Progression of acoustic, phonemic, lexical and sentential neural features emerge during speech listening

I.M Dushyanthi Karunathilake, Christian Brodbeck, Shohini Bhattasali, Philip Resnik, Jonathan Z. Simon

University of Maryland, College Park
University of Connecticut

ARO, Feb 13,2023

Introduction

How do our brains perceive and construct meaning from continuous sounds?

- Acoustic inputs are turned into a meaning via intermediate stages.

  – Intermediate stages can perform

    ‣ Acoustic analysis (acoustic envelope)

    ‣ Phonological analysis (phonemes)

    ‣ Lexical processing of word (words)

    ‣ Superlexical processing of sentence (sentential)

- Not all processing is straight bottom-up

  – Top-down mechanisms

    ‣ Selective attention

   ‣ Secondary processing upon 'error' correction

    ‣ Prediction
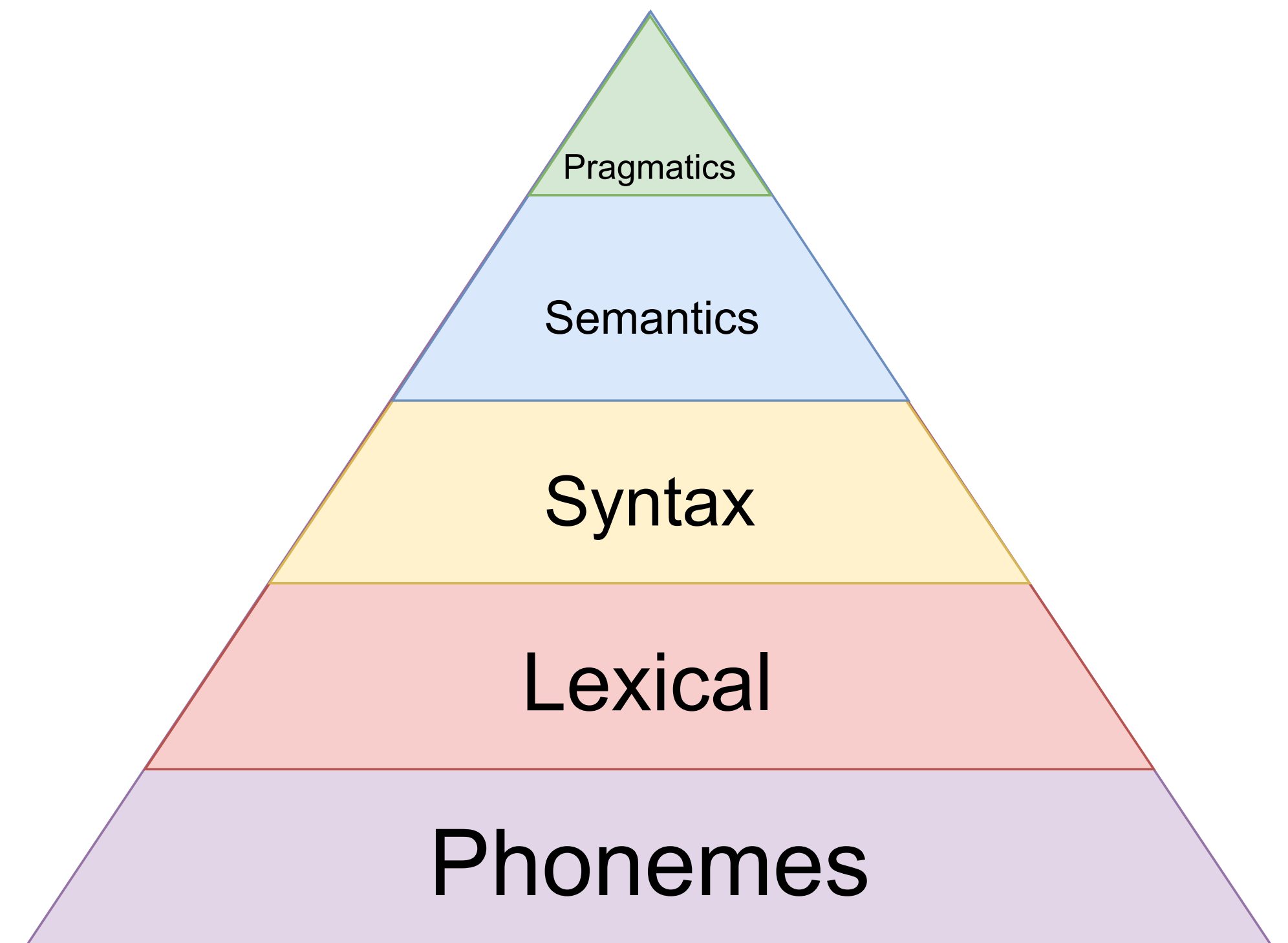
Background

- fMRI based research has investigated which brain regions process speech and language features [Xu et al. 2004]

- EEG/MEG research has shown, which different speech features are represented in the brain
[Gillis et al. 2021, Brodbeck et al. 2018]

Here we investigate

- the progression and representation of different speech features along the speech and language hierarchy

- how speech features emerge under different speech conditions

- which, how and when features are driven by bottom-up and top-down mechanisms

High level meaning

Lexical

Sub-lexical

Acoustic

Speech Stimuli
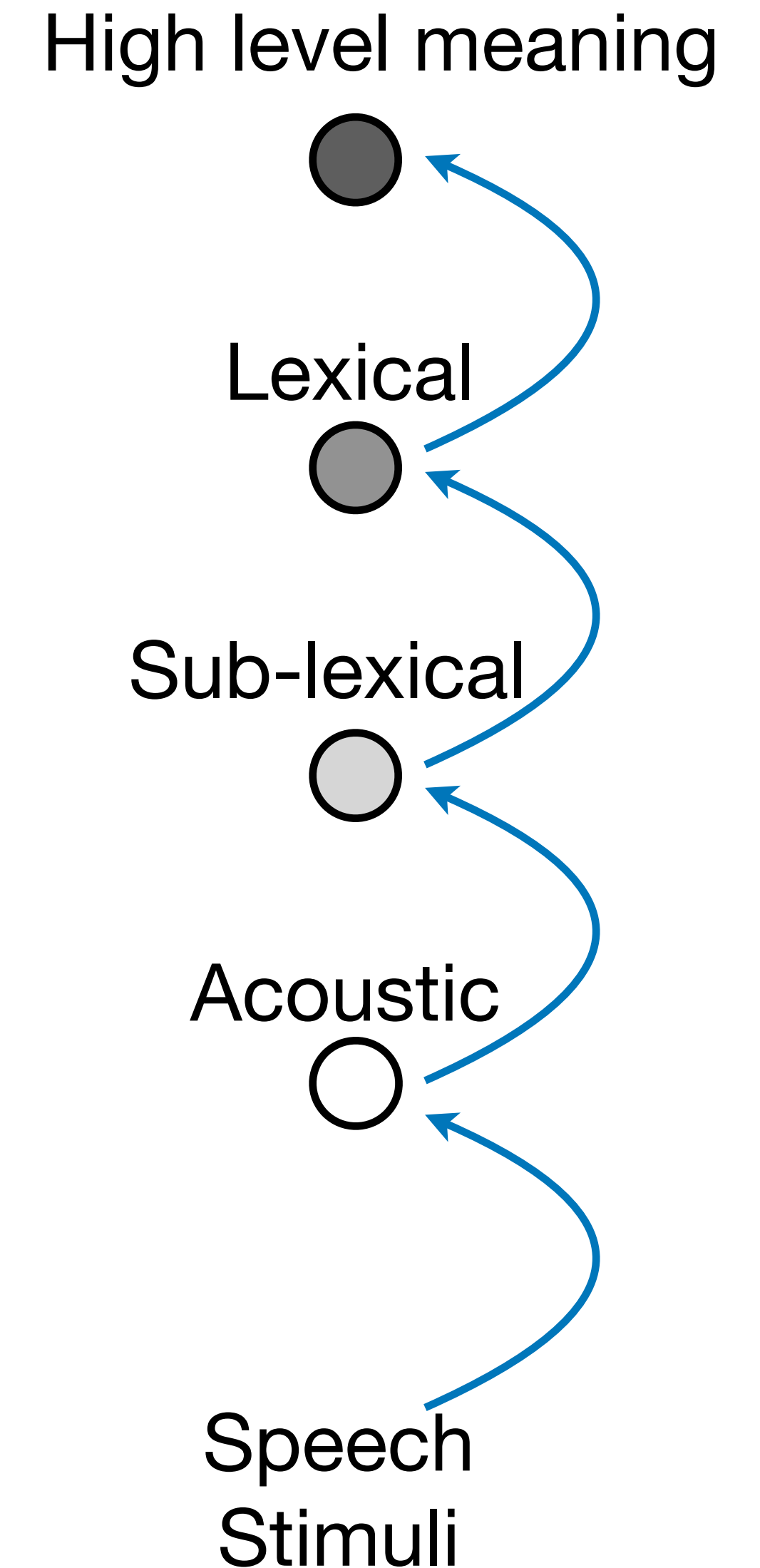
Experimental Design

Participants
- 30 younger adults (15-female, 18 - 30 yrs)
- Native English Speakers

Data
- Magnetoencephalography (MEG)

Task
- Listening to 1-minute long passages (1-minute x 4 passage types x 2 repetitions)
    - Non fiction story (The Botany of Desire, by Michael Pollan)
- 4 passage types
    - Speech modulated noise
    - Non-words
    - Scrambled words
    - Narrative

Stimuli
- Speech materials were synthesized using google text to speech (gTTS) synthesizer

Experimental Design (Stimuli)

