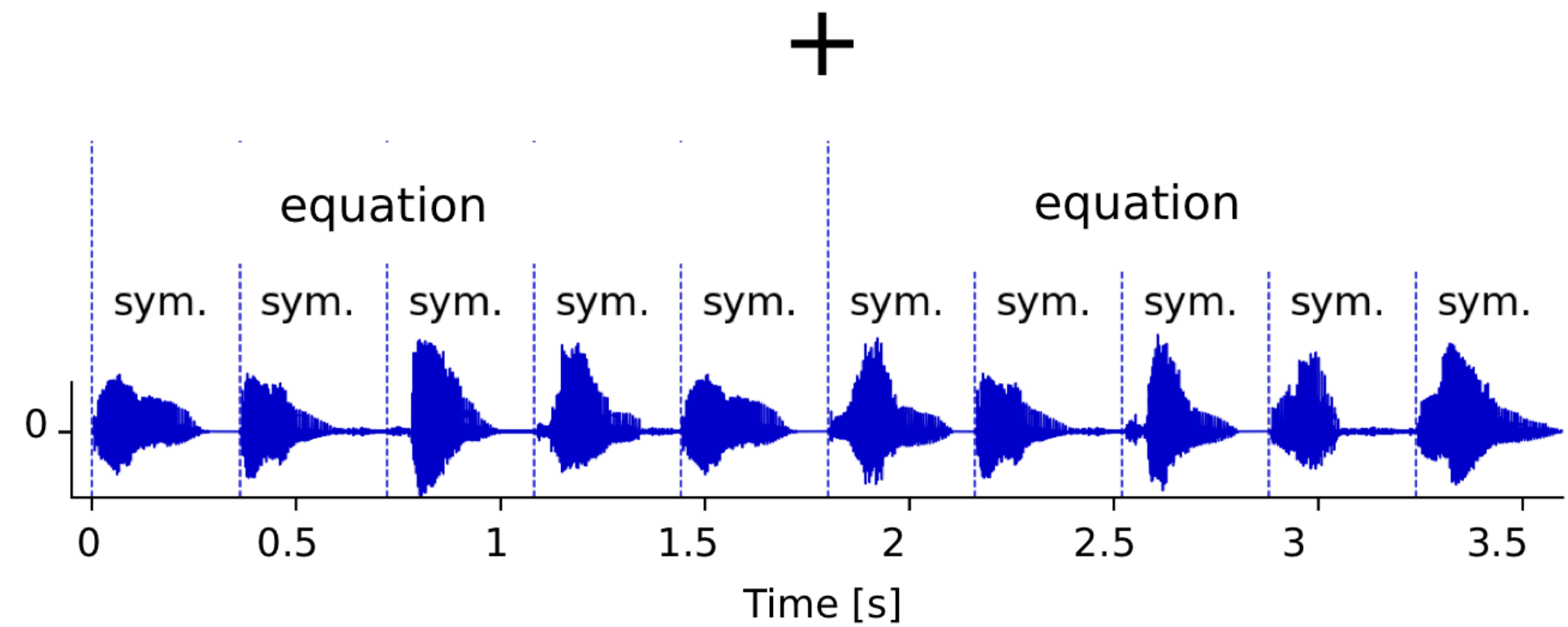
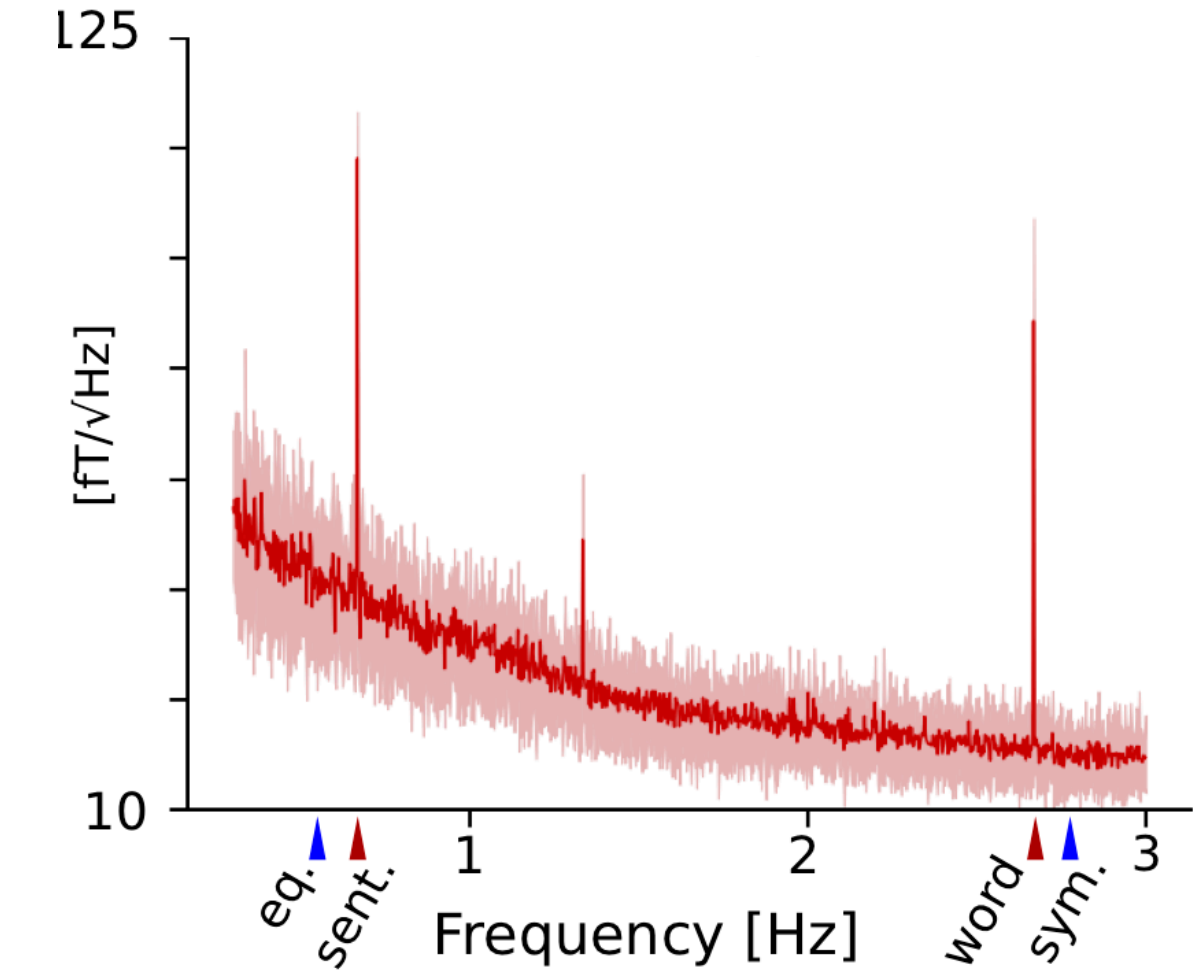
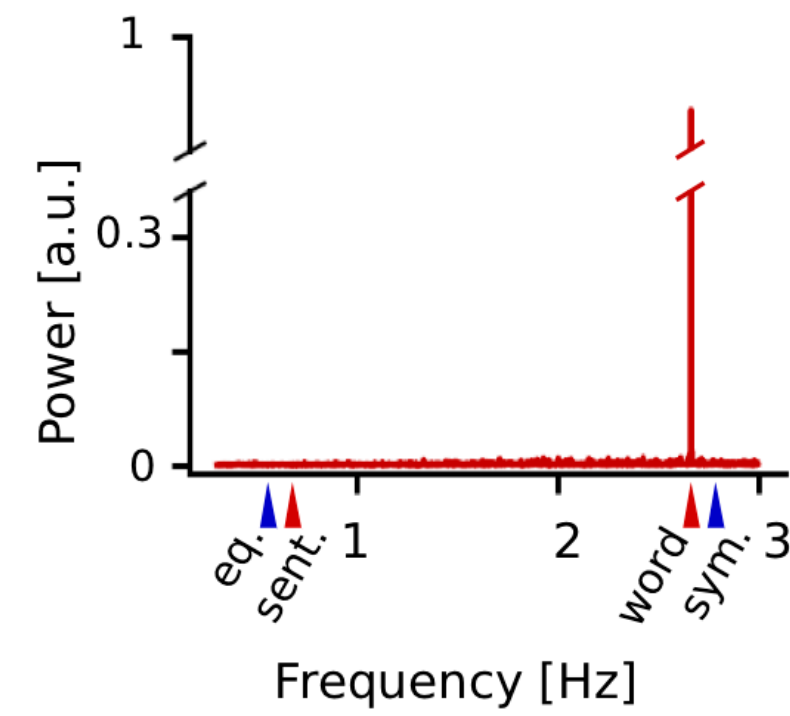
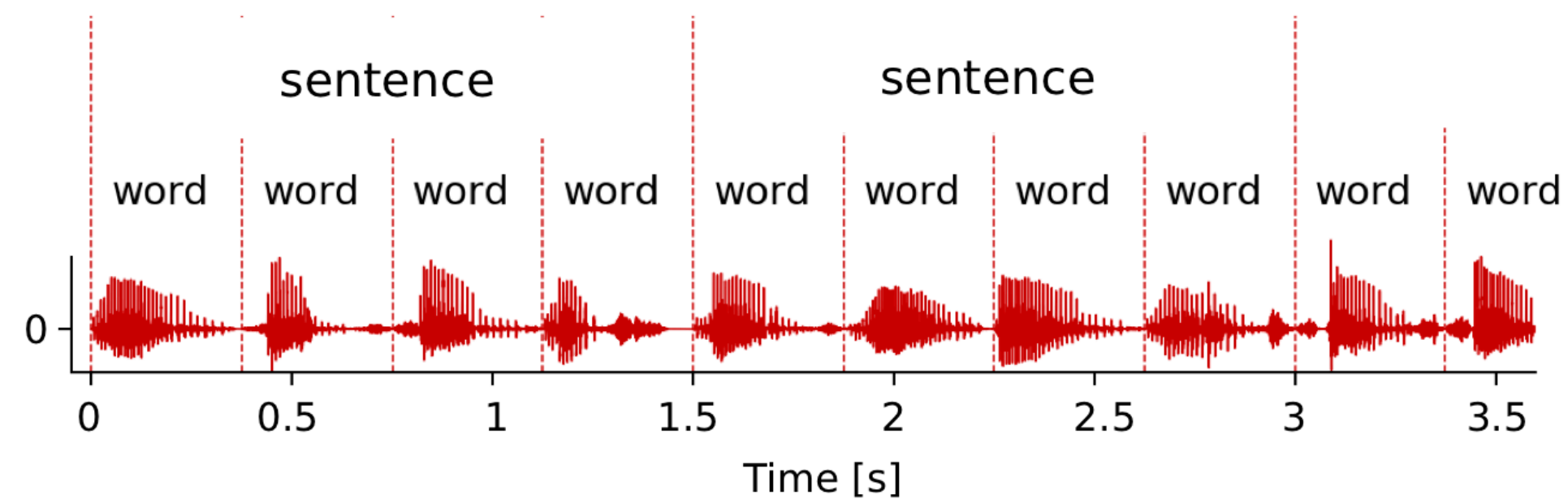
# Isochronous Cocktail Party



?



# Isochronous Cocktail Party

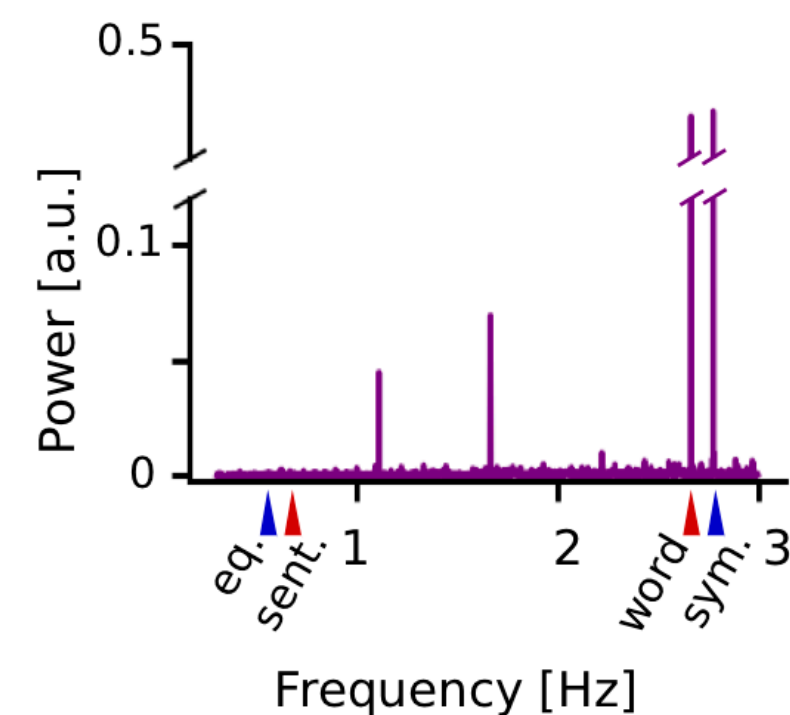
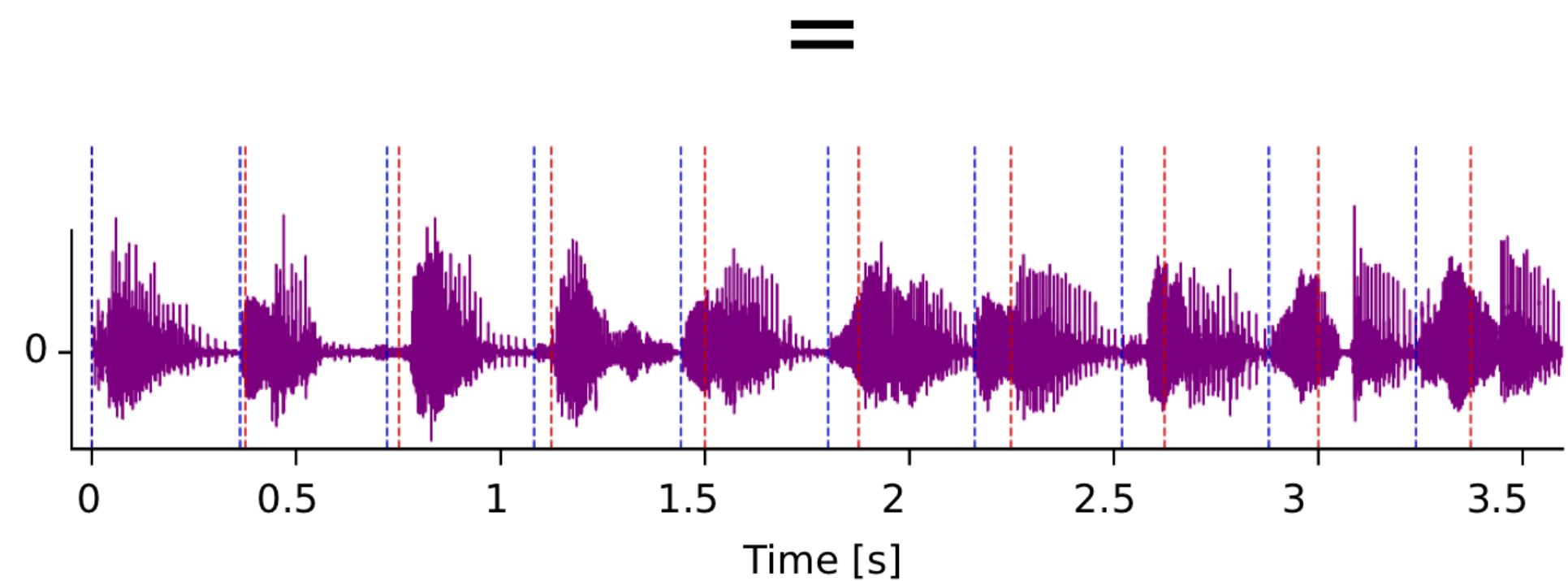
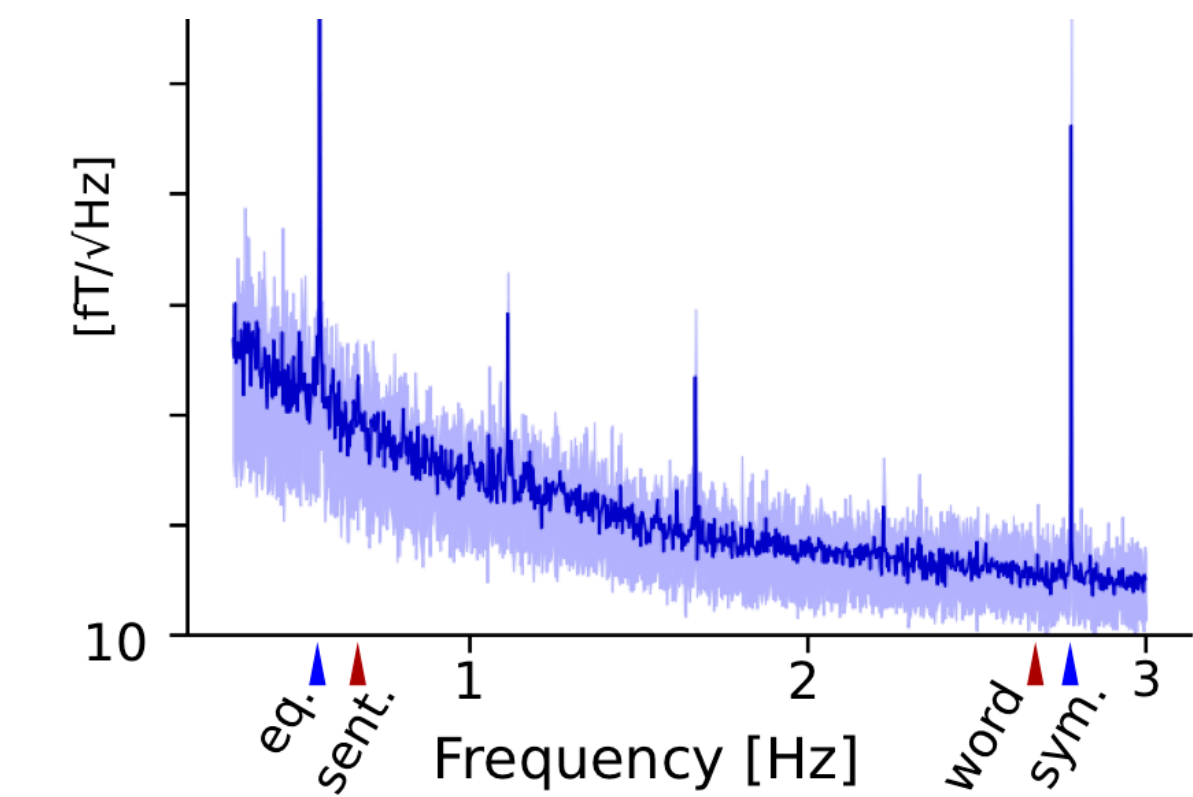
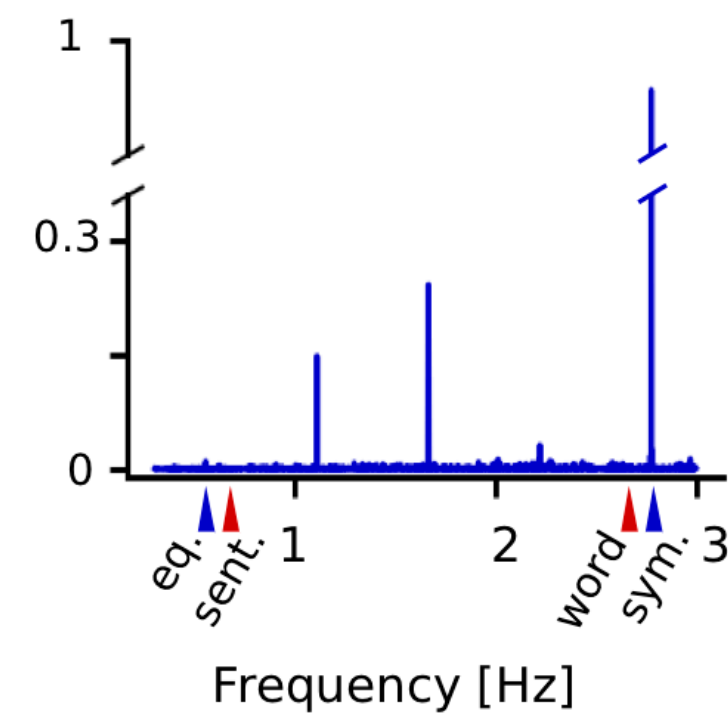
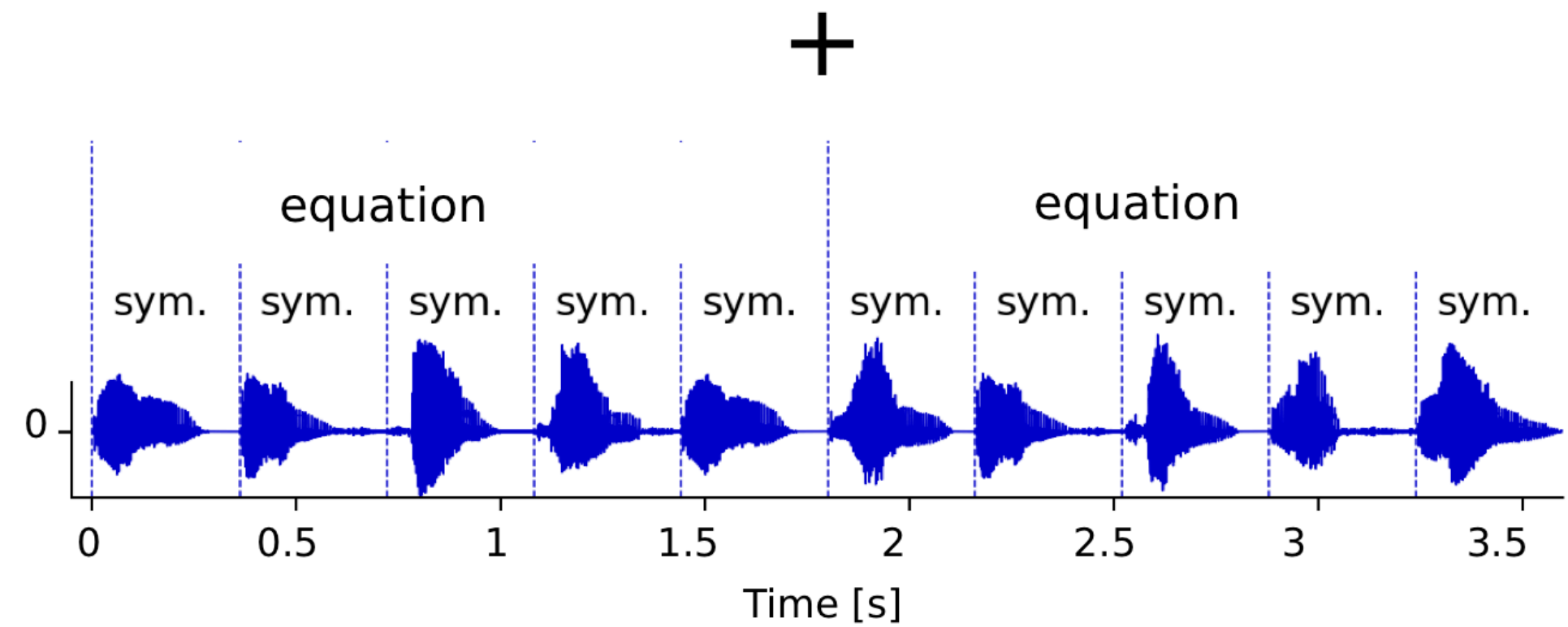
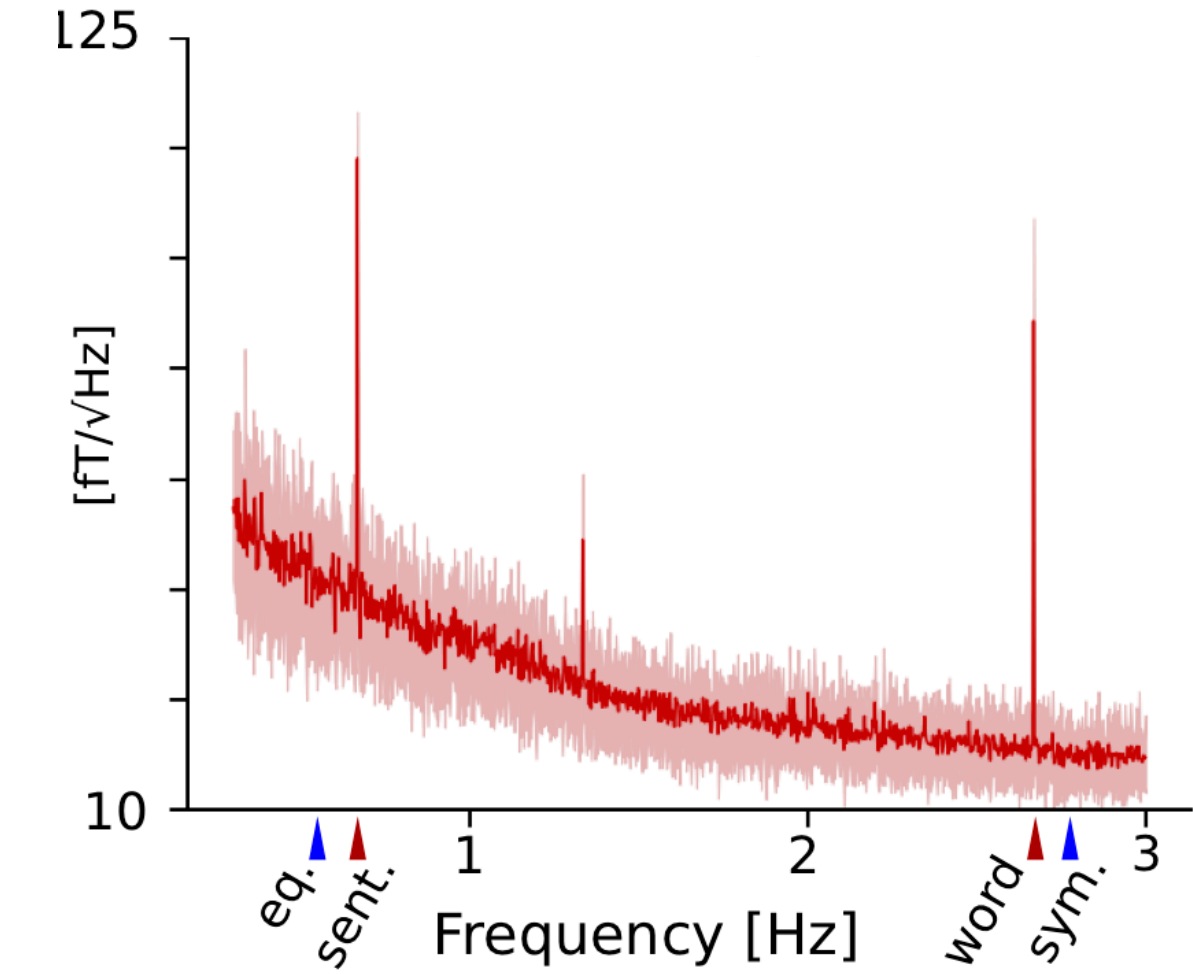
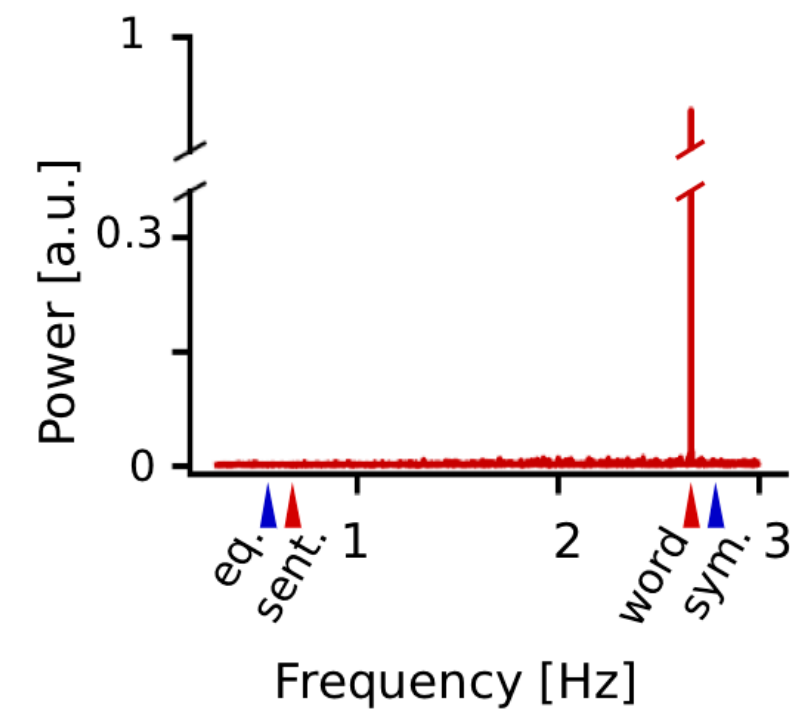
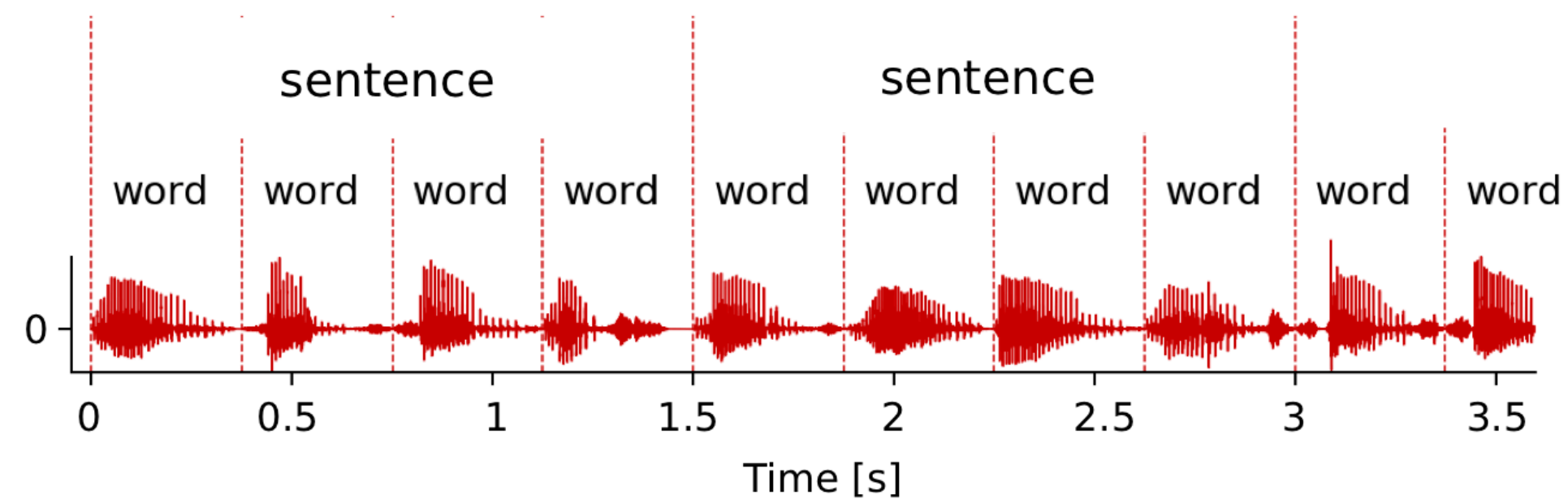


?





# Isochronous Cocktail Party

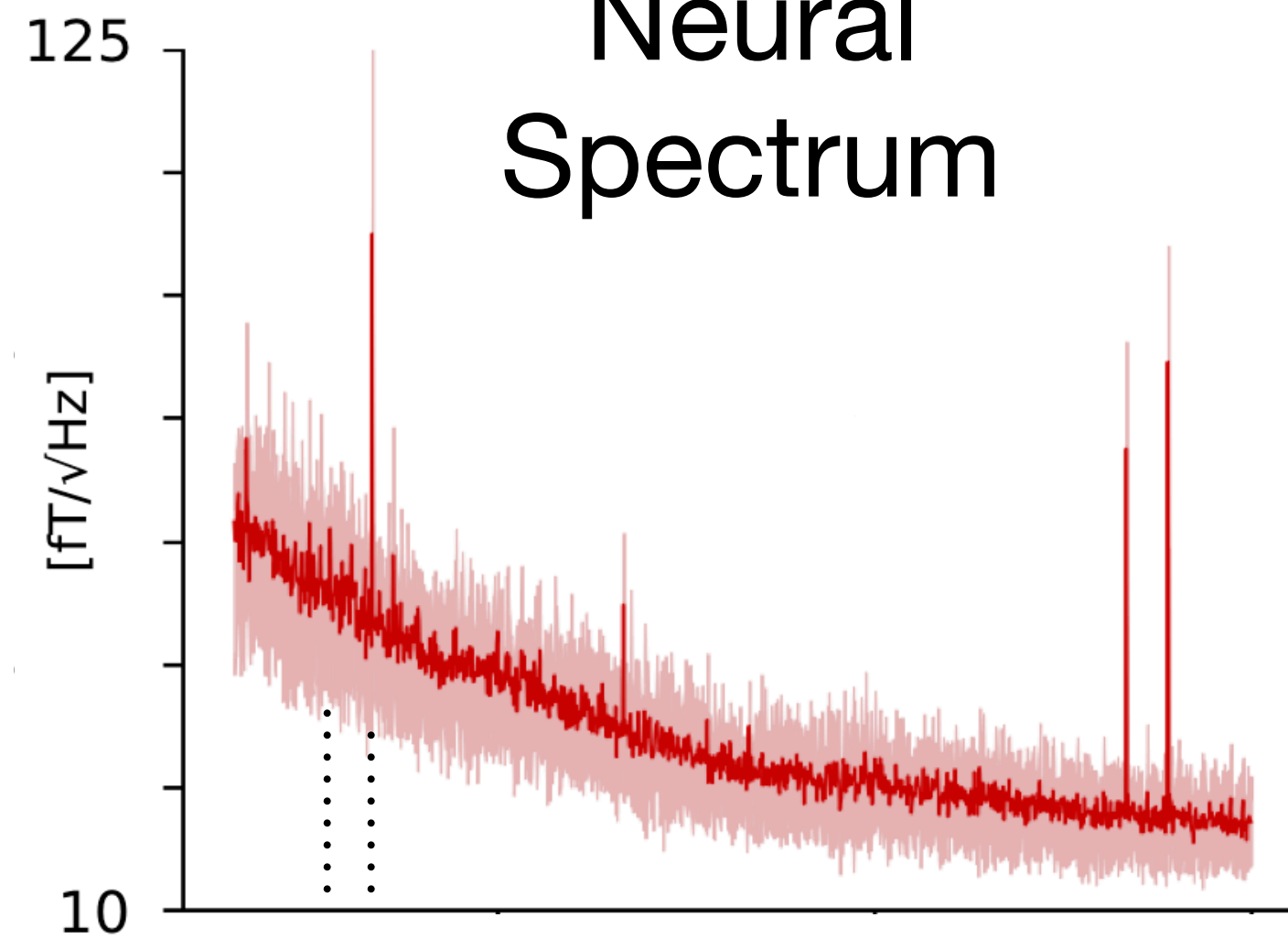


?