Speech-envelope Modulated Noise

Speech

Speech Spectrum

0.01

Speech-Envelope Modulated Noise

Speech-Envelope Modulated Noise Spectrum

0.01

0 5 s

0 2 4 kHz
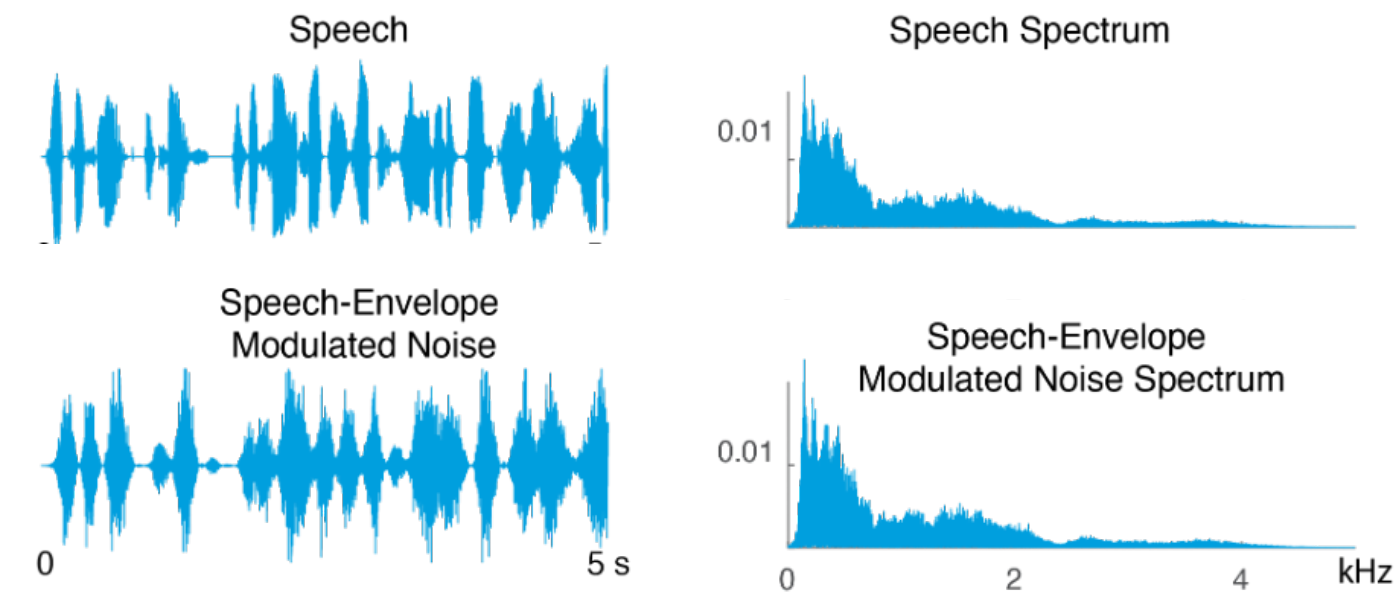
Non-words

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 ……
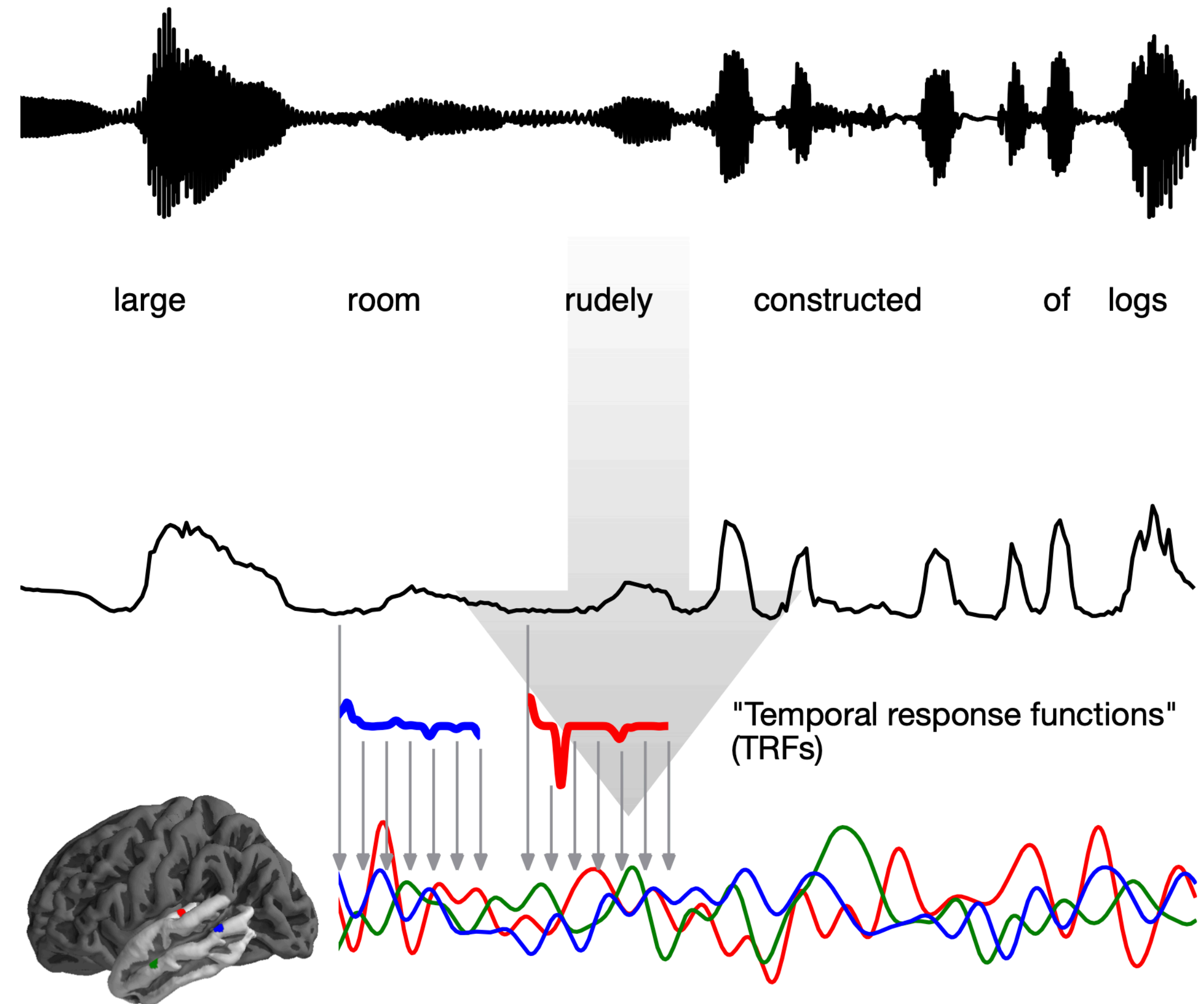
Narrative

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 …..

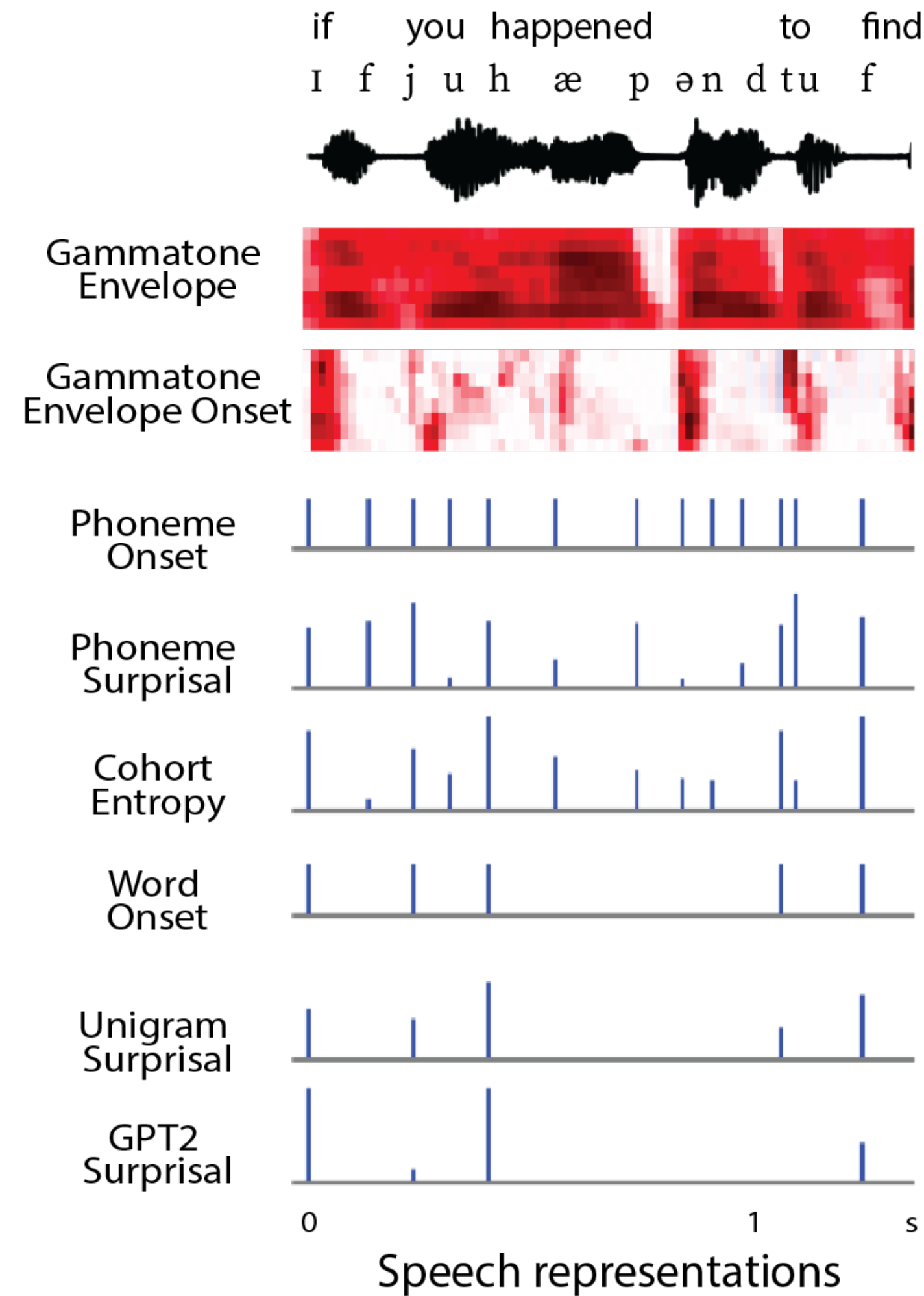Speech-like prosody and rhythm,

Male speaker

4

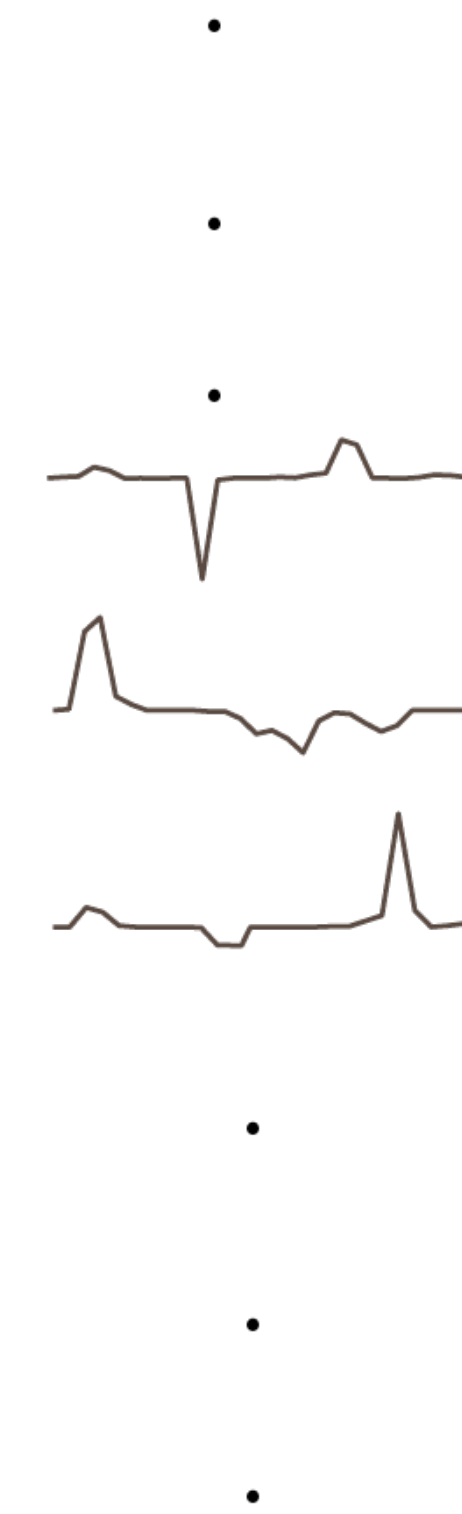# Analysis - Temporal Response Functions

- Predicting neural response from speech

- Why TRFs?
  - ‣ TRFs analogues to evoked response
  - ‣ Peak amplitude ≈ processing intensity
  - ‣ Peak Latency ≈ source location

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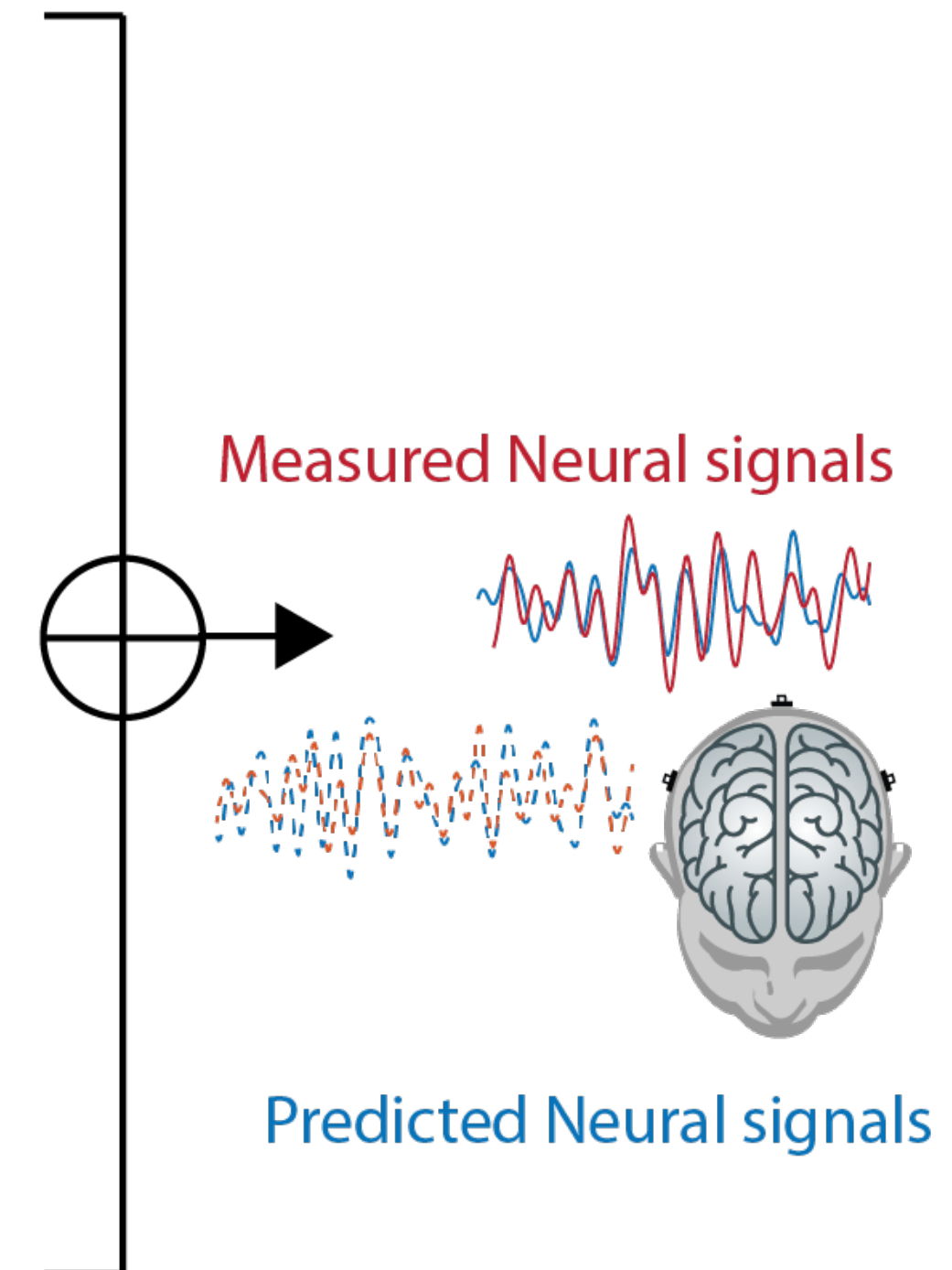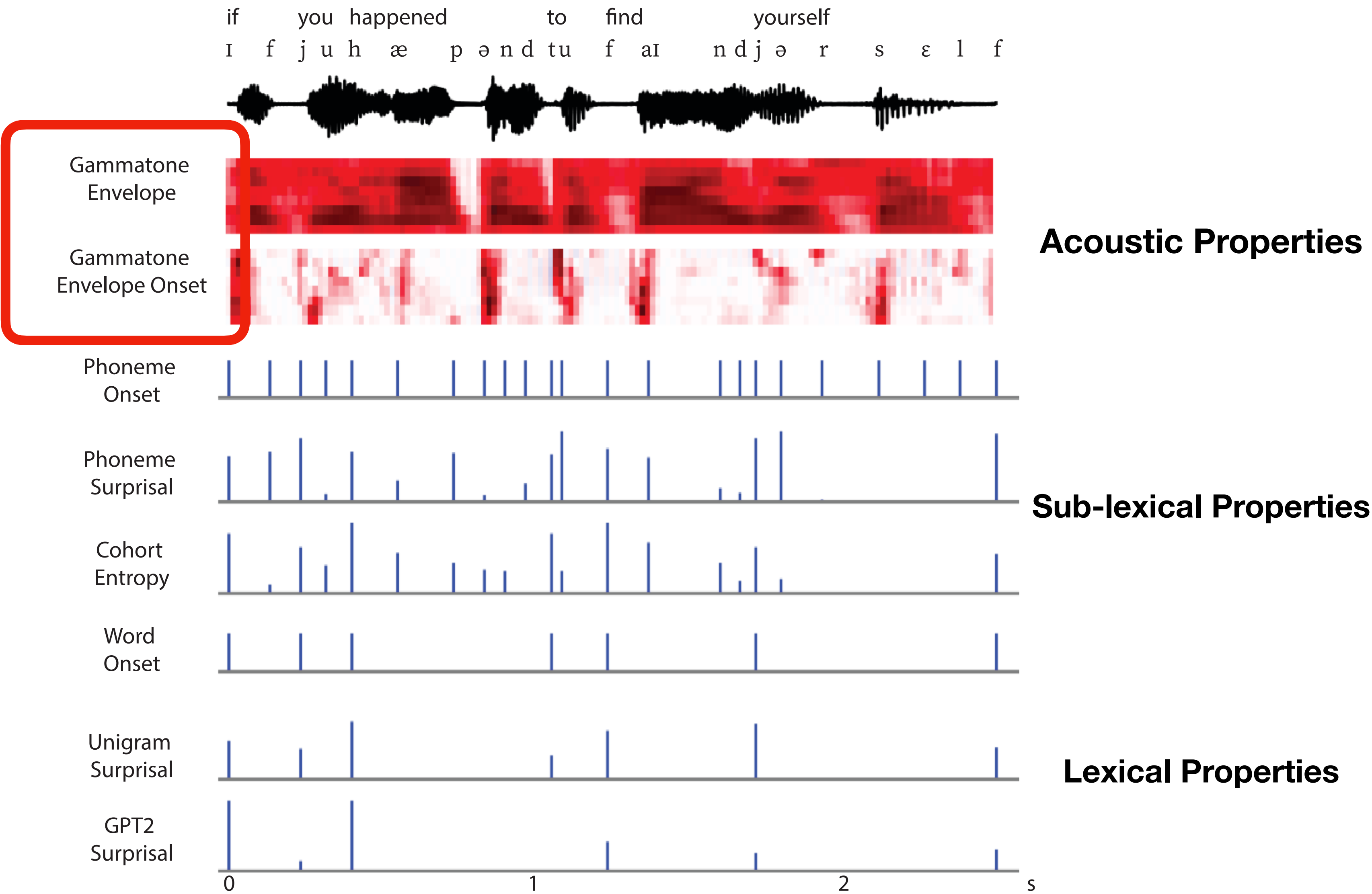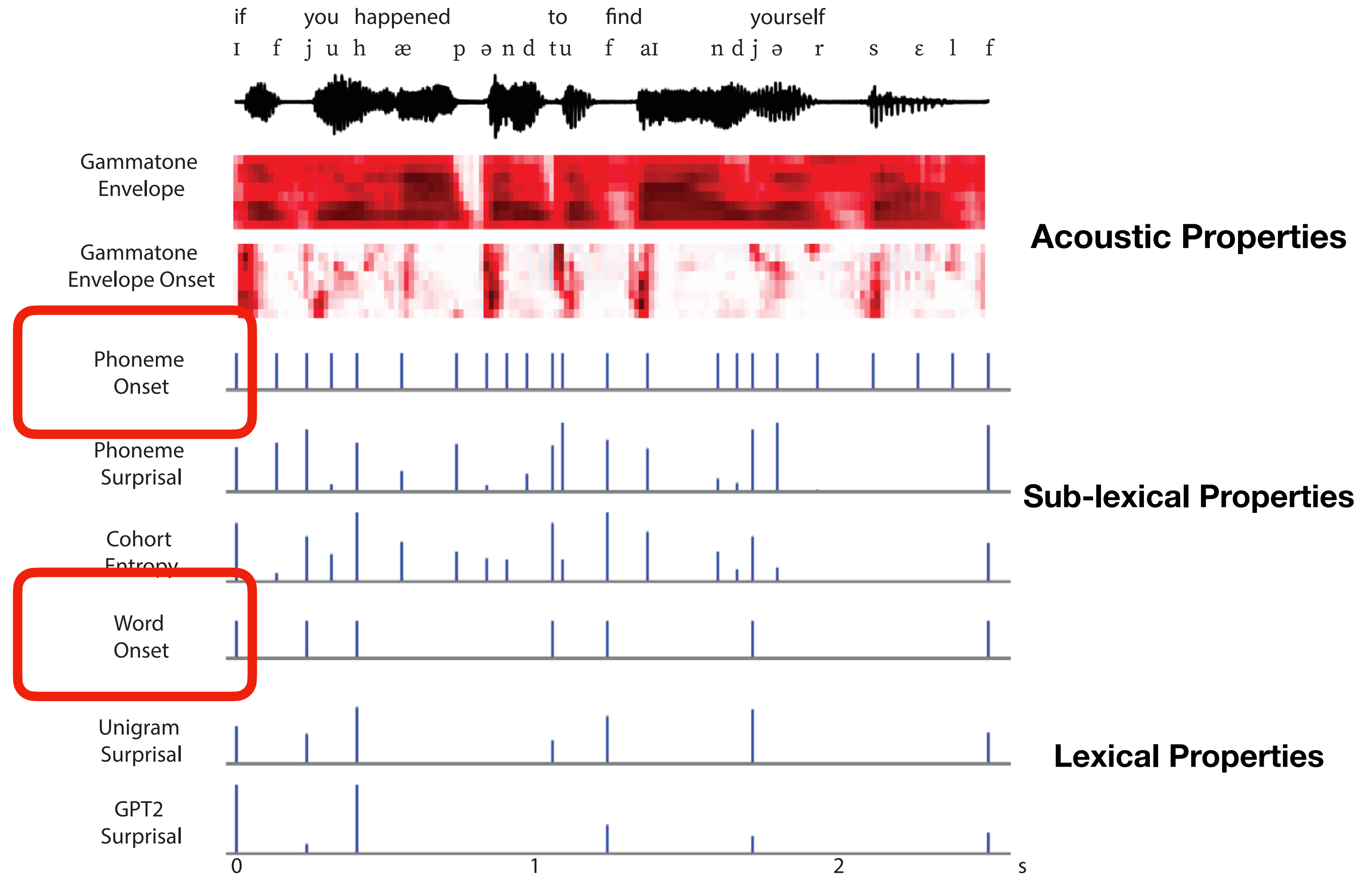 - ‣ compete to explain variance (advantage over evoked response)