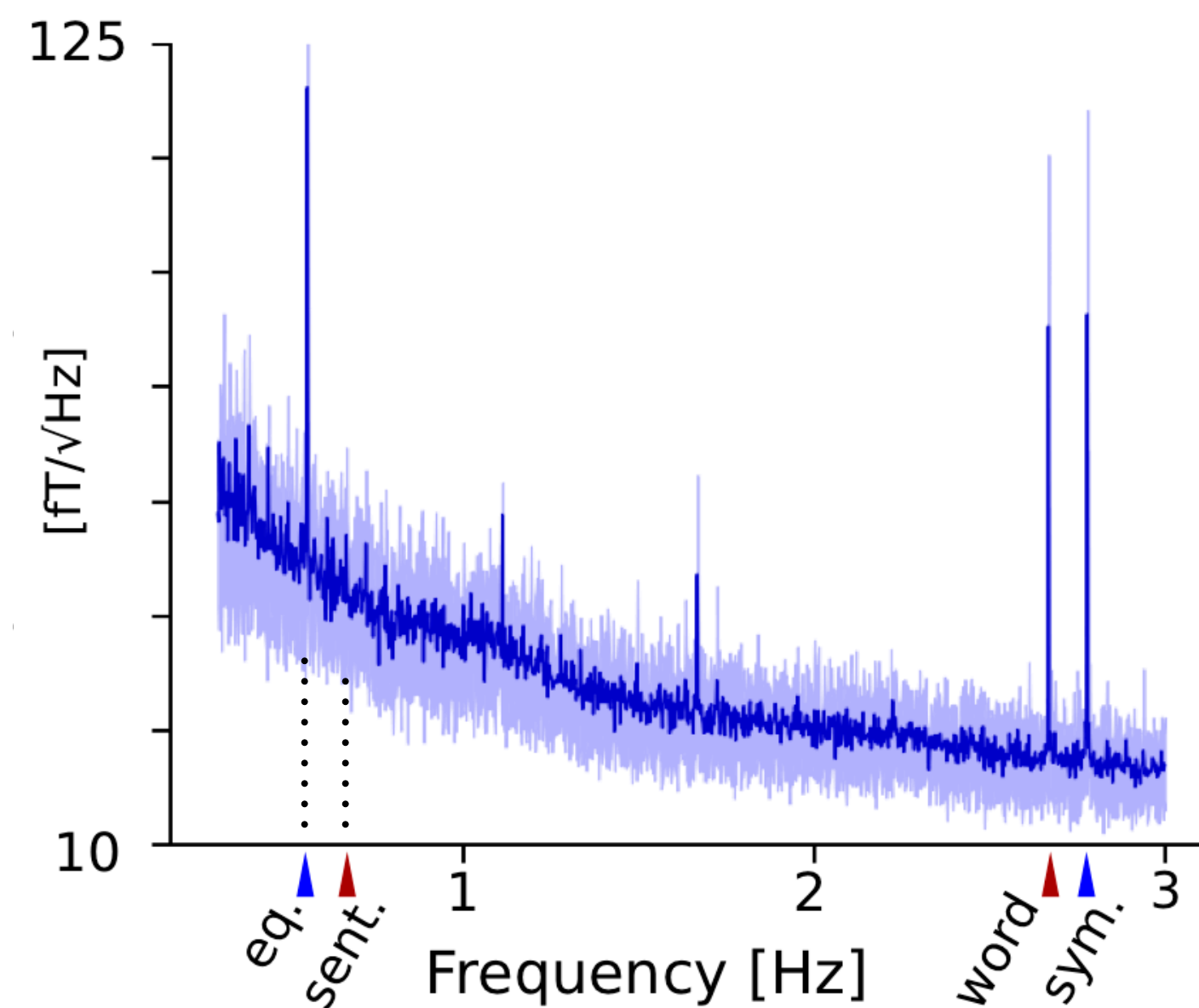


# Isochronous Cocktail Party

Neural  
Spectrum



Attend to  
Sentences

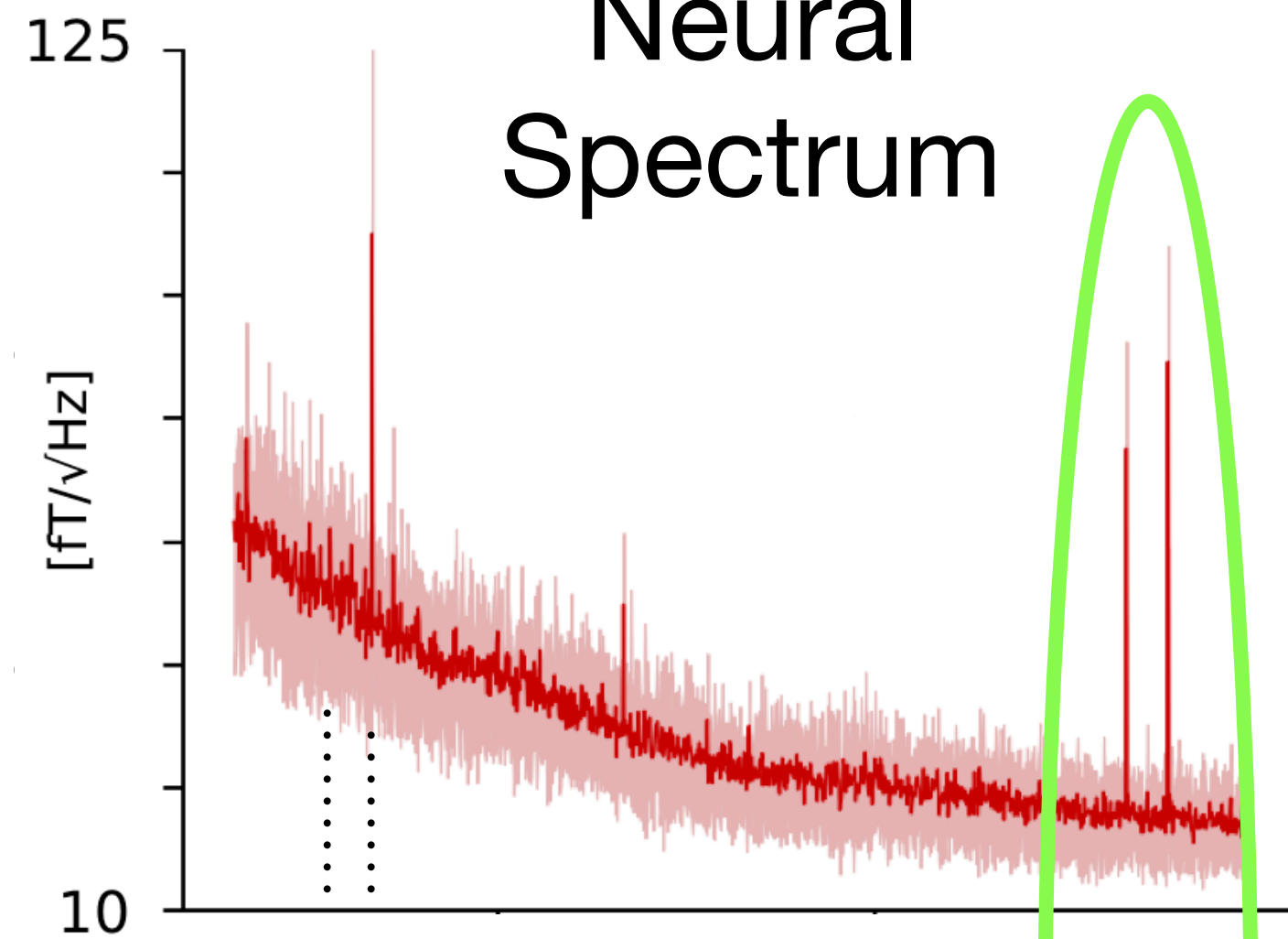


Attend to  
Equations

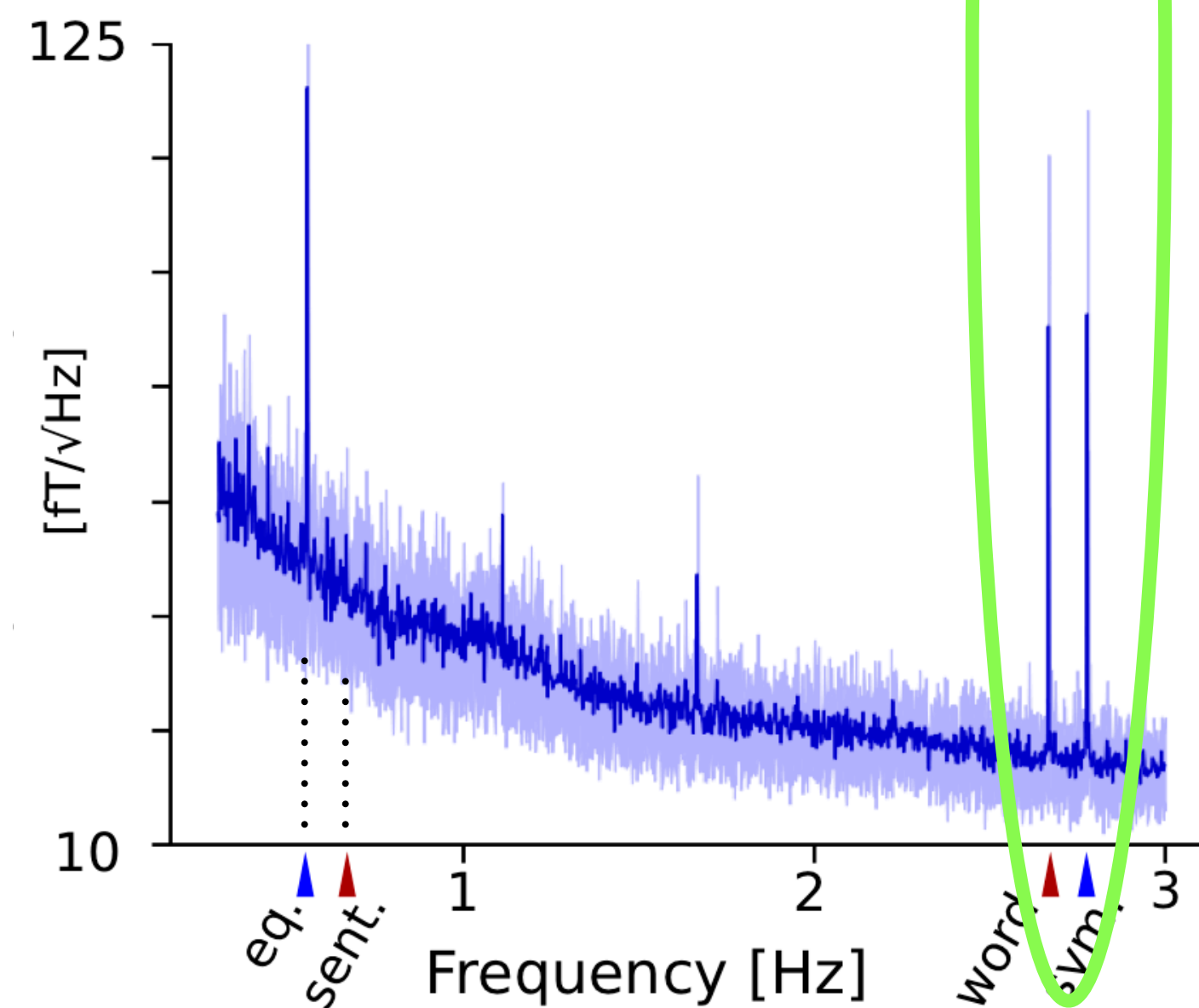
# Isochronous Cocktail Party

Neural  
Spectrum

Attend to  
Sentences



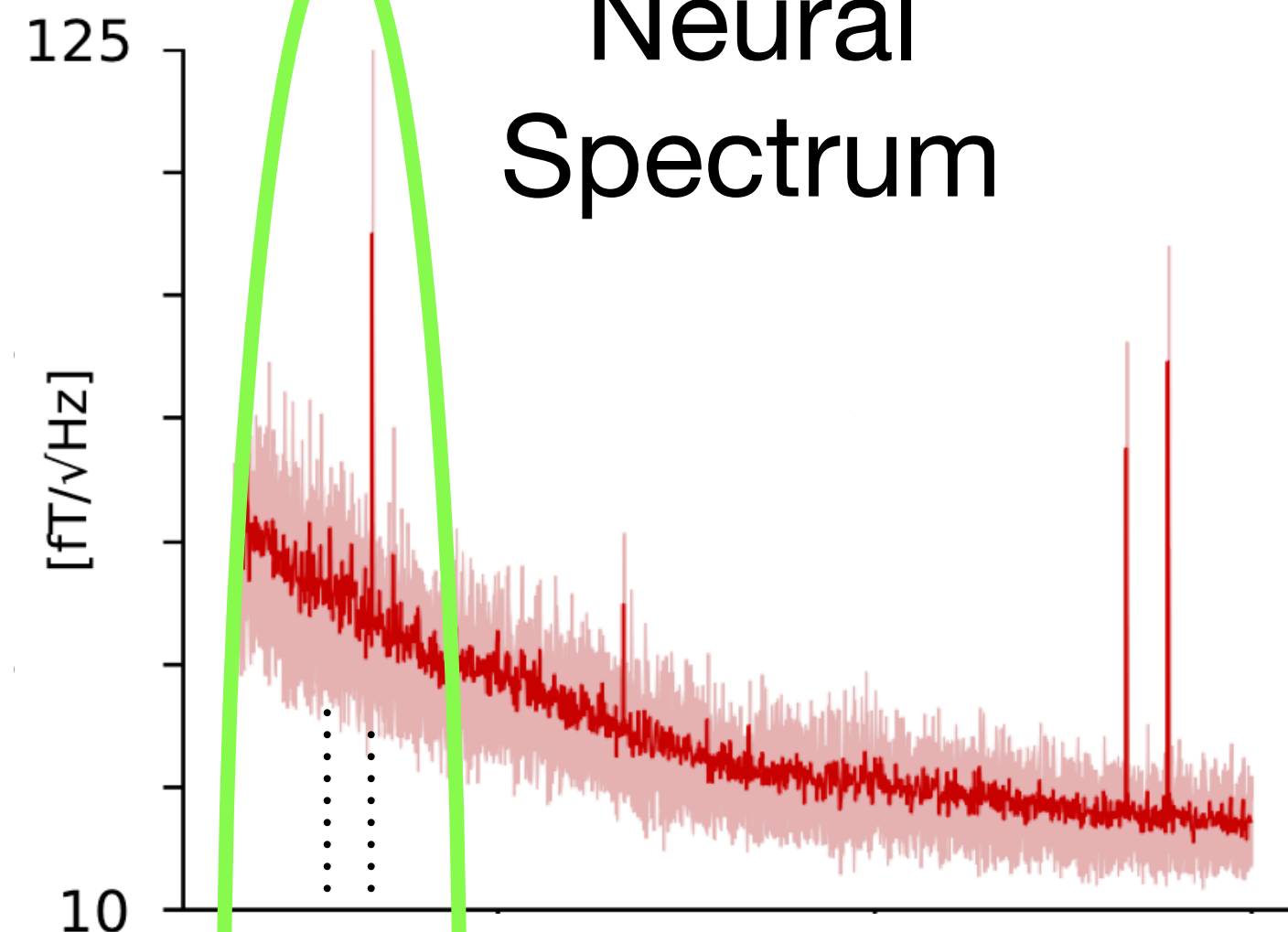
Attend to  
Equations



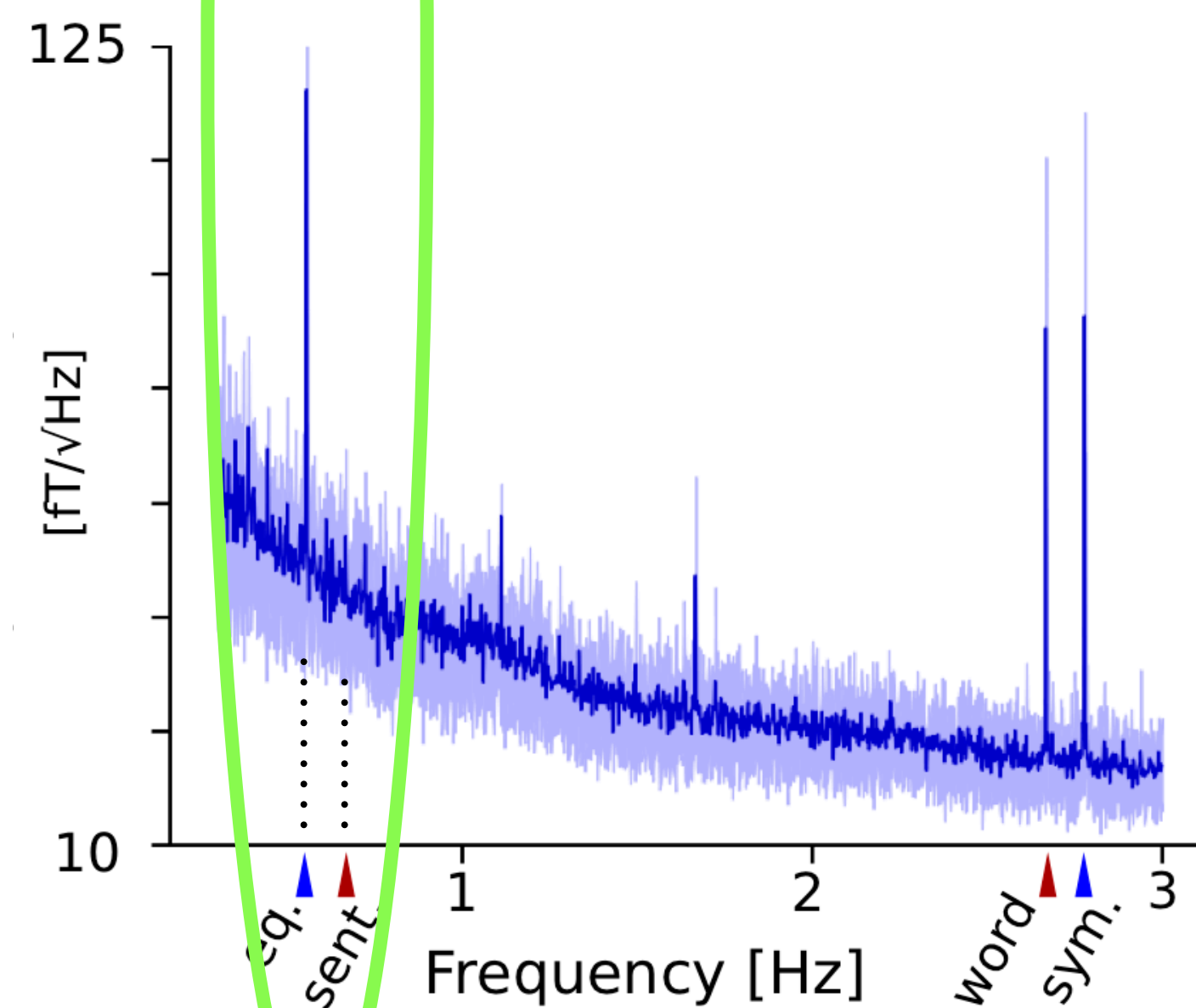
# Isochronous Cocktail Party

Neural  
Spectrum

Attend to  
Sentences



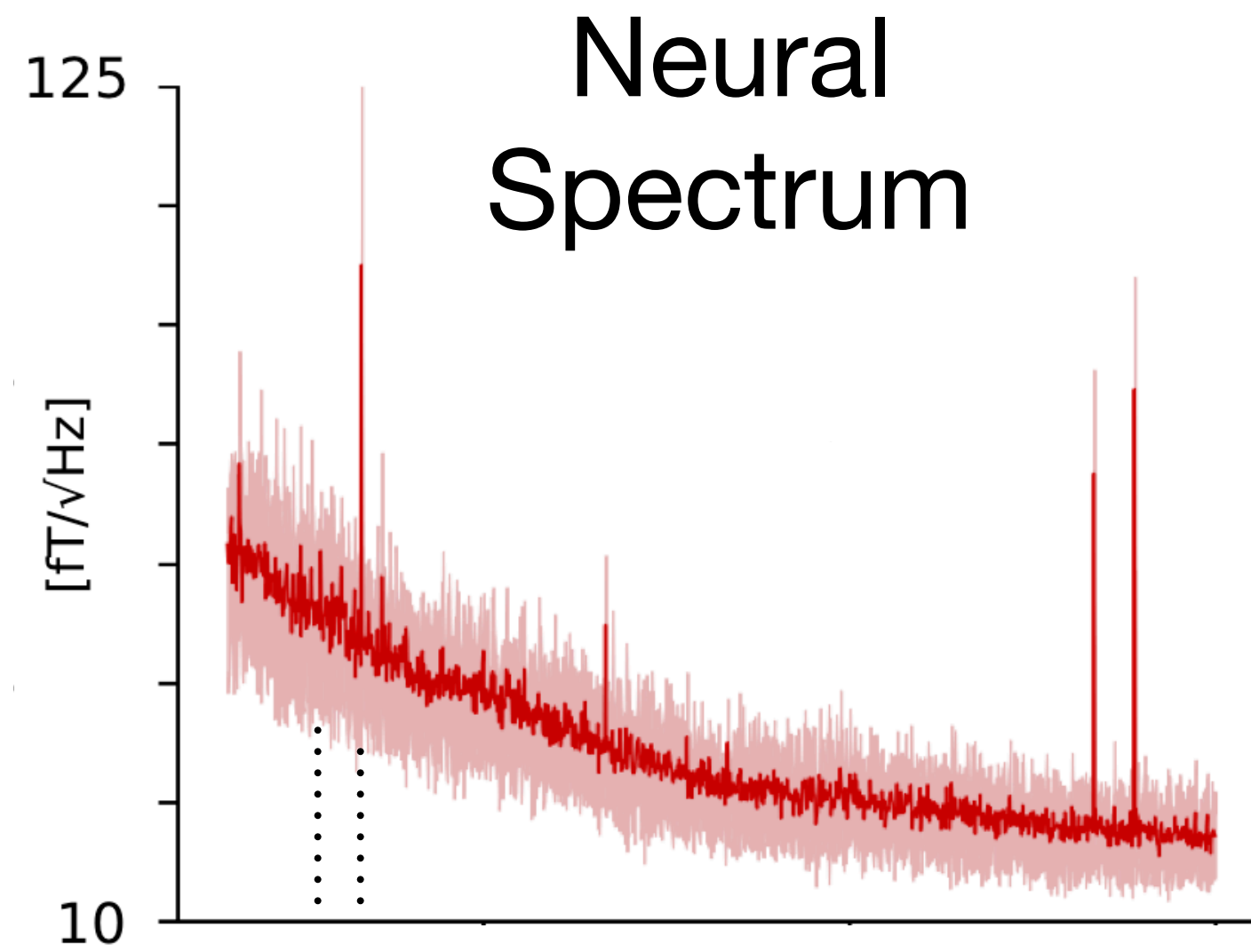
Attend to  
Equations



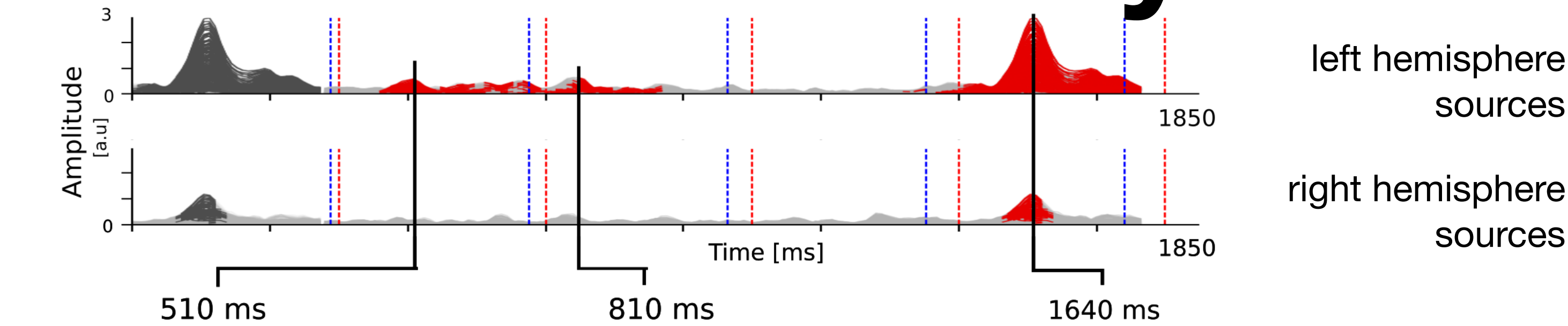
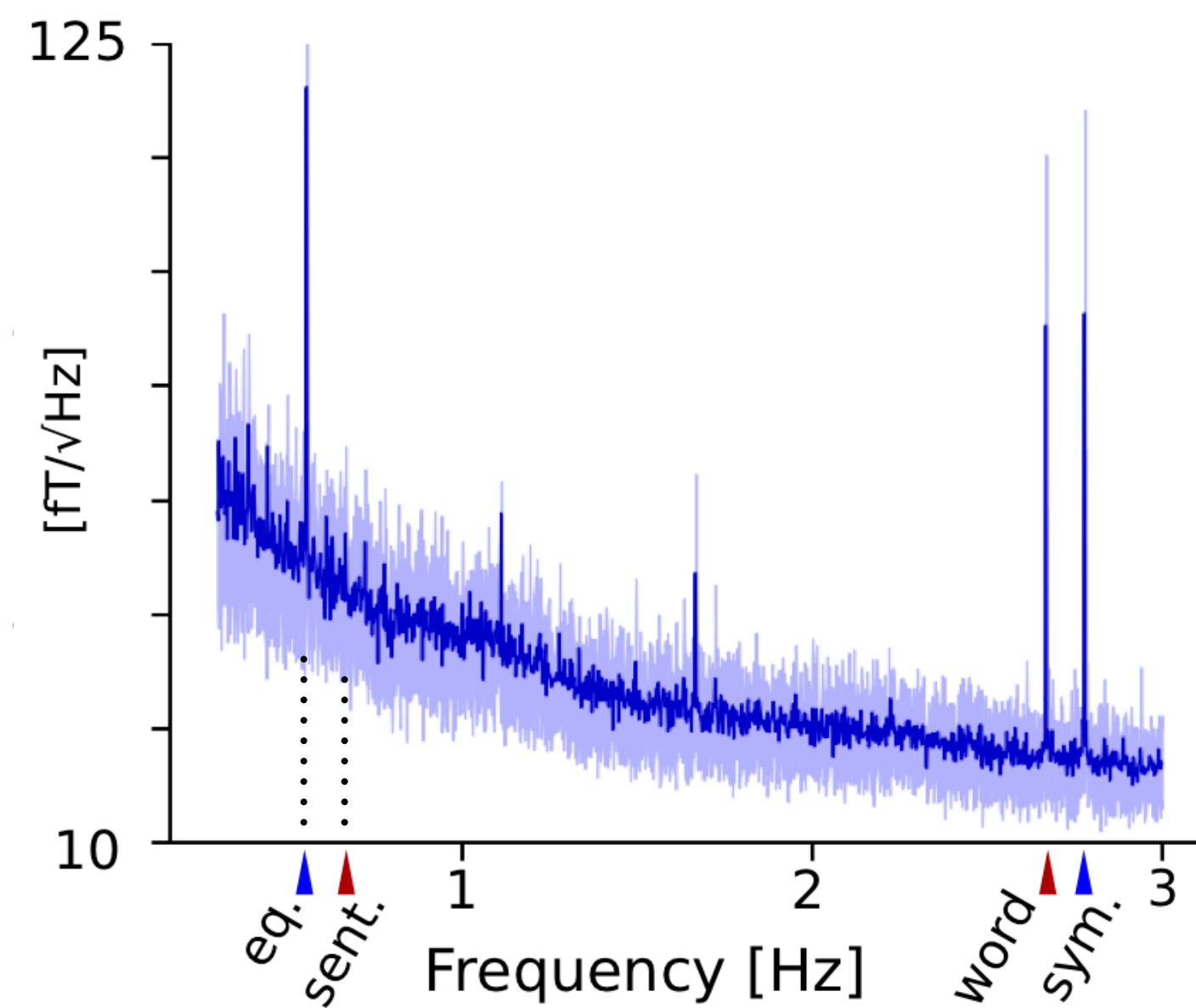


# Isochronous Cocktail Party

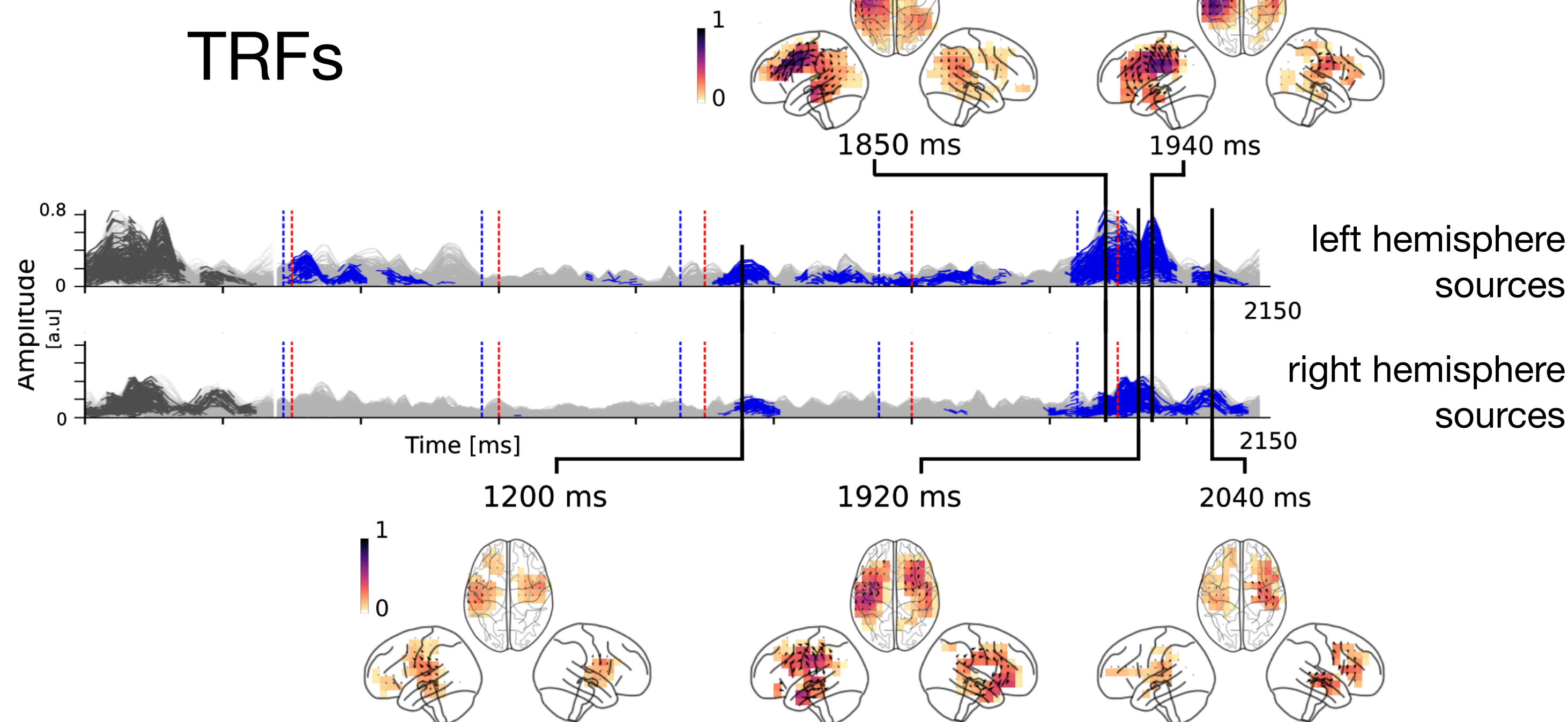
Attend to Sentences



Attend to Equations



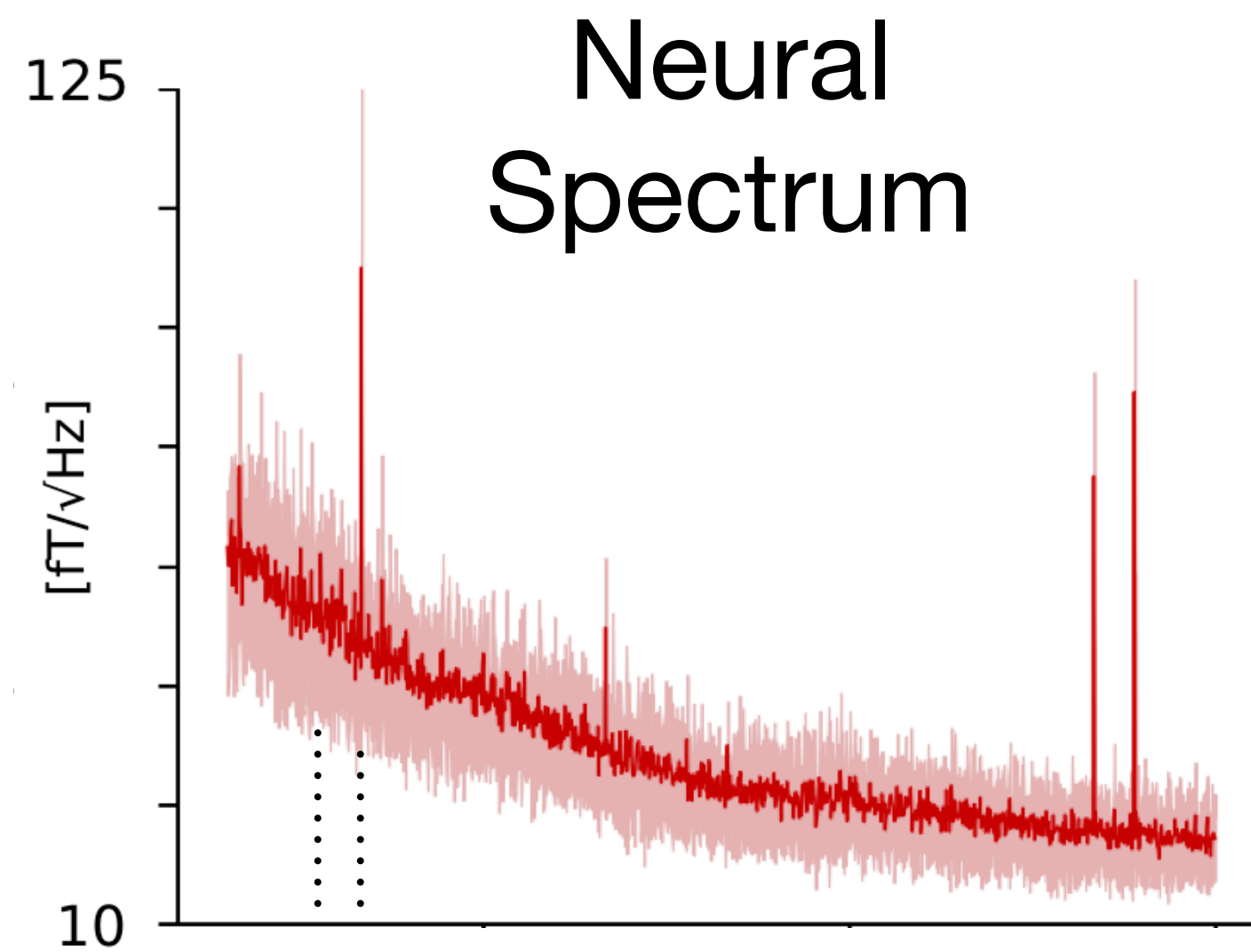
TRFs



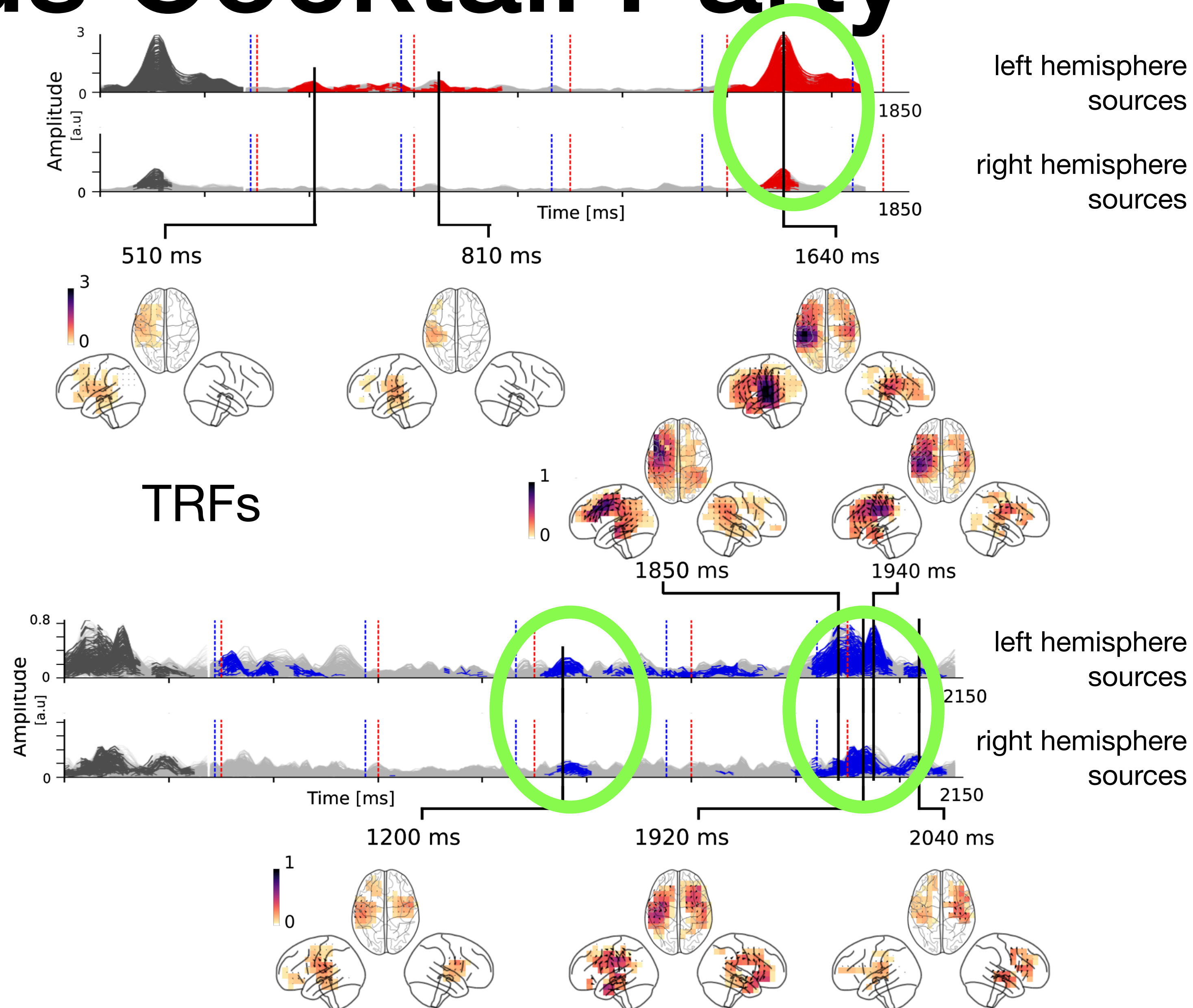
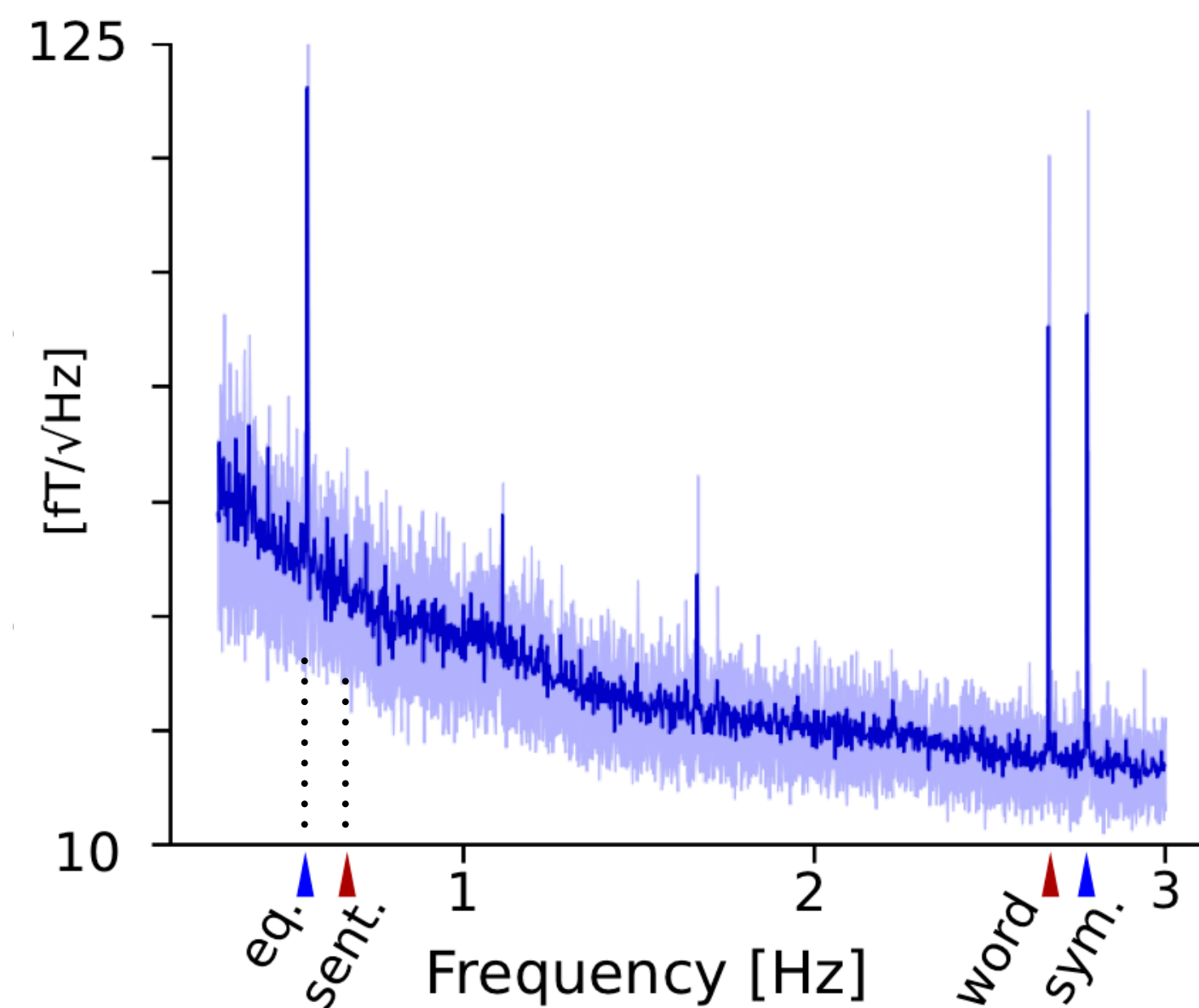


# Isochronous Cocktail Party

Attend to Sentences



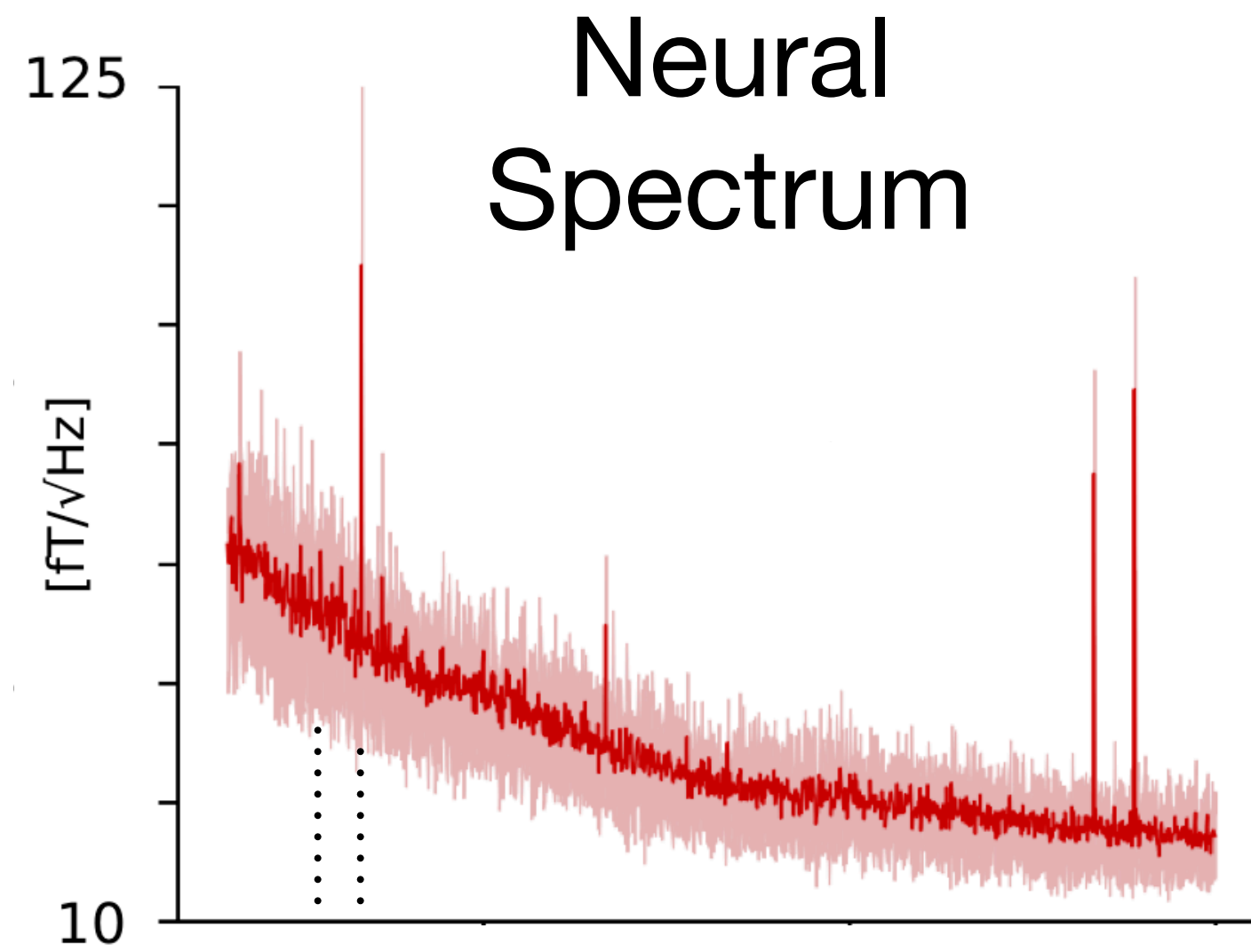
Attend to Equations



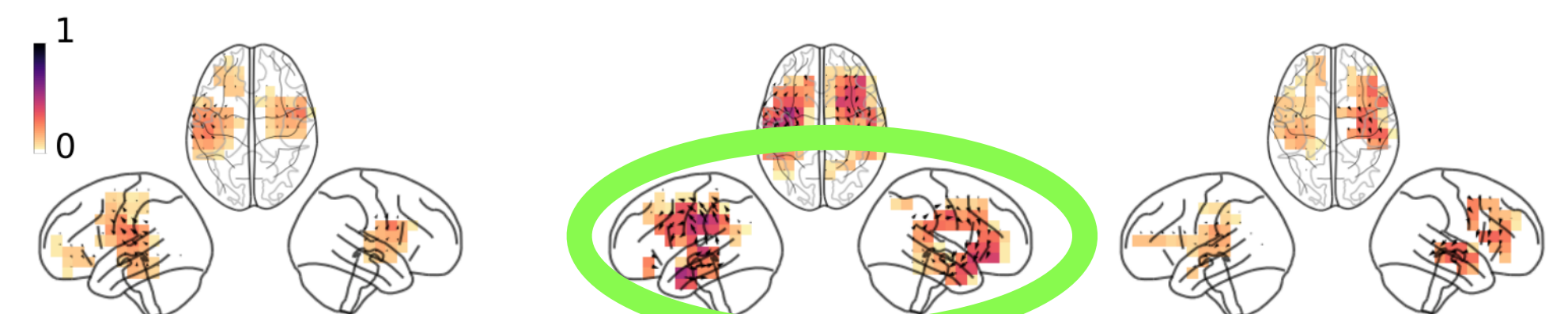
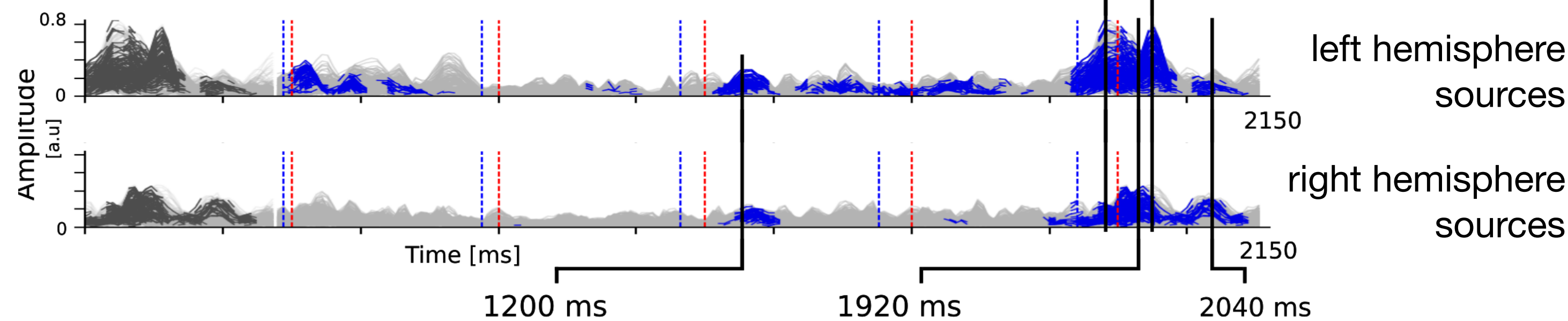
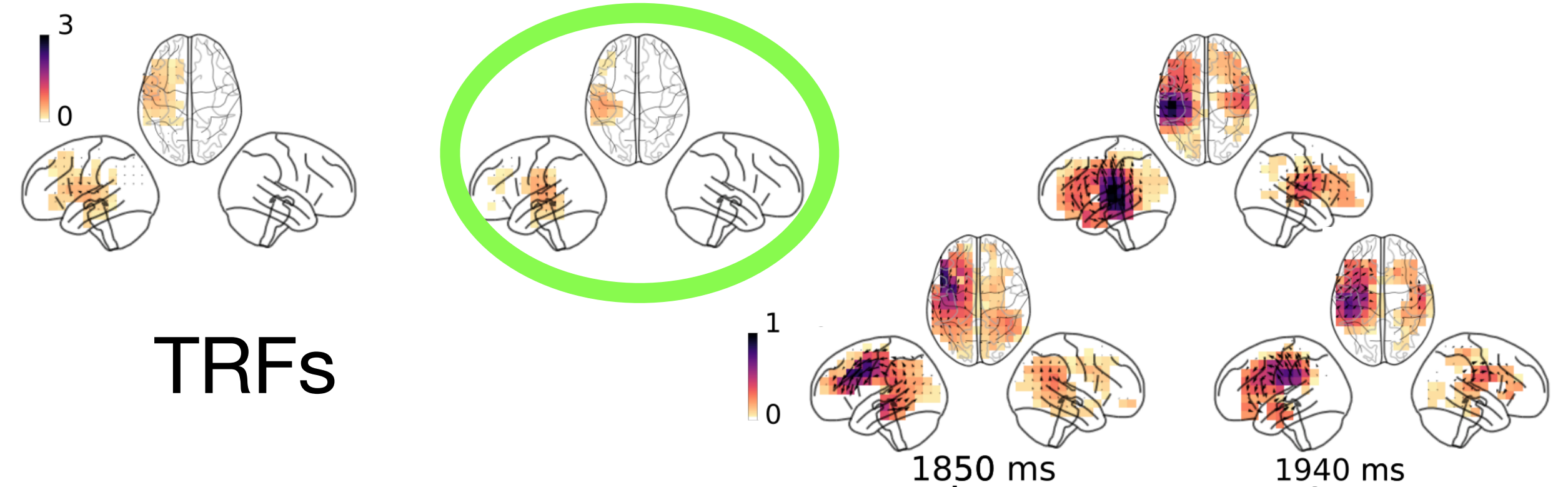
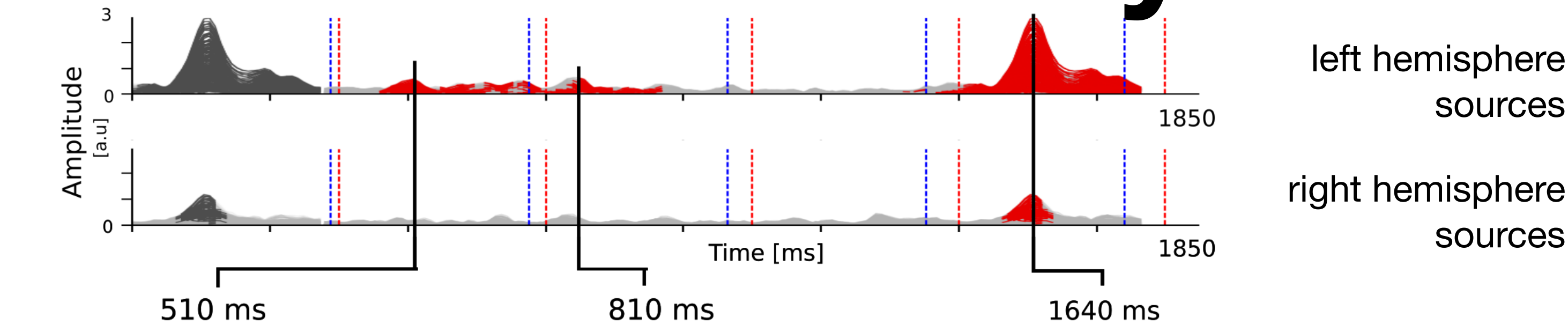
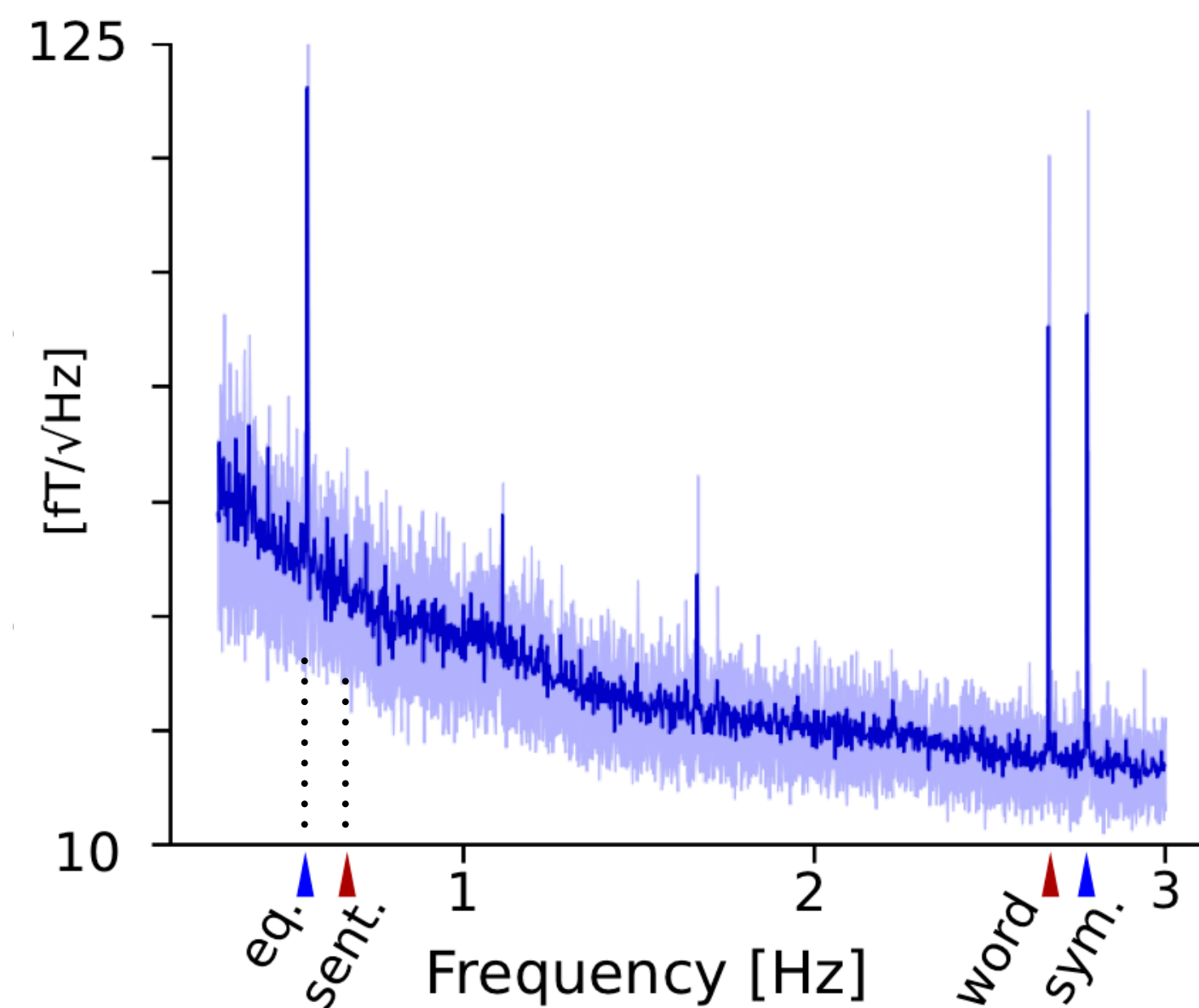


# Isochronous Cocktail Party

Attend to Sentences

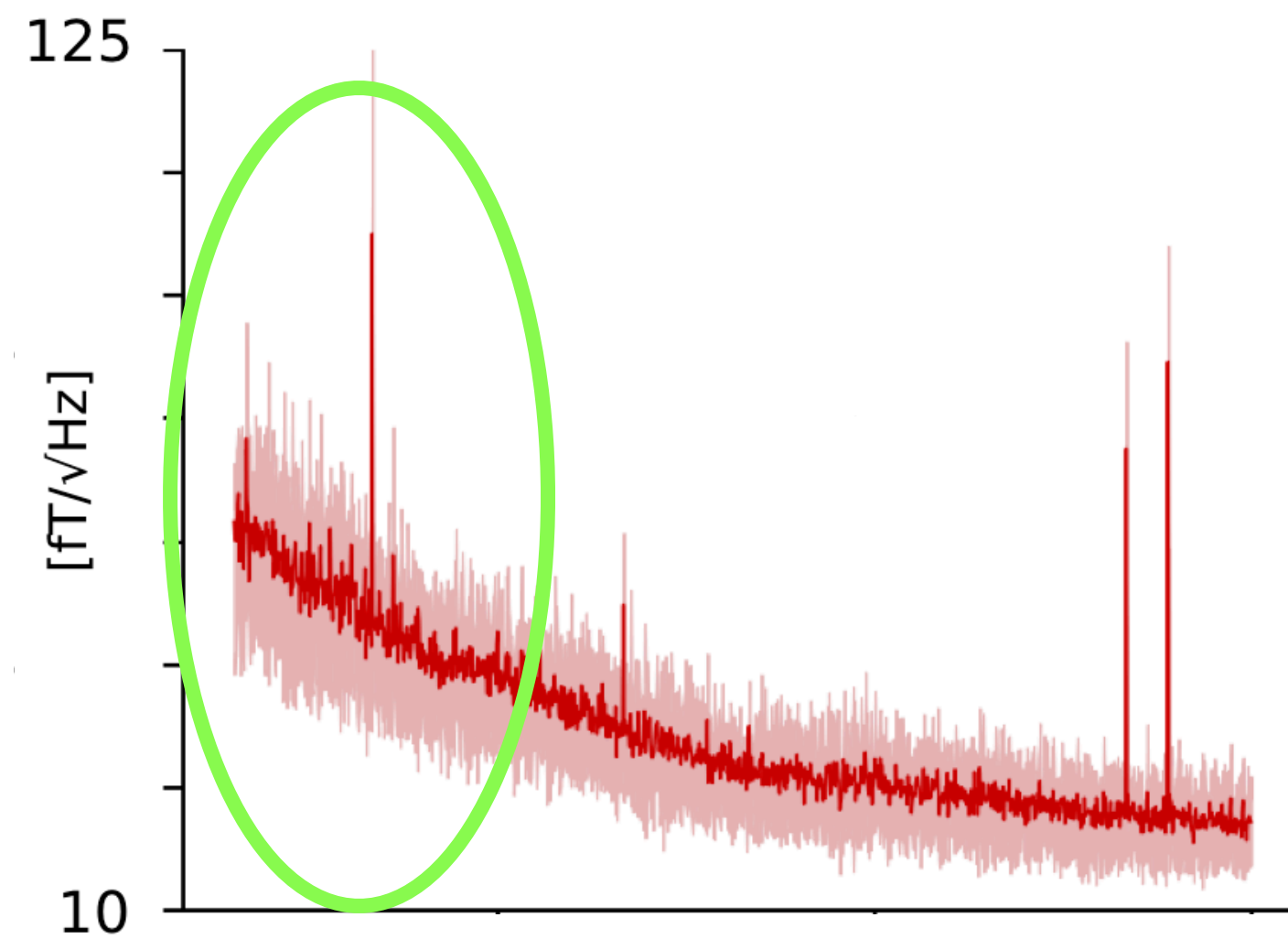


Attend to Equations

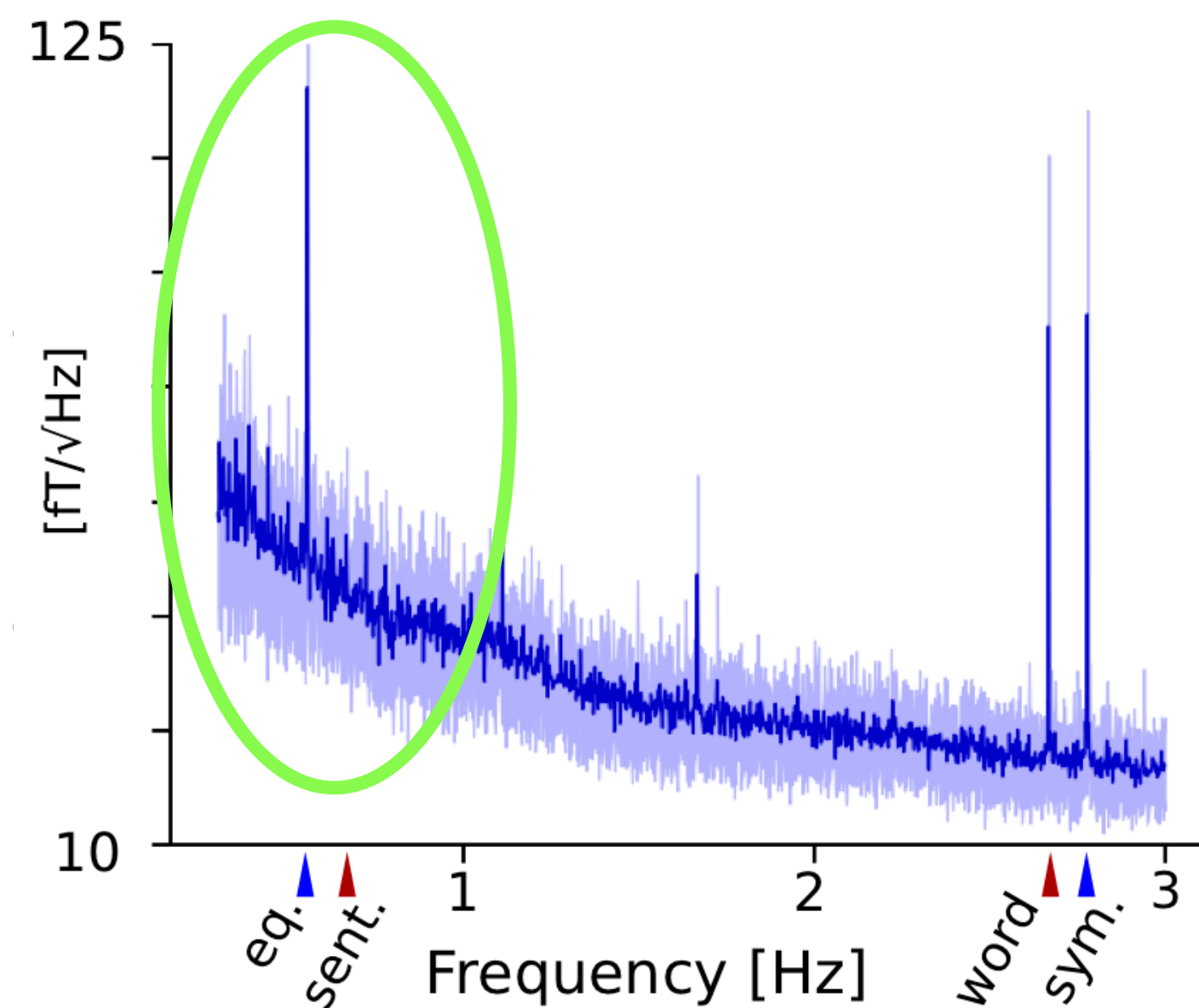


# Representations of Understanding

Attend to  
Sentences



Attend to  
Equations

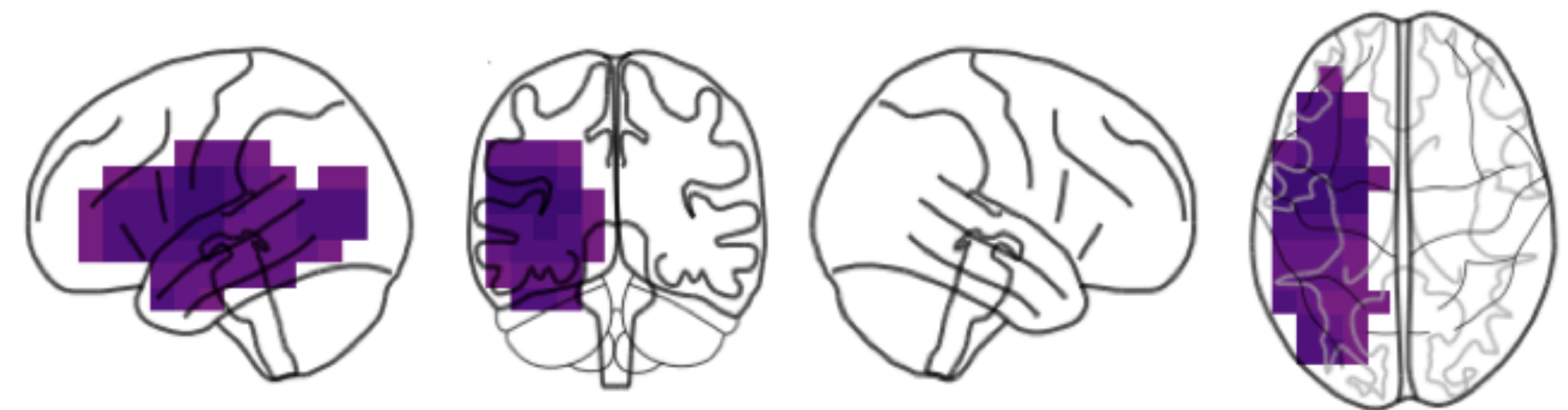
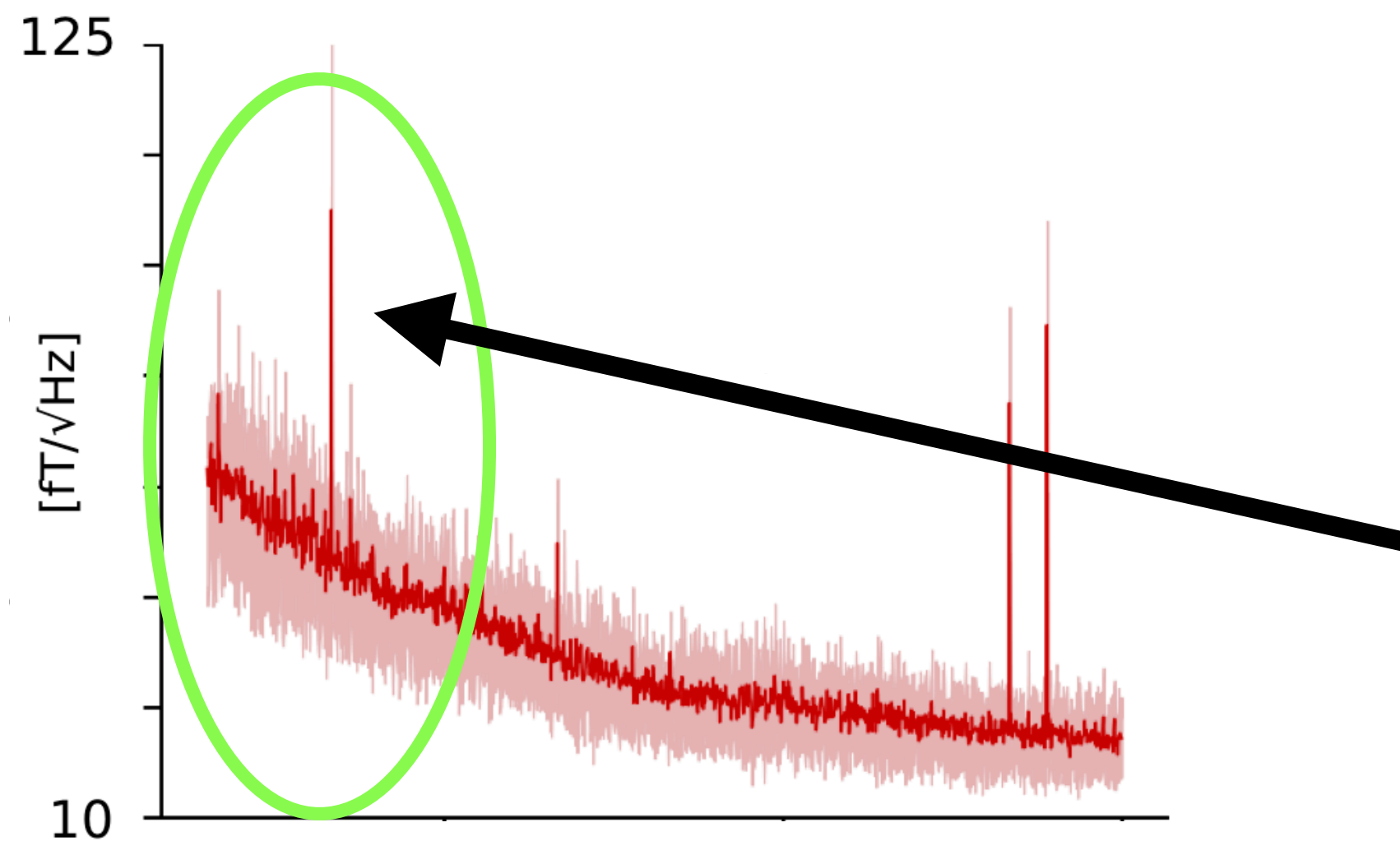




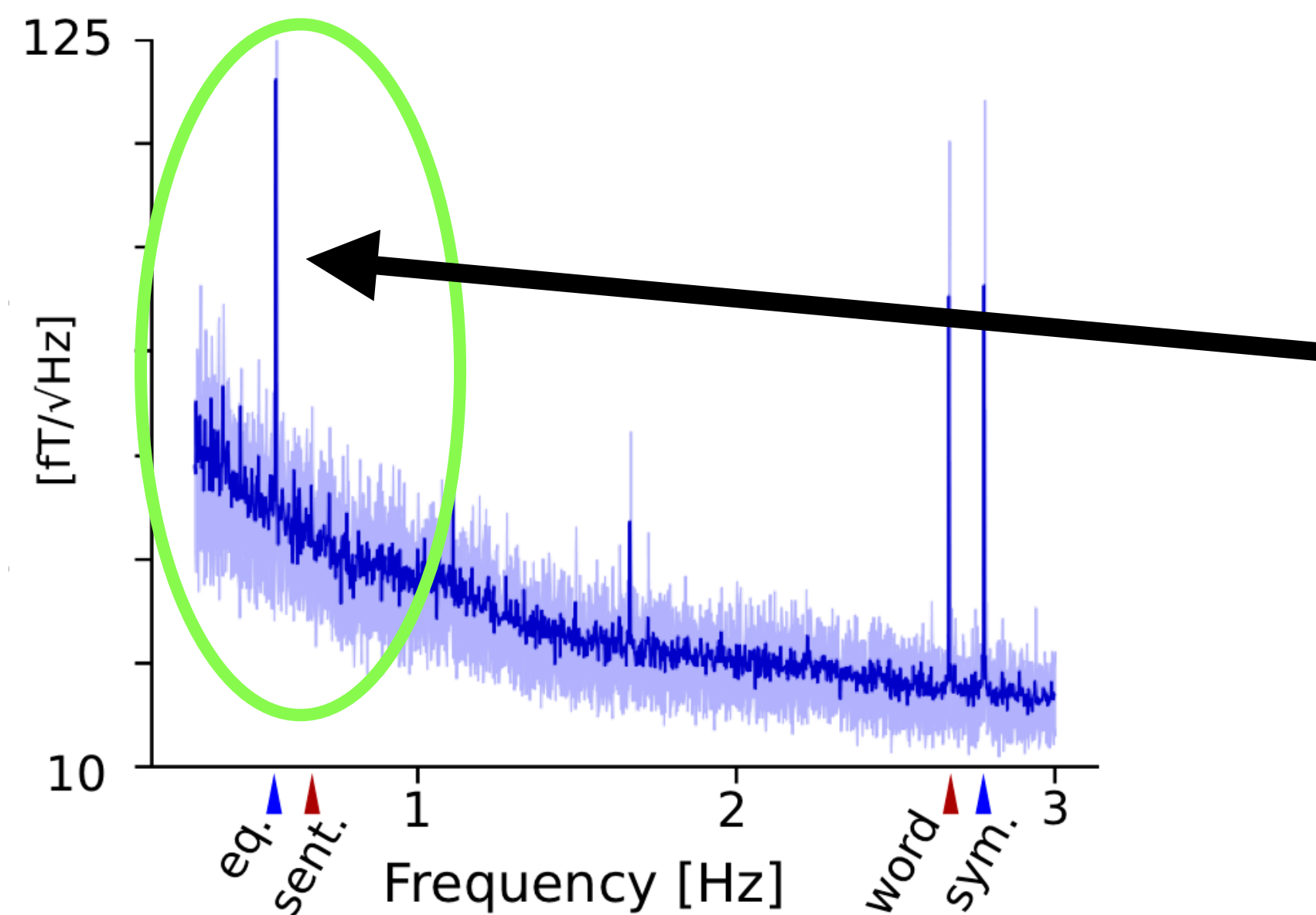
# Representations of Understanding

## Neural Correlation with Behavior

Attend to Sentences



Attend to Equations





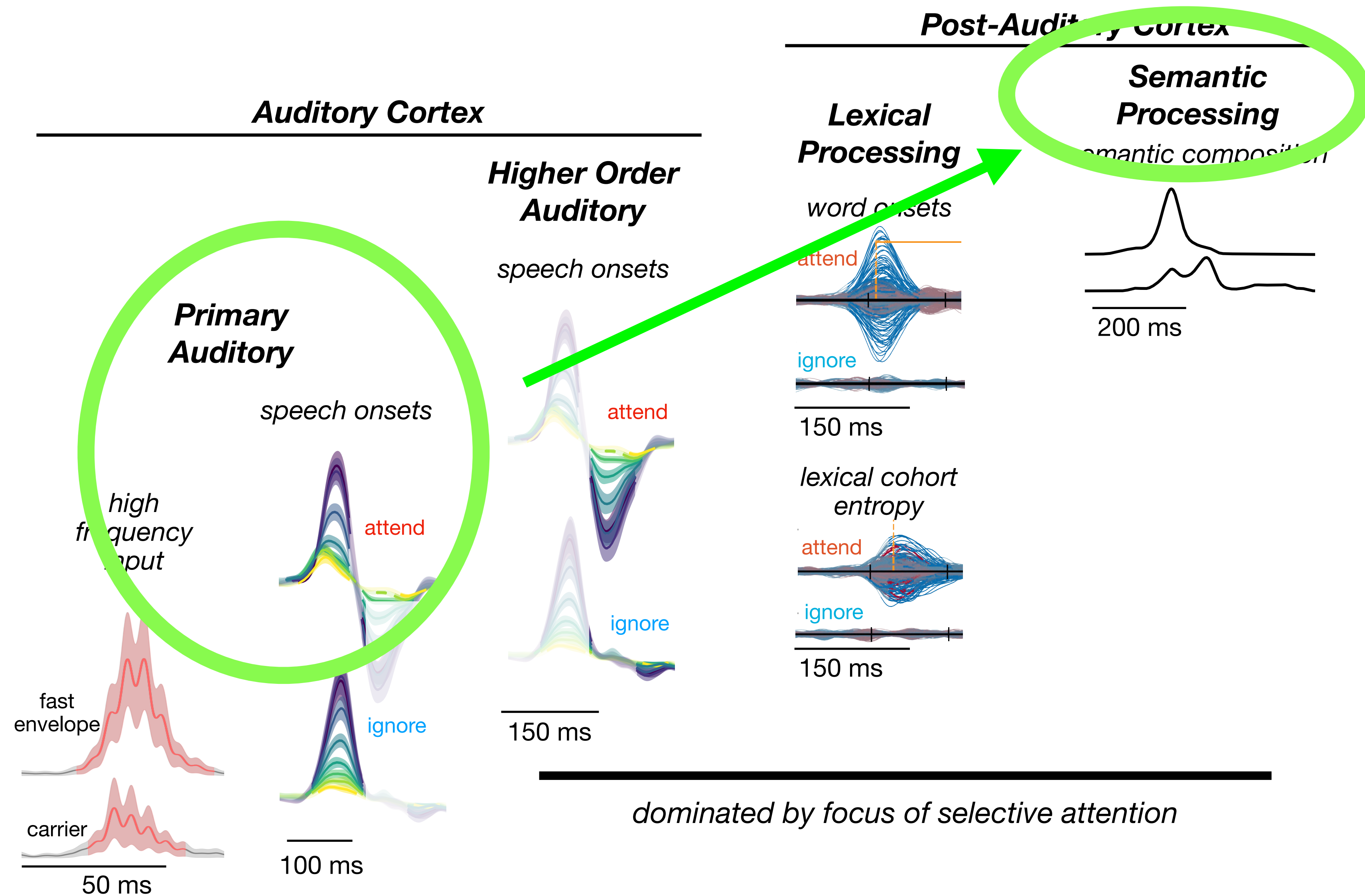
# Neural Markers of Comprehension

- 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

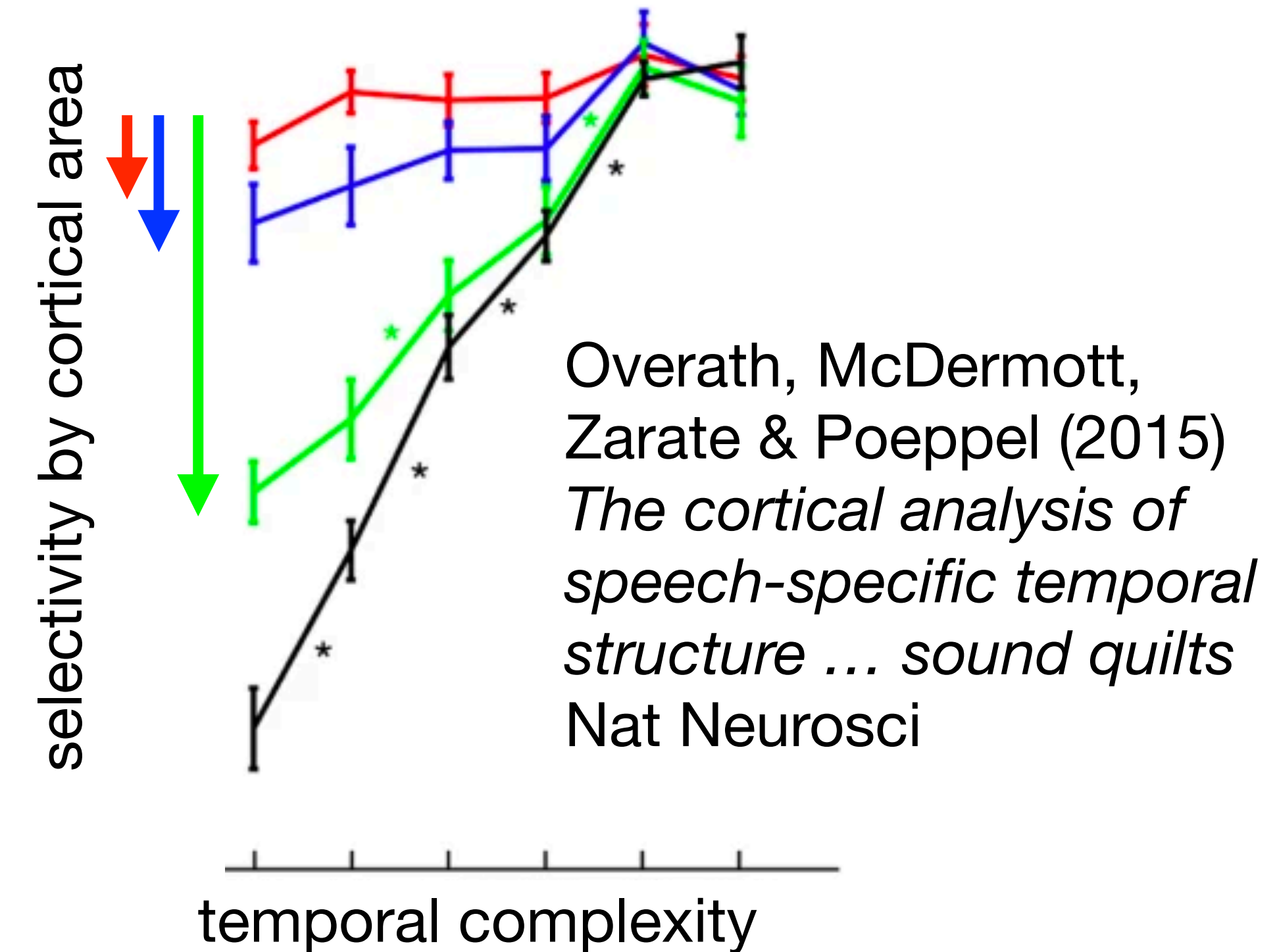
- Introduction—Cortical representations of continuous speech
- *Early & fast* cortical representation of continuous speech
- Cortical representations of speech *meaning*
- ***Progression*** of representations of continuous speech through cortex (bottom-up and top-down)
- Objective measures of speech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

# Cortical Representations Across Cortex



# Progression of Speech Representations

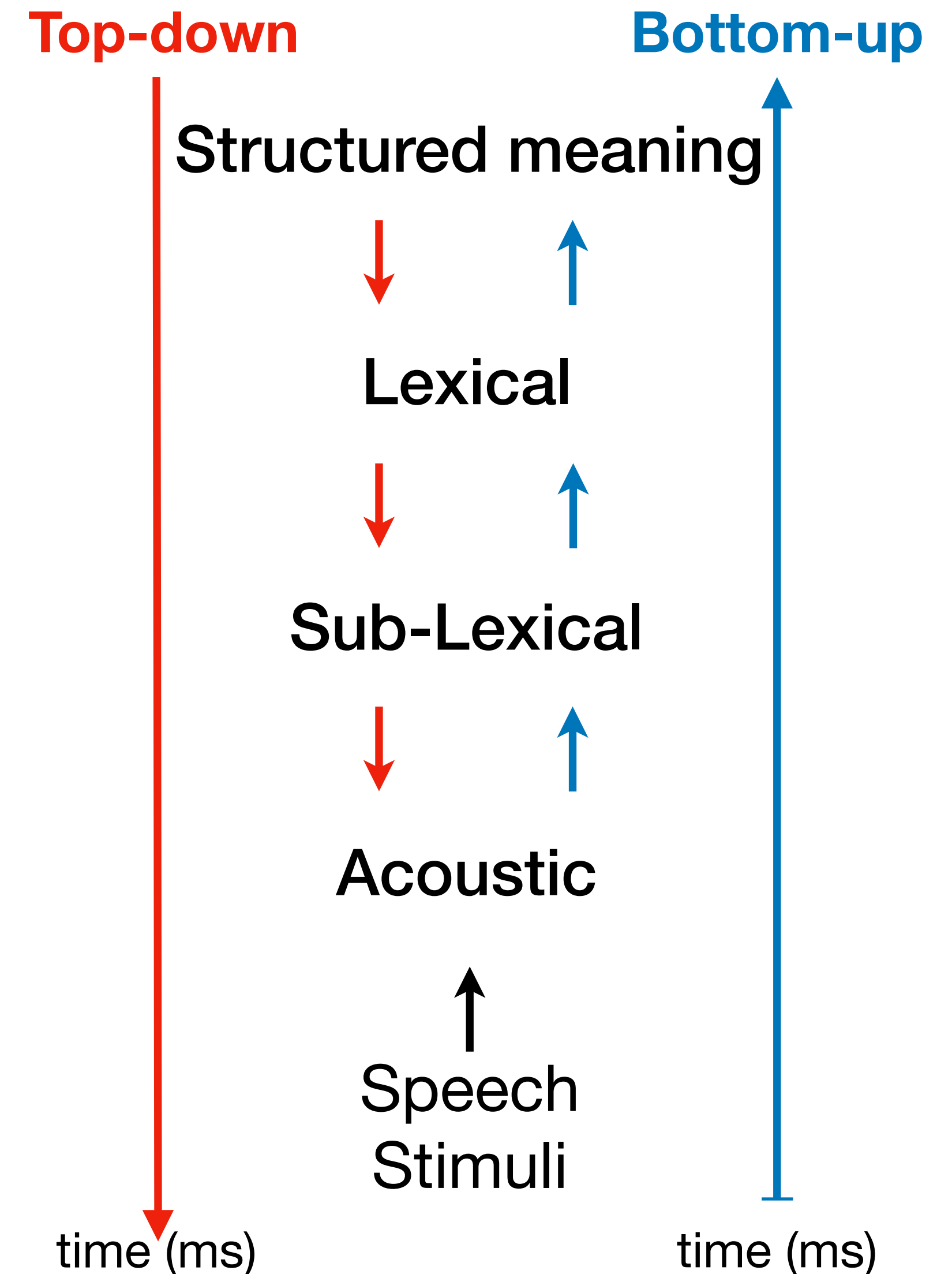
- Previous fMRI research on which brain regions process which speech and language features
- Progression of feature-based (bottom-up) levels
  - complex auditory stimulus, to
  - speech sounds, to
  - linguistic information via speech sounds
- Not all processing is straight bottom up
  - selective attention
  - secondary processing upon “error” detection
- MEG & EEG excel at showing temporal (i.e., latency) progression of processing





# 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



# Experimental Design

## Task

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

## Stimuli

4 passage types

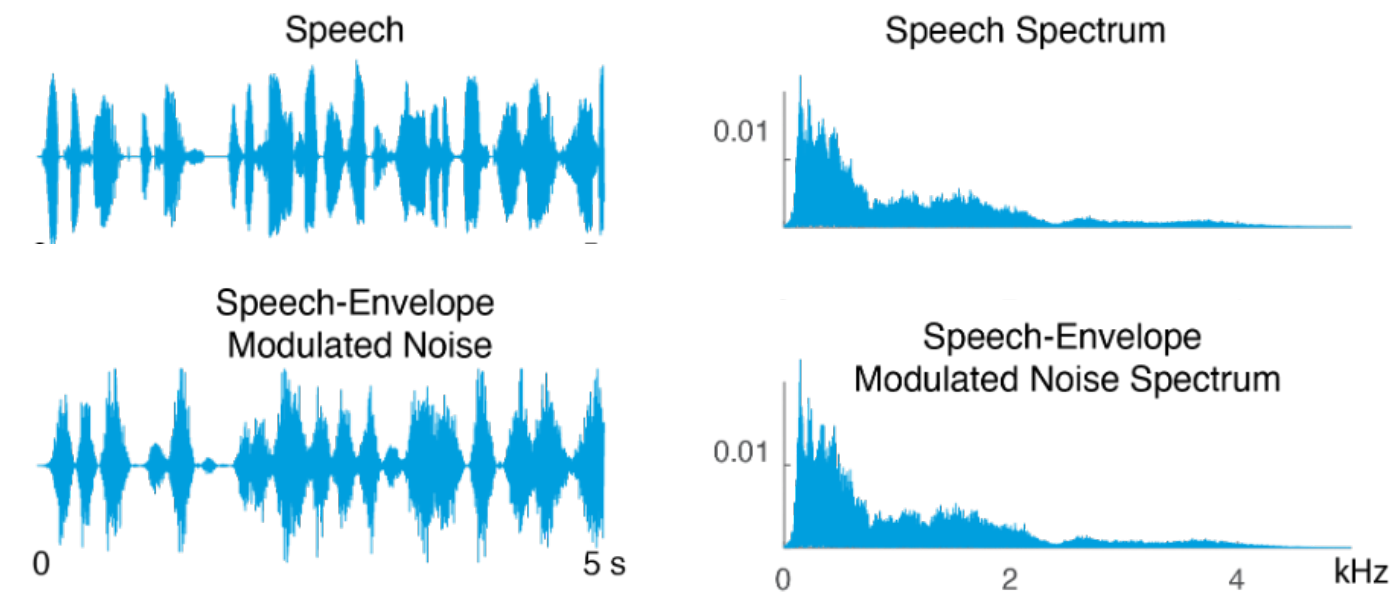
- Speech modulated noise
- Non-words
- Scrambled words
- Narrative

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



# 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 .....

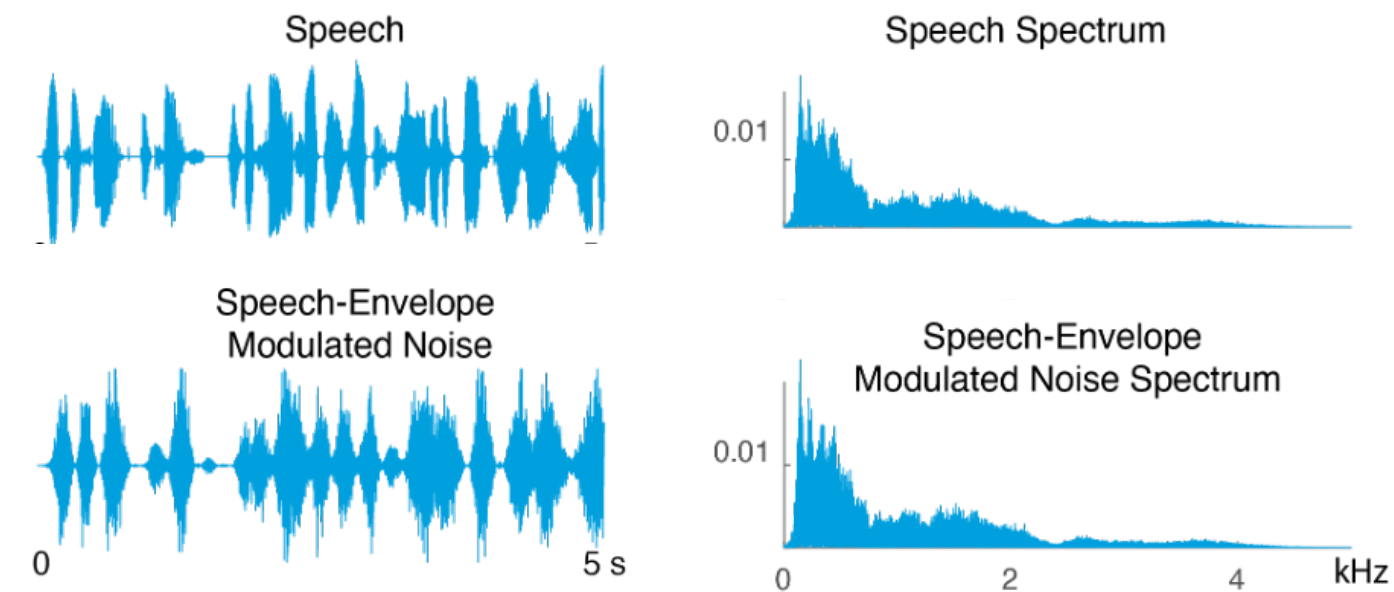
continuous-  
speech-like  
prosody and  
rhythm





# 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 .....

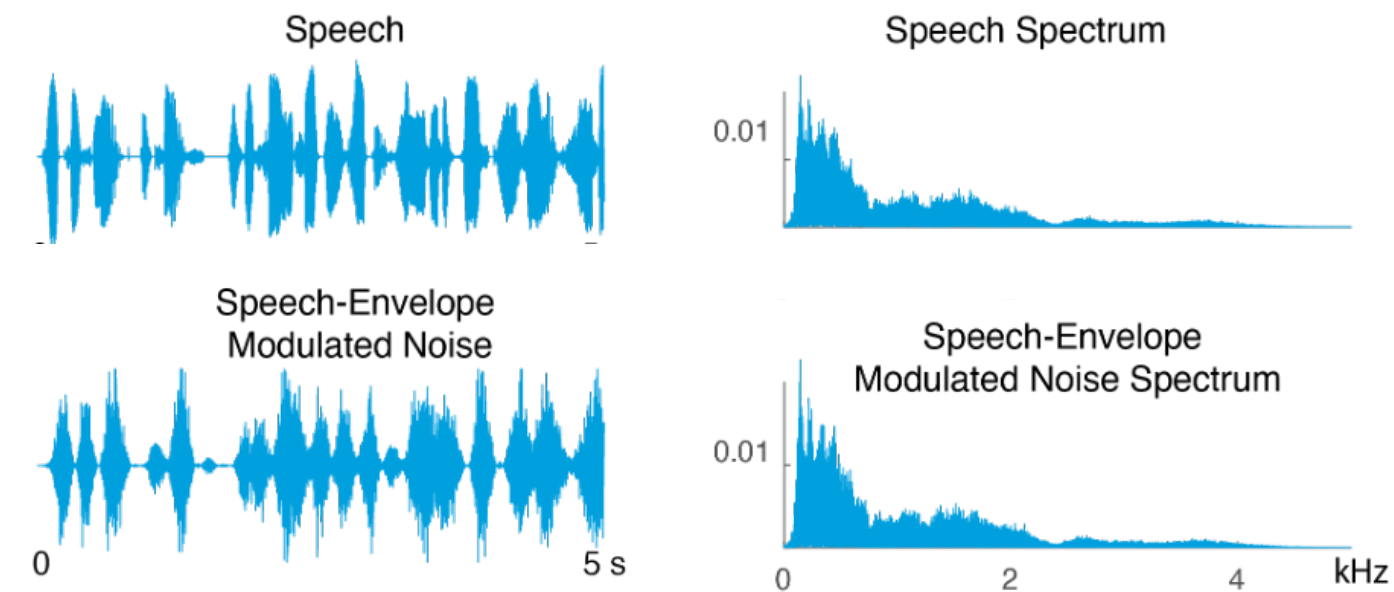
continuous-  
speech-like  
prosody and  
rhythm





# 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

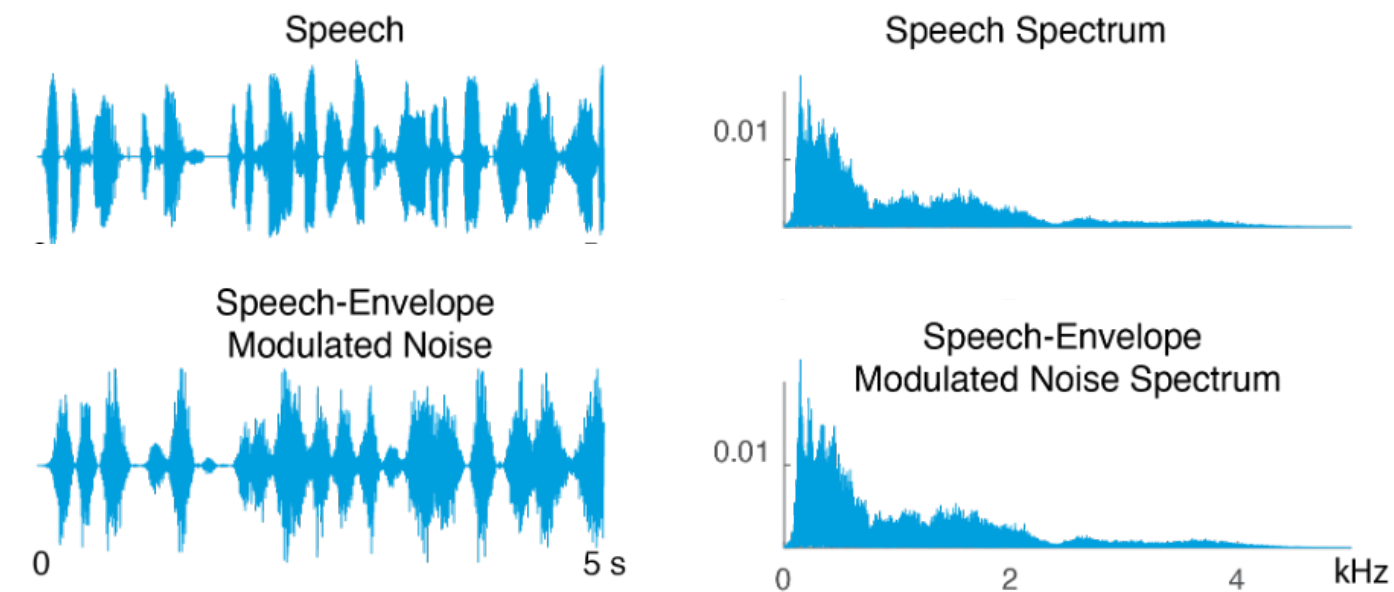
If you happened to find yourself on the banks of the Ohio River on a  
particular afternoon in the spring of 1806-somewhere just to the north of  
Wheeling, West Virginia, say, you would probably have noticed a strange  
makeshift craft drifting lazily down the river. At the time, this particular .....

continuous-  
speech-like  
prosody and  
rhythm



# 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 .....

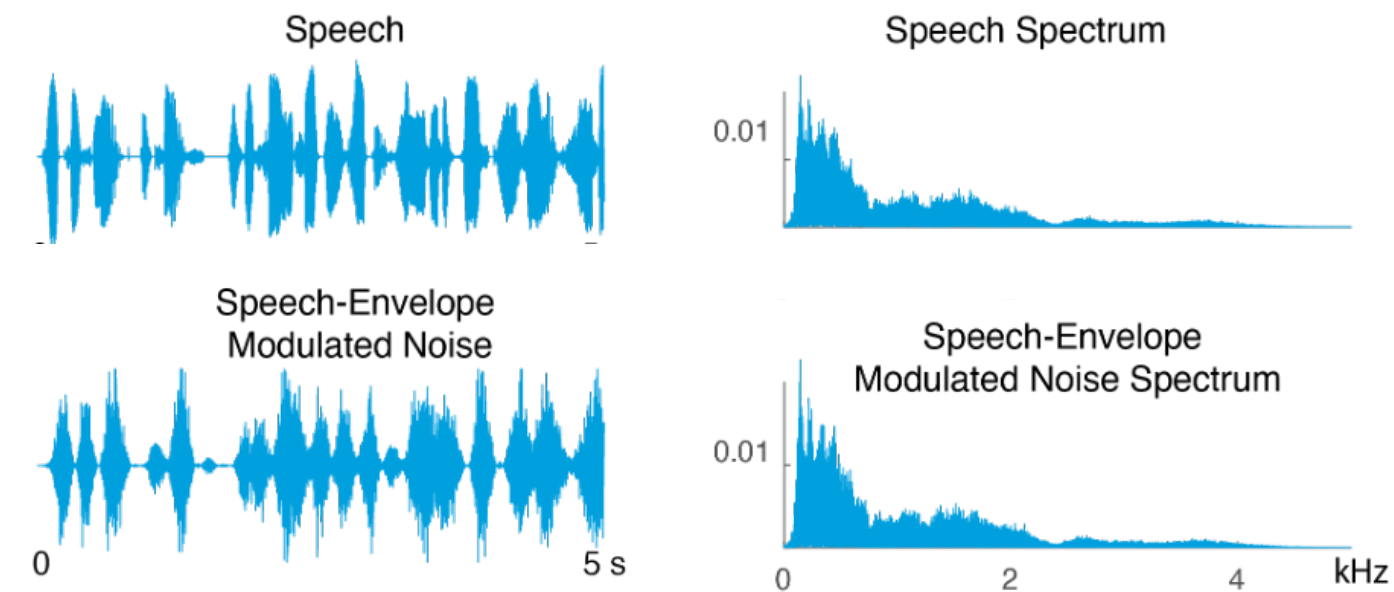
continuous-  
speech-like  
prosody and  
rhythm





# 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

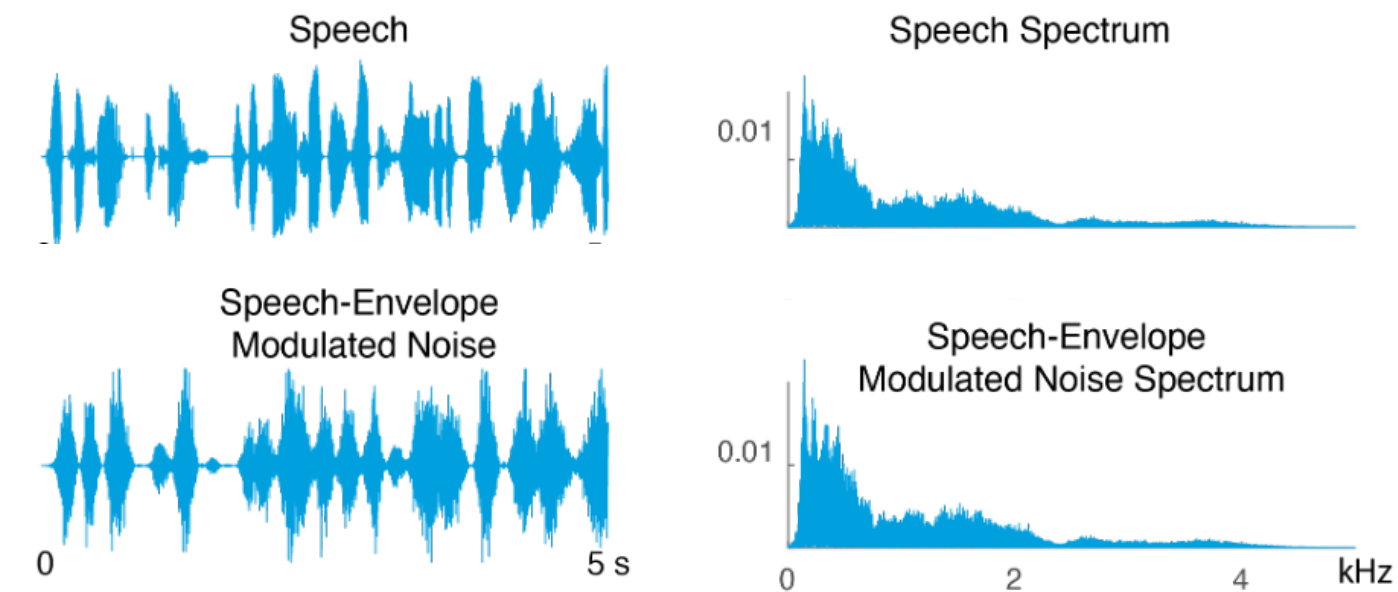
If you happened to find yourself on the banks of the Ohio River on a  
particular afternoon in the spring of 1806-somewhere just to the north of  
Wheeling, West Virginia, say, you would probably have noticed a strange  
makeshift craft drifting lazily down the river. At the time, this particular .....

continuous-  
speech-like  
prosody and  
rhythm



# 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 .....

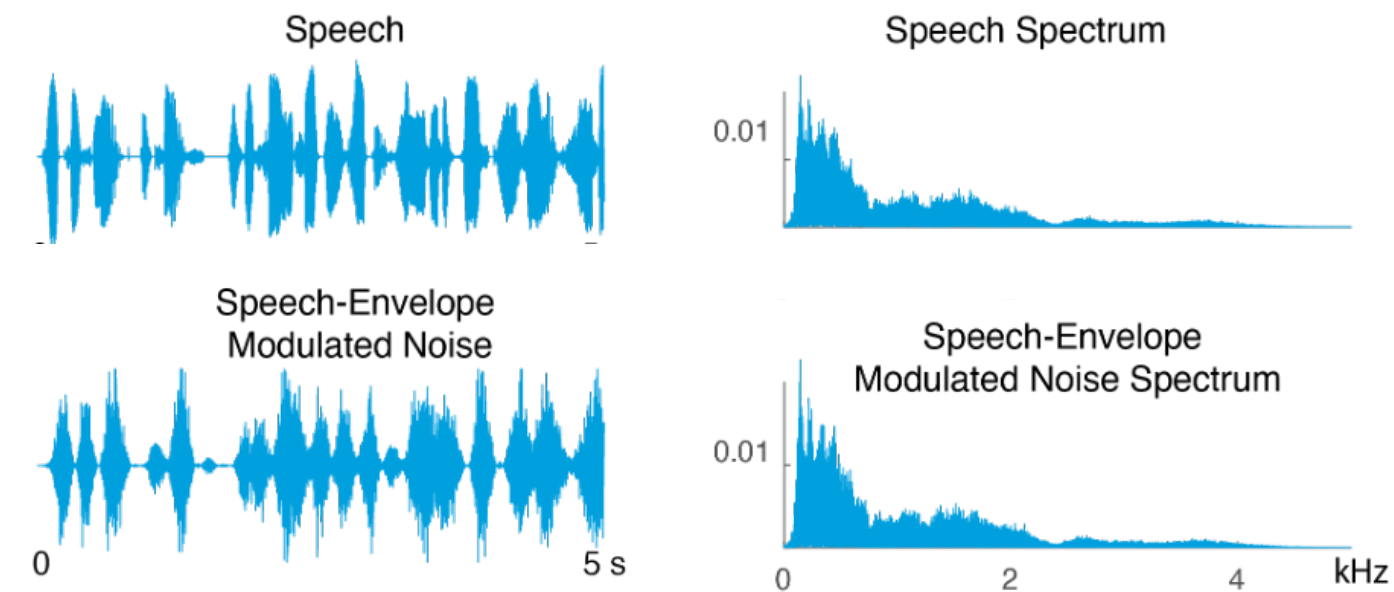
continuous-  
speech-like  
prosody and  
rhythm





# 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

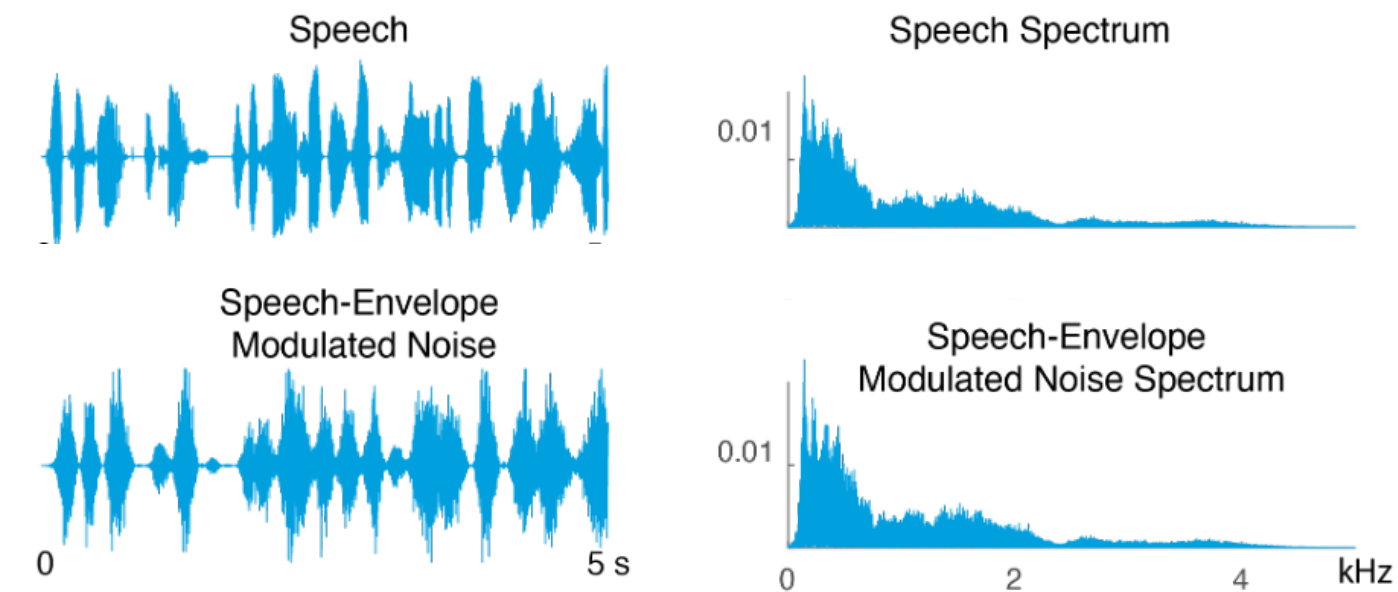
If you happened to find yourself on the banks of the Ohio River on a  
particular afternoon in the spring of 1806-somewhere just to the north of  
Wheeling, West Virginia, say, you would probably have noticed a strange  
makeshift craft drifting lazily down the river. At the time, this particular .....

continuous-  
speech-like  
prosody and  
rhythm



# 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 .....

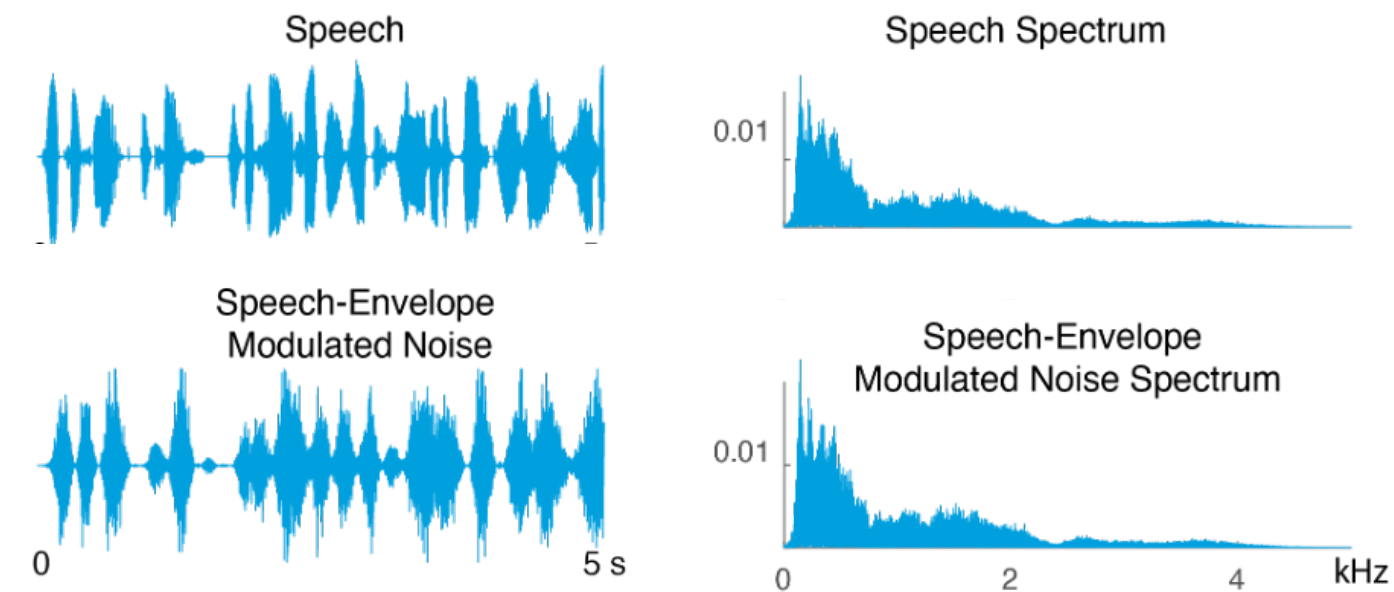
continuous-  
speech-like  
prosody and  
rhythm