large   room   rudely   constructed   of   logs

"Temporal response functions" (TRFs)

Analysis - Temporal Response Functions

- Predicting neural response from speech

- Why TRFs?
  - ‣ TRFs analogues to evoked response
  - ‣ Peak amplitude ≈ processing intensity
  - ‣ Peak Latency ≈ source location

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

Speech Representation

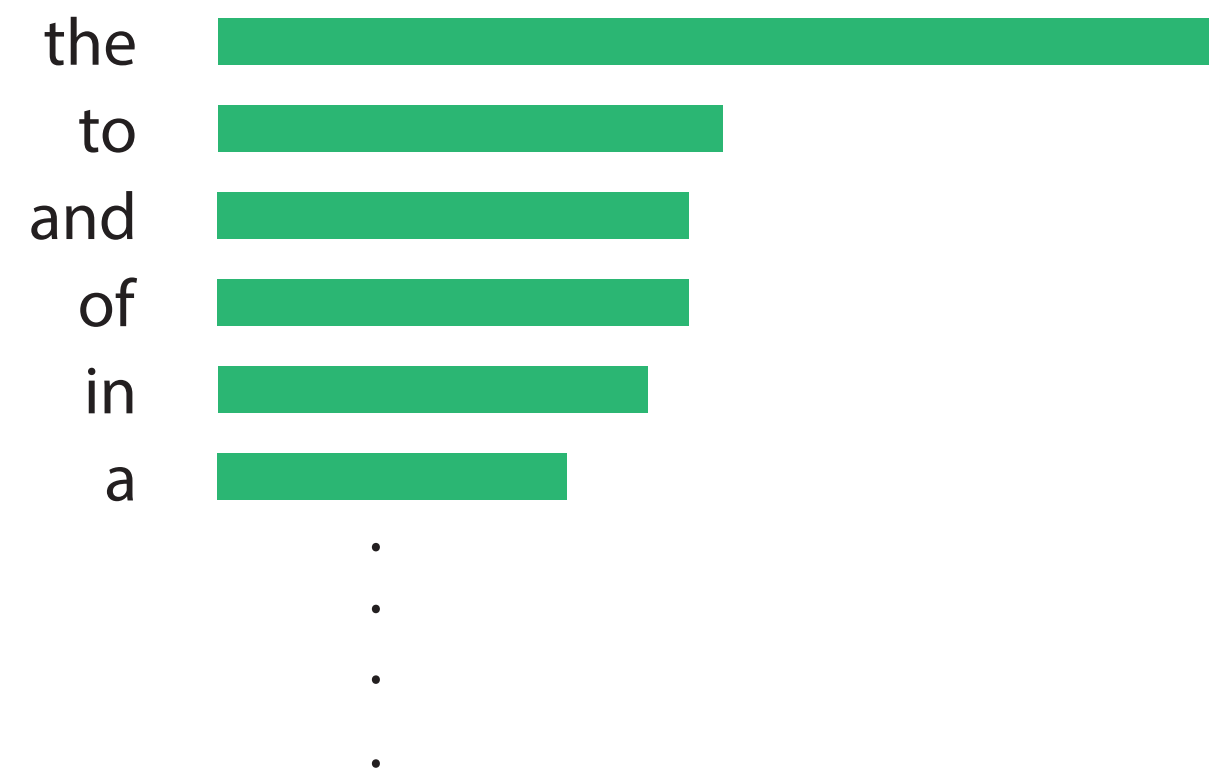if   you happened   to   find   yourself
ɪ f j u h   æ   p ə n d tu   f aɪ   n d j ə   r   s   ɛ   l   f

Gammatone Envelope

Gammatone Envelope Onset

Acoustic Properties

Phoneme Onset

Phoneme Surprisal

Cohort Entropy

Word Onset

Sub-lexical Properties

Unigram Surprisal

GPT2 Surprisal

Lexical Properties

0   1   2   s

5

Speech Representation

Speech Representation

Frequency of words based on SUBTLEX

$$Surprisal = -log(Probability)$$

if   you happened   to   find   yourself

Acoustic Properties

Sub-lexical Properties

Lexical Properties

Speech Representation

if you happened to find

yourself
a
out
the
it
that
one
your
.
.
.