# 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

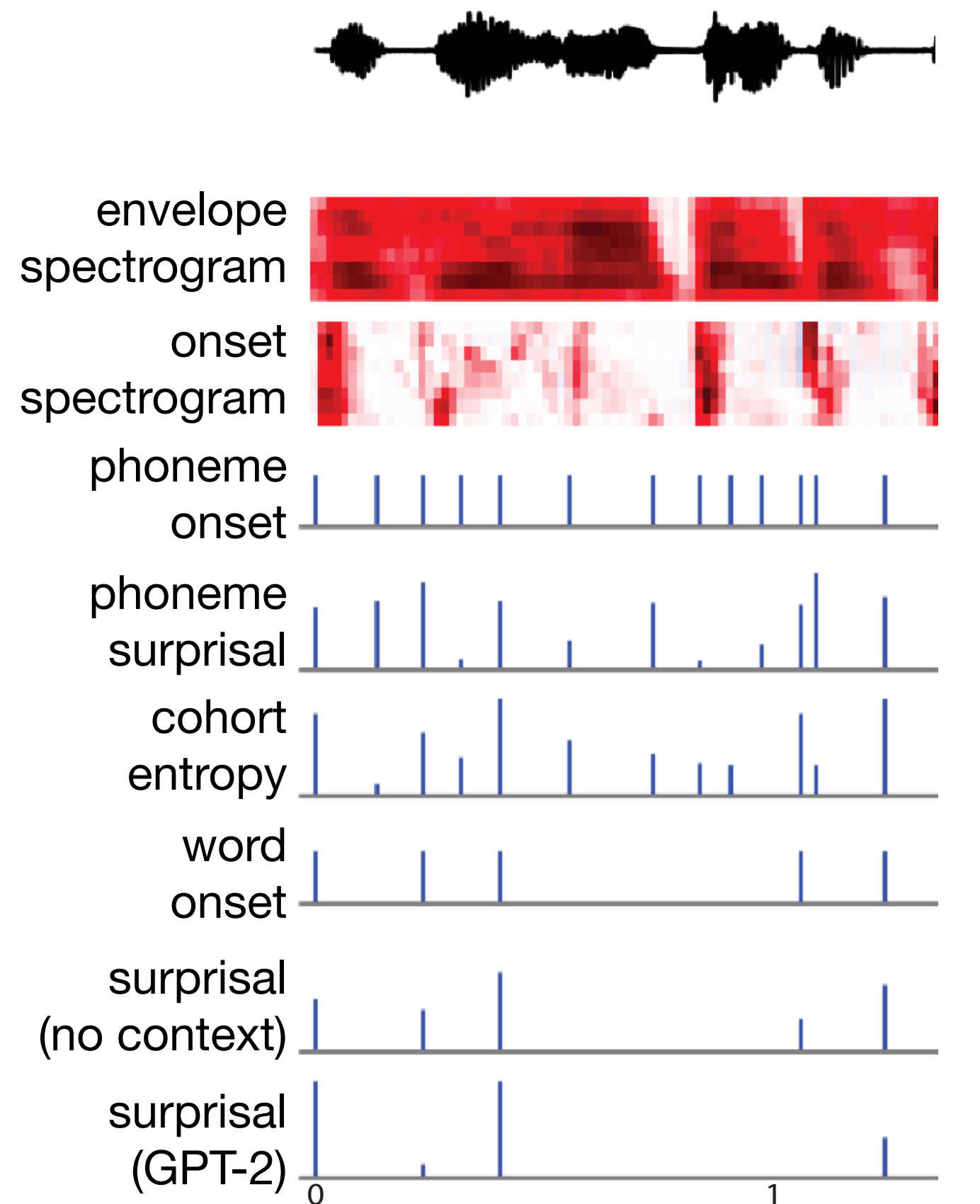
If you happened to find yourself on the banks of the Ohio River on a  
particular afternoon in the spring of 1806-somewhere just to the north of  
Wheeling, West Virginia, say, you would probably have noticed a strange  
makeshift craft drifting lazily down the river. At the time, this particular .....

continuous-  
speech-like  
prosody and  
rhythm

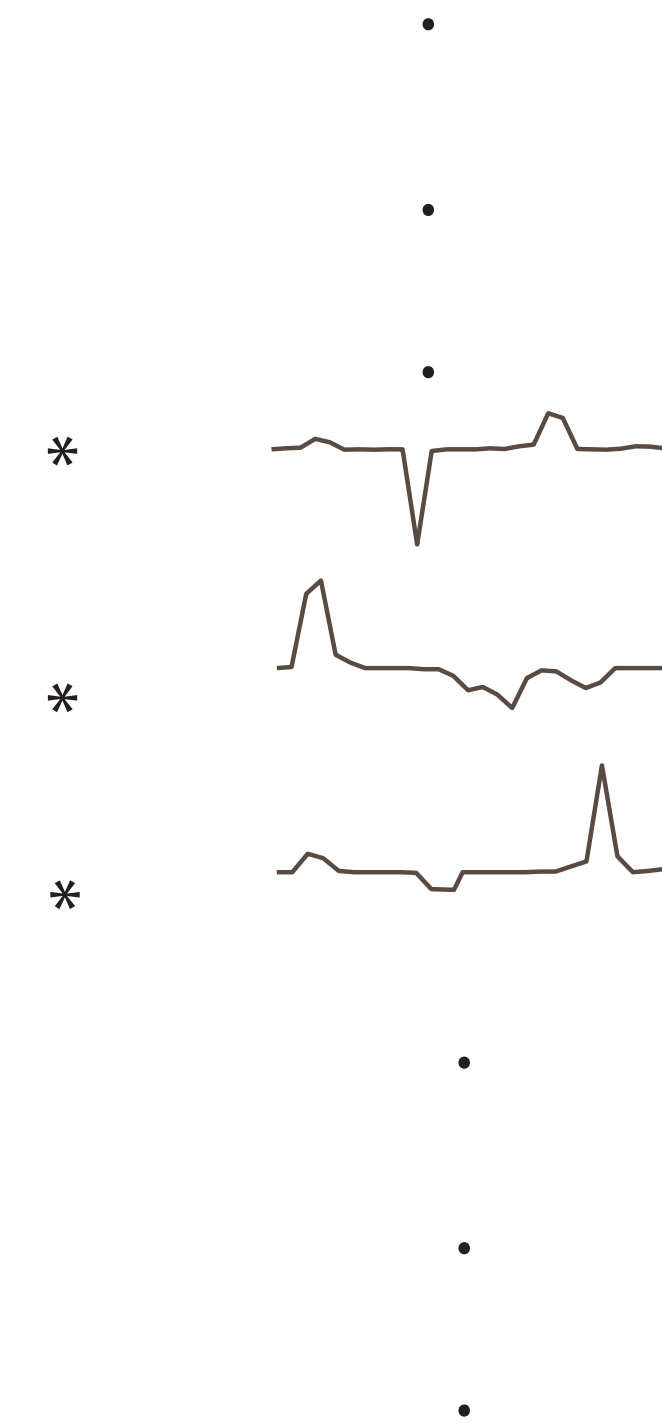


# Simultaneous Temporal Response Functions

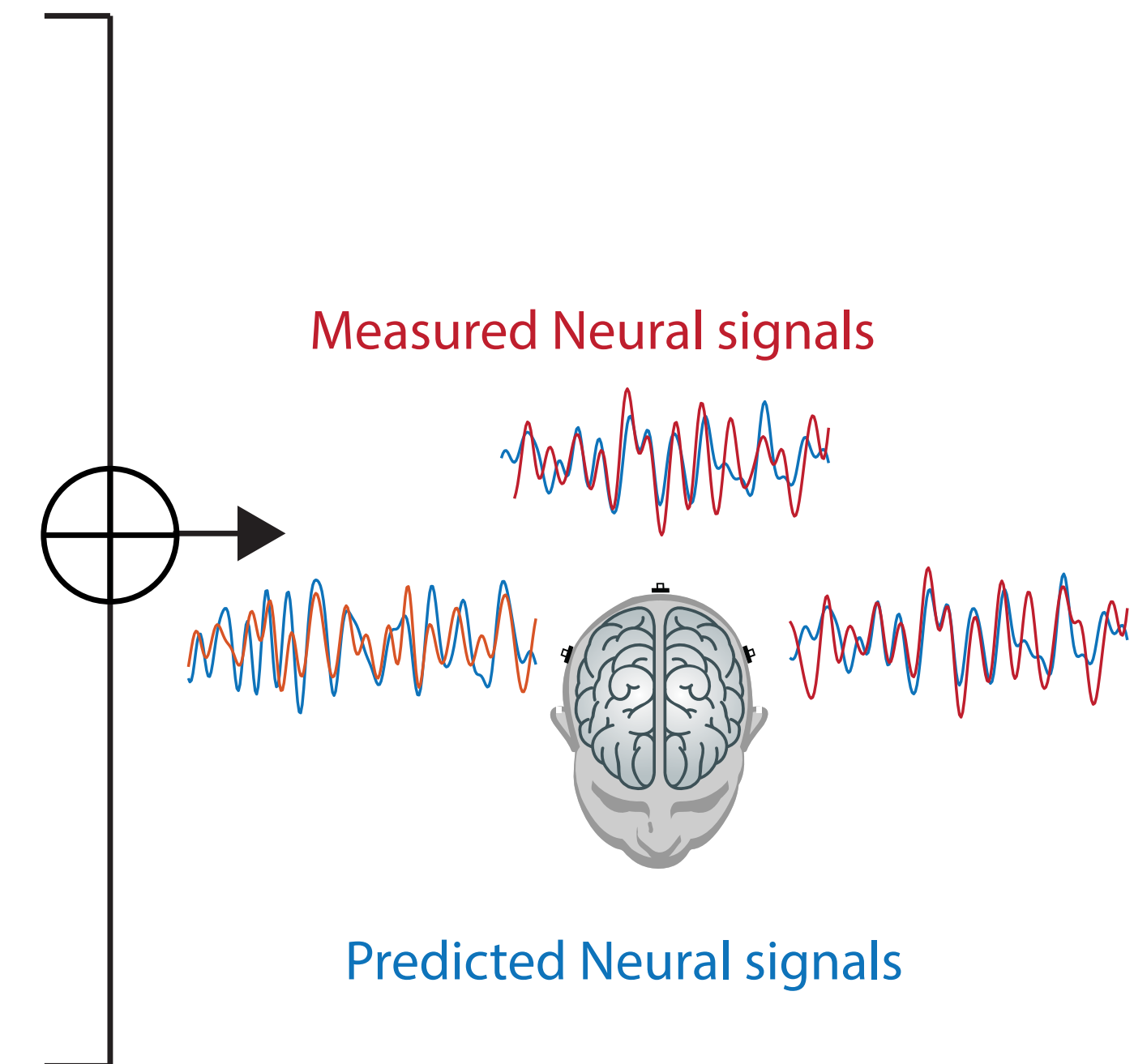
- TRFs predict neural response to speech
  - Analogous to evoked response
  - Peak amplitude  $\approx$  processing intensity
  - Peak Latency  $\approx$  source location
- Multiple TRFs estimated simultaneously
  - compete to explain variance (advantage over evoked response)



Speech Representations



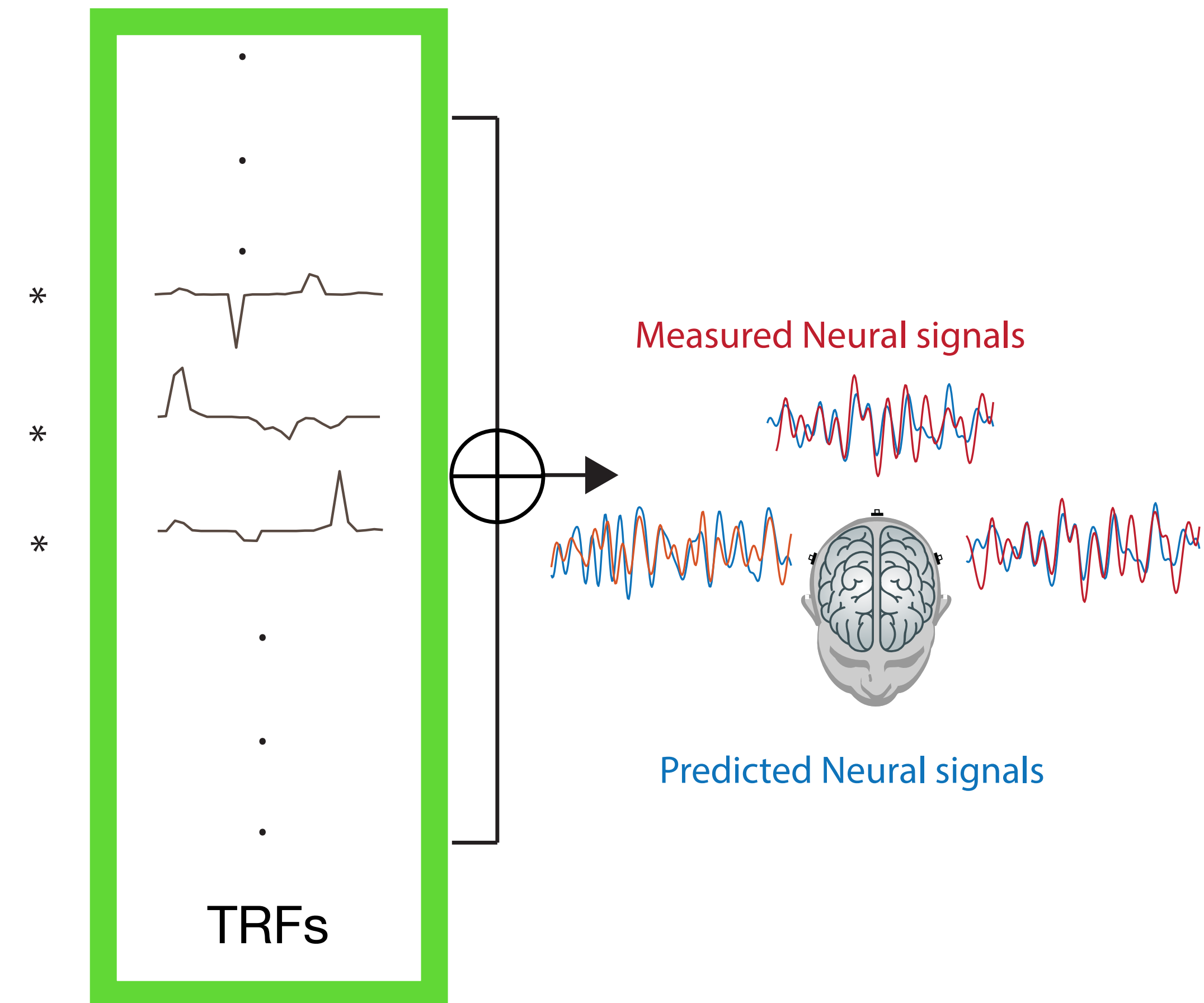
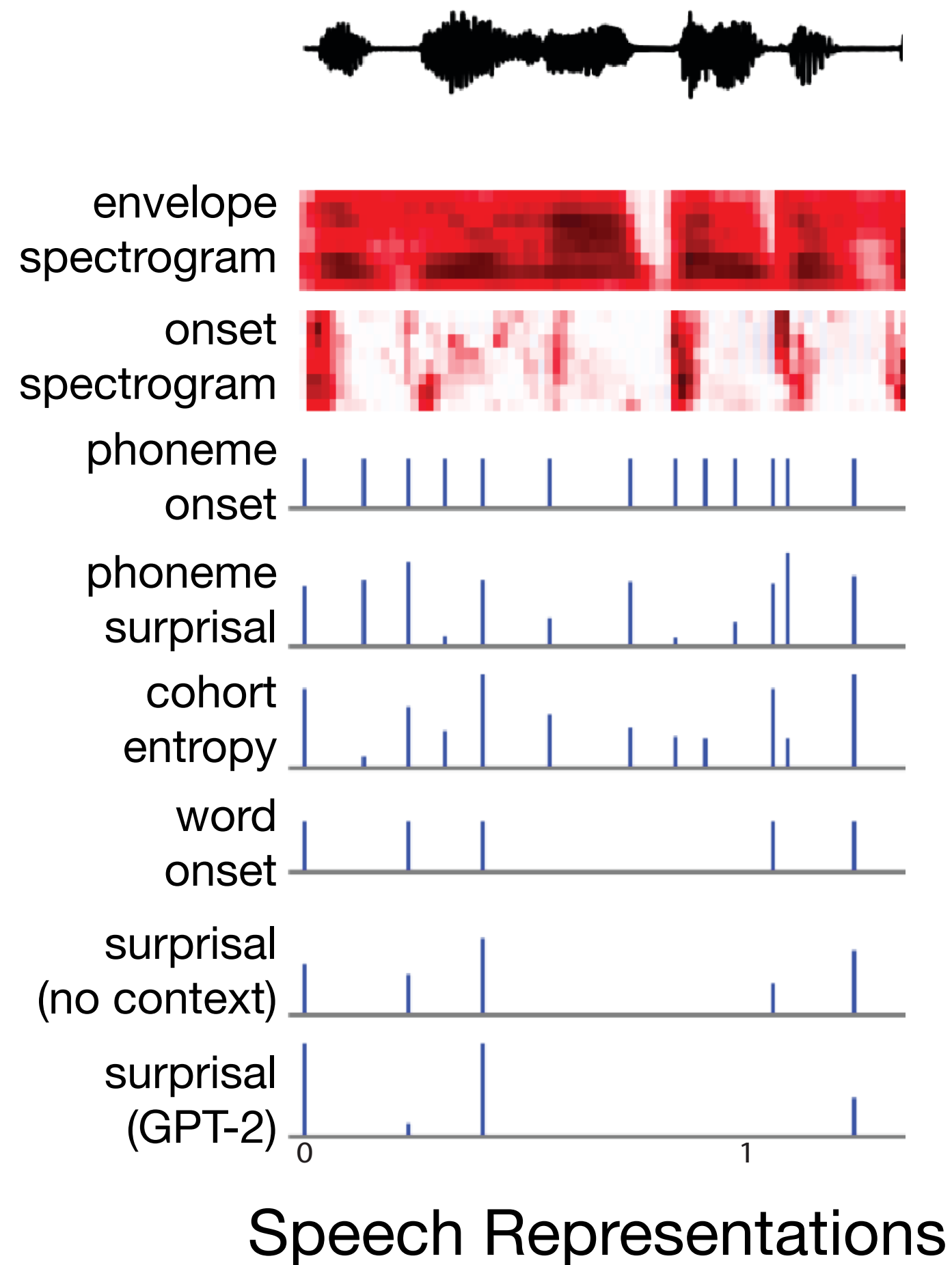
TRFs



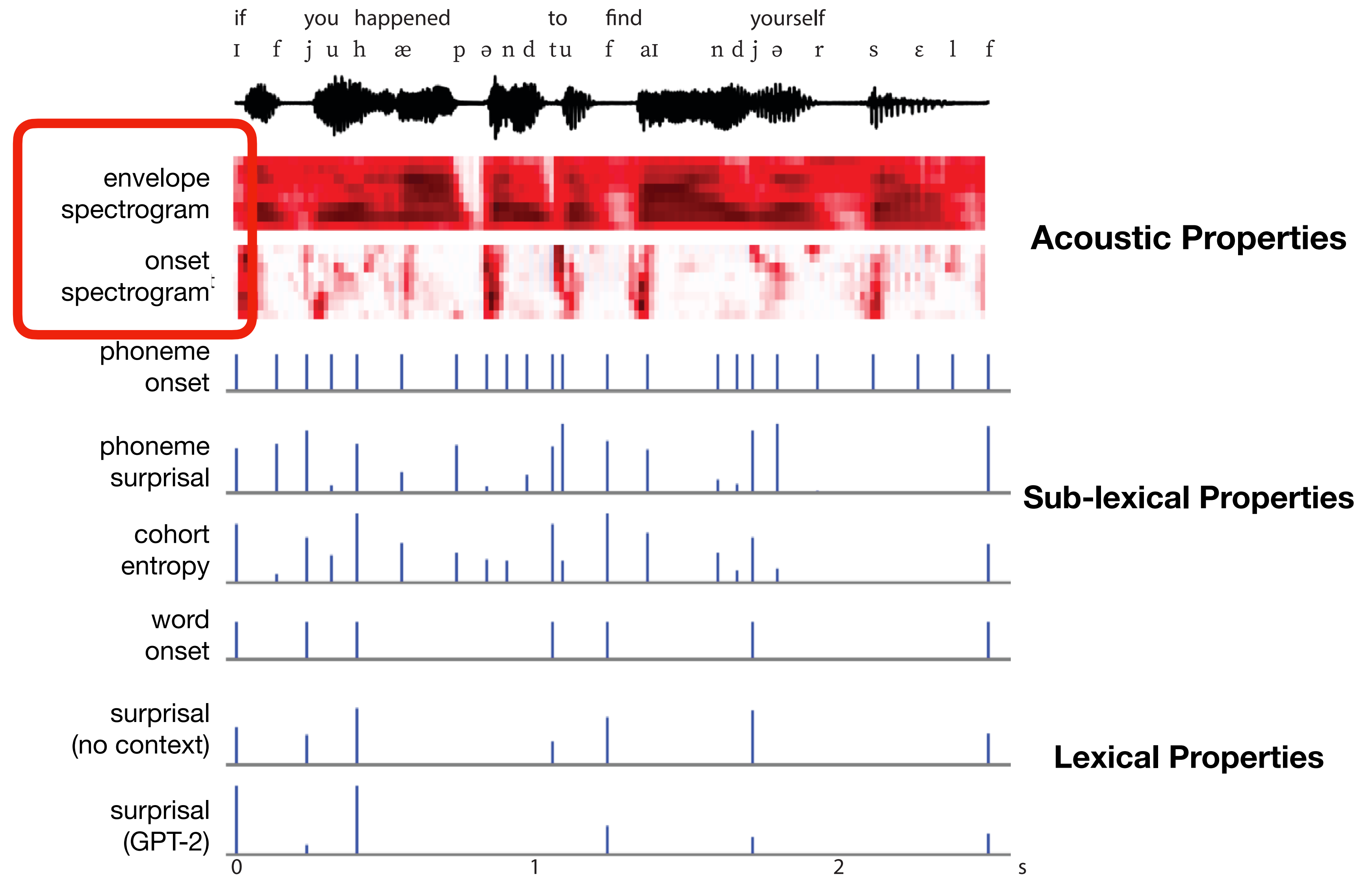


# Simultaneous Temporal Response Functions

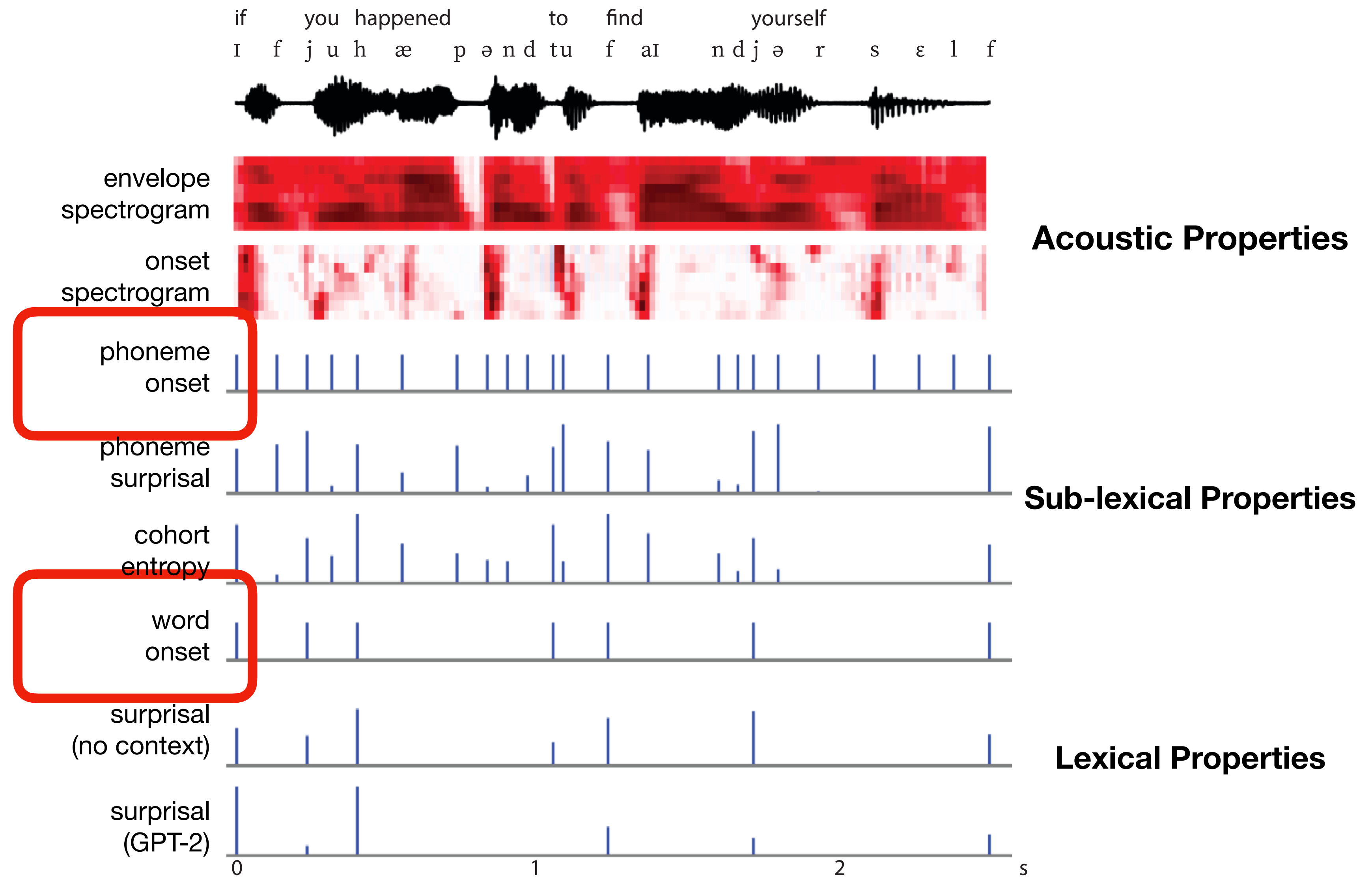
- TRFs predict neural response to speech
  - Analogous to evoked response
  - Peak amplitude  $\approx$  processing intensity
  - Peak Latency  $\approx$  source location
- Multiple TRFs estimated simultaneously
  - compete to explain variance (advantage over evoked response)



# Speech Representations



# Speech Representations



# Speech Representations

if you happened to find yourself  
ɪ f j u h æ p ɒ n d t u f aɪ n d j ə r s ε l f



envelope  
spectrogram



onset  
spectrogram



**Acoustic Properties**

phoneme  
onset



phoneme  
surprisal



**Sub-lexical Properties**

cohort  
entropy



word  
onset



surprisal  
(no context)



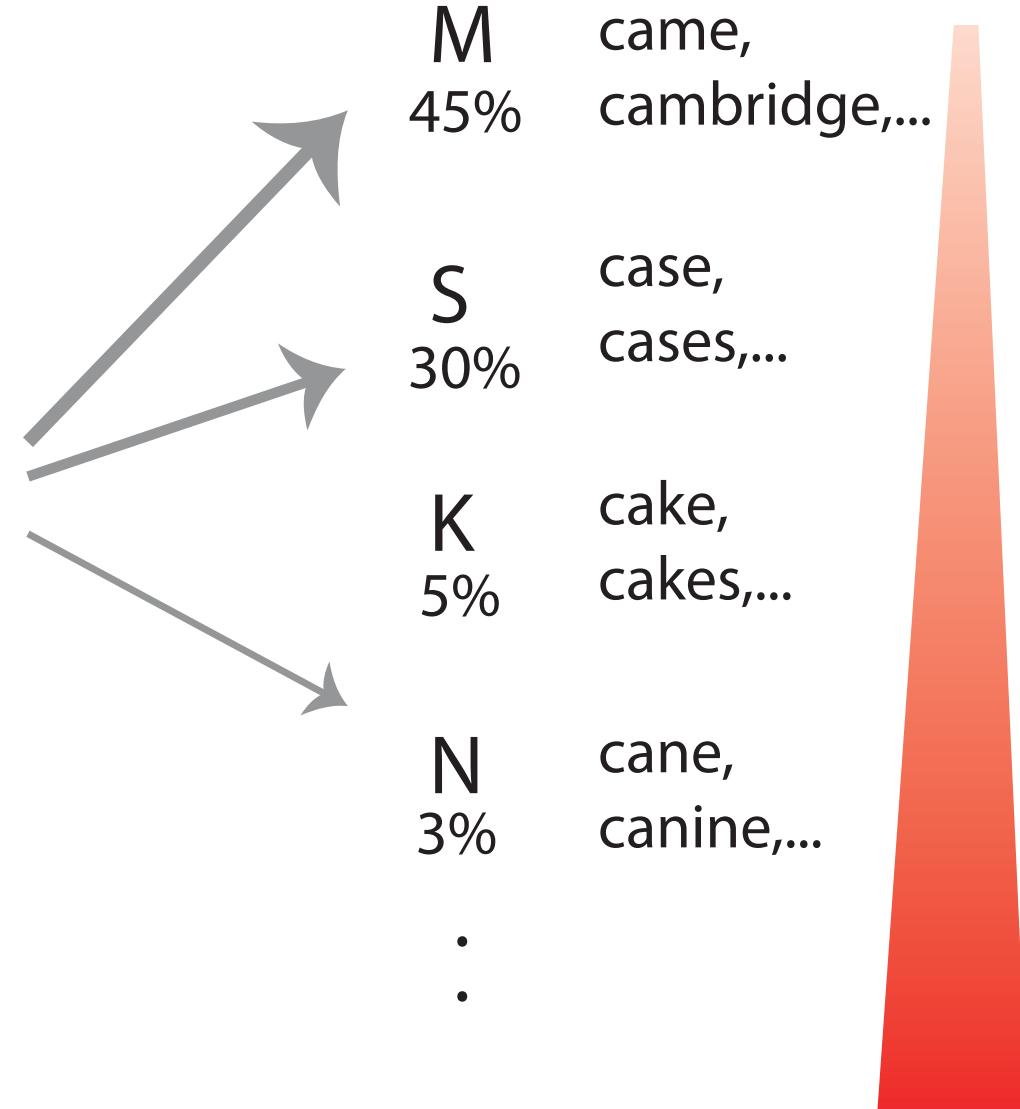
**Lexical Properties**

surprisal  
(GPT-2)



0 1 2 s

KEY —



M  
45%  
came,  
cambridge,...

S  
30%  
case,  
cases,...

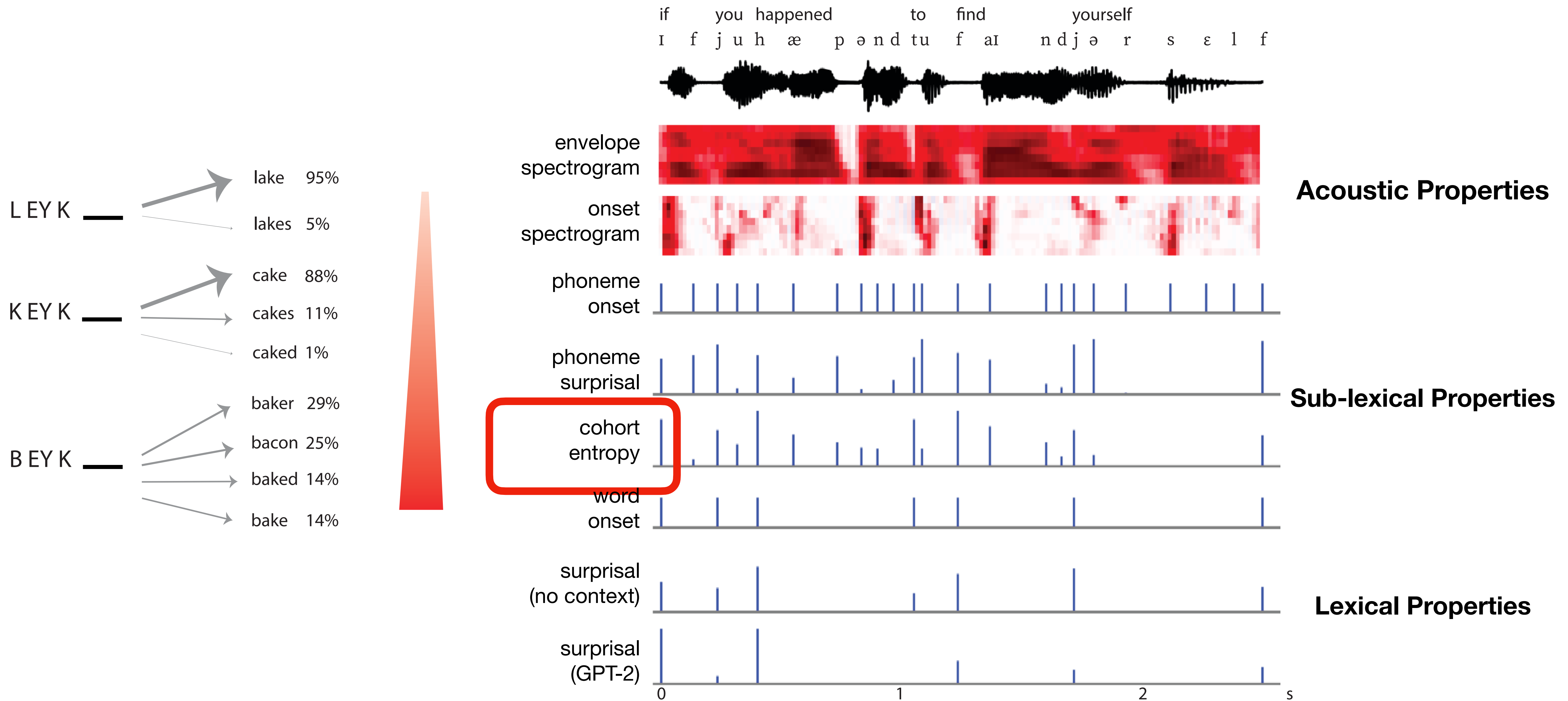
K  
5%  
cake,  
cakes,...

N  
3%  
cane,  
canine,...

:



# Speech Representations



# Speech Representations

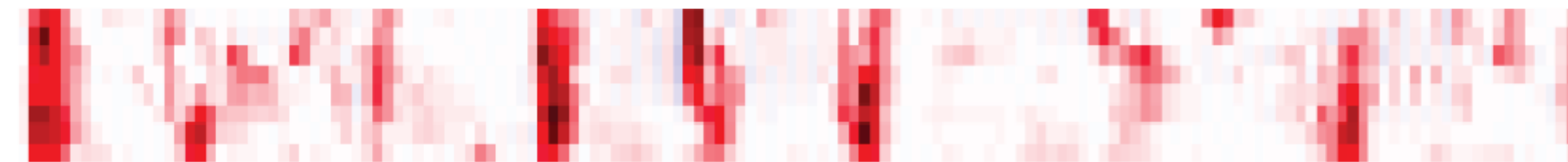
if you happened to find yourself  
ɪ f j u h æ p ə n d t u f aɪ n d j ə r s ε l f



envelope  
spectrogram

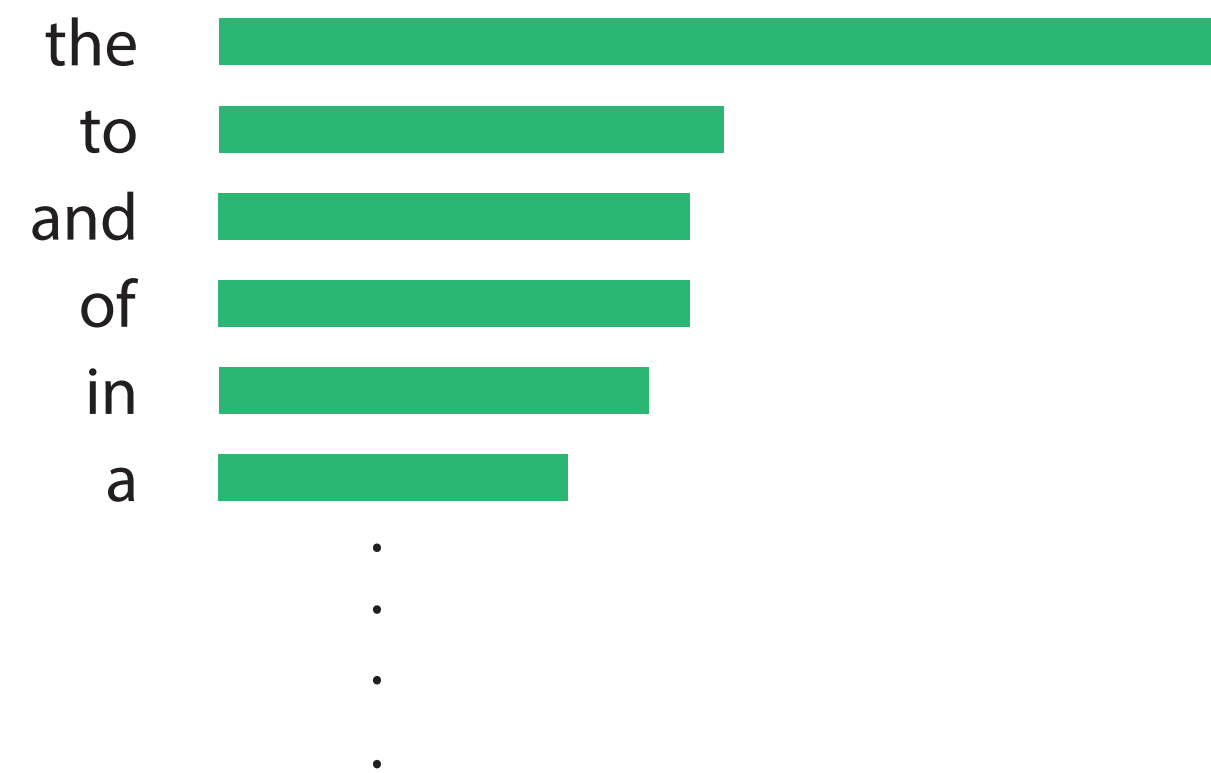


onset  
spectrogram



**Acoustic Properties**

Frequency of words based on SUBTLEX



phoneme  
onset



phoneme  
surprisal



**Sub-lexical Properties**

cohort  
entropy



word  
onset



surprisal  
(no context)



**Lexical Properties**

surprisal  
(GPT-2)



0 1 2 s

# Speech Representations

if you happened to find yourself  
ɪ f j u h æ p ɒ n d t u f aɪ n d j ə r s ε l f



envelope  
spectrogram



onset  
spectrogram



**Acoustic Properties**

phoneme  
onset



phoneme  
surprisal



cohort  
entropy



word  
onset



surprisal  
(no context)

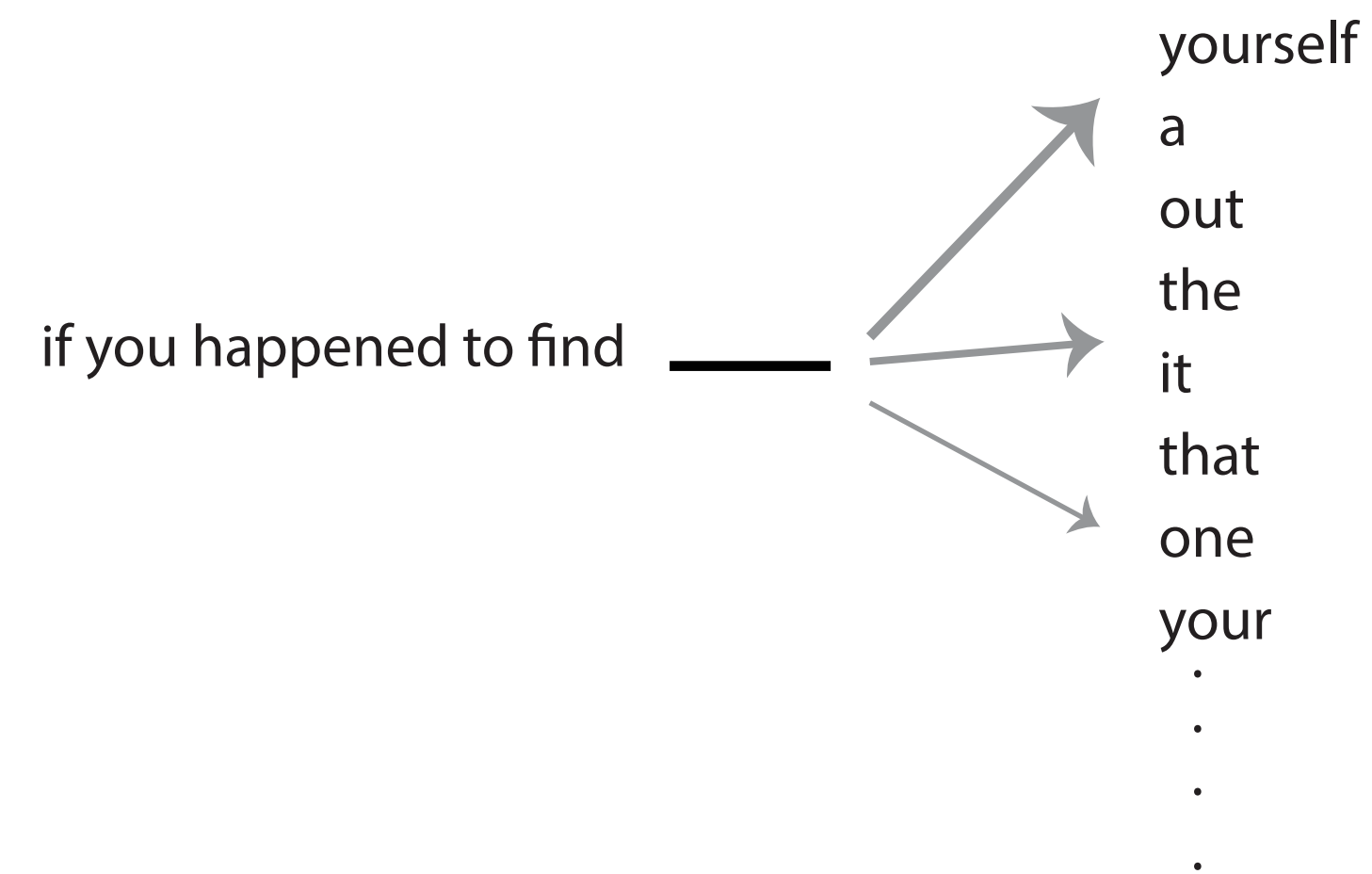


**Lexical Properties**

surprisal  
(GPT-2)

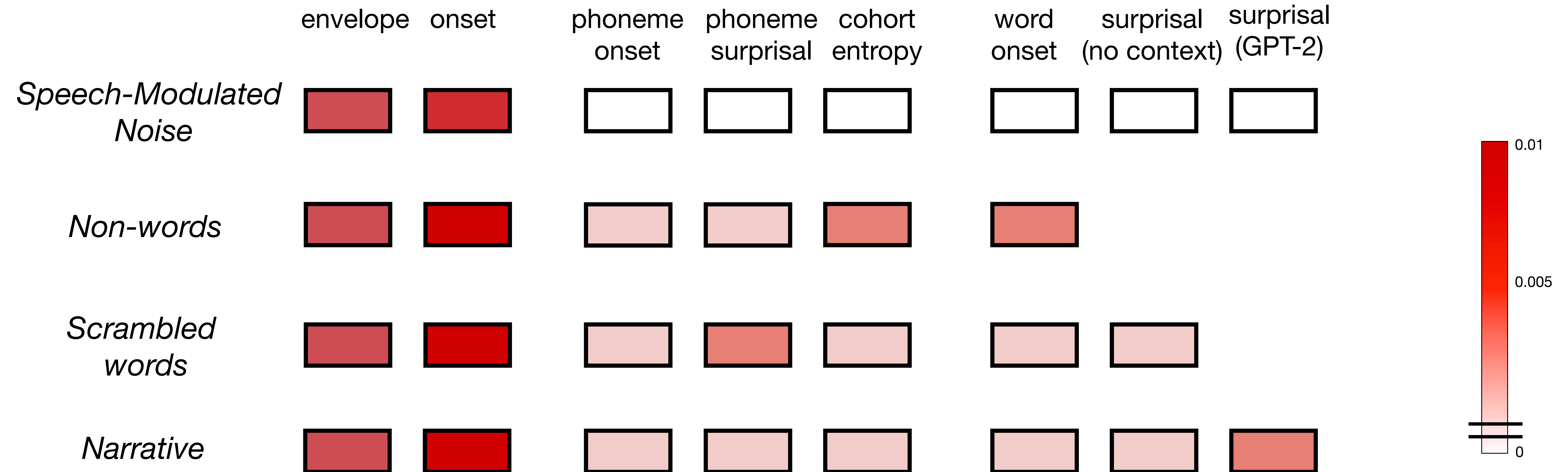


0 1 2 s



# Neural Prediction Results

## Emergence of neural features as the incremental processing occur

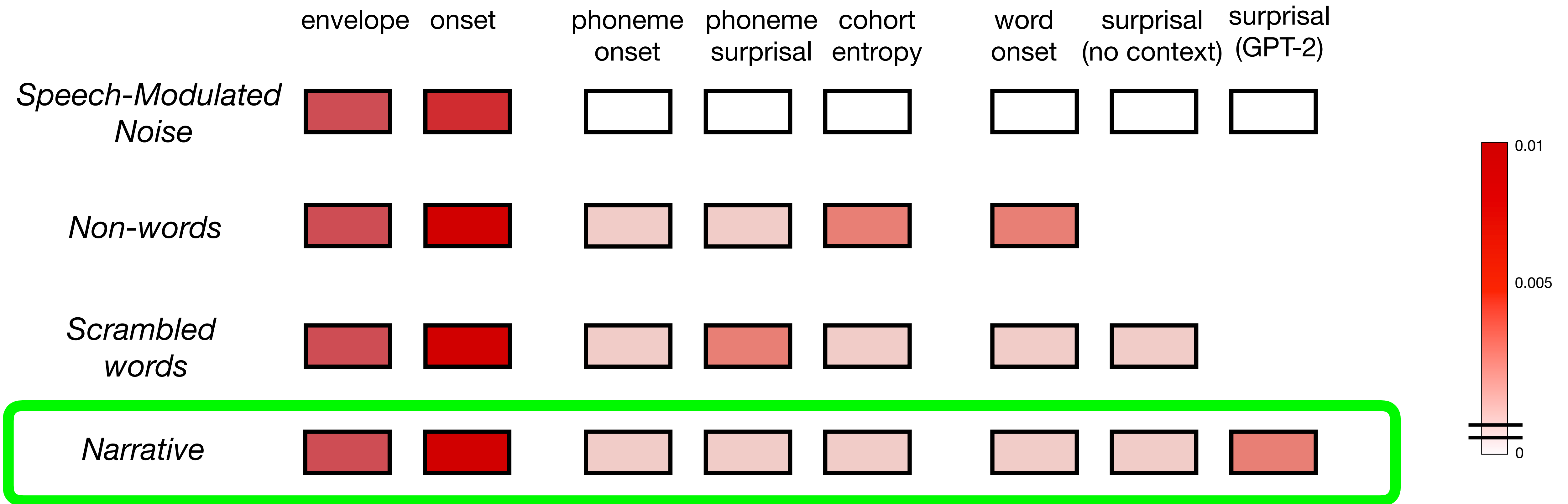


- Acoustic features are encoded for both non-speech and speech stimuli
- (Sub)-lexical features are encoded only when (sub)-lexical boundaries are intelligible
- Context based word surprisal emerges for narrative passage
- When context supports, context based surprisal is better tracked compared to naive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



# Neural Prediction Results

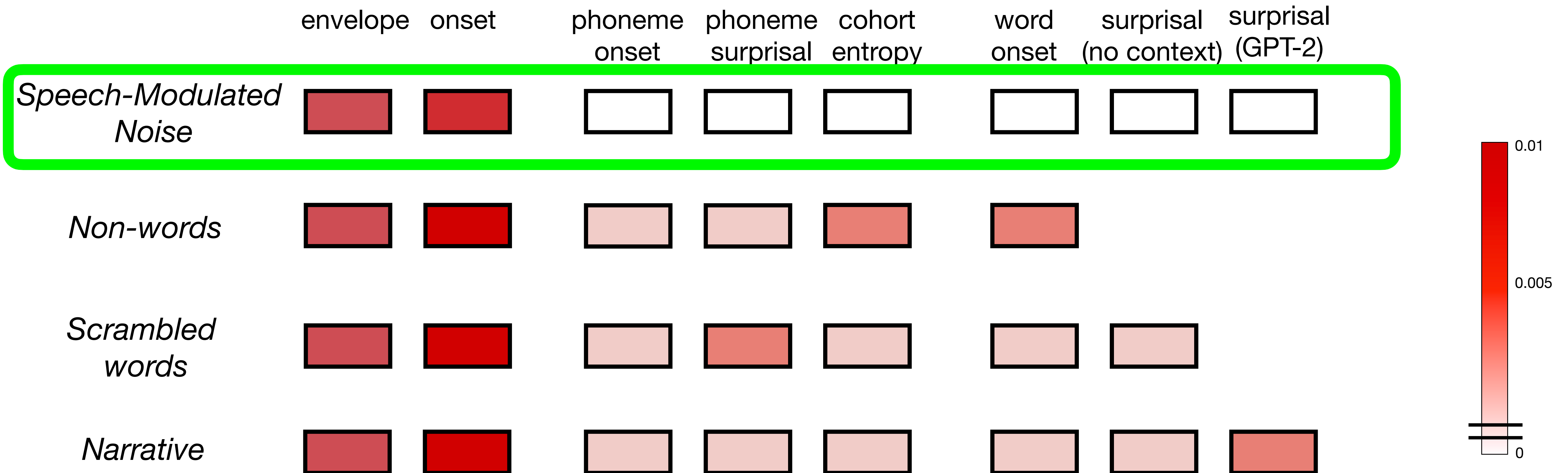
## Emergence of neural features as the incremental processing occur