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   ɛ l f

Gammatone
Envelope

Gammatone
Envelope Onset

Acoustic Properties

Phoneme
Onset

Phoneme
Surprisal

Cohort
Entropy

Word
Onset

Sub-lexical Properties

Unigram
Surprisal

Lexical Properties

GPT2
Surprisal

0   1   2   s

5
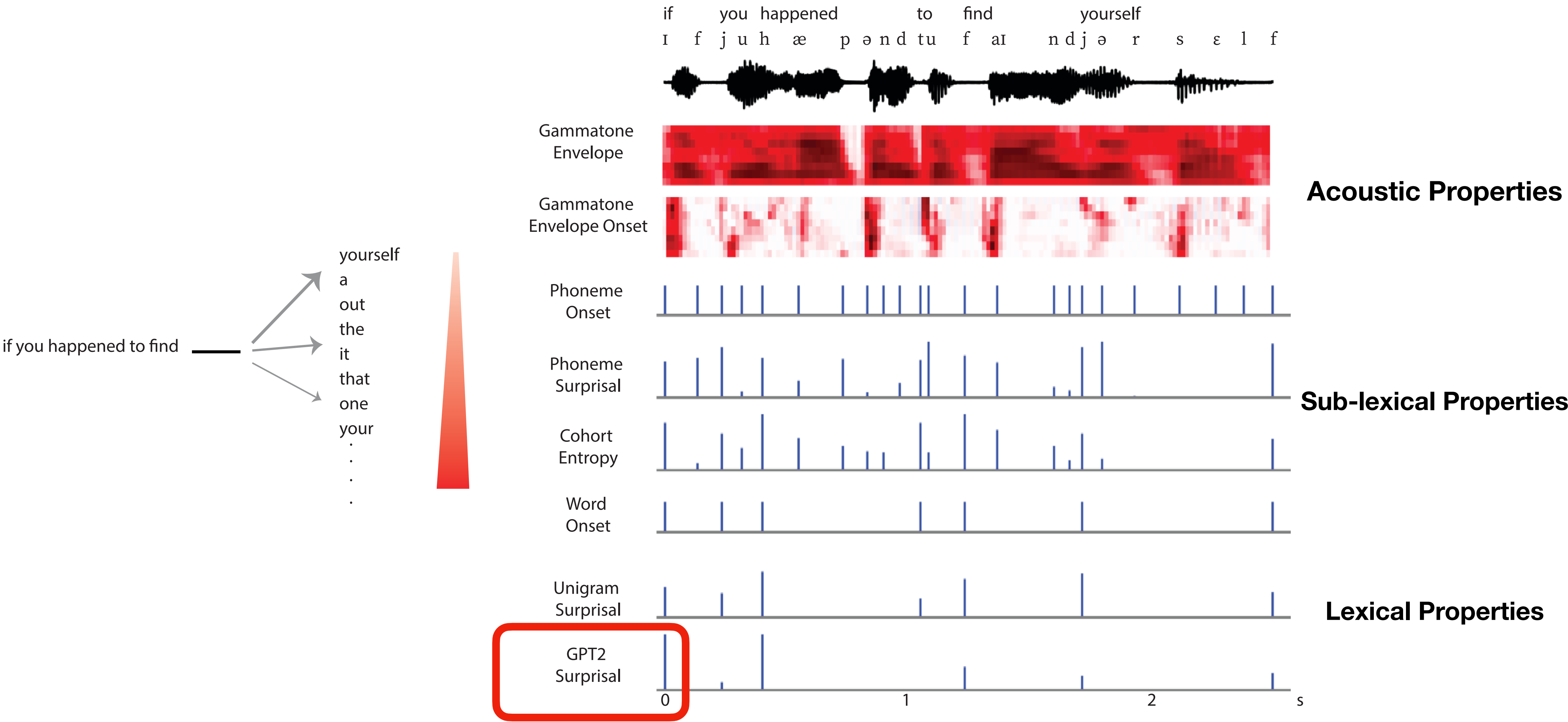
Speech Representation
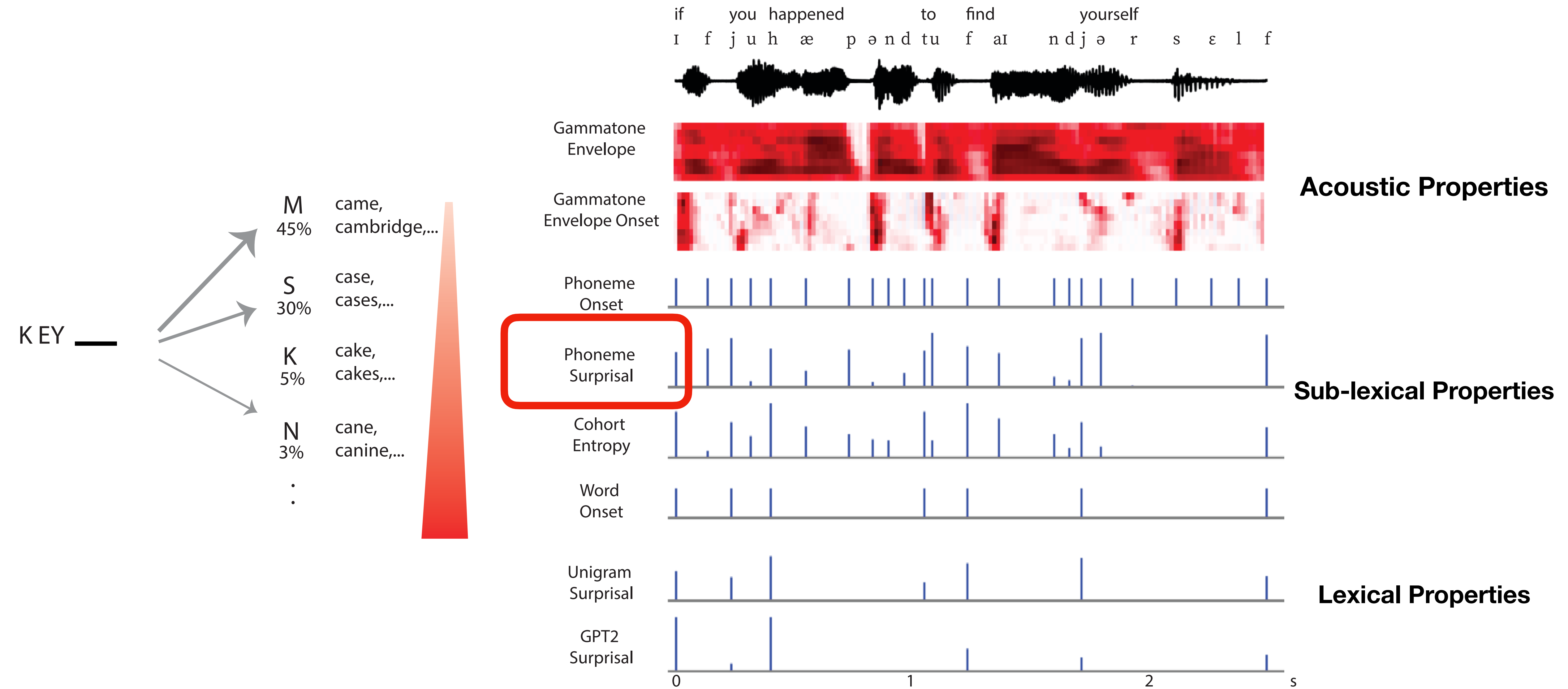
Speech Representation
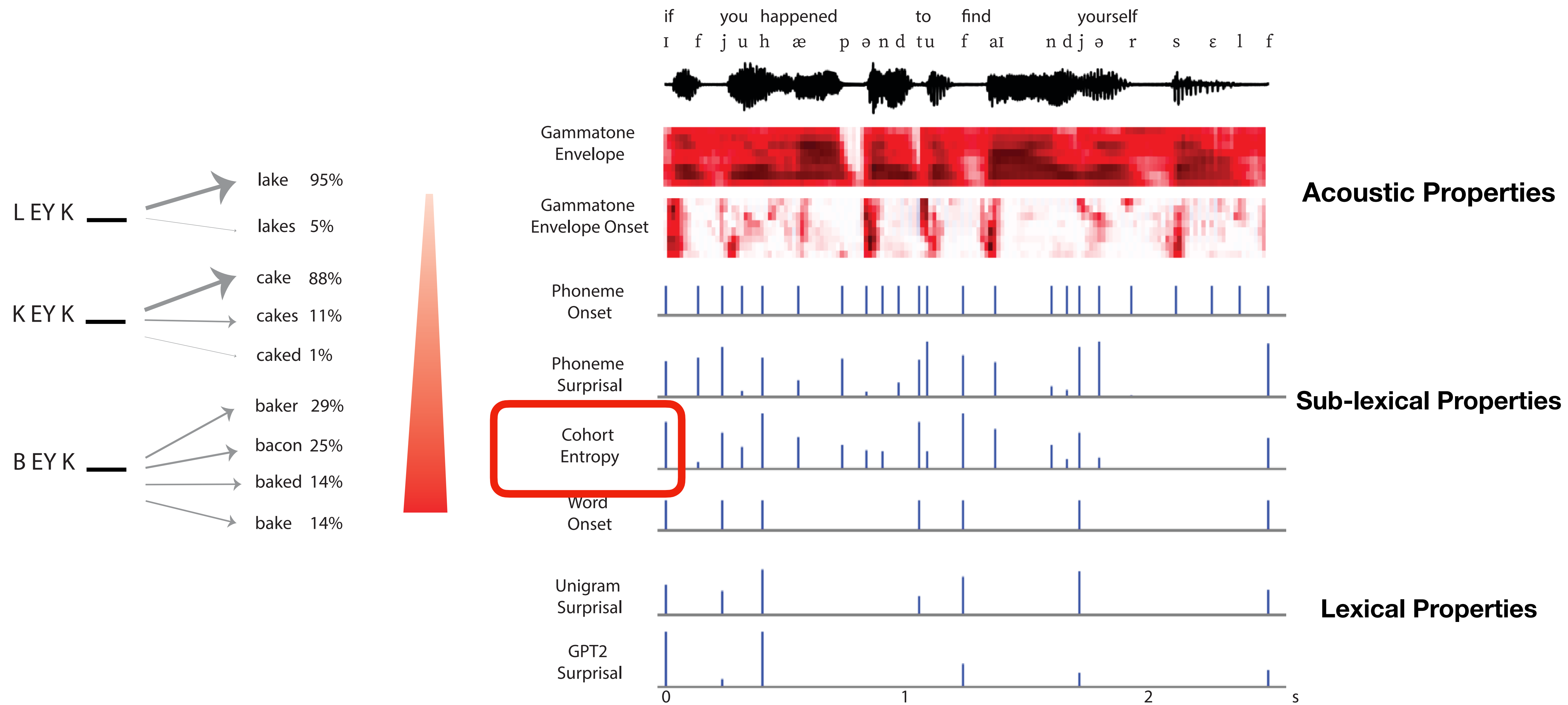
Neural Prediction Results

Emergence of neural features as the incremental processing occurs
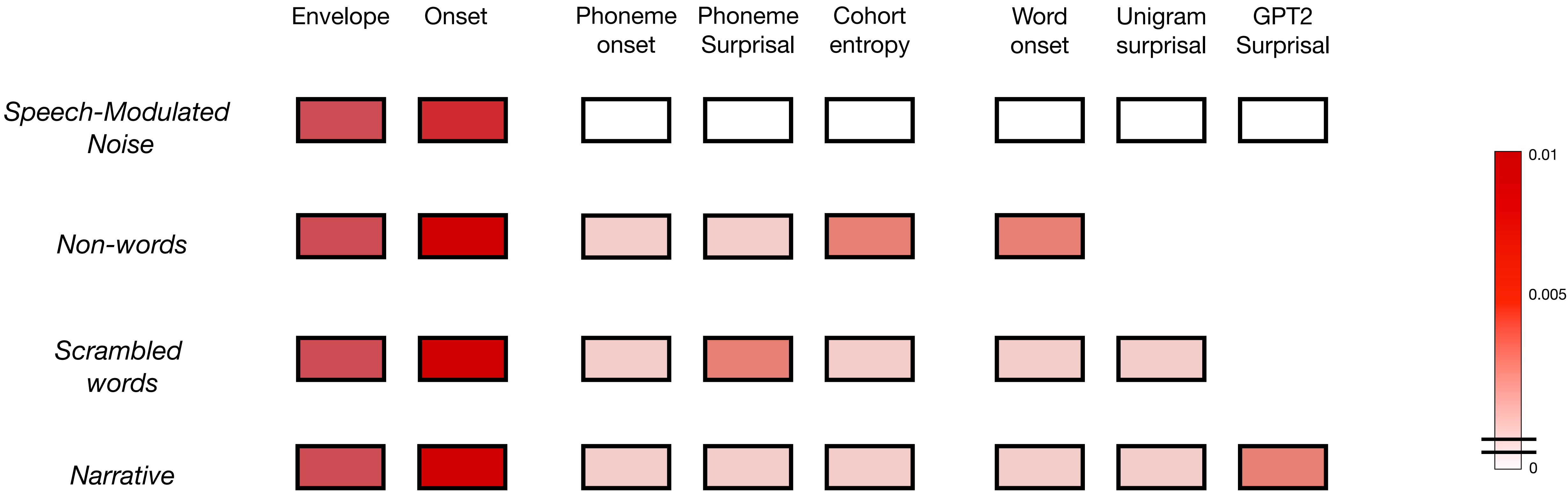
- Acoustic features are encoded for both non-speech and speech stimuli

- (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 unigram surprisal

6

Hemispheric Lateralization Results

Speech feature processing lateralization

Note: Lateralization may be task dependent!

Acoustic TRF Results

**Gammatone
Envelope Onset**

**Gammatone
Envelope**
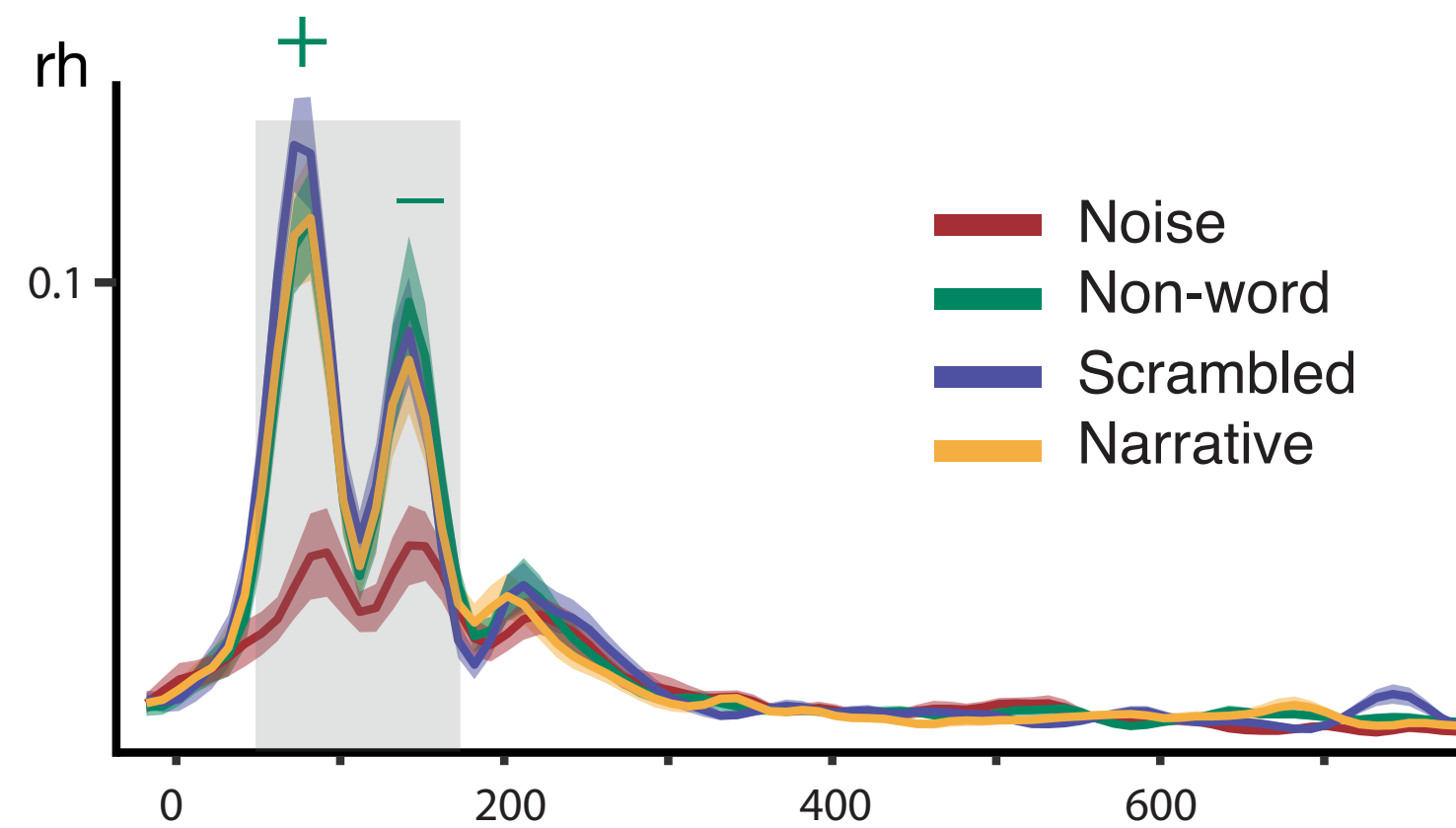
- Speech responses > Noise response
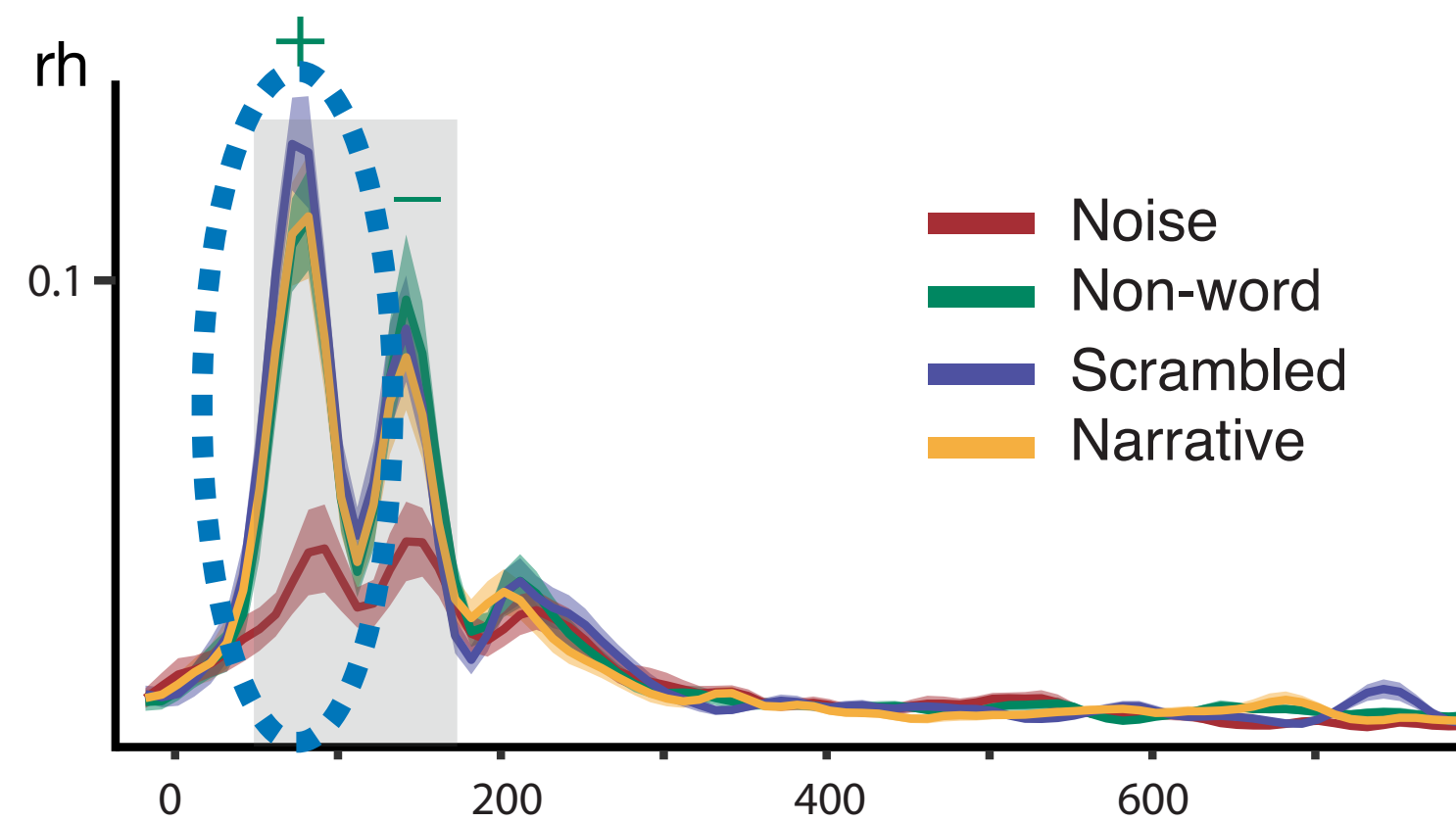
- Responses similar for all speech conditions
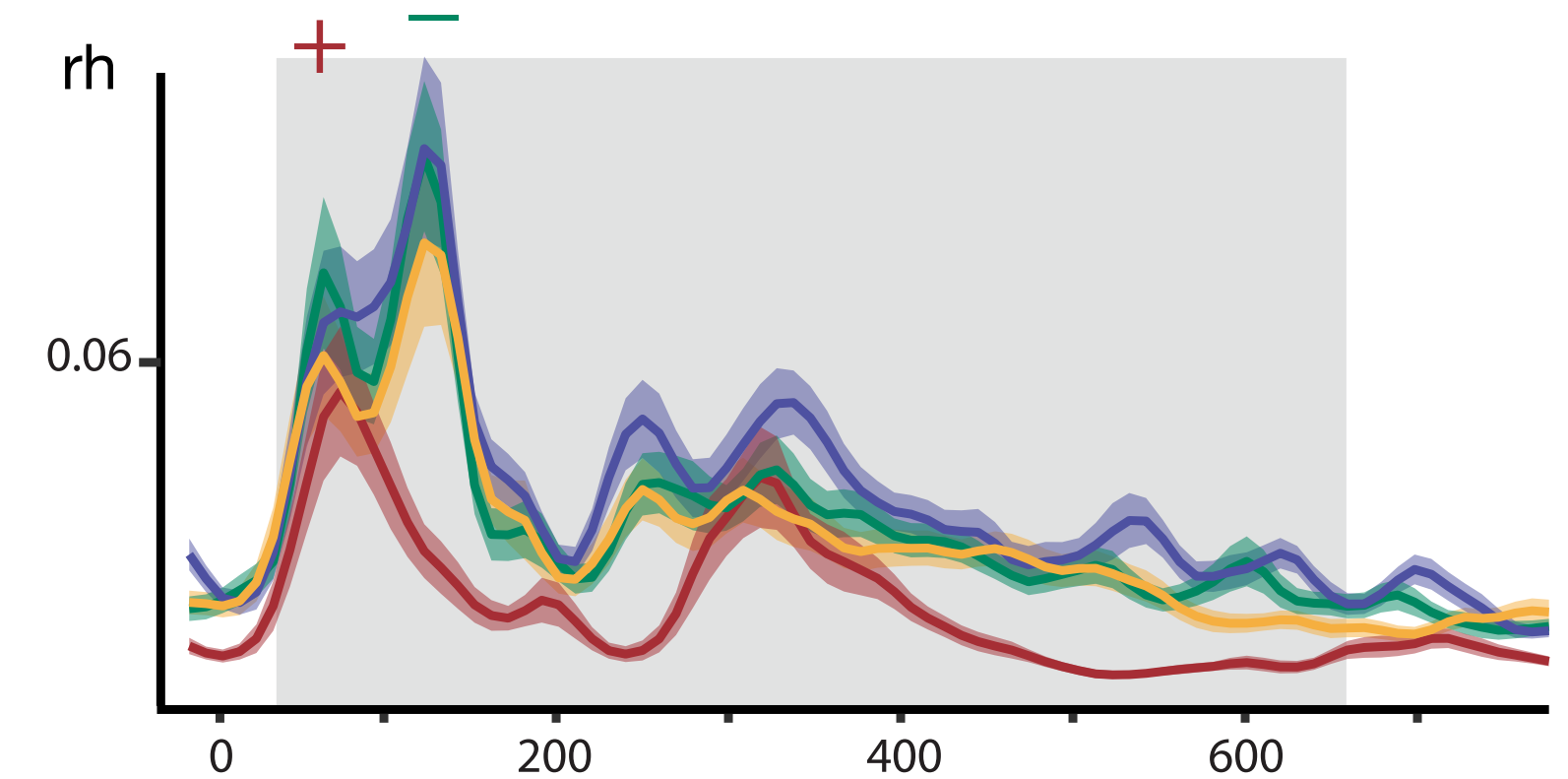
- Speech responses > Noise response