- 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

# Neural Prediction Results

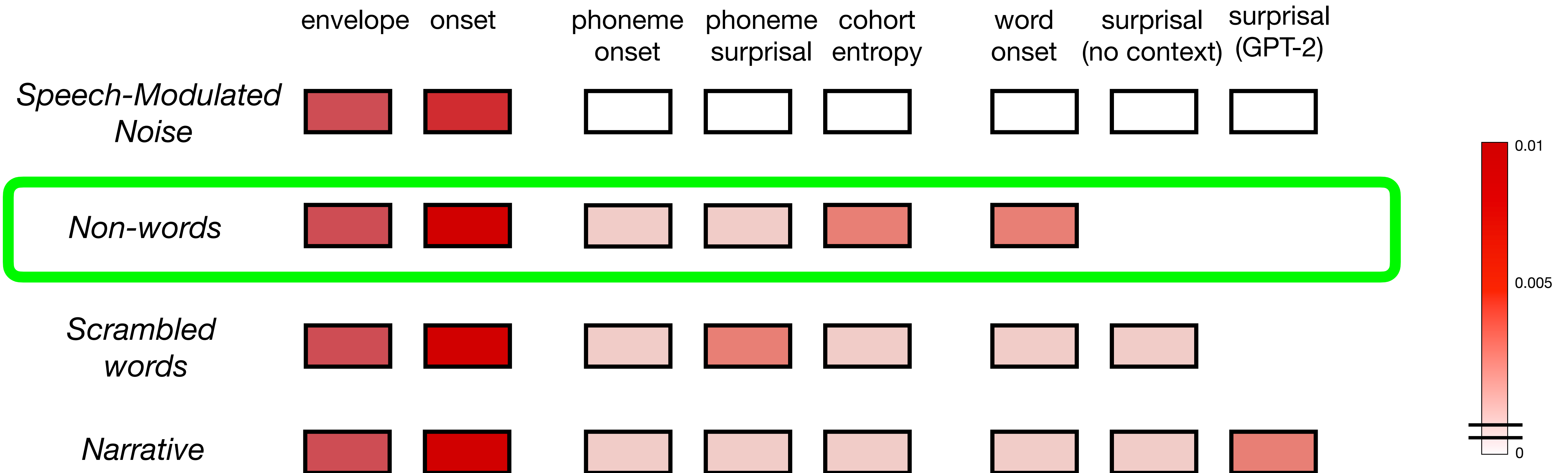
## Emergence of neural features as the incremental processing occur



- 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

# Neural Prediction Results

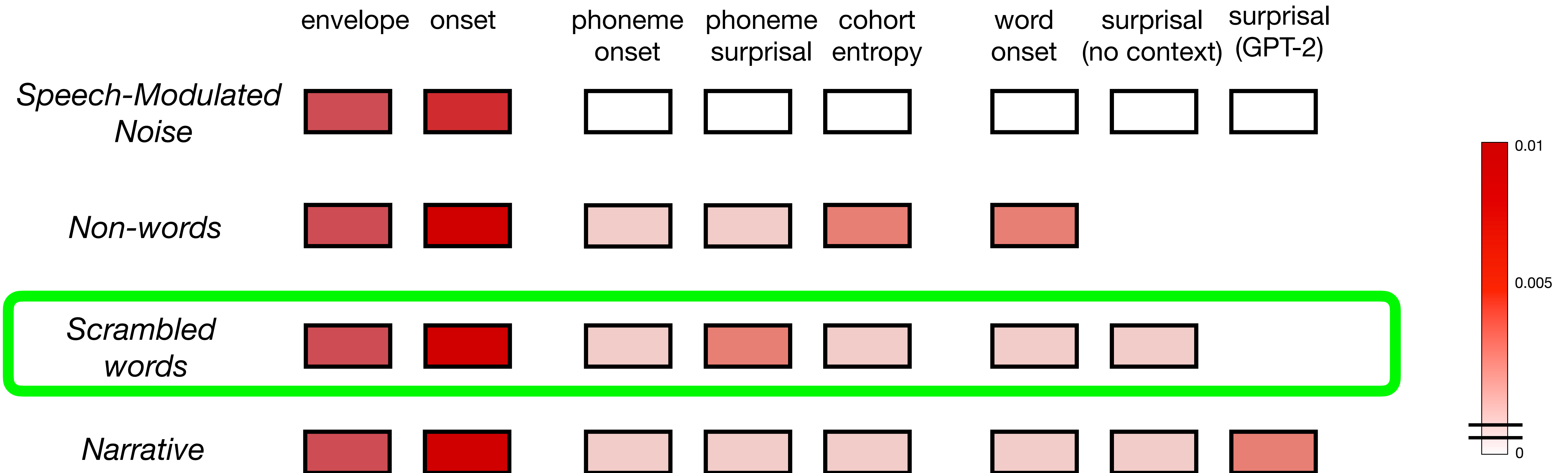
## Emergence of neural features as the incremental processing occur



- 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

# Neural Prediction Results

## Emergence of neural features as the incremental processing occur

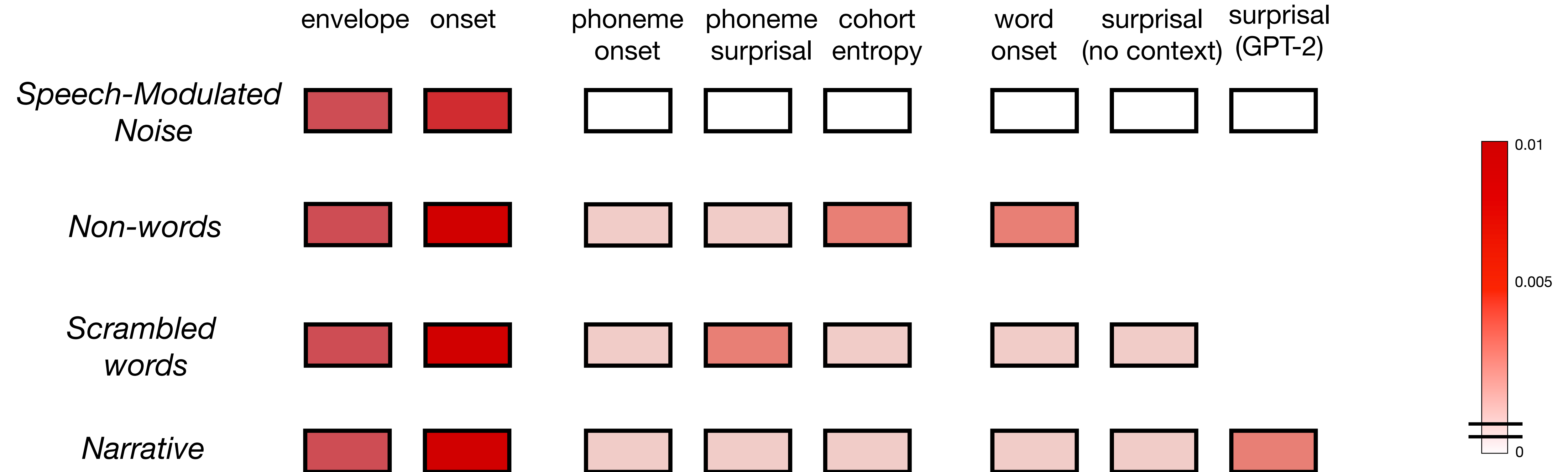


- Acoustic features are encoded for both non-speech and speech stimuli
- (Sub)-lexical features are encoded only when (sub)-lexical boundaries are intelligible
- Context based word surprisal emerges for narrative passage
- When context supports, context based surprisal is better tracked compared to naive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



# 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

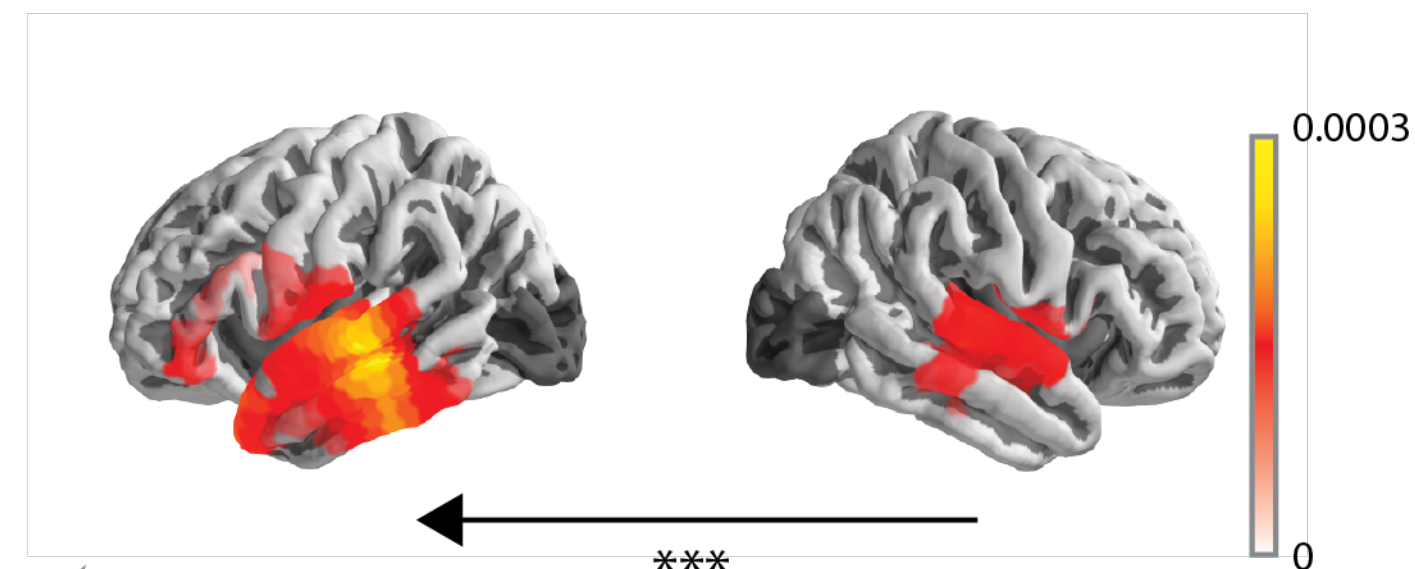
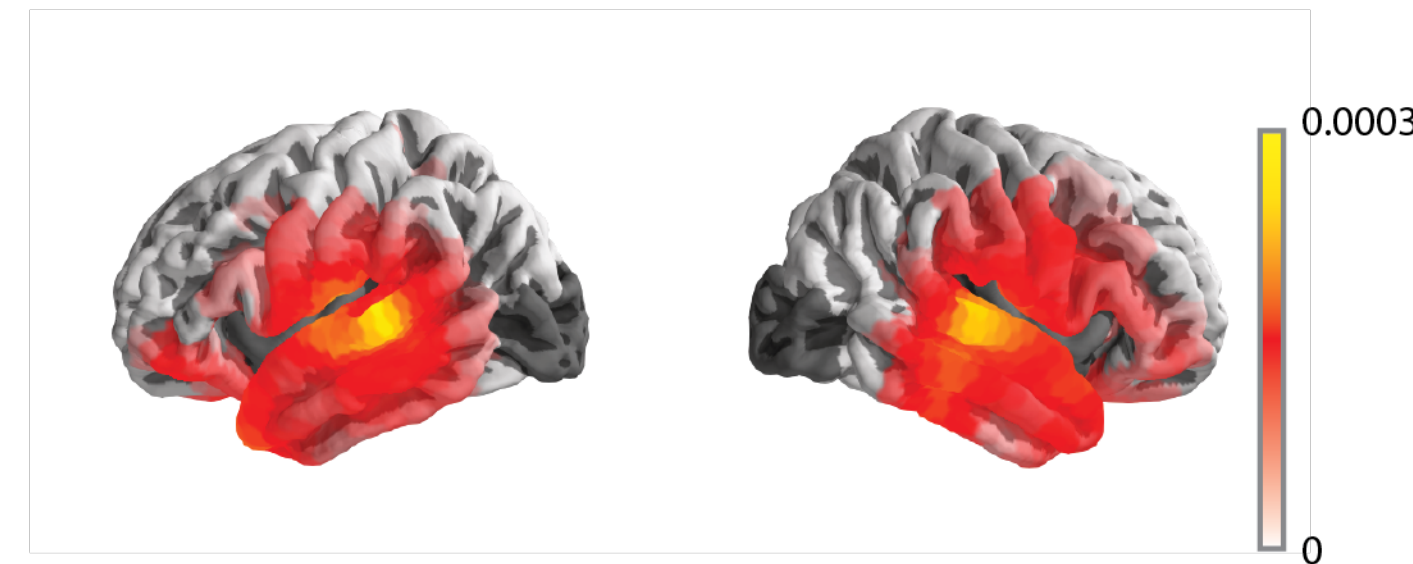
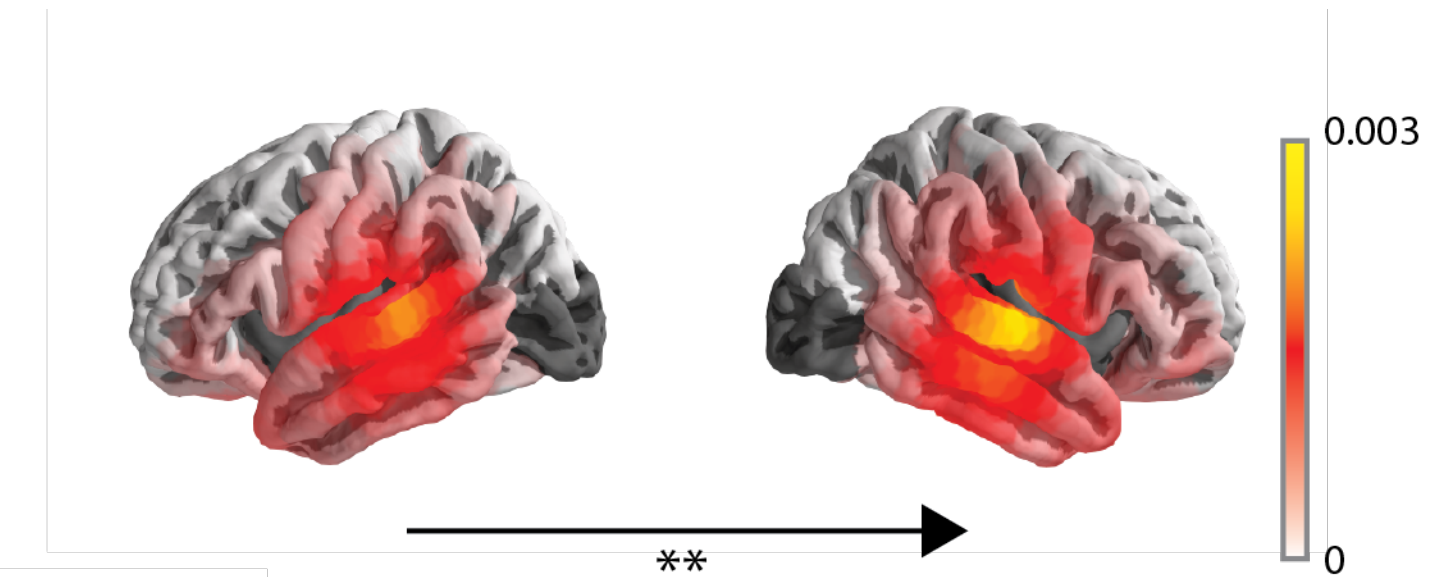
**Phoneme  
Surprisal**

Cohort Entropy

Word Onset

Unigram  
Surprisal

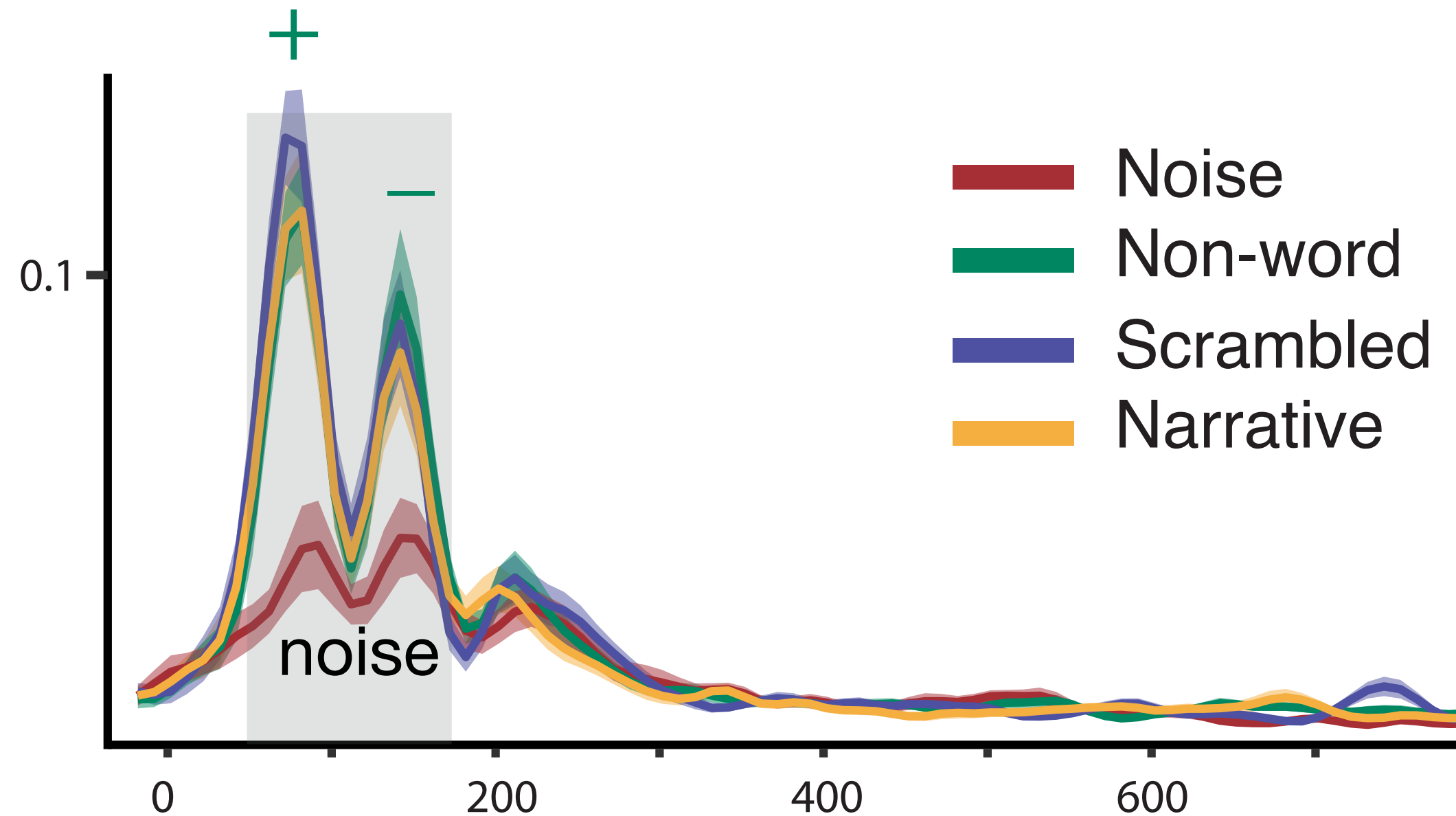
**GPT2  
Surprisal**



Note: lateralization results can be task dependent

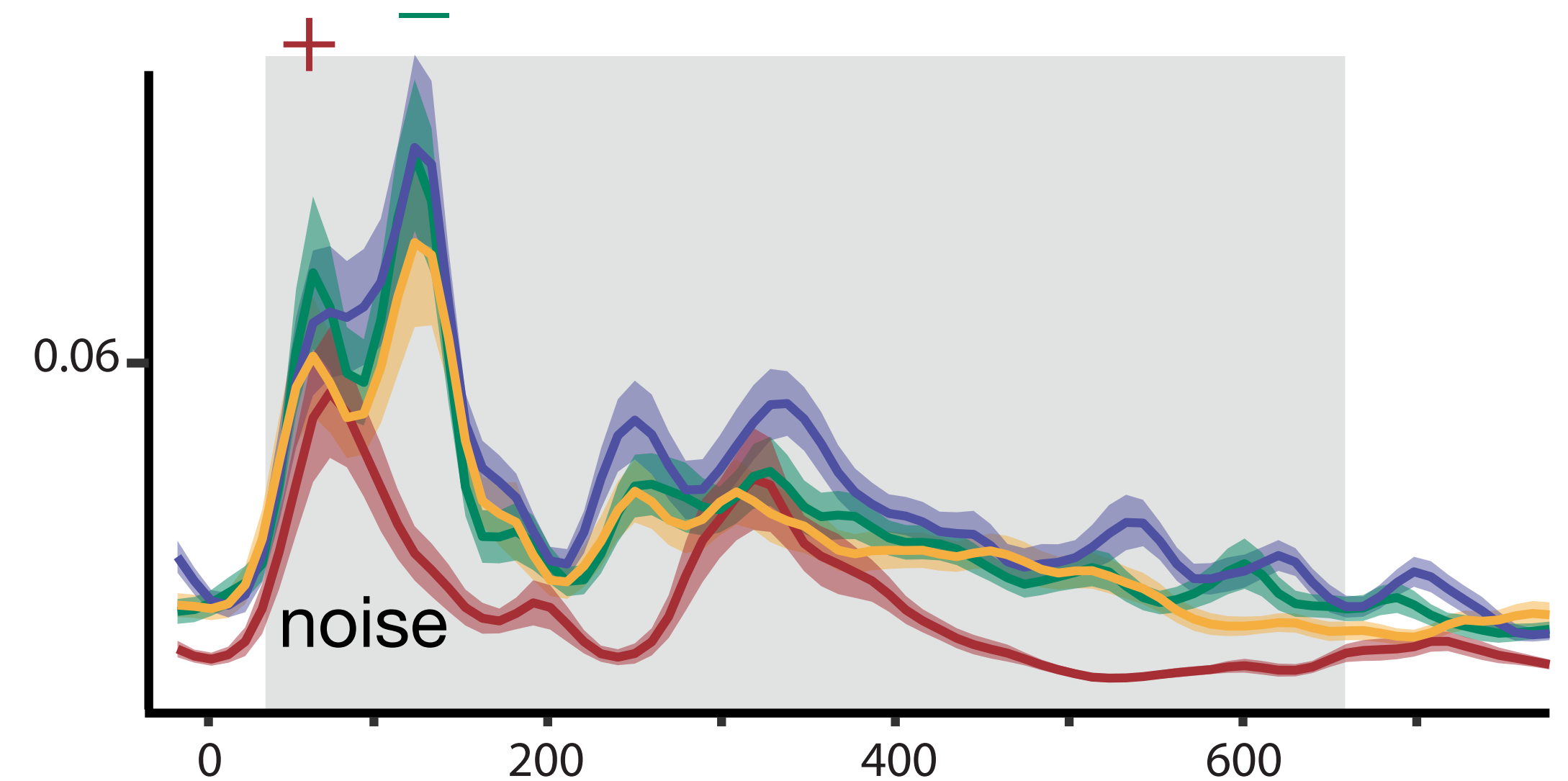
# Acoustic TRF Results

acoustic onsets



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

acoustic envelope

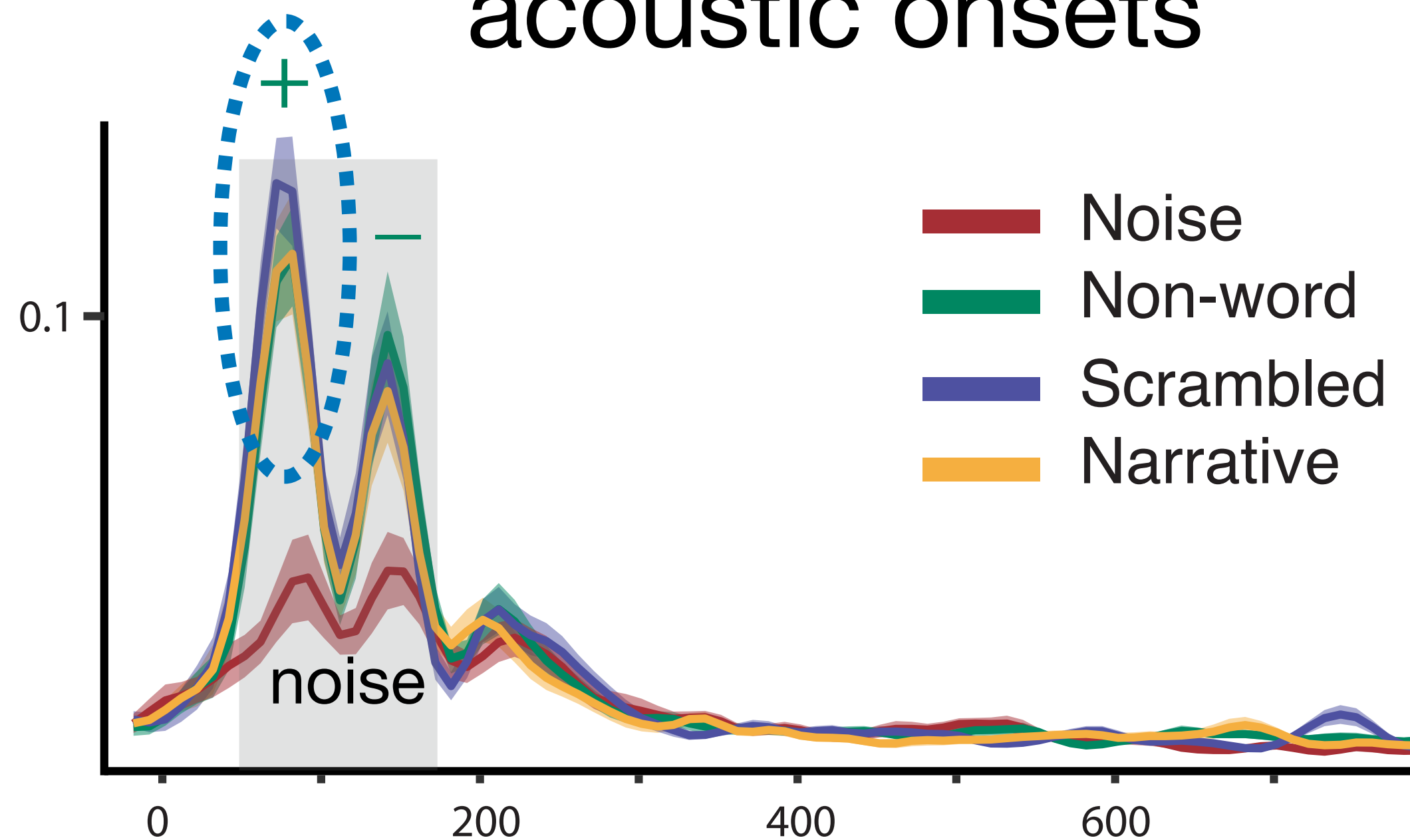


- Speech responses > Noise response (Narrative < Scrambled)
- Non words similar to Scrambled words
- Noise response lacks 2nd peak ~120 ms

right hemisphere shown  
condition based differences similar in left

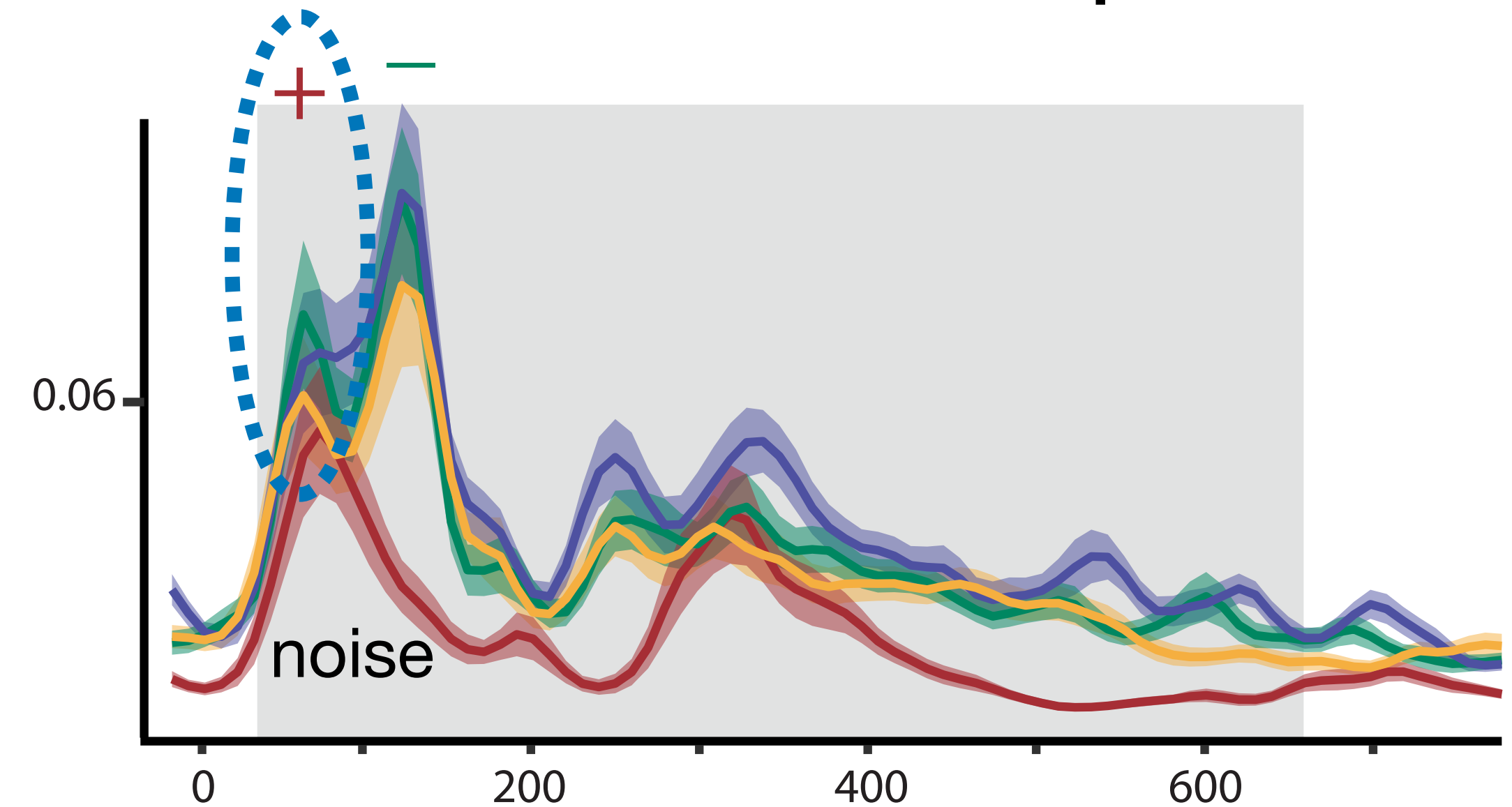
# Acoustic TRF Results

acoustic onsets



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

acoustic envelope



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

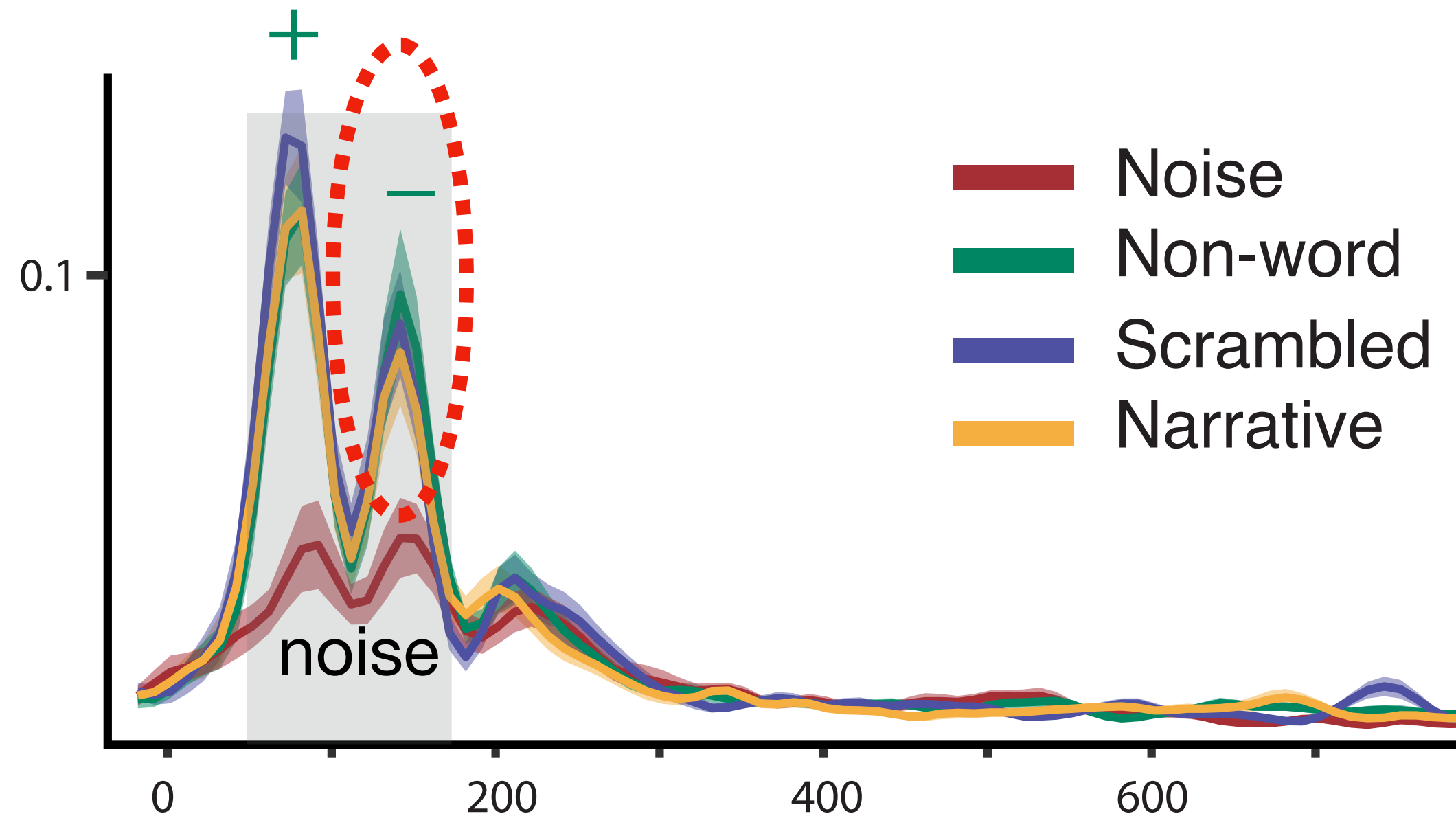
60 ms: acoustic bottom-up processing

right hemisphere shown  
condition based differences similar in left



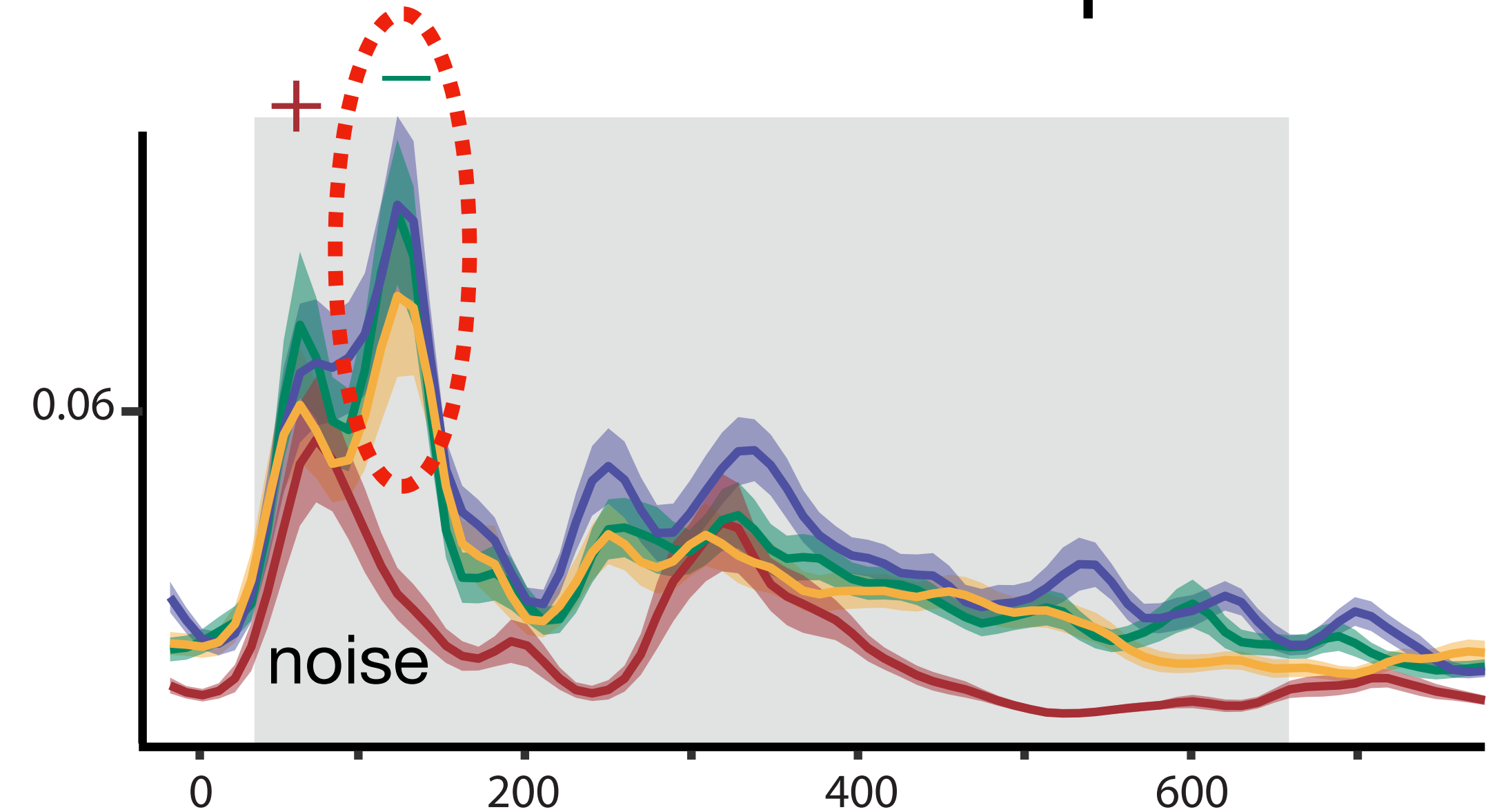
# Acoustic TRF Results

acoustic onsets



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

acoustic envelope



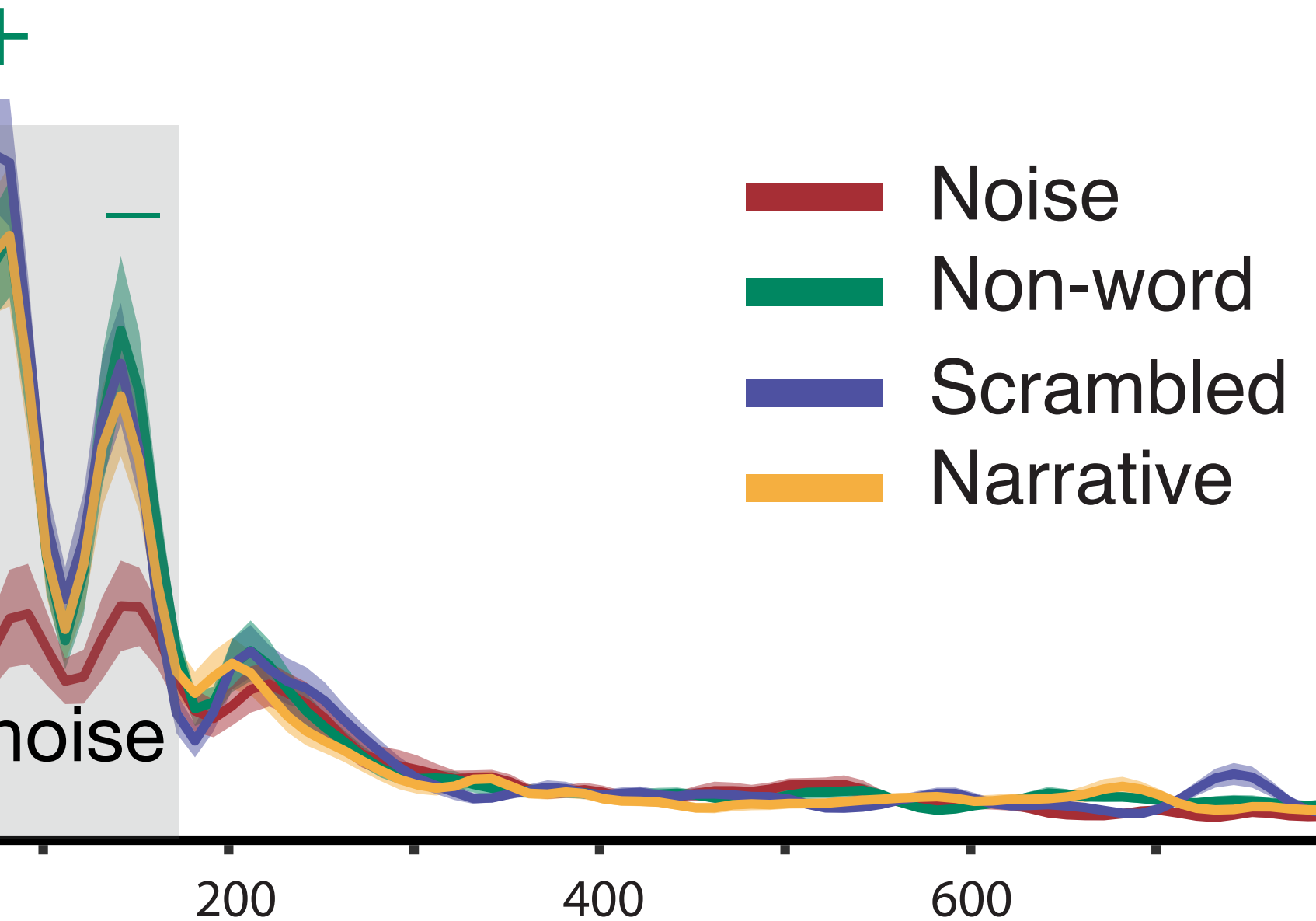
- Speech responses > Noise response (Narrative < Scrambled)
- Non words similar to Scrambled words
- Noise response lacks 2nd peak ~120 ms

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

right hemisphere shown  
condition based differences similar in left

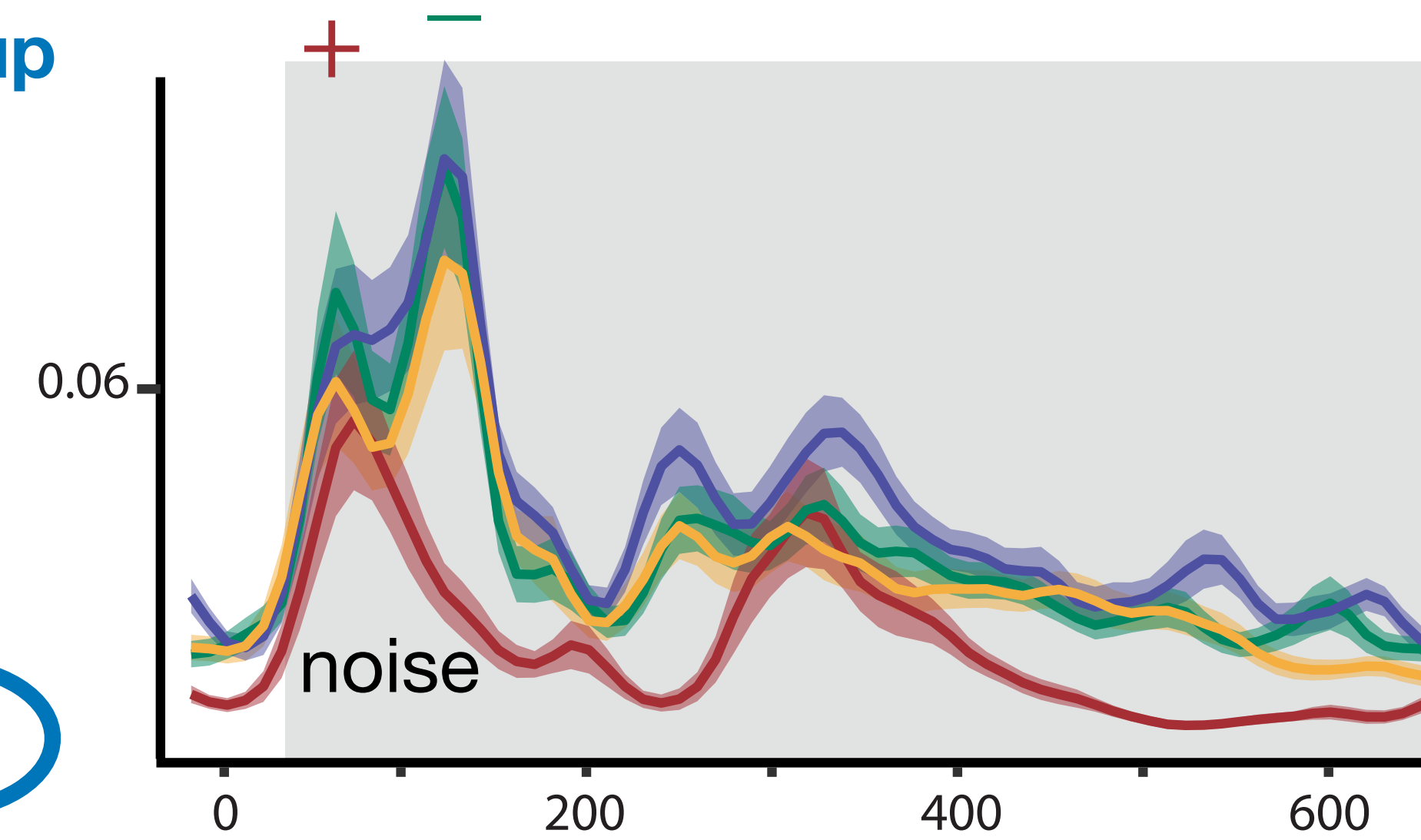
# Acoustic TRF Results

acoustic onsets



h responses > Noise response  
speech roughly equal)

acoustic envelope



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

right hemisphere shown  
condition based differences similar in left

Top-down

Bottom-up

High level meaning

Acoustic

Speech  
Stimuli

120

0

time (ms)

60

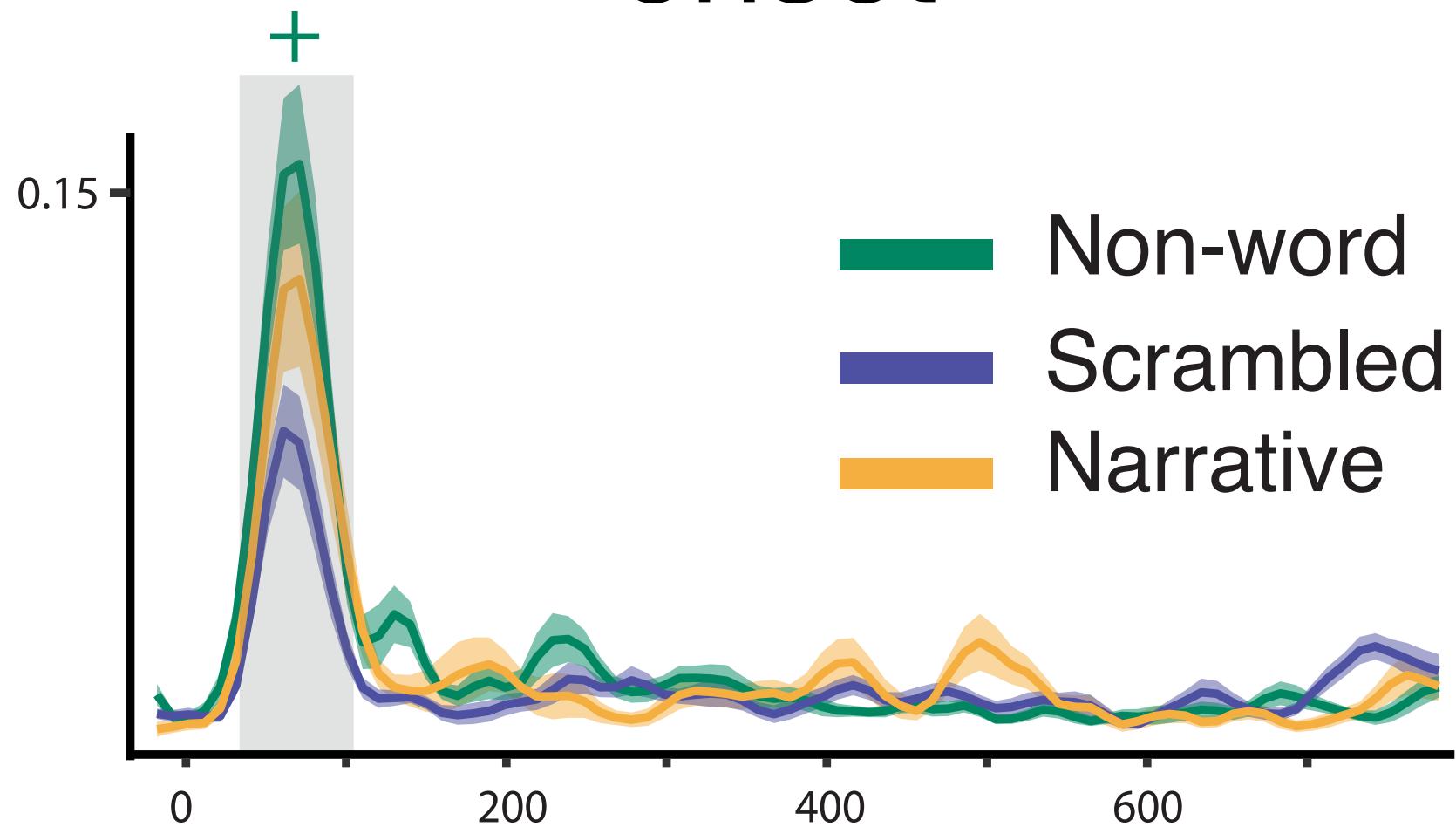
0

time (ms)