- Narrative response < scrambled words

- Non words similar to scrambled words

- Noise response lacks second peak ~120 ms

Only right hemisphere shown,
condition based differences similar in left

Acoustic TRF Results

Gammatone Envelope Onset

Gammatone Envelope

Bottom-up

Legend:
- Noise
- Non-word
- Scrambled
- Narrative

Acoustic

Speech Stimuli

60

0

time (ms)

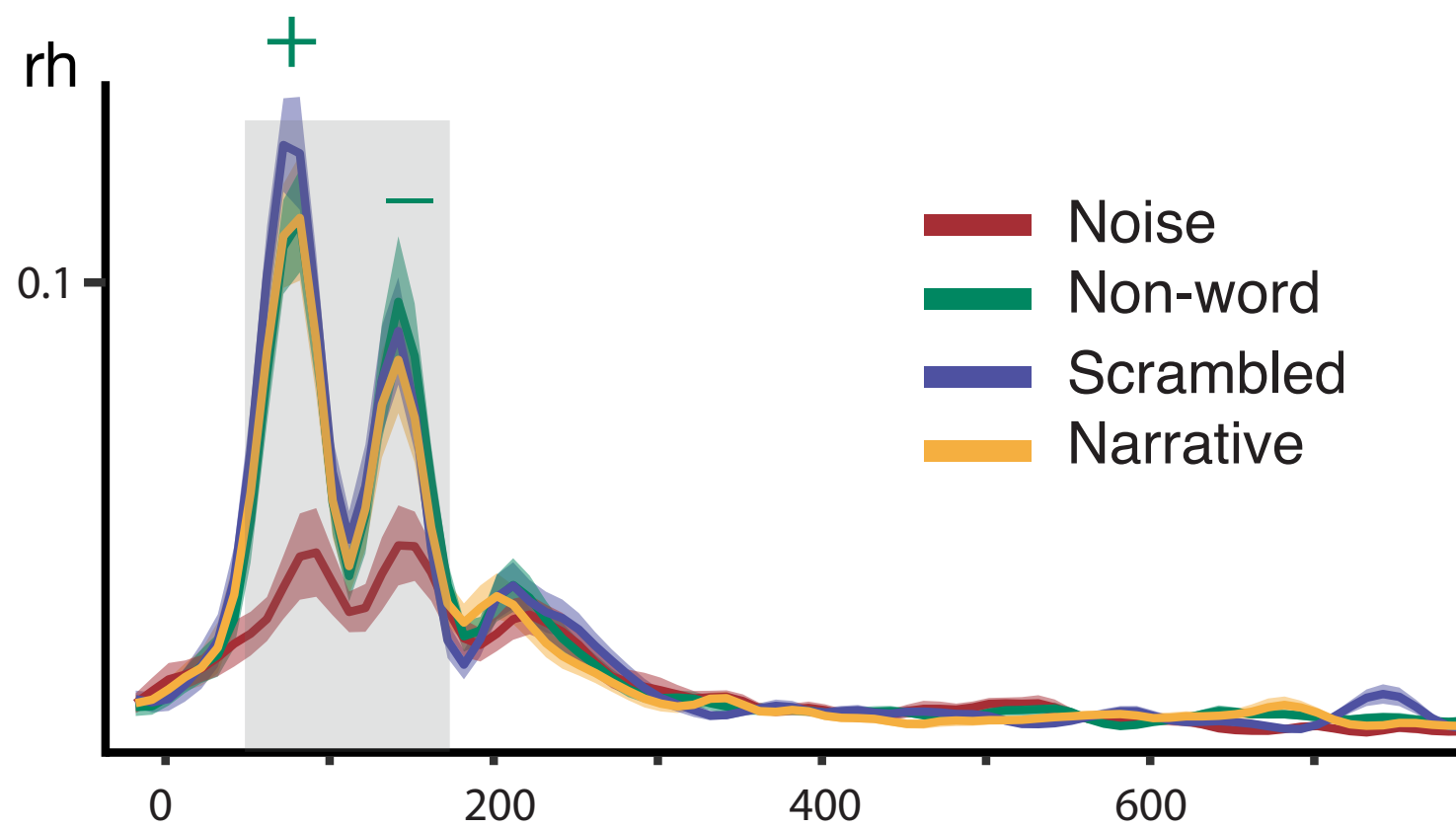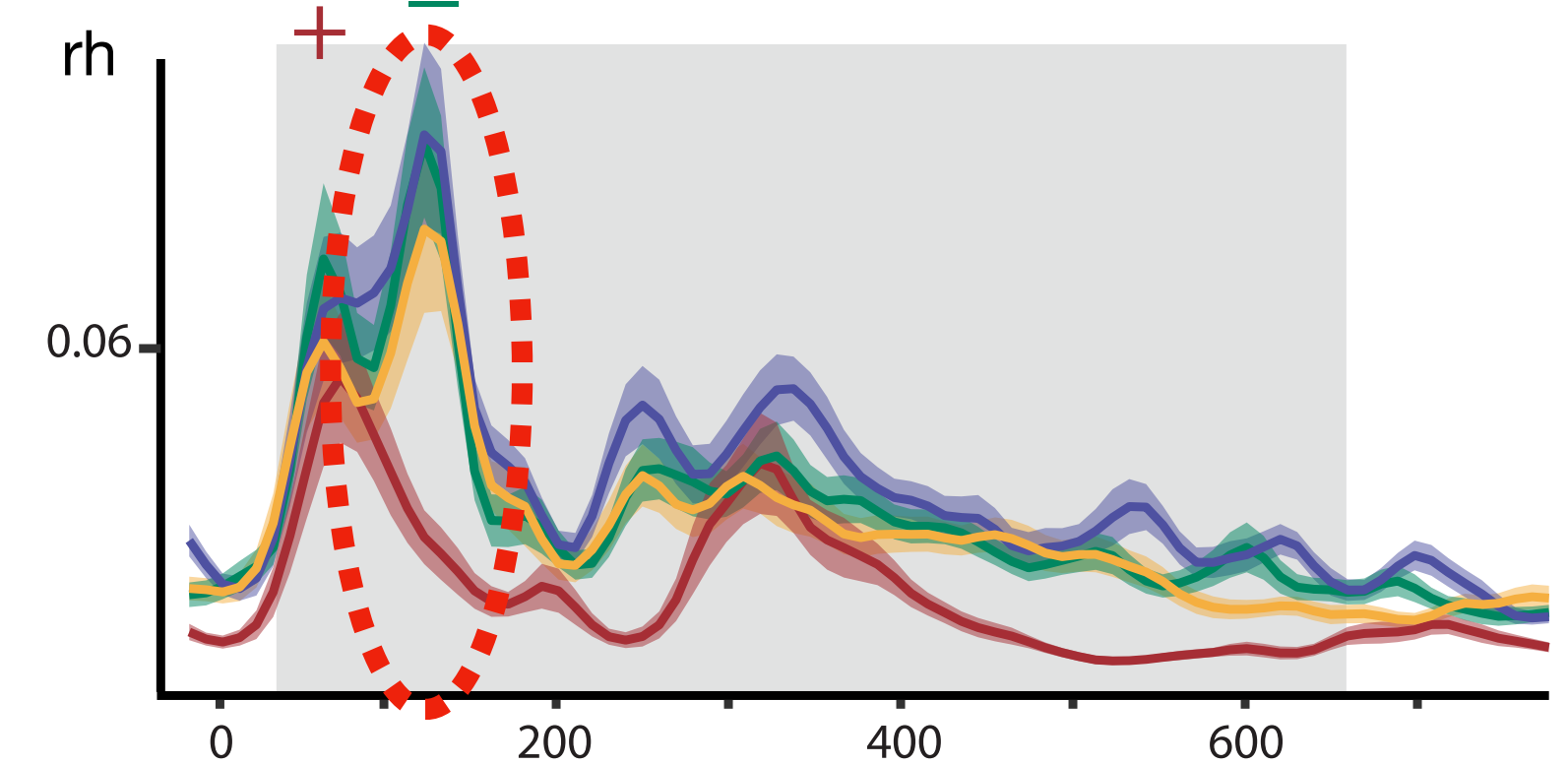- Speech response > Noise response
- Responses similar for all speech conditions