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

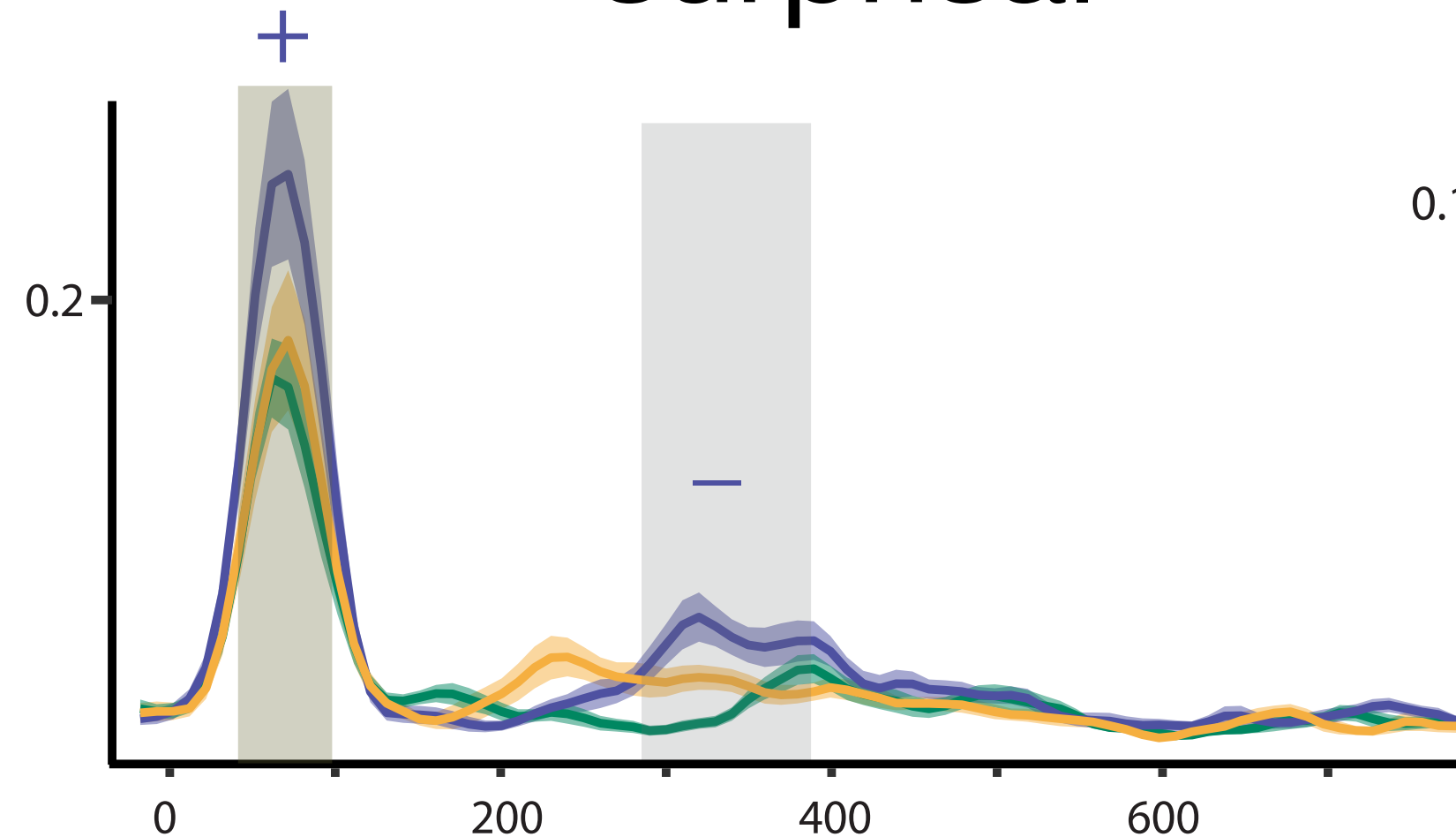
# Phonemic TRF Results

phoneme  
onset



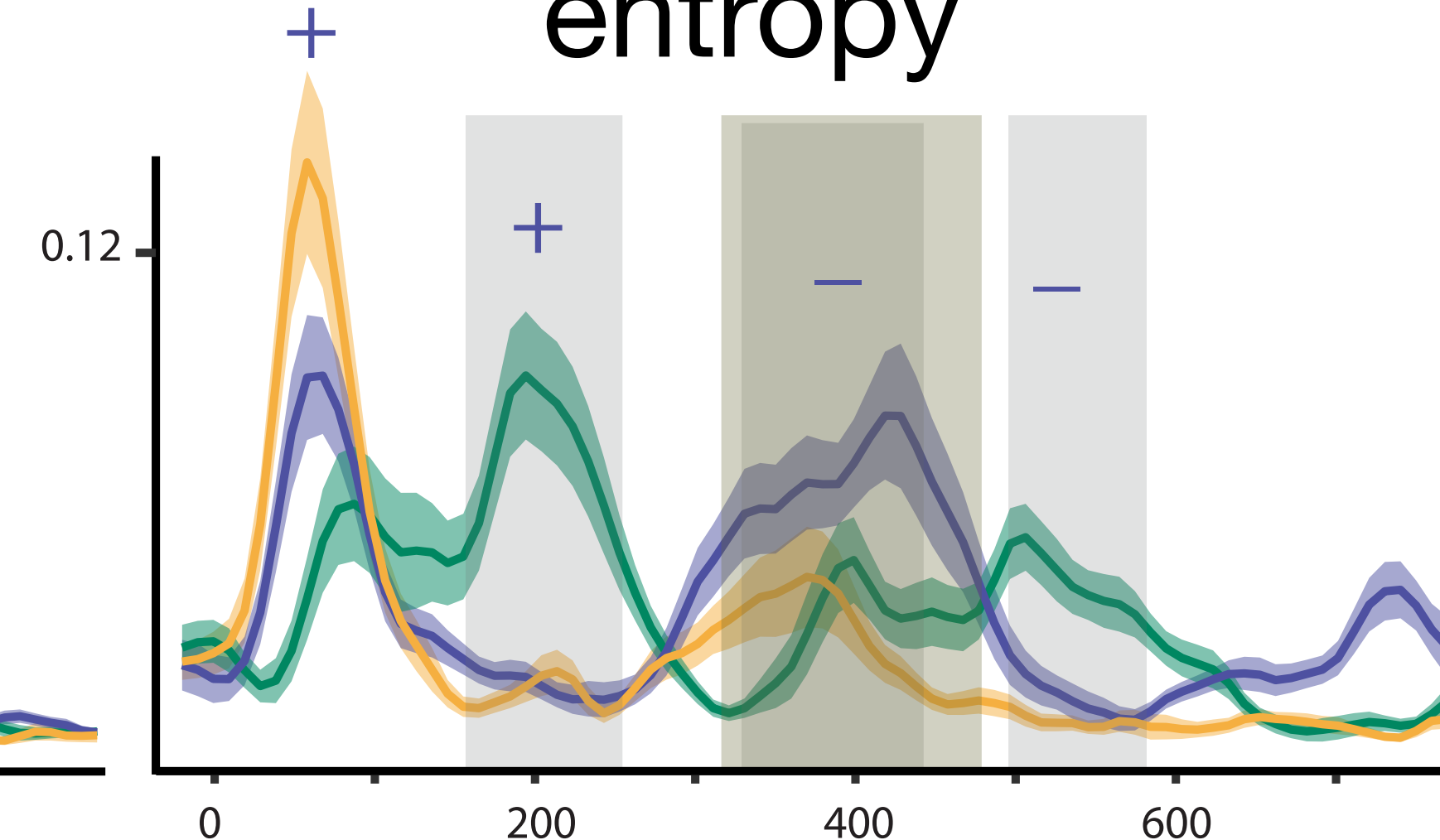
- Non-words largest
- No later processing

phoneme  
surprisal



- Early phone processing ~80 ms (scrambled > narrative)
- Late phone processing ~350 ms (words > non-words)

cohort  
entropy

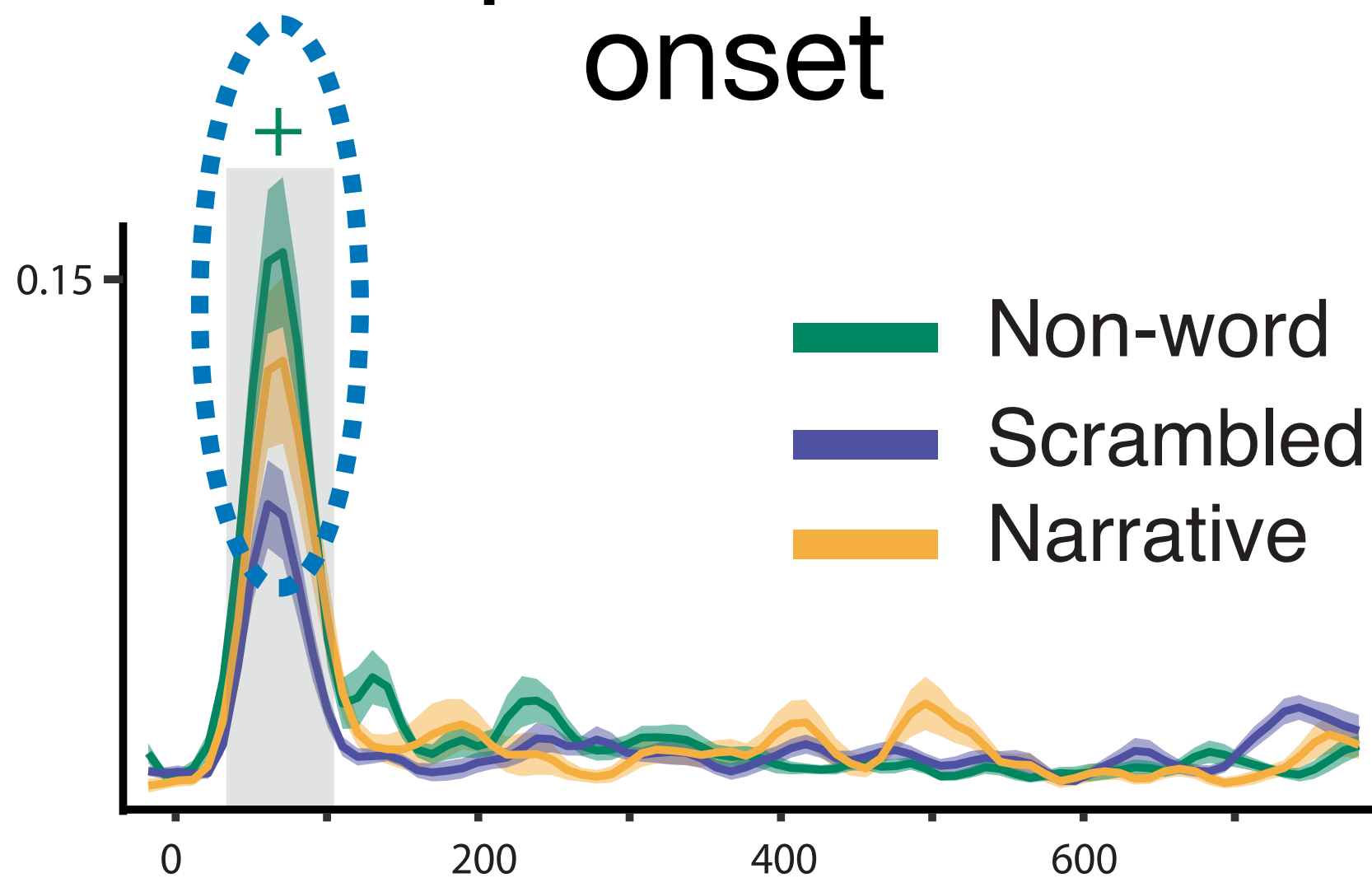


- Late context processing
- N400-like response (reduced for narrative)
- Additional/delayed peaks in non-words (difference in stimulus distributions)

left hemisphere shown (right similar)

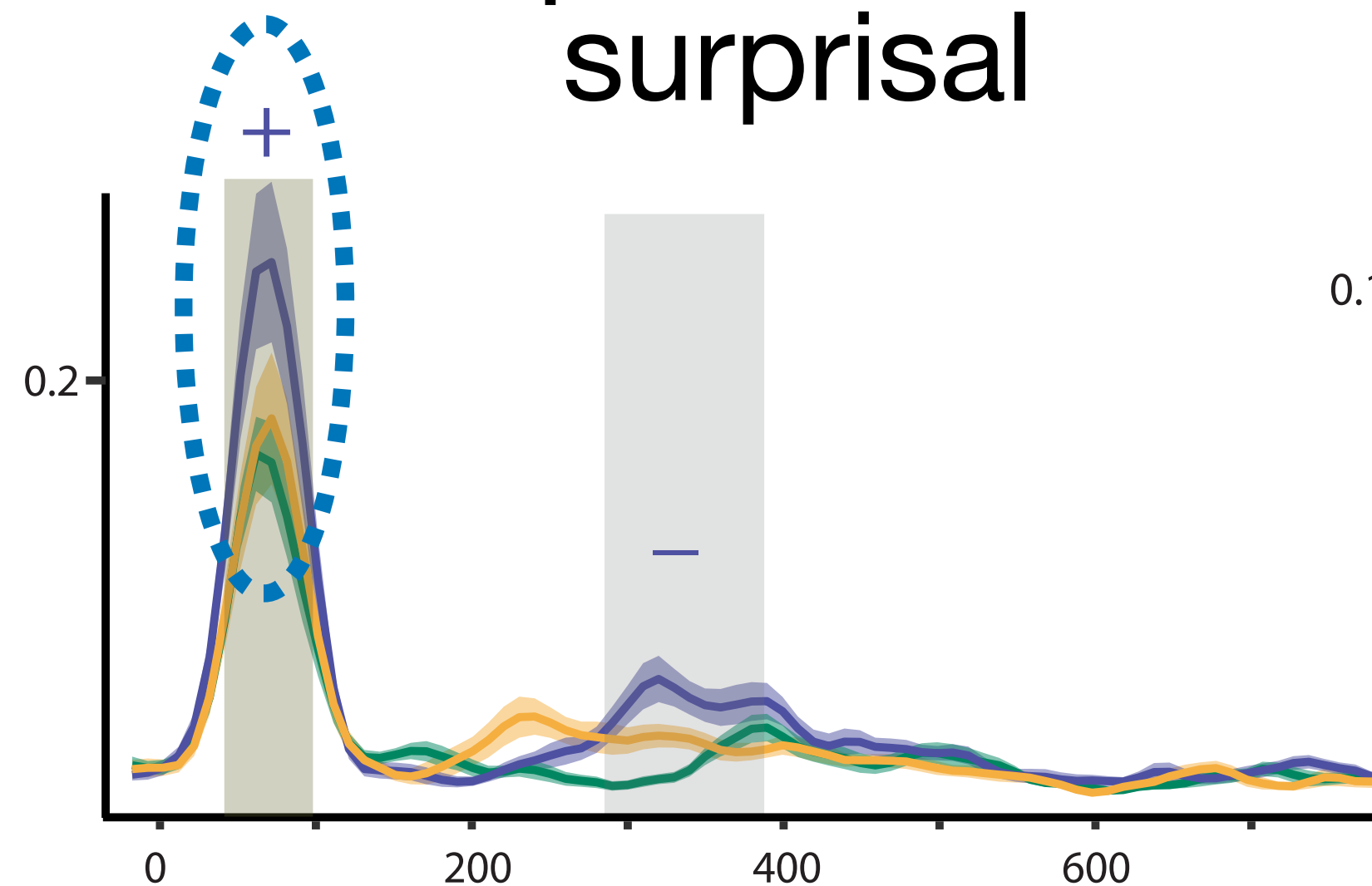
# Phonemic TRF Results

phoneme  
onset



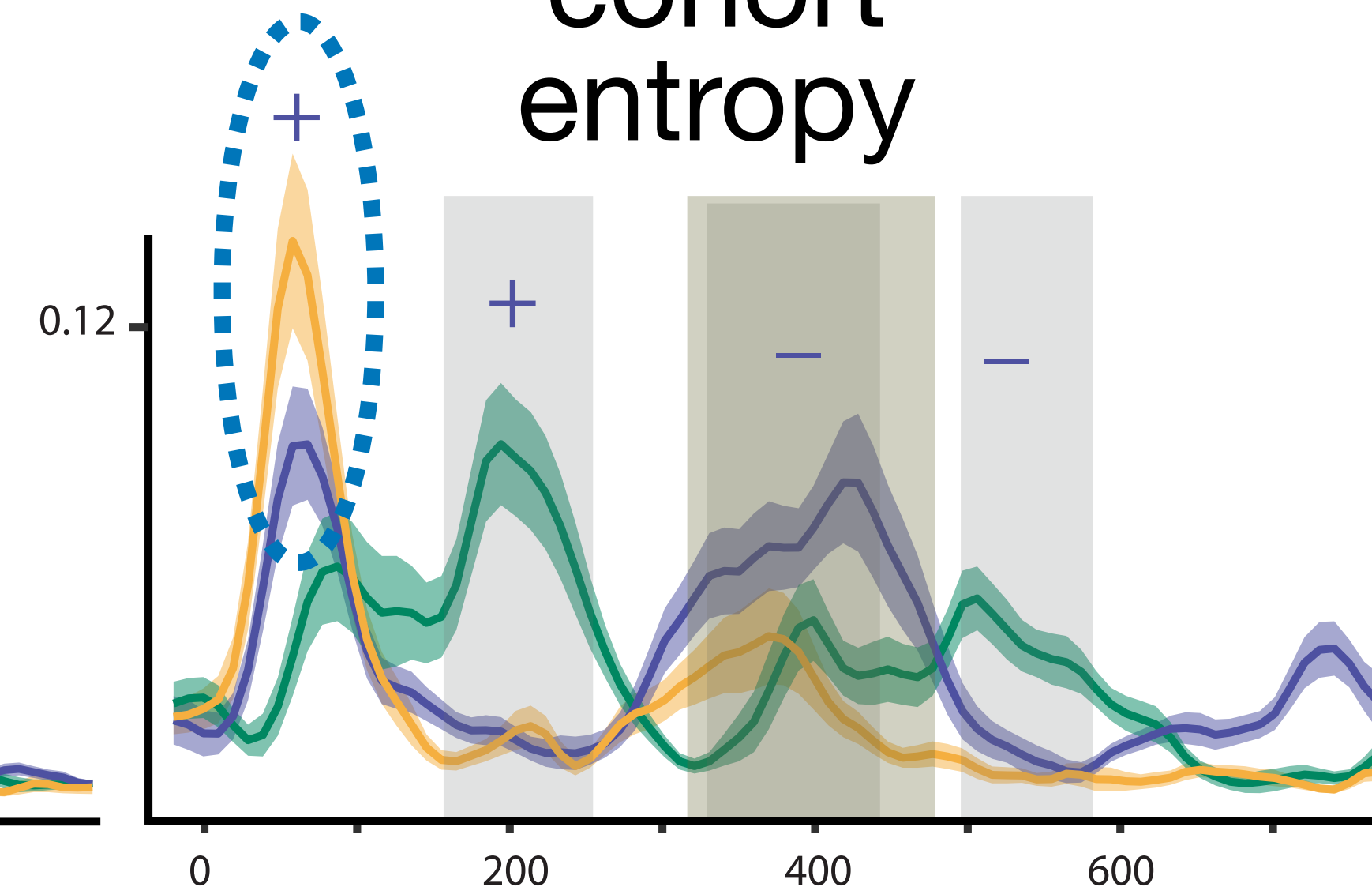
- Non-words largest
- No later processing

phoneme  
surprisal



- Early phone processing ~80 ms (scrambled > narrative)
- Late phone processing ~350 ms (words > non-words)

cohort  
entropy



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

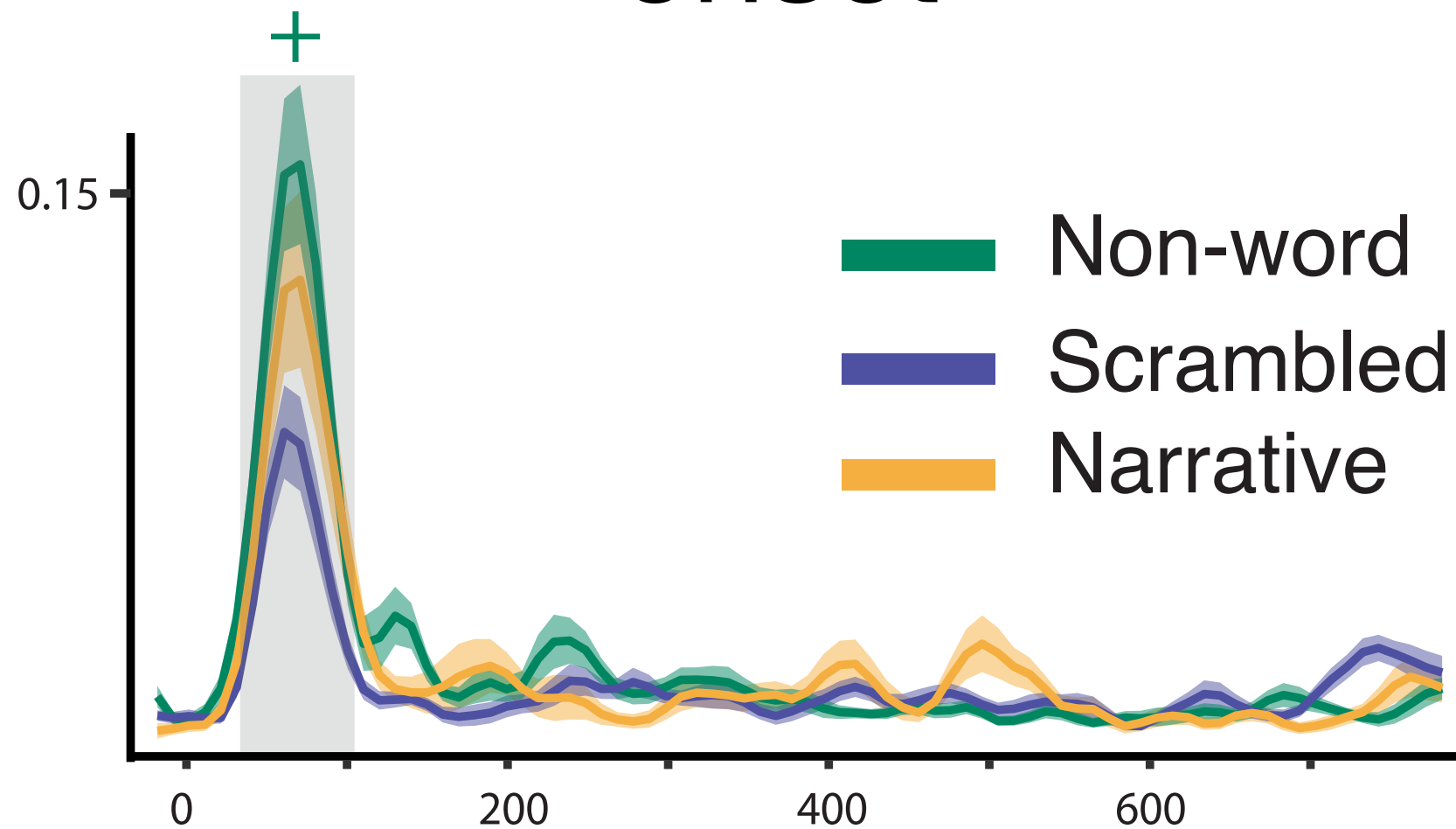
80 ms: simple phoneme processing

left hemisphere shown (right similar)



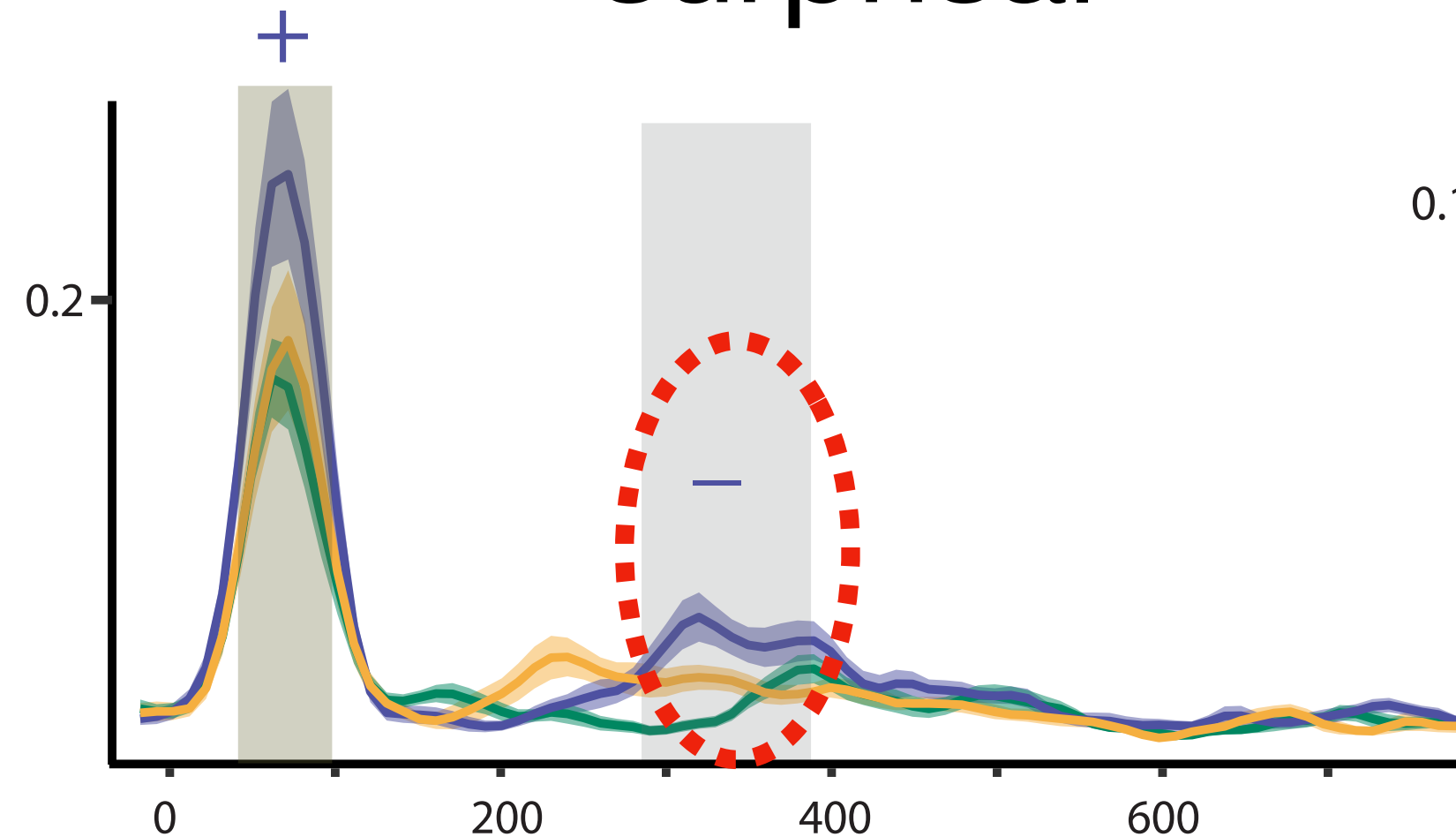
# Phonemic TRF Results

phoneme  
onset



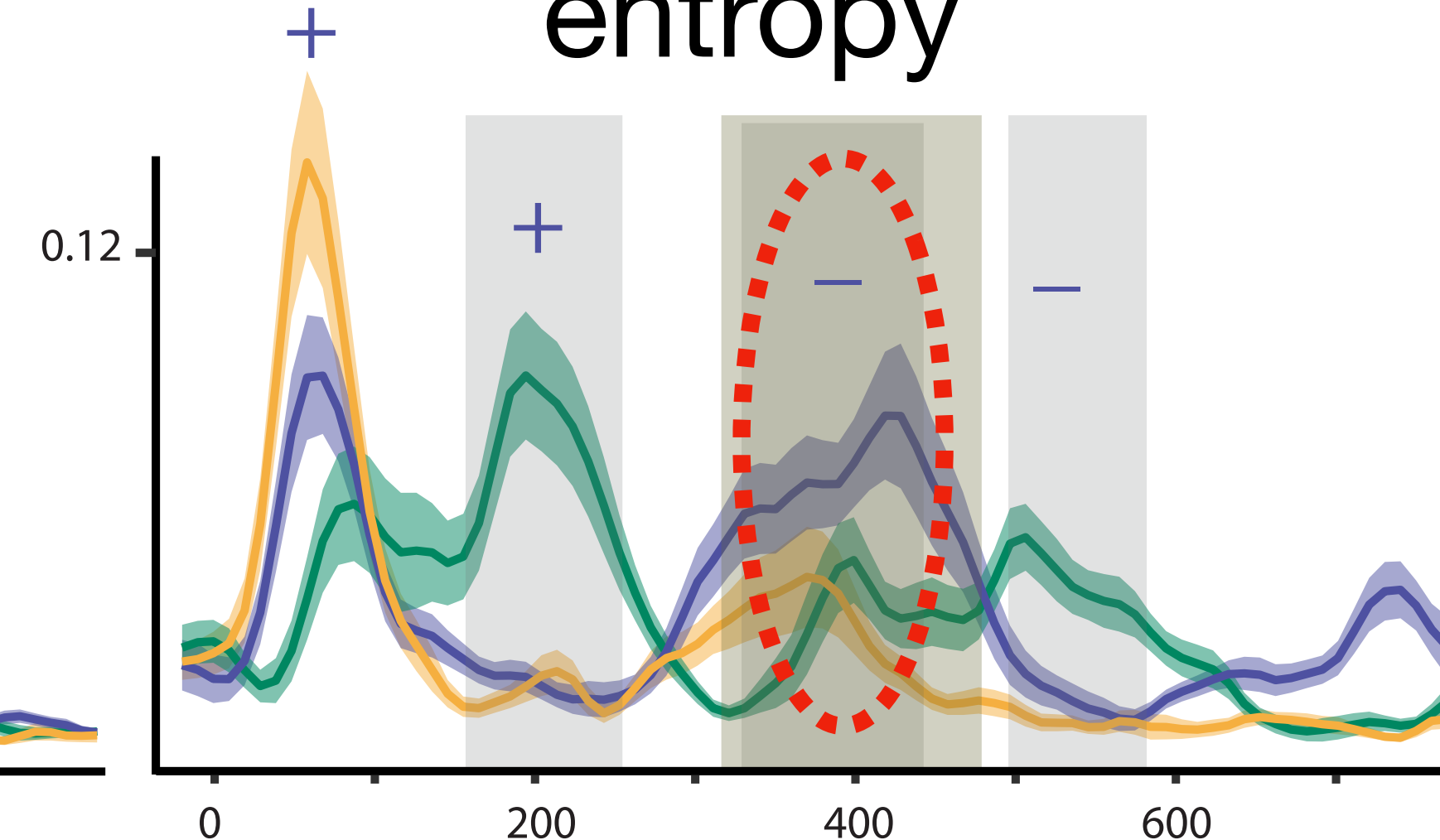
- Non-words largest
- No later processing

phoneme  
surprisal



- Early phone processing ~80 ms (scrambled > narrative)
- Late phone processing ~350 ms (words > non-words)

cohort  
entropy

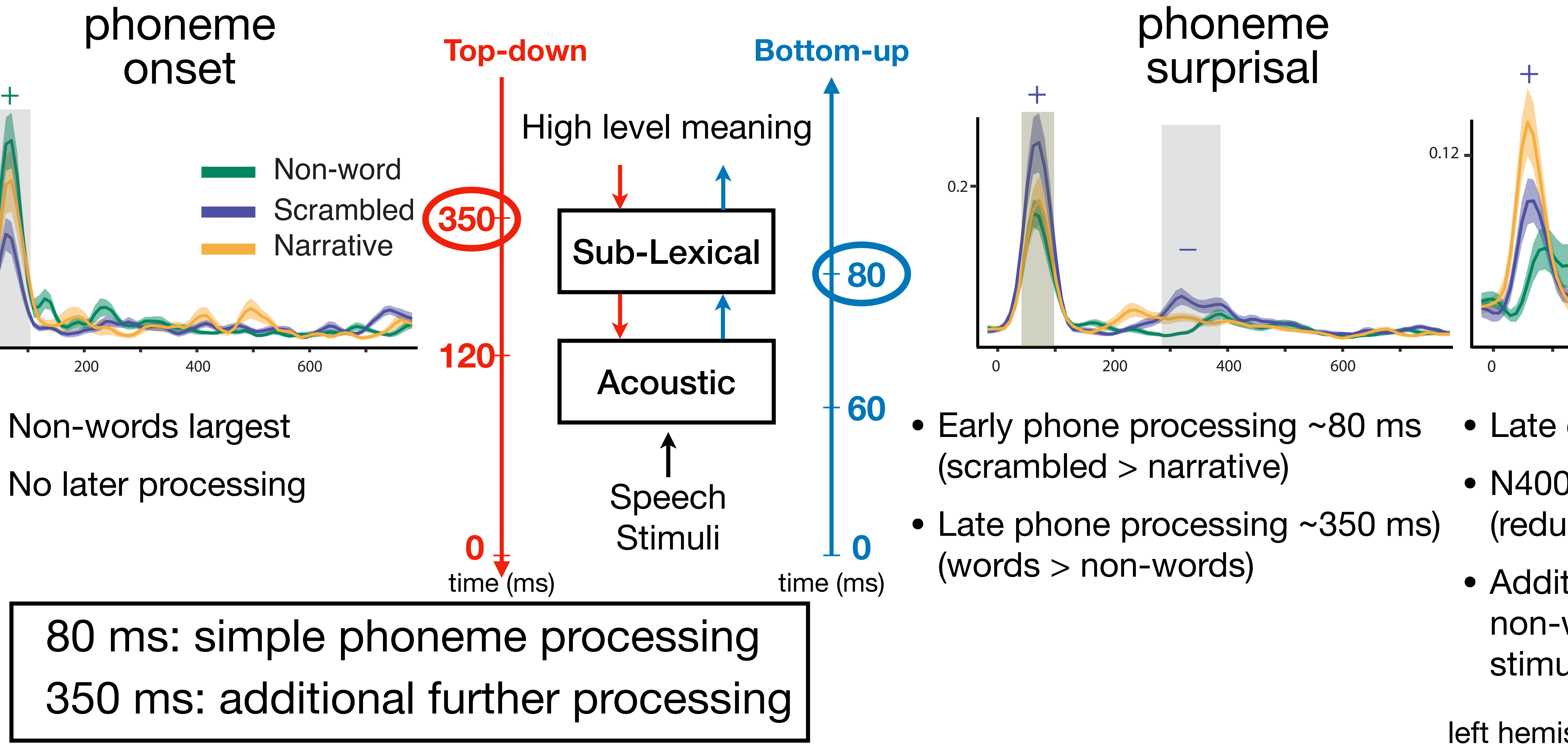


- Late context processing
- N400-like response (reduced for narrative)
- Additional/delayed peaks in non-words (difference in stimulus distributions)

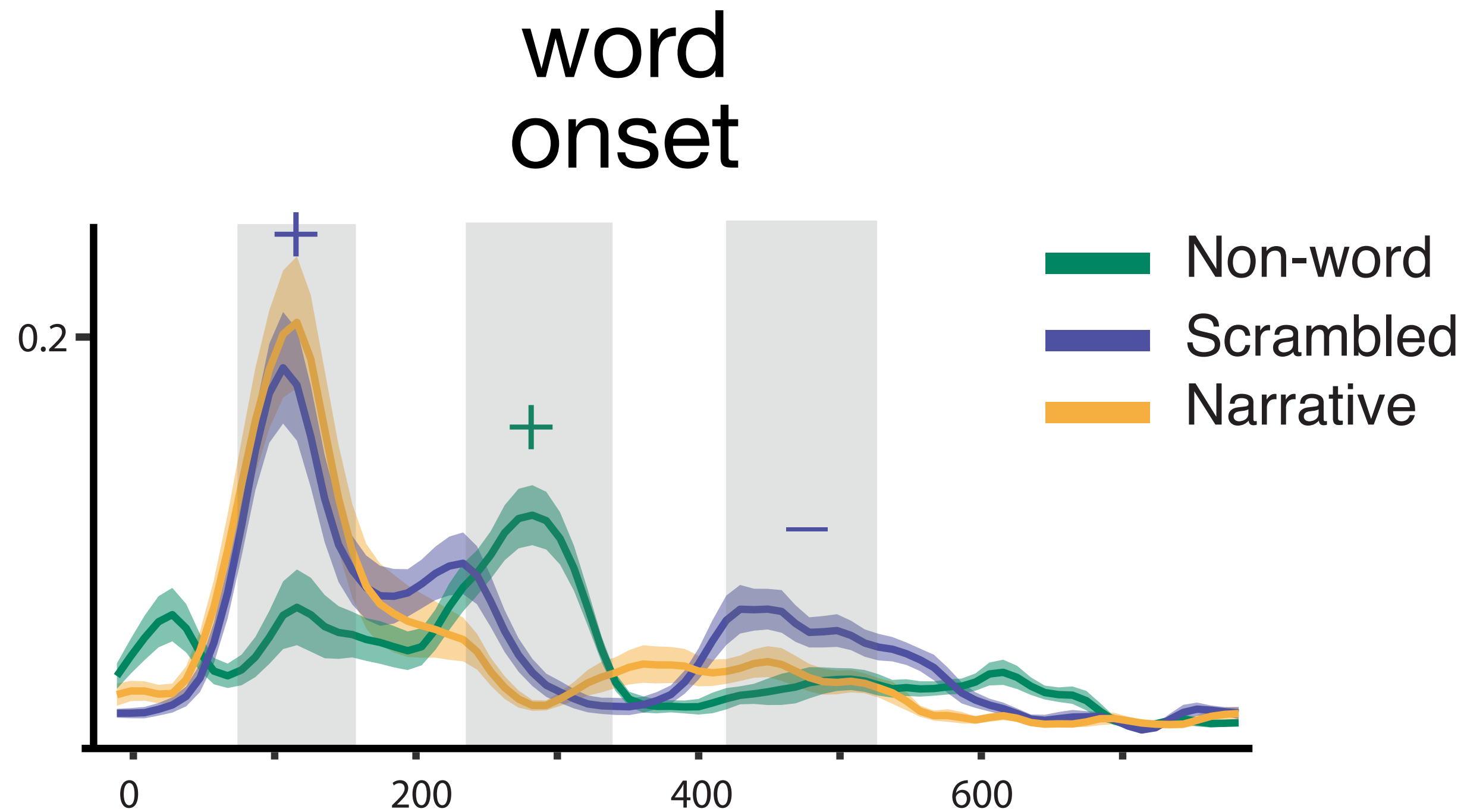
80 ms: simple phoneme processing  
350 ms: additional further processing

left hemisphere shown (right similar)

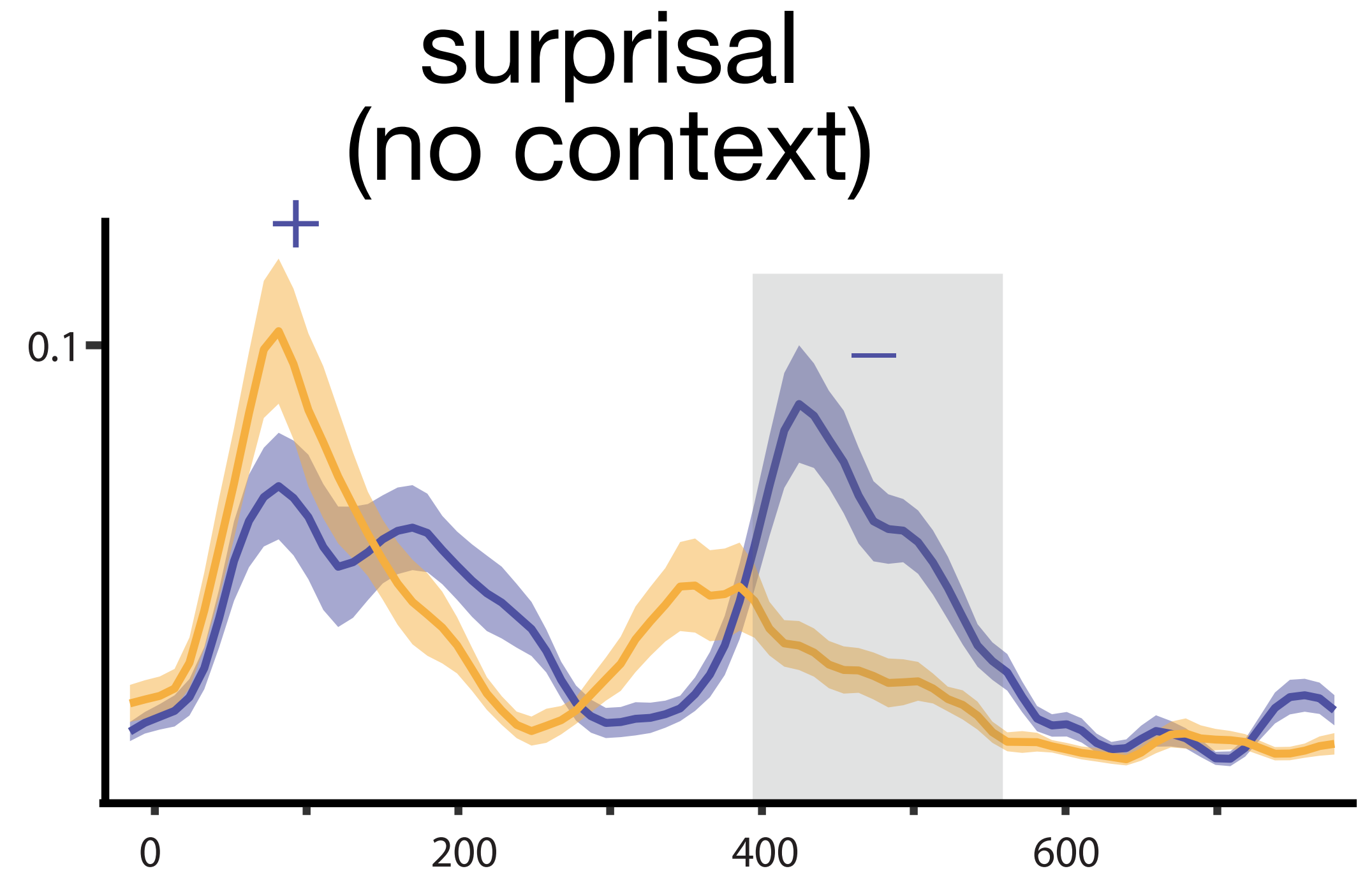
# Phonemic TRF Results



# Word-based TRF Results



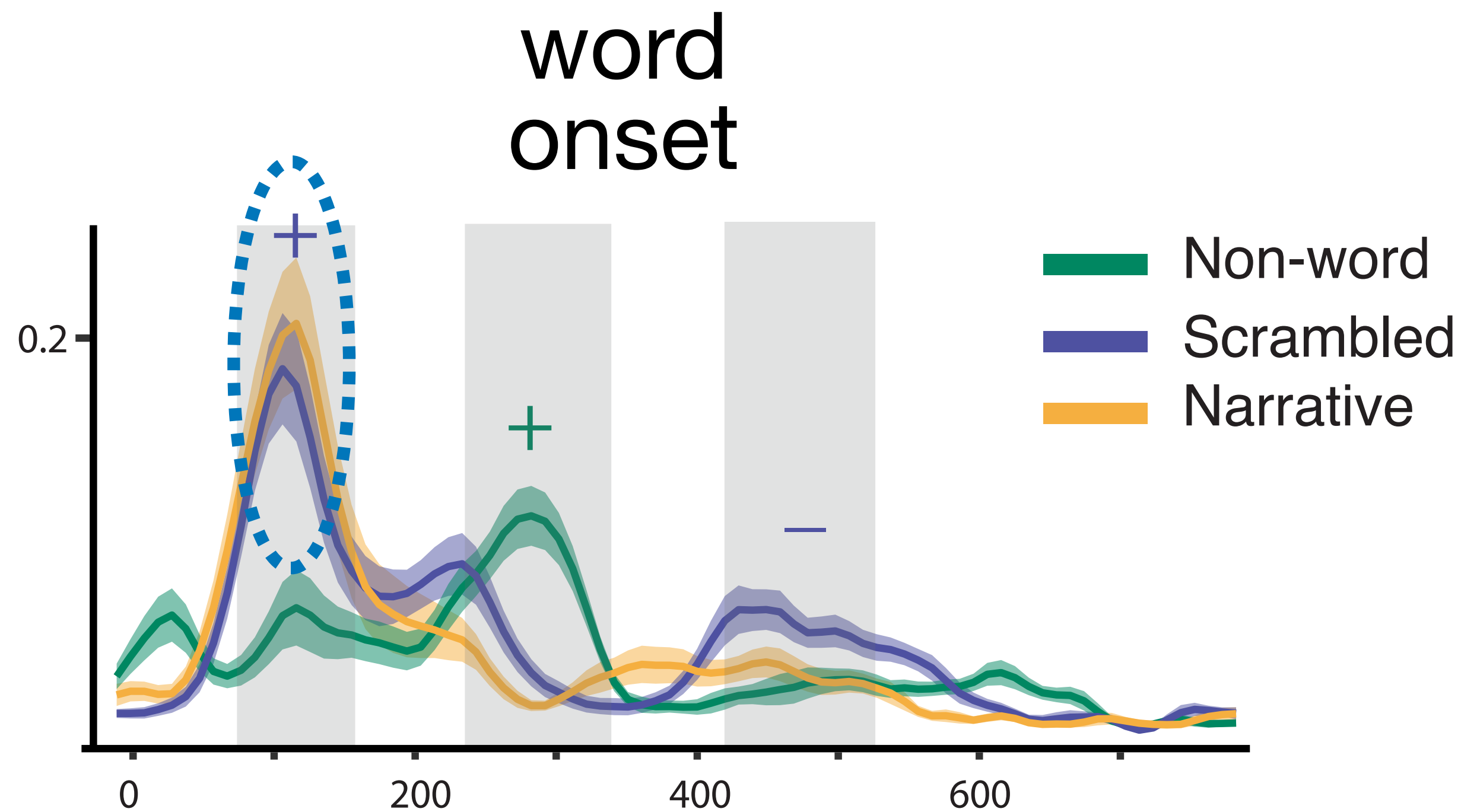
- Scrambled  $\approx$  narrative for rapid processing
- Scrambled words  $>$  narrative at  $\sim 450$  ms
- words: Left hemi  $>$  Right (non-words: L  $\approx$  R)



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

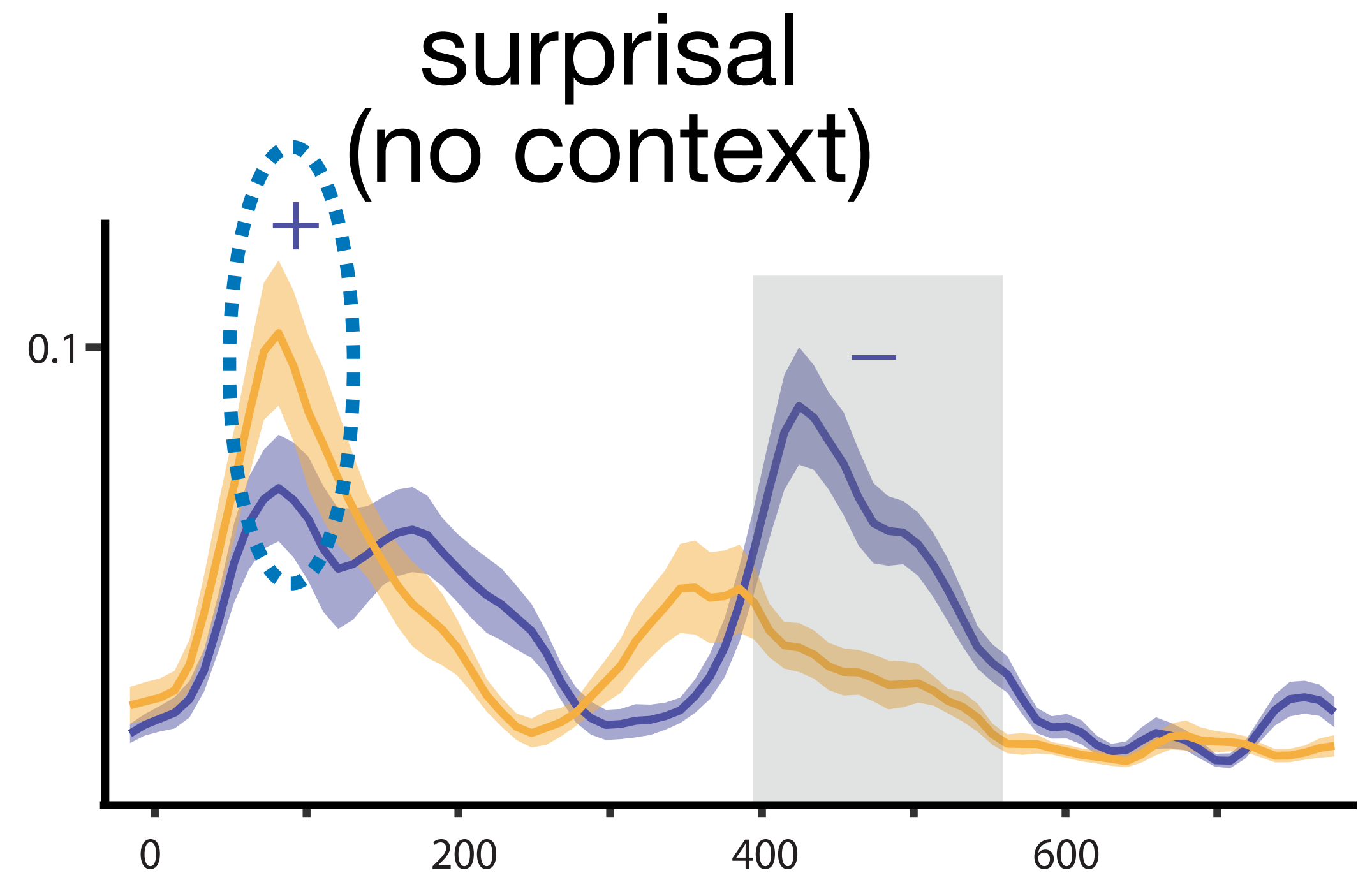
left hemisphere shown  
(right much weaker except for non-word onset)

# Word-based TRF Results



- Scrambled  $\approx$  narrative for rapid processing
- Scrambled words  $>$  narrative at  $\sim 450$  ms
- words: Left hemi  $>$  Right (non-words: L  $\approx$  R)

100 ms: simple word processing



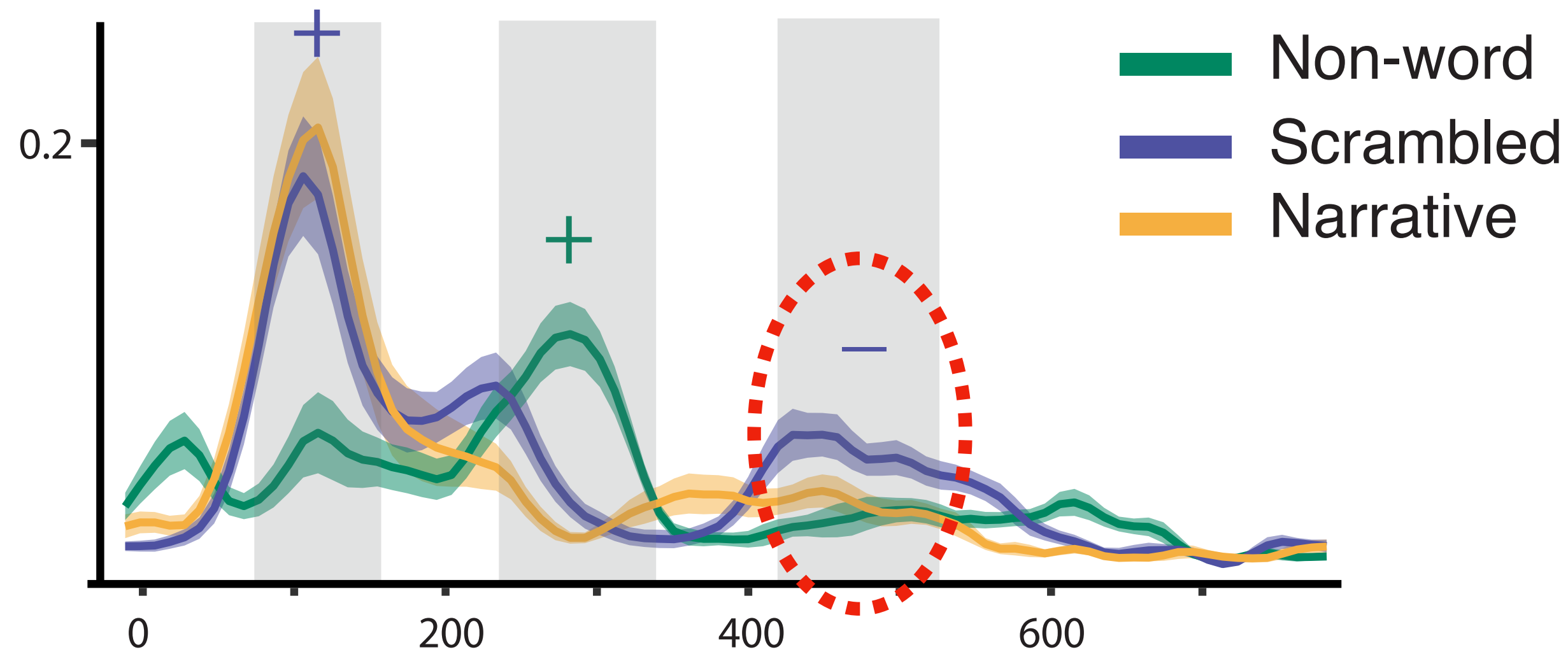
- N400 like response
- Reduction in surprisal when context
- Left hemi  $>$  Right hemi
- Right hemisphere: Scrambled  $\approx$  Narrative

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



# Word-based TRF Results

word  
onset

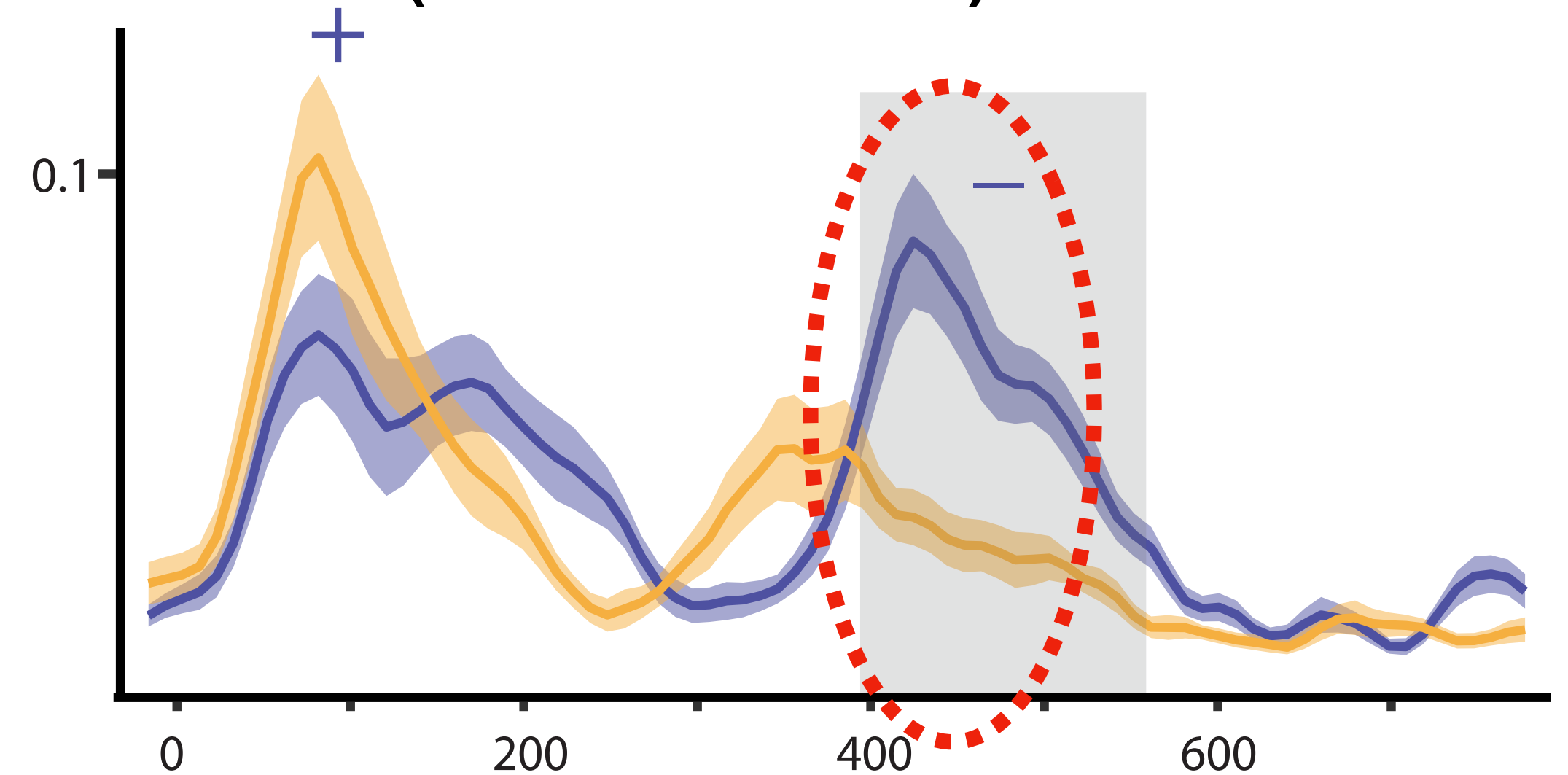


- Scrambled  $\approx$  narrative for rapid processing
- Scrambled words  $>$  narrative at  $\sim 450$  ms
- words: Left hemi  $>$  Right (non-words: L  $\approx$  R)

100 ms: simple word processing

450 ms: “error” correction processing

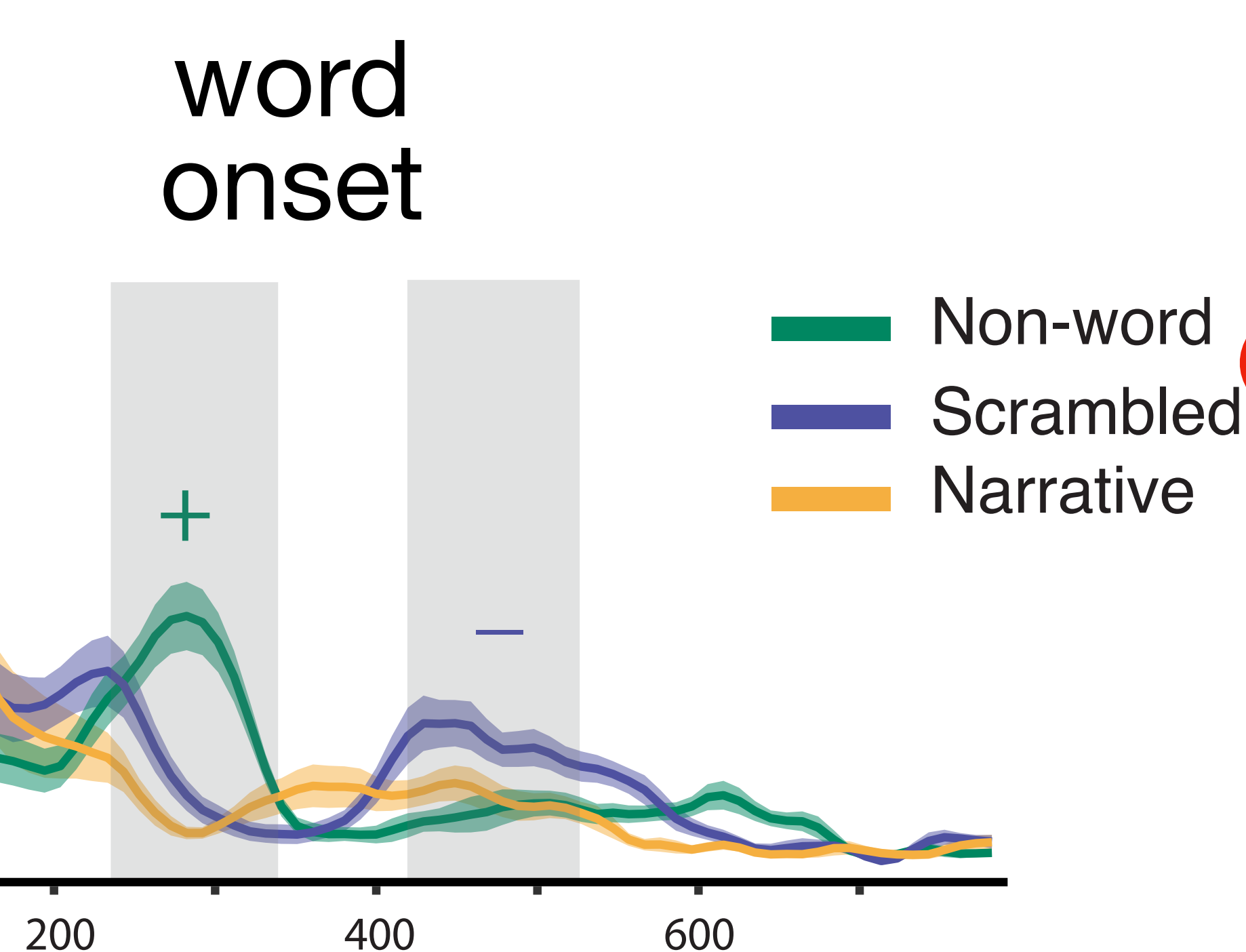
surprisal  
(no context)



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

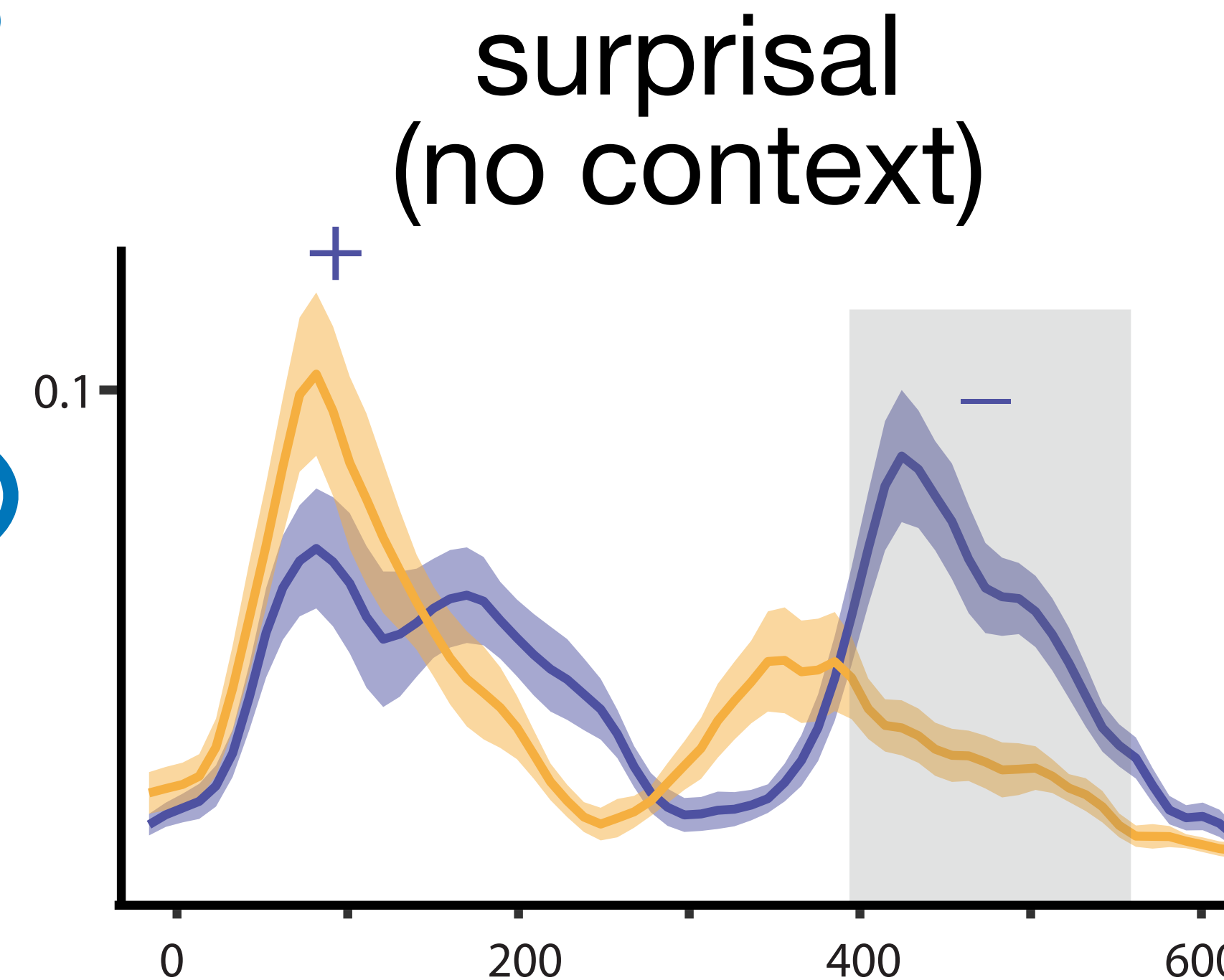
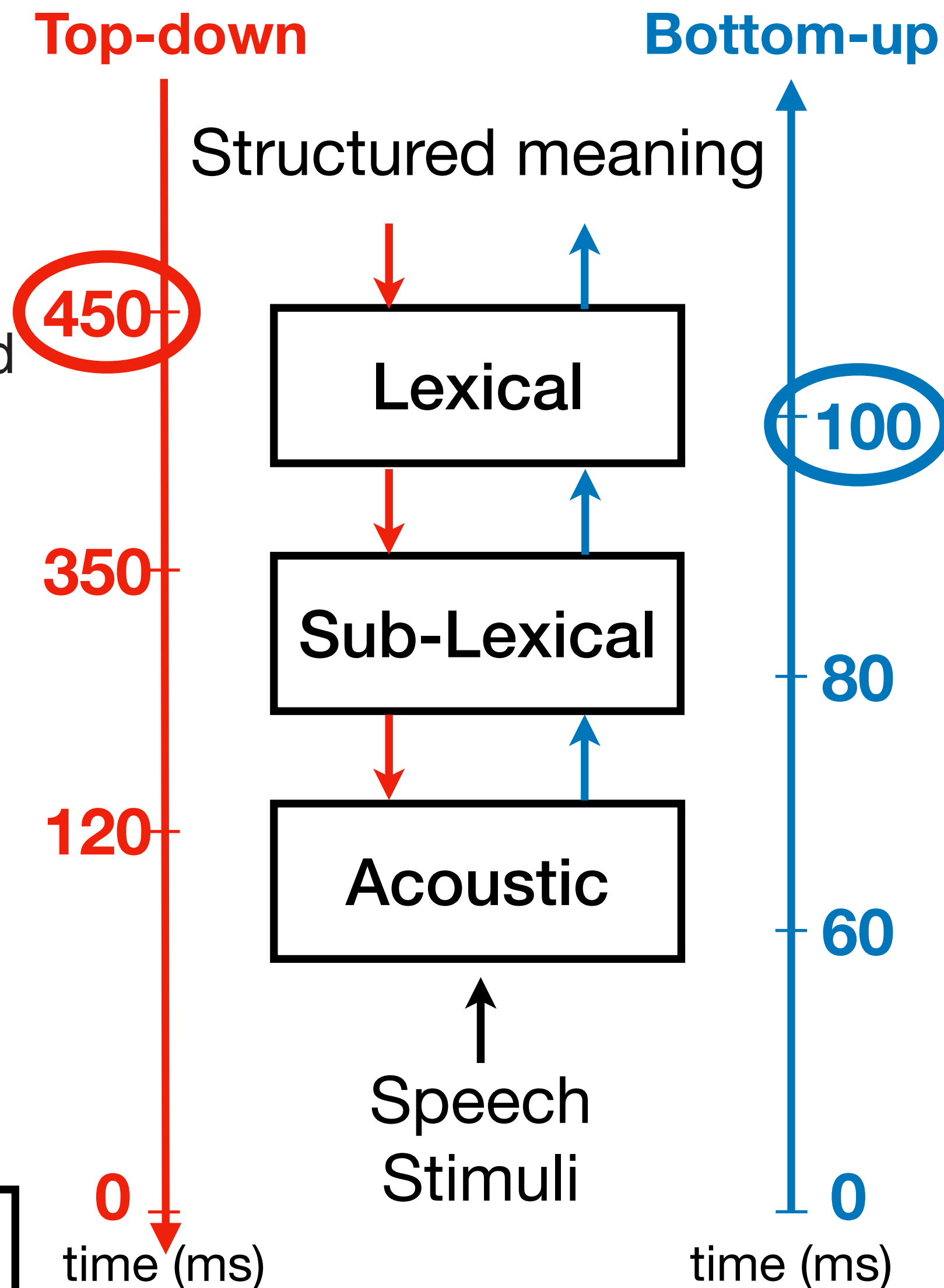
left hemisphere shown  
(right much weaker except for non-word onset)

# Word-based TRF Results



ed  $\approx$  narrative for rapid processing  
 ed words  $>$  narrative at  $\sim 450$  ms  
 left hemi  $>$  Right (non-words: L  $\approx$  R)

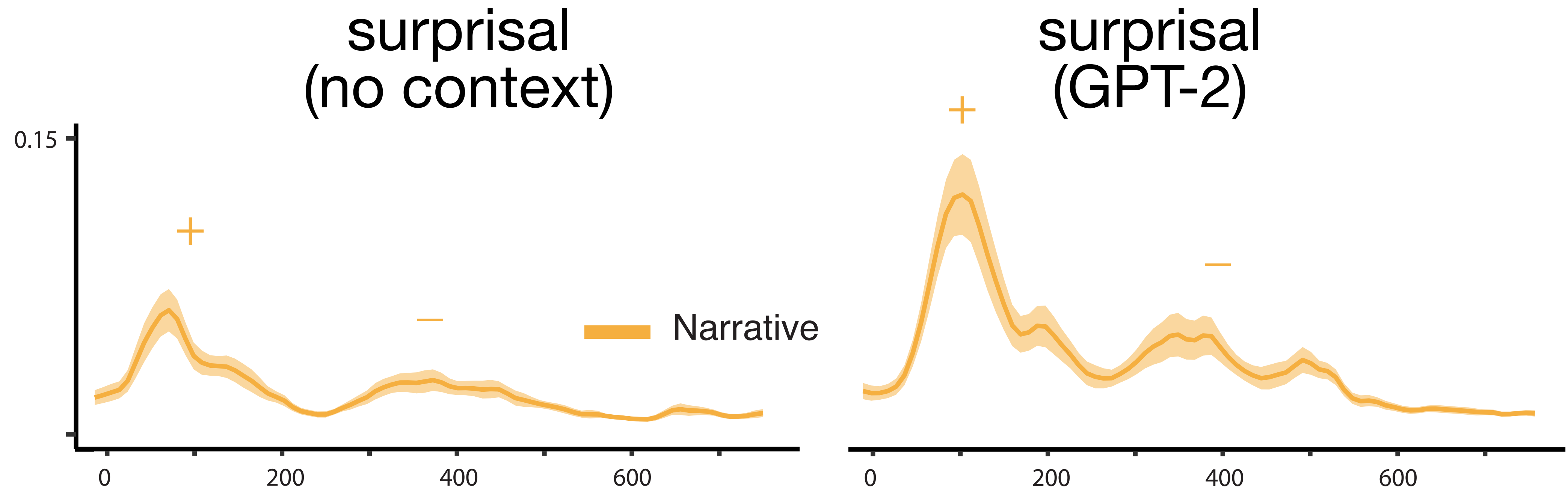
simple word processing  
 or" correction processing



- N400 like response
- Reduction in surprisal when co
- Left hemi  $>$  Right hemi
- Right hemisphere: Scrambled

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

# Contextual Word Surprisal Results

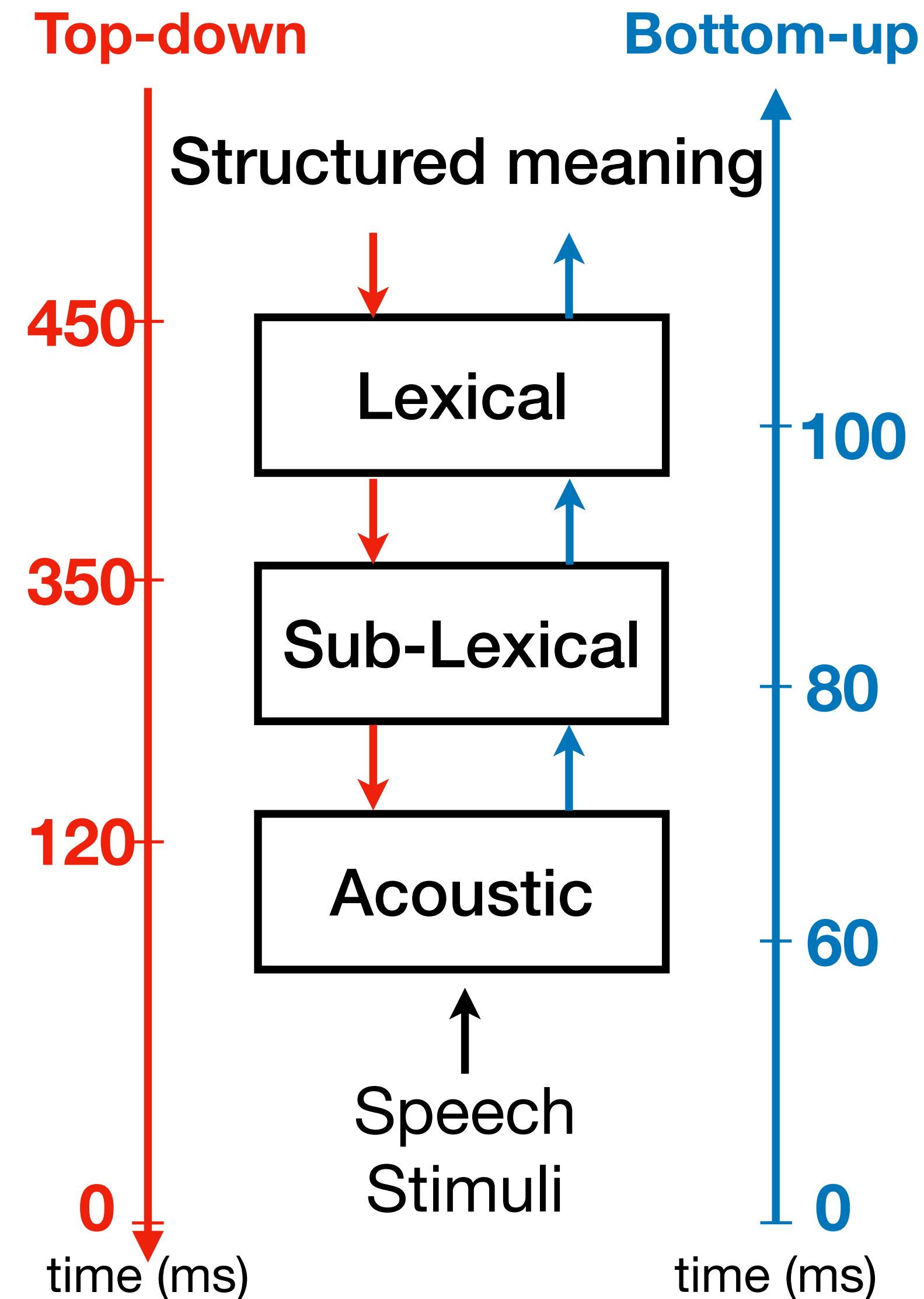


- When context helps, context-based surprisal is better tracked than raw surprisal
- N400 like response in both predictors

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

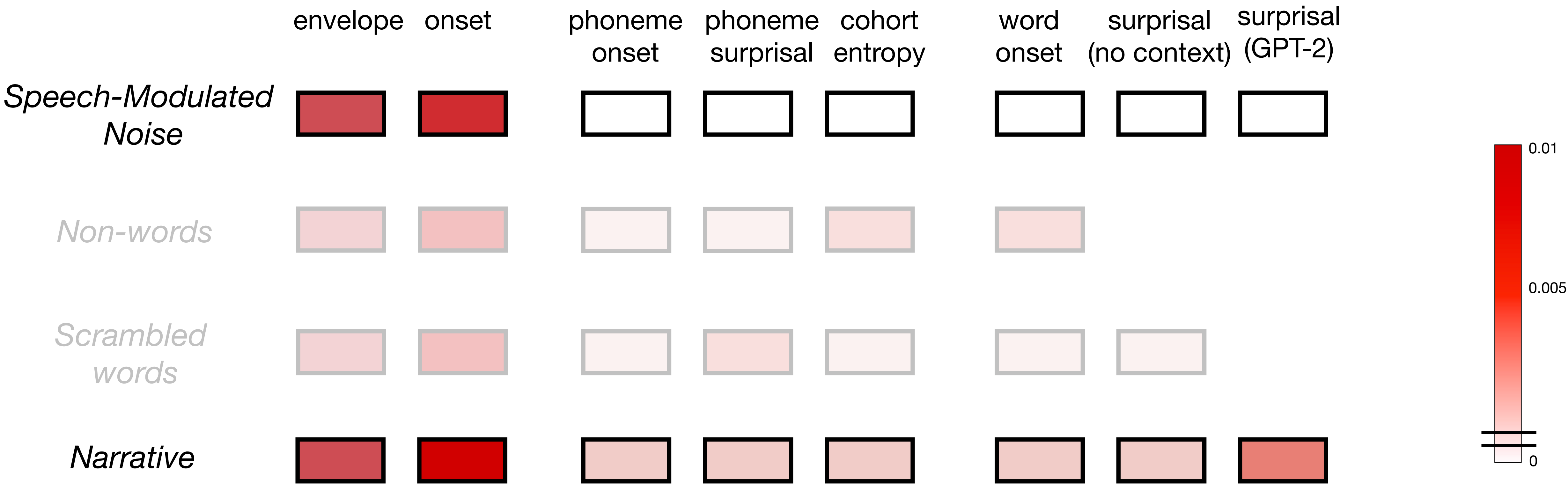




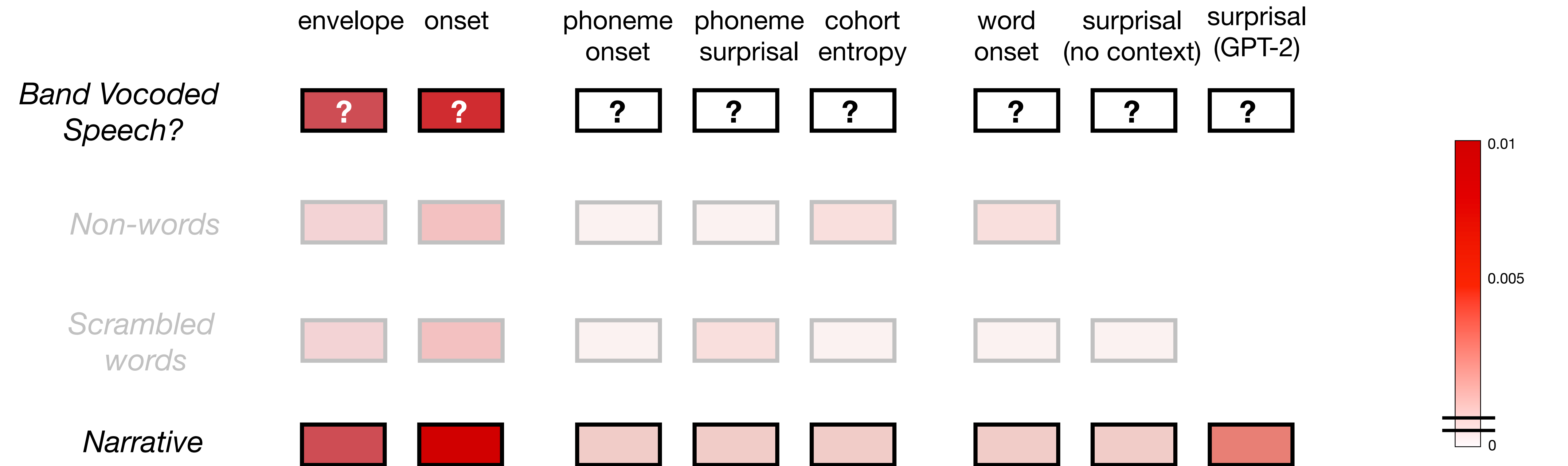
# Outline

- 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*

# Previous Neural Prediction Results

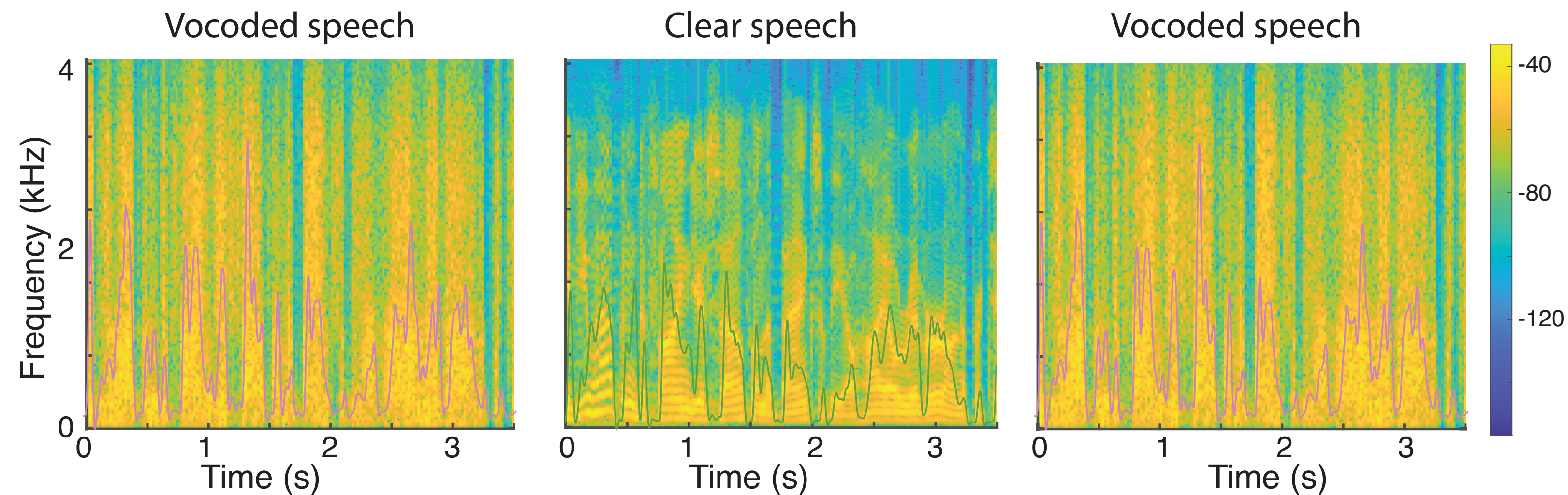
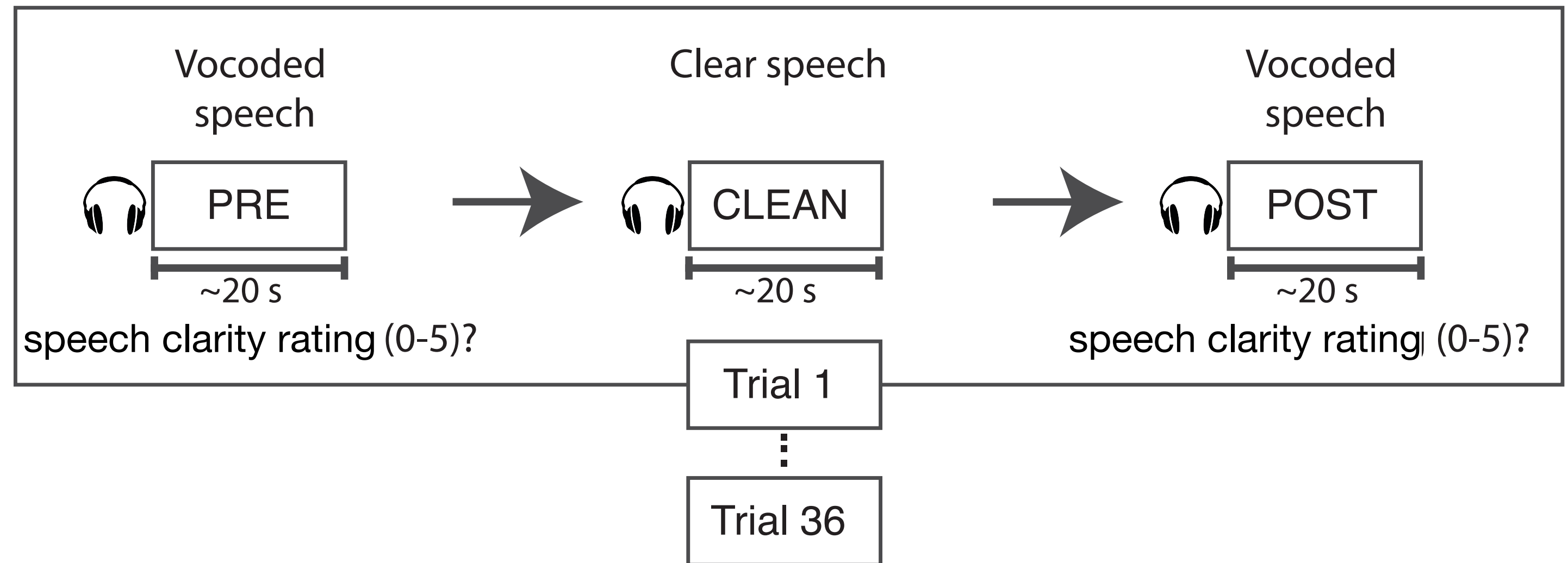


# Possible Neural Prediction Results



# Intelligibility Experimental Design

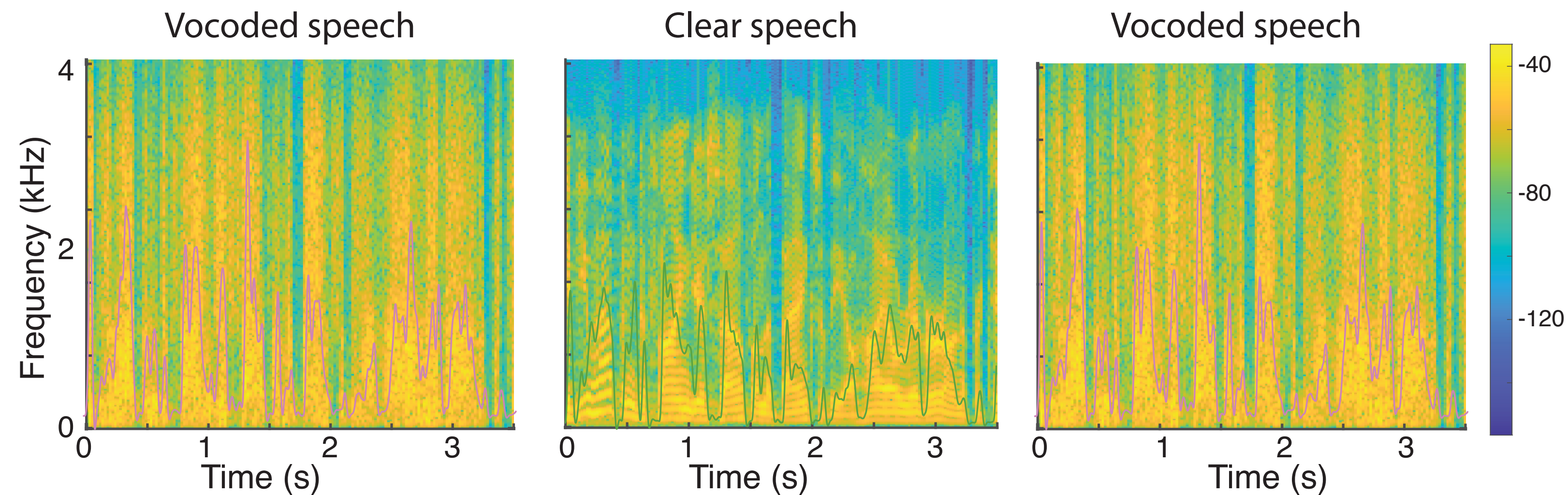
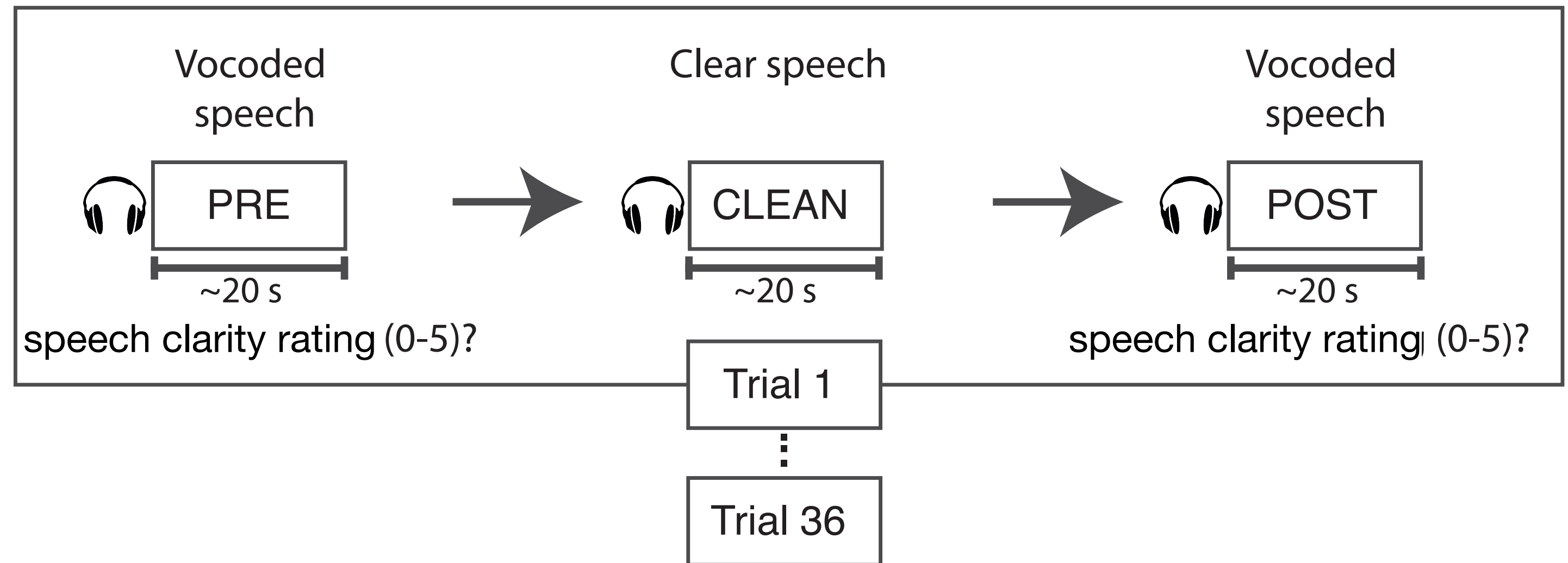
- Manipulate intelligibility but keep acoustics unchanged
  - Speech acoustics: three-band noise-vocoded speech
  - Intelligibility manipulated via priming
- Hypothesized intelligibility measure(s)
  - word boundaries





# Intelligibility Experimental Design

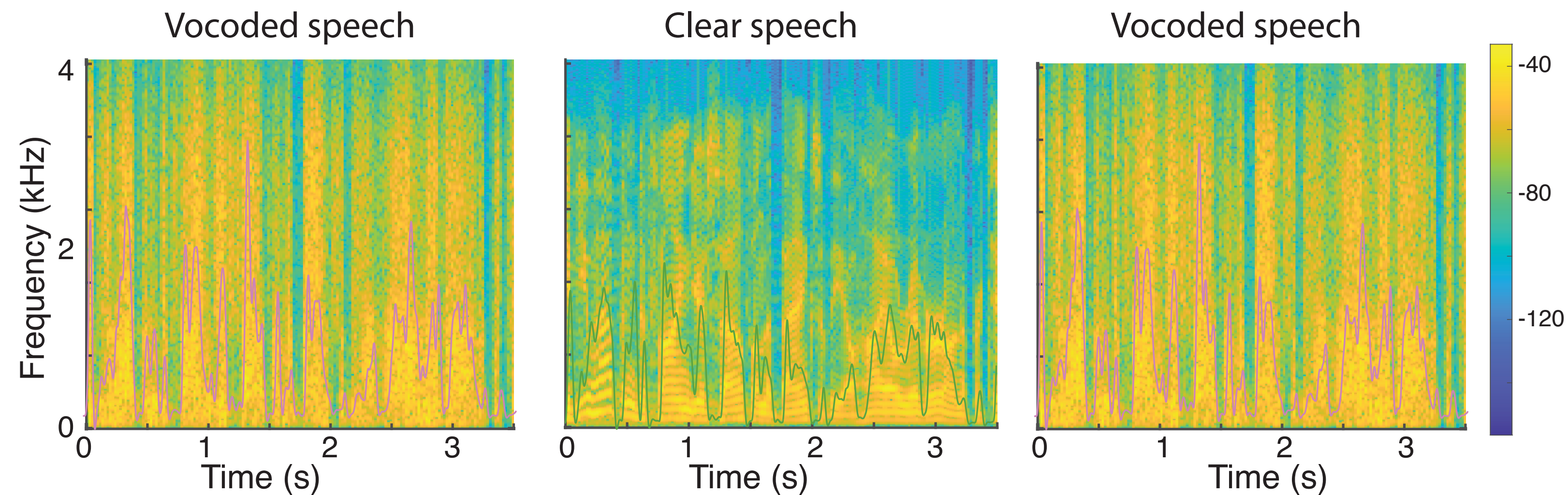
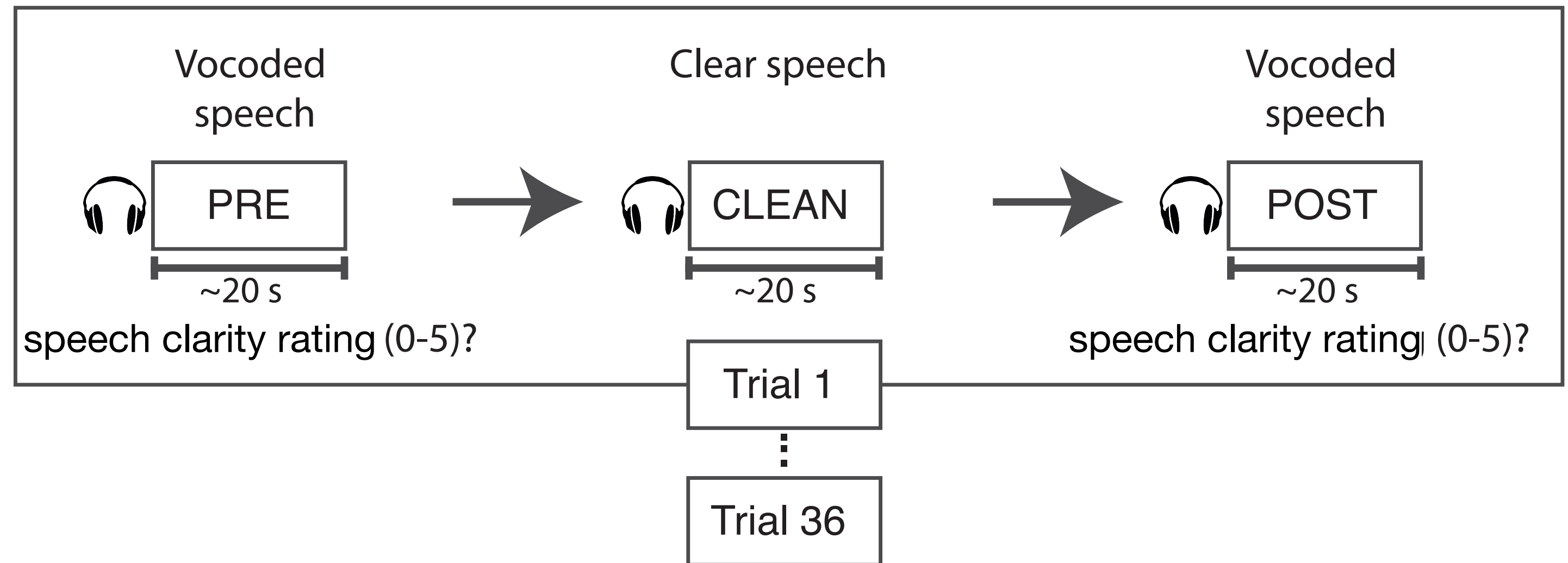
- Manipulate intelligibility but keep acoustics unchanged
  - Speech acoustics: three-band noise-vocoded speech
  - Intelligibility manipulated via priming
- Hypothesized intelligibility measure(s)
  - word boundaries





# Intelligibility Experimental Design

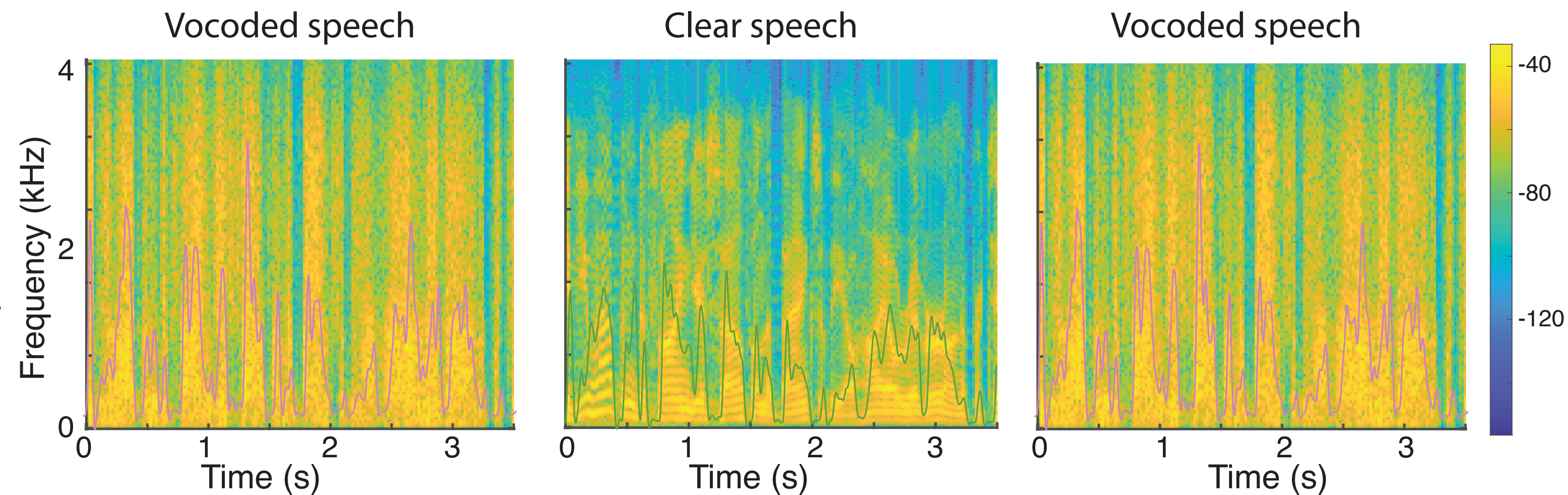
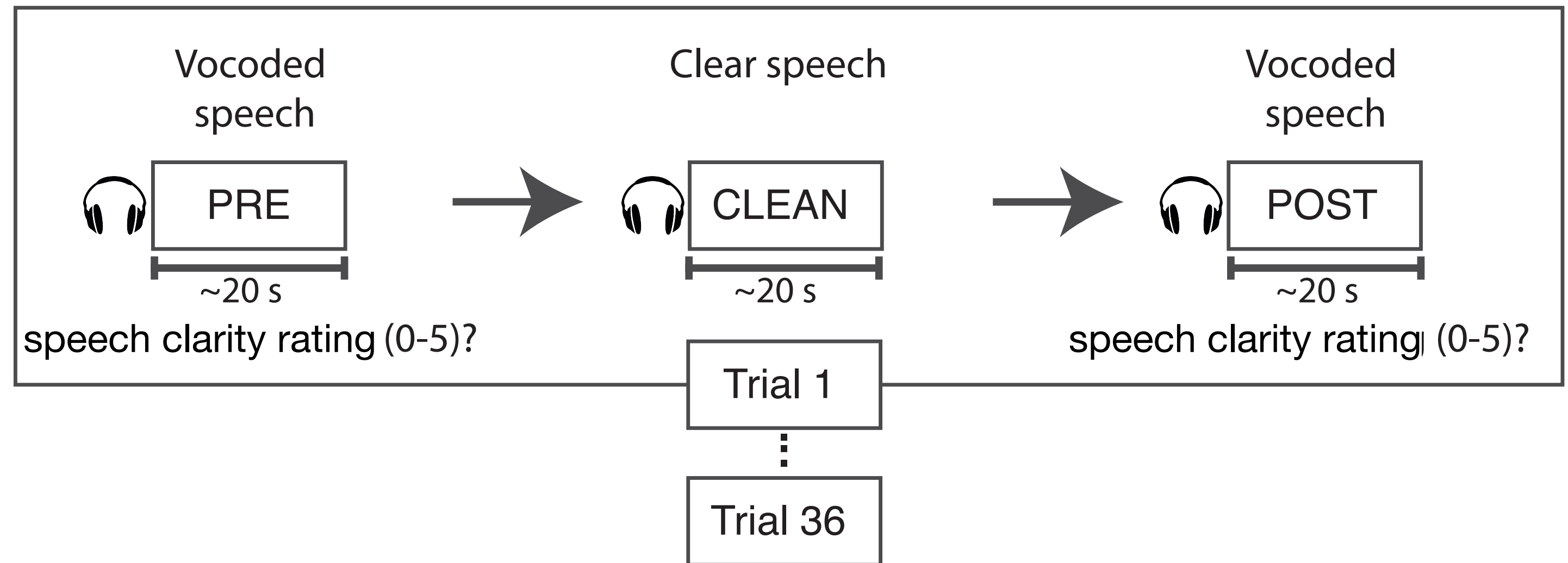
- Manipulate intelligibility but keep acoustics unchanged
  - Speech acoustics: three-band noise-vocoded speech
  - Intelligibility manipulated via priming
- Hypothesized intelligibility measure(s)
  - word boundaries





# 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

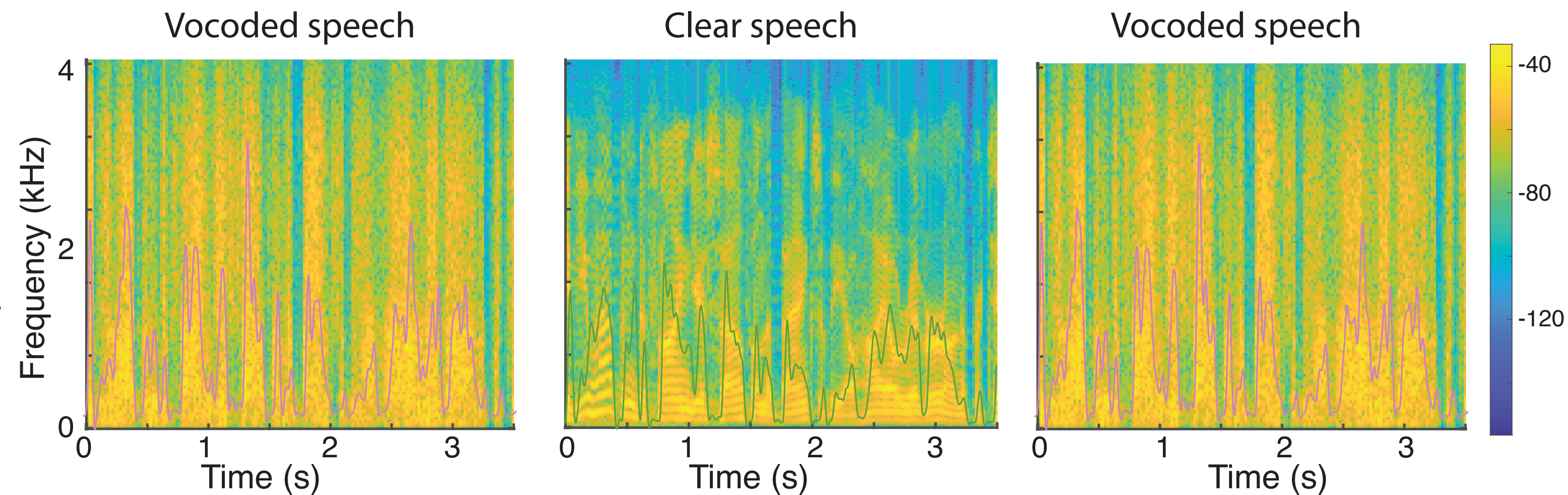
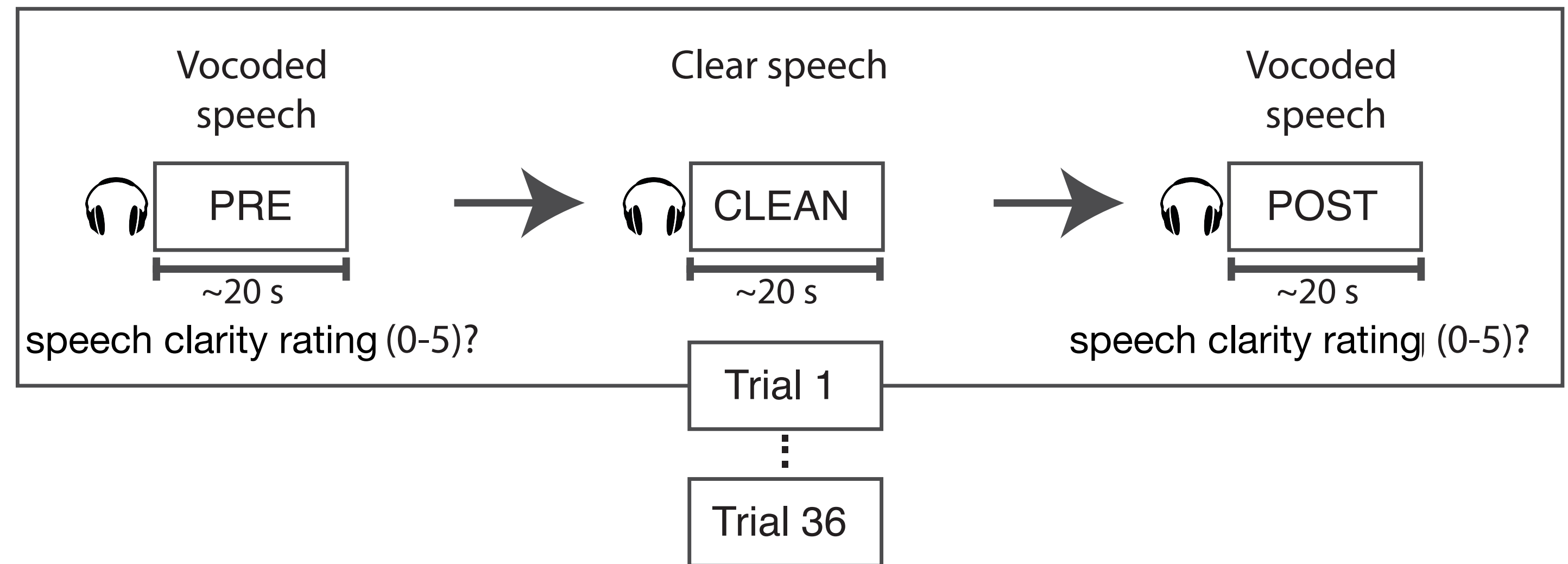


*“Slice an apple through at its equator, and you will find five small chambers arrayed in a perfectly symmetrical starburst—a pentagram.”*



# 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

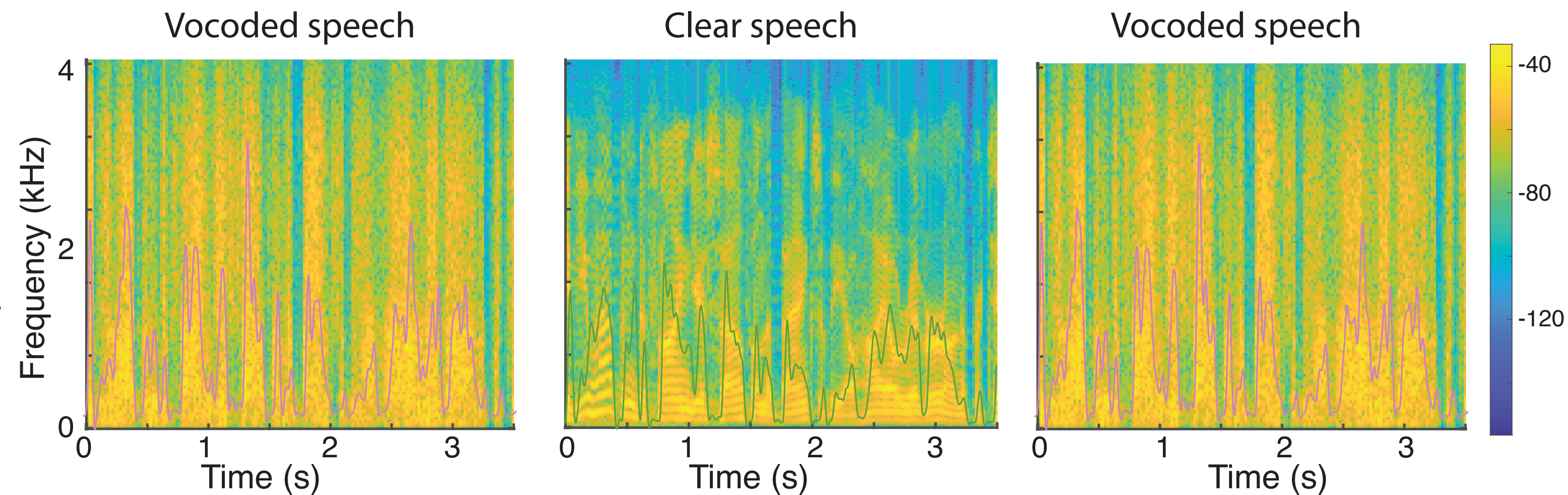
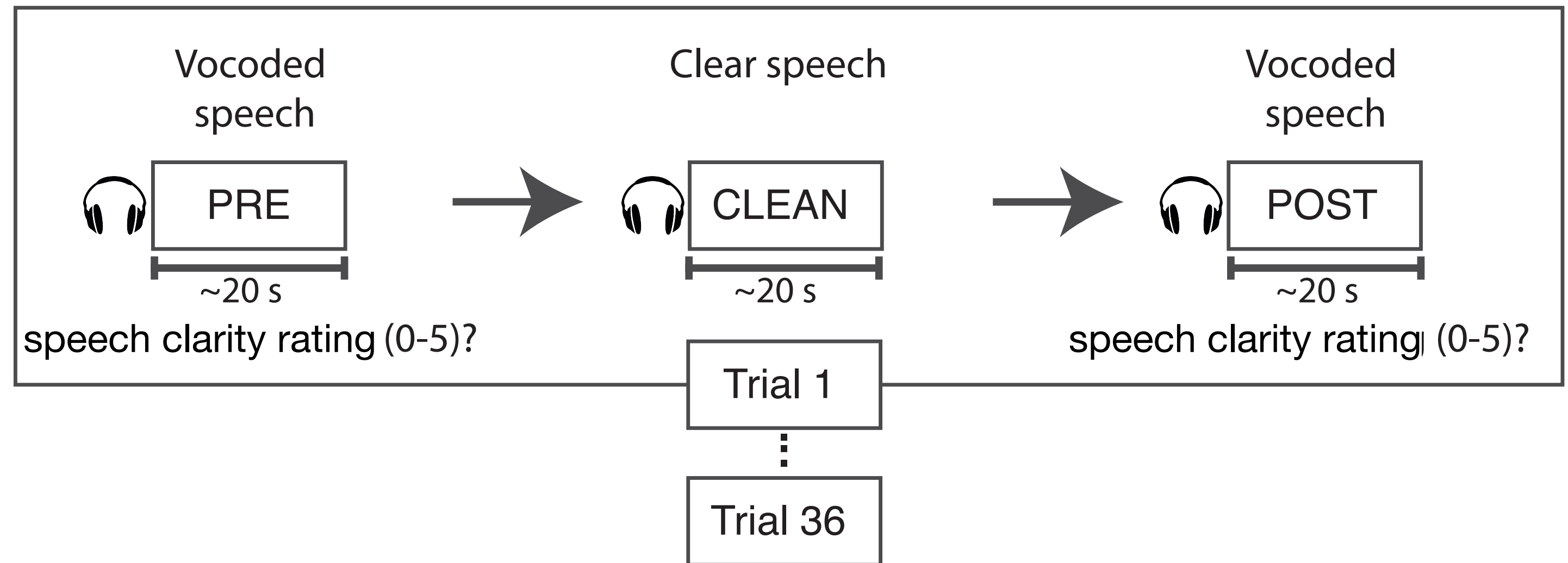


*“Slice an apple through at its equator, and you will find five small chambers arrayed in a perfectly symmetrical starburst—a pentagram.”*



# 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

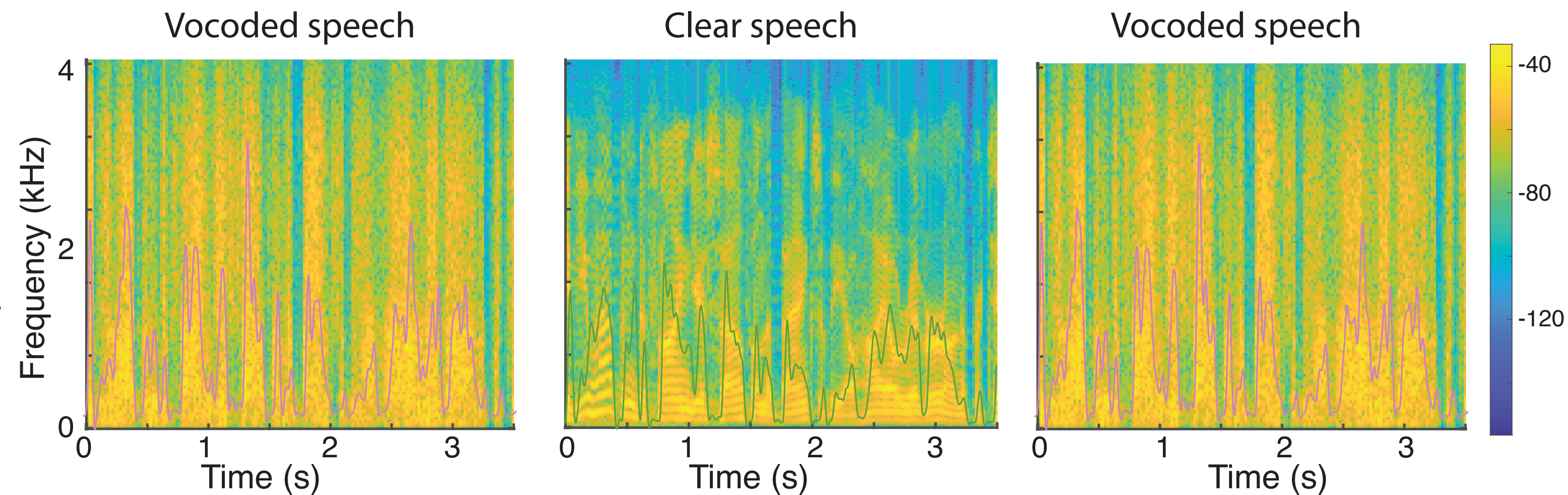
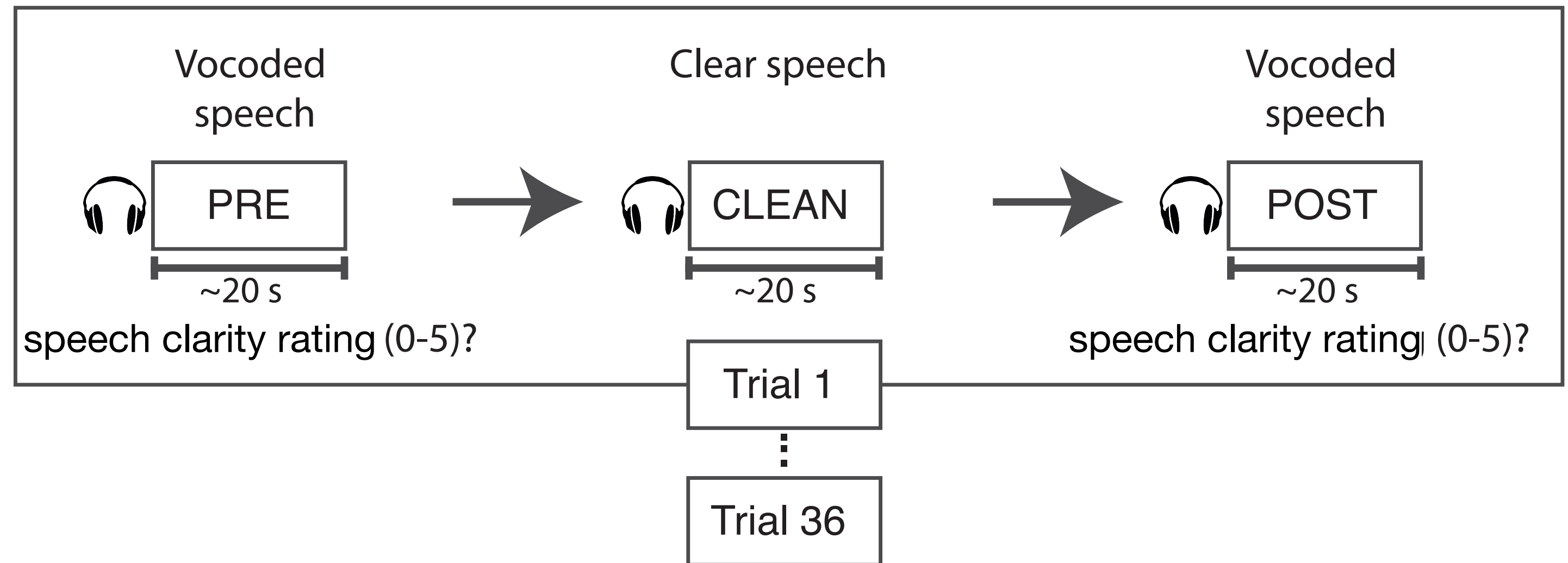


*“Slice an apple through at its equator, and you will find five small chambers arrayed in a perfectly symmetrical starburst—a pentagram.”*



# Intelligibility Experimental Design

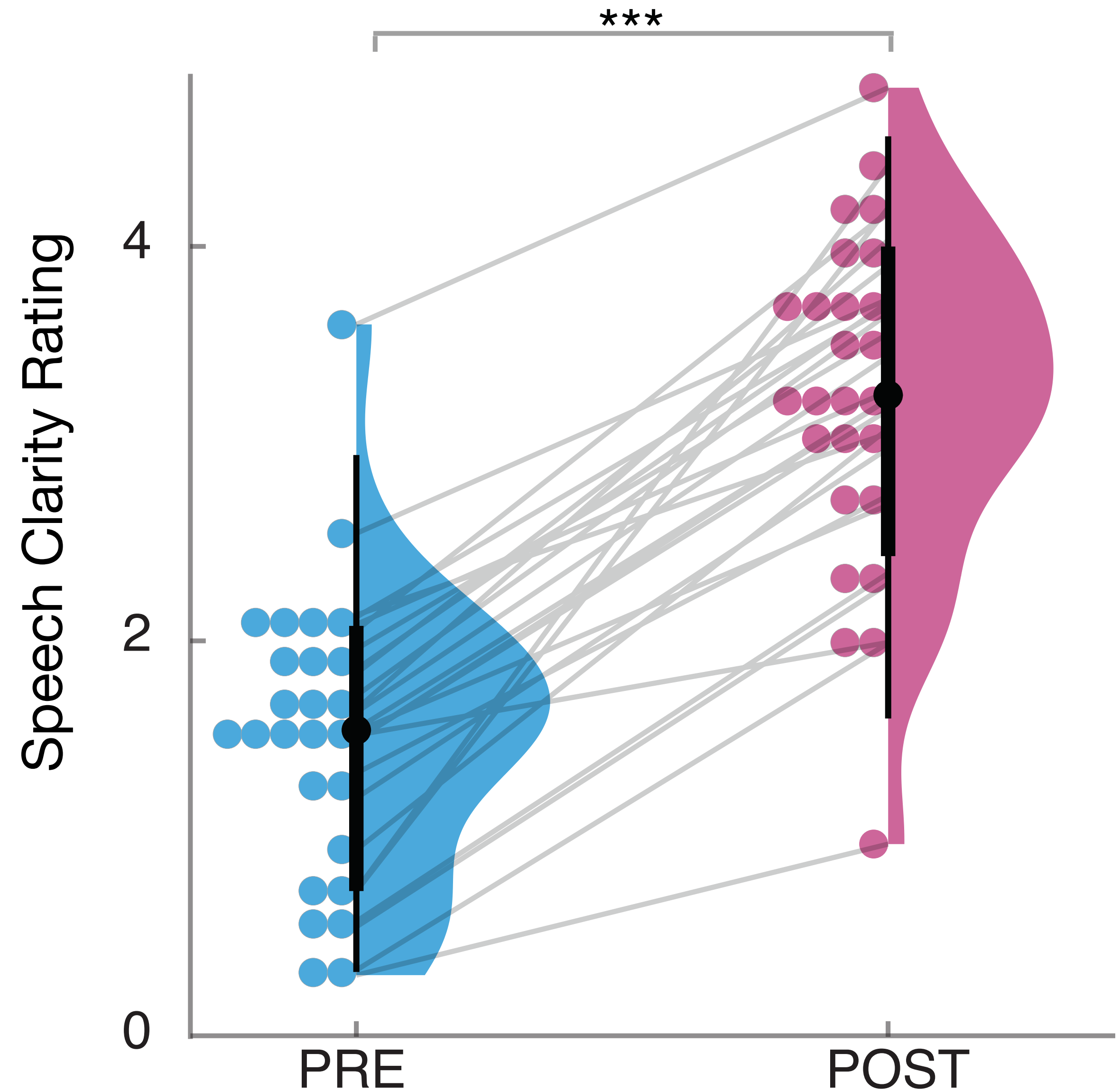
- Manipulate intelligibility but keep acoustics unchanged
  - Speech acoustics: three-band noise-vocoded speech
  - Intelligibility manipulated via priming
- Hypothesized intelligibility measure(s)
  - word boundaries



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

# Intelligibility Behavioral Results

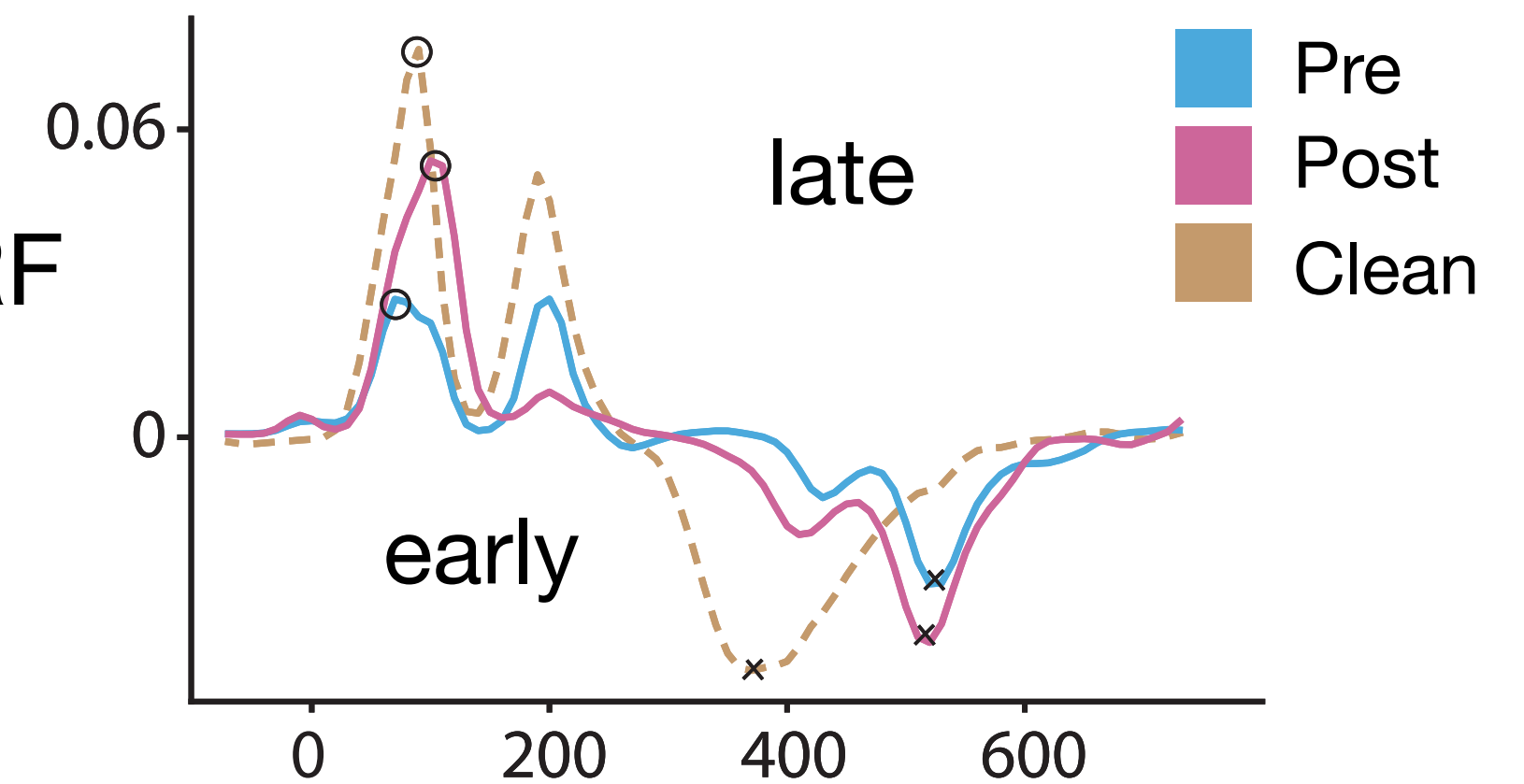
Speech Clarity **increases**  
from PRE condition  
to POST condition



# Intelligibility Neural Results

- Word onset TRF shows both early (+) and late (-) processing stages

Word onset TRF  
*single subject*

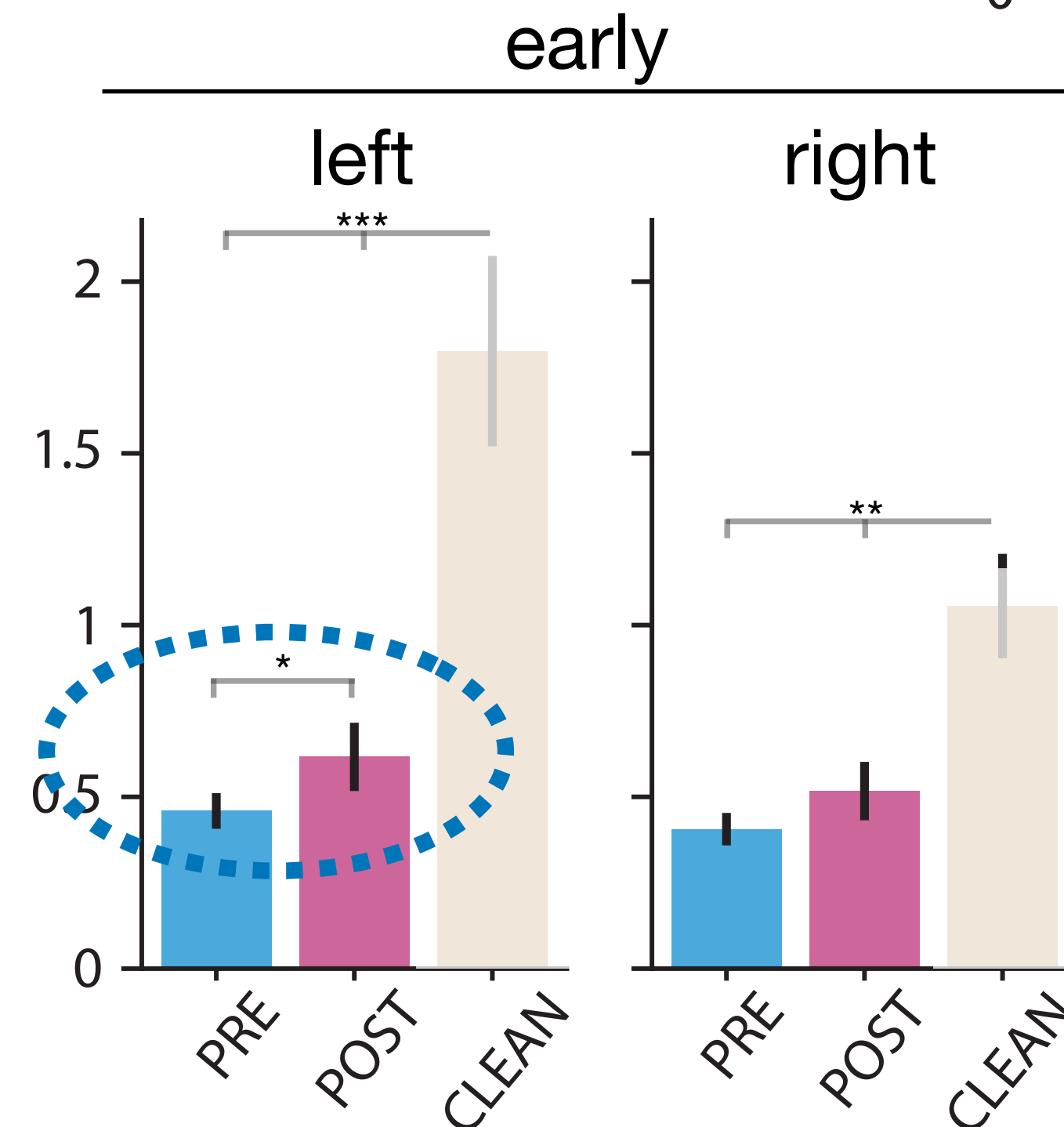
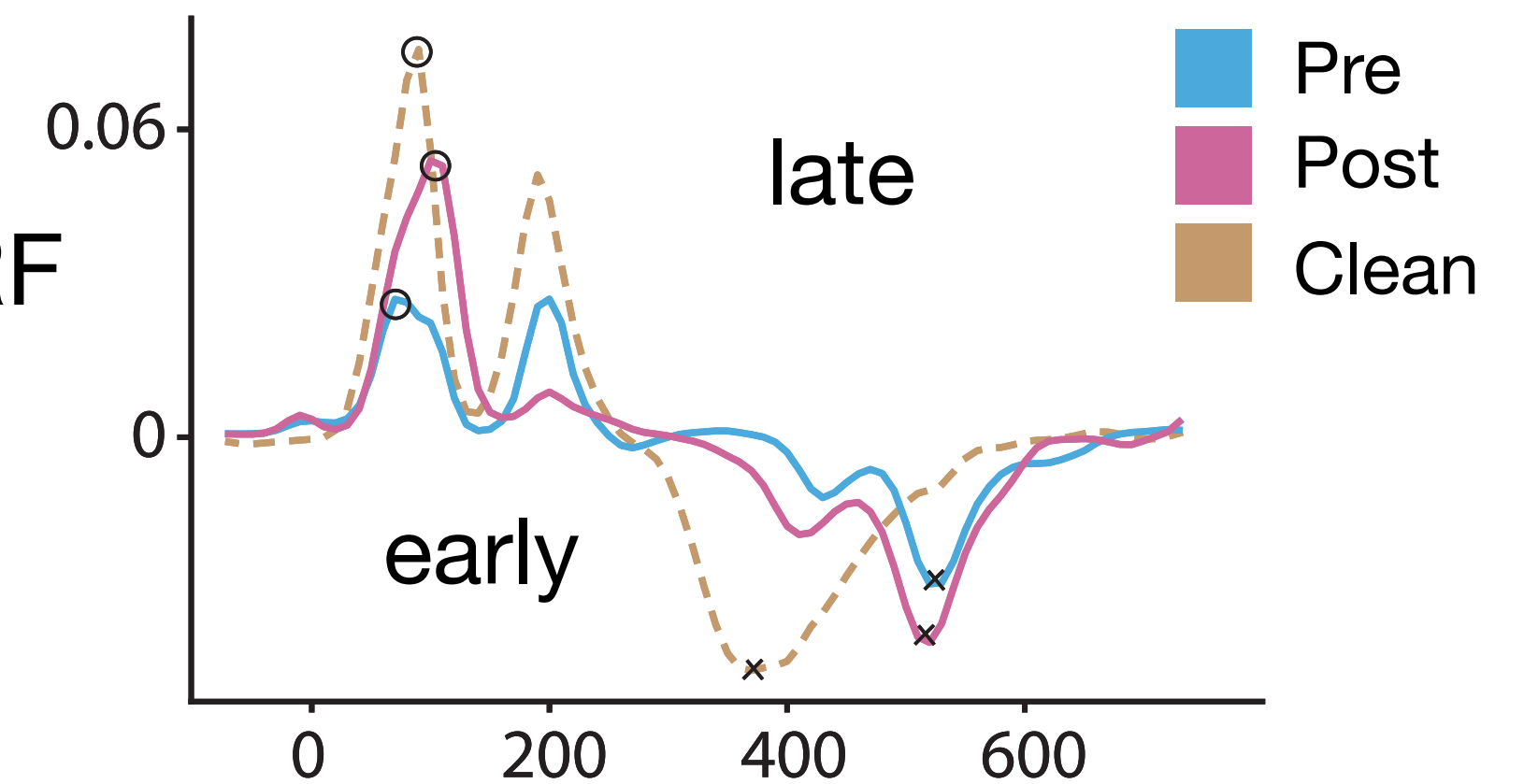




# Intelligibility Neural Results

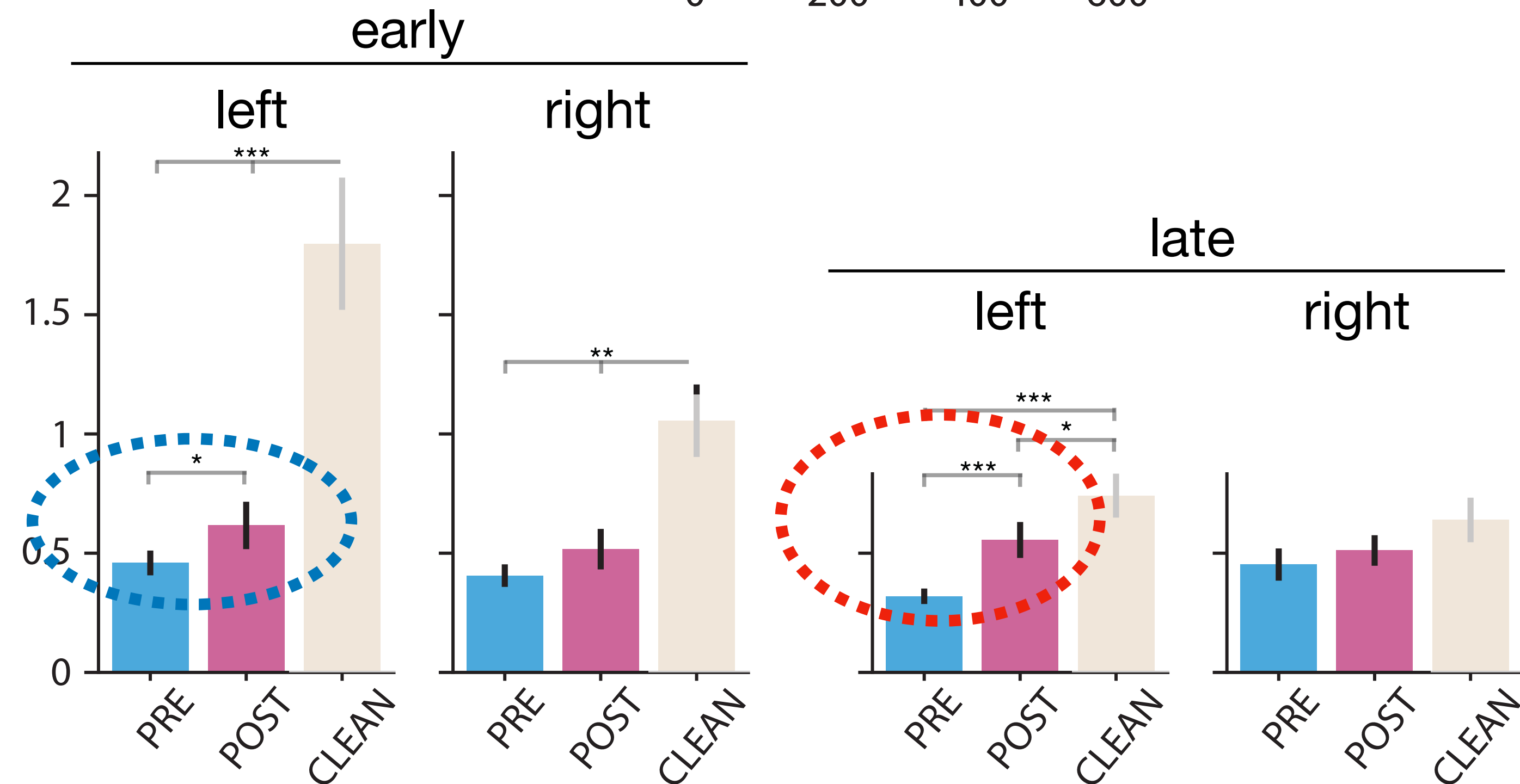
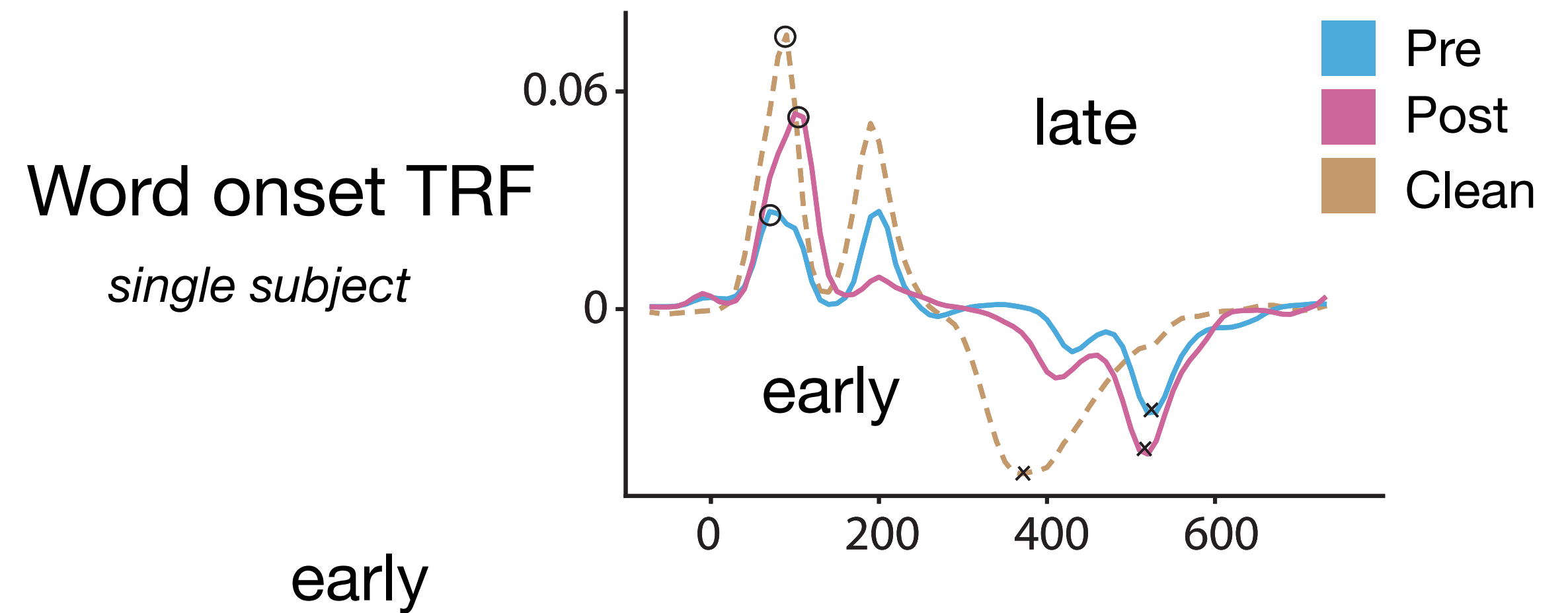
- Word onset TRF shows both early (+) and late (-) processing stages
- Response increases Pre→Post
  - Only in left hemisphere

Word onset TRF  
*single subject*



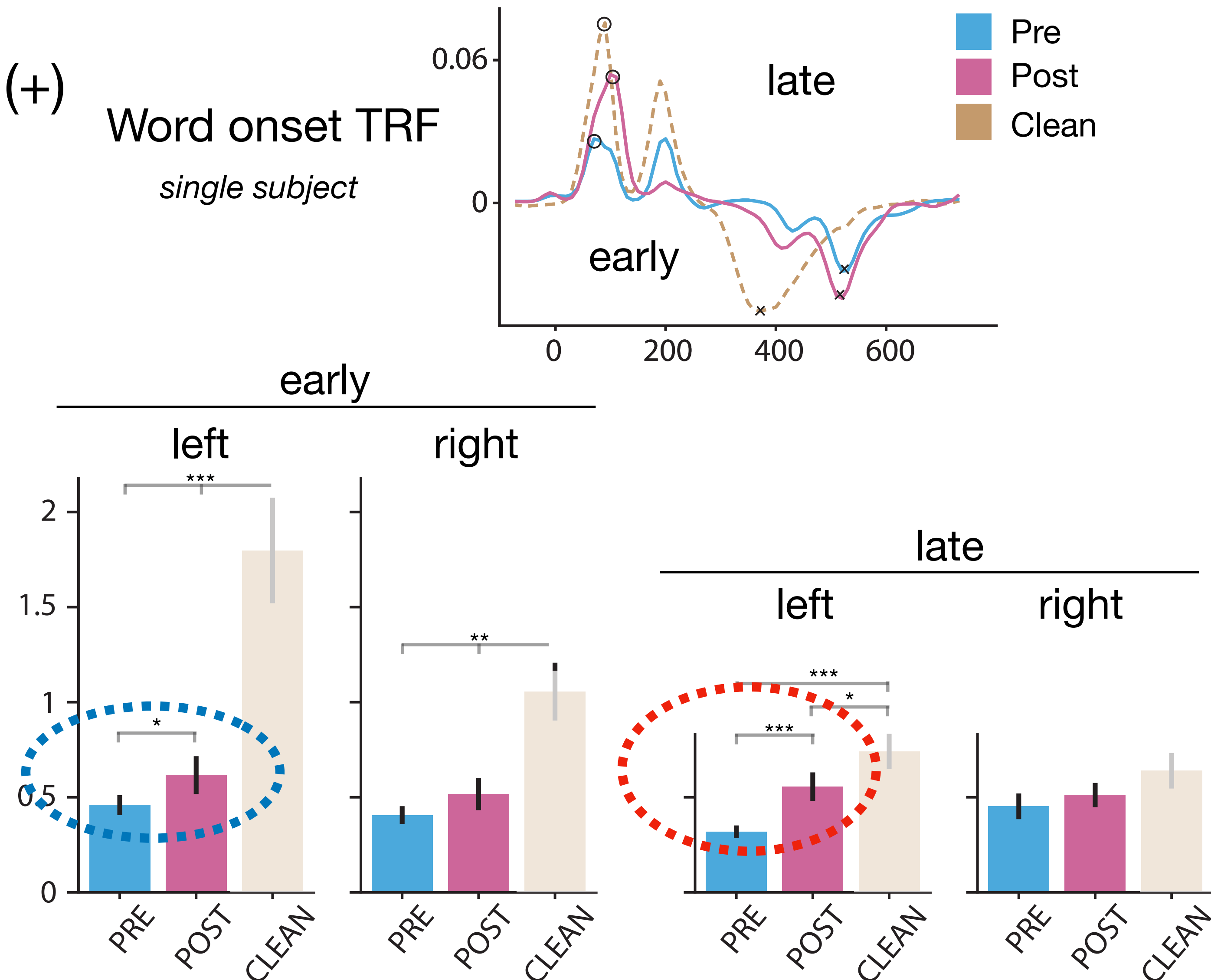
# Intelligibility Neural Results

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



# Intelligibility Neural Results

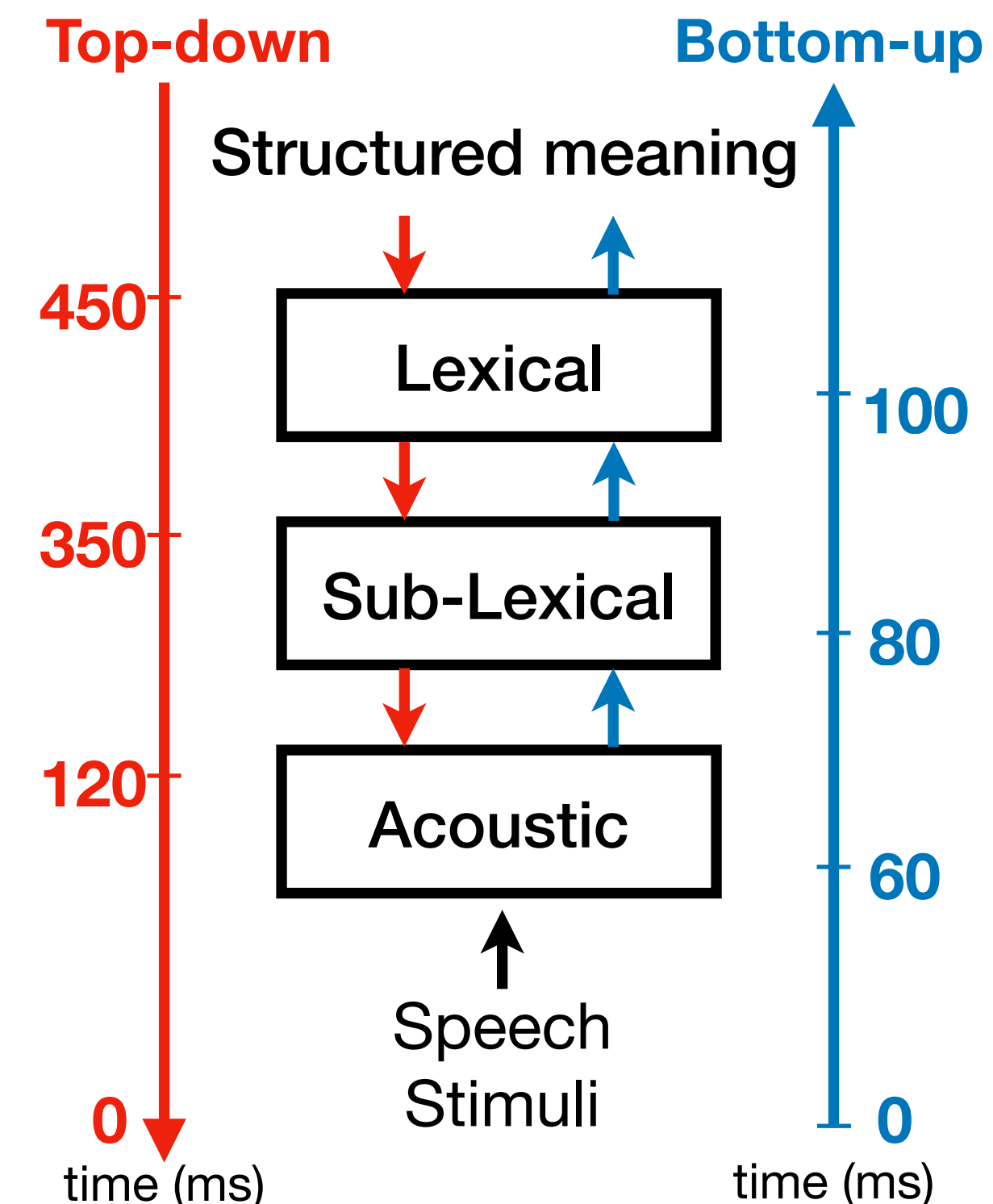
- Word onset TRF shows both early (+) and late (-) processing stages
- Response increases Pre→Post
  - Only in left hemisphere
  - Late processing stage shows larger change than early
- Response to Word Onset:
  - Objective measure of intelligibility*
  - Acoustic responses: no change
  - Response to Word Surprisal:
    - Additional intelligibility measure*



# Final Summary

*temporal patterns in **speech acoustics***  
*temporal **neural** patterns  $\rightleftharpoons$  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](http://ter.ps/simonpubs)



Mastodon: [@jzsimon@fediscience.org](https://mastodon.social/@jzsimon)

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