- Speech responses > Noise response
- Narrative response < scrambled words
- Non words similar to scrambled words
- Noise response lacks second peak ~120 ms

8

Acoustic TRF Results

Gammatone Envelope Onset

Gammatone Envelope

Top-down

Bottom-up

High level meaning

Acoustic

Speech Stimuli

120

60

0

0

time (ms)

time (ms)

rh

0.1

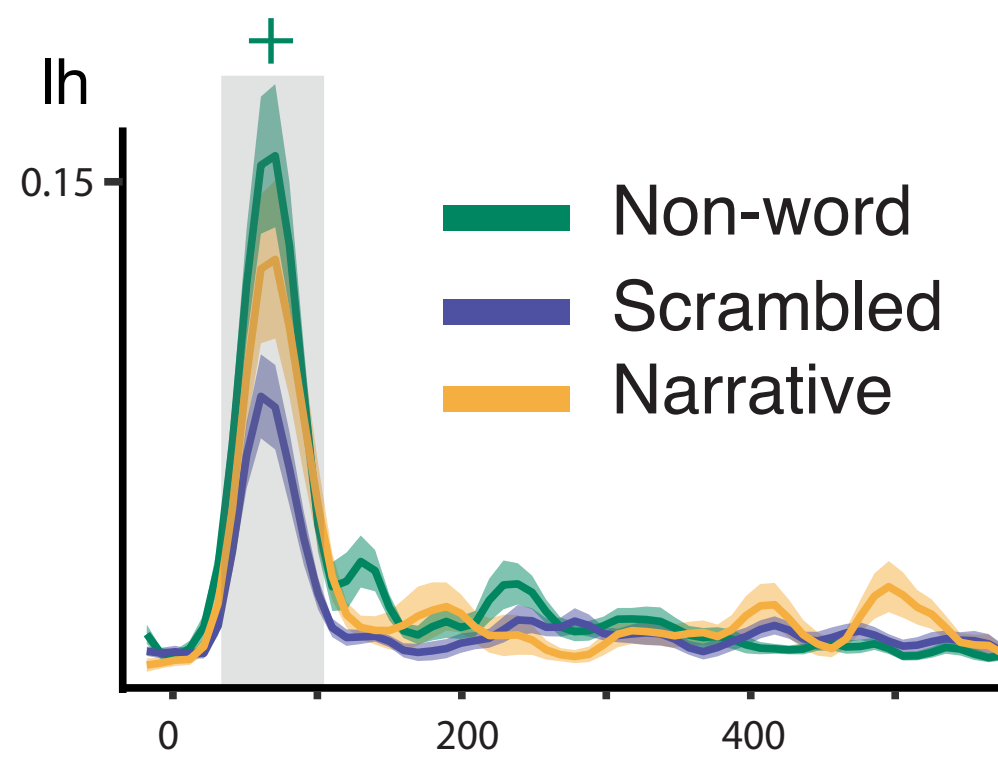Noise
Non-word
Scrambled
Narrative

rh

0.06

- Speech response > Noise response
- Responses similar for all speech conditions

- Speech responses > Noise response
- Narrative response < scrambled words
- Non words similar to scrambled words
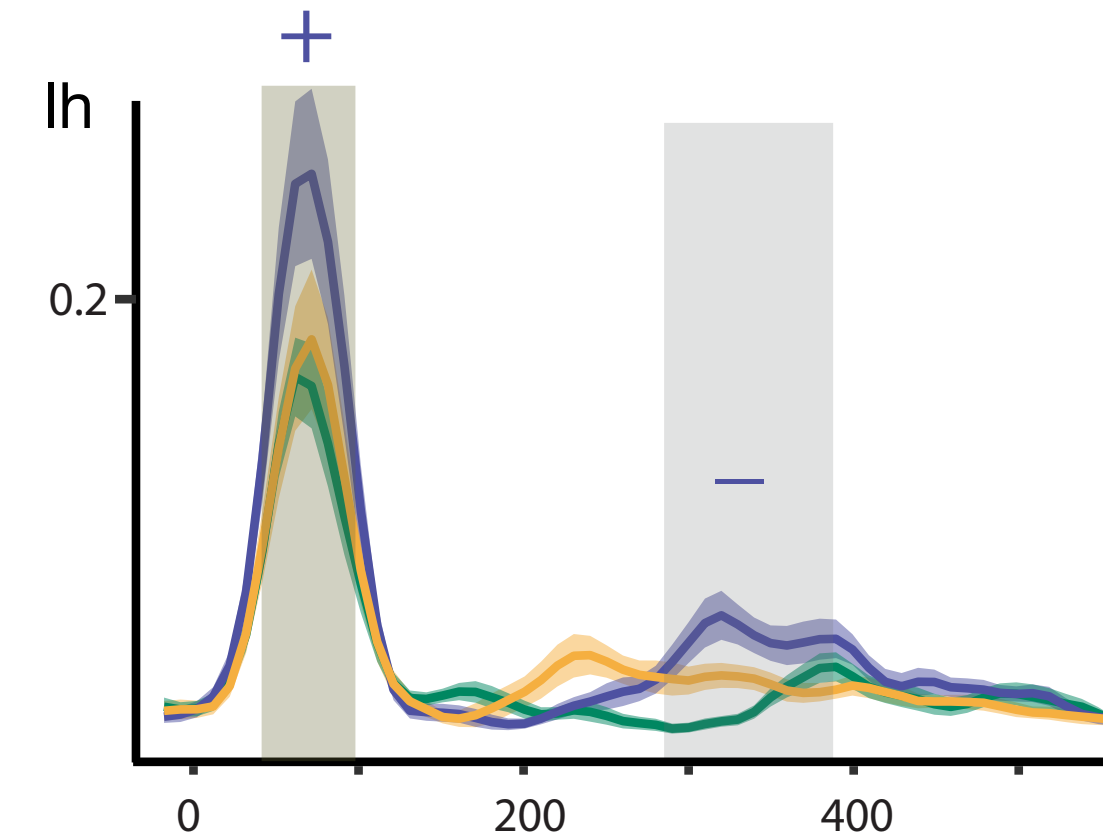- Noise response lacks second peak ~120 ms

Phonemic TRF Results

Phoneme Onset

Phoneme Surprisal

Cohort Entropy

- Non-words response largest

- No later processing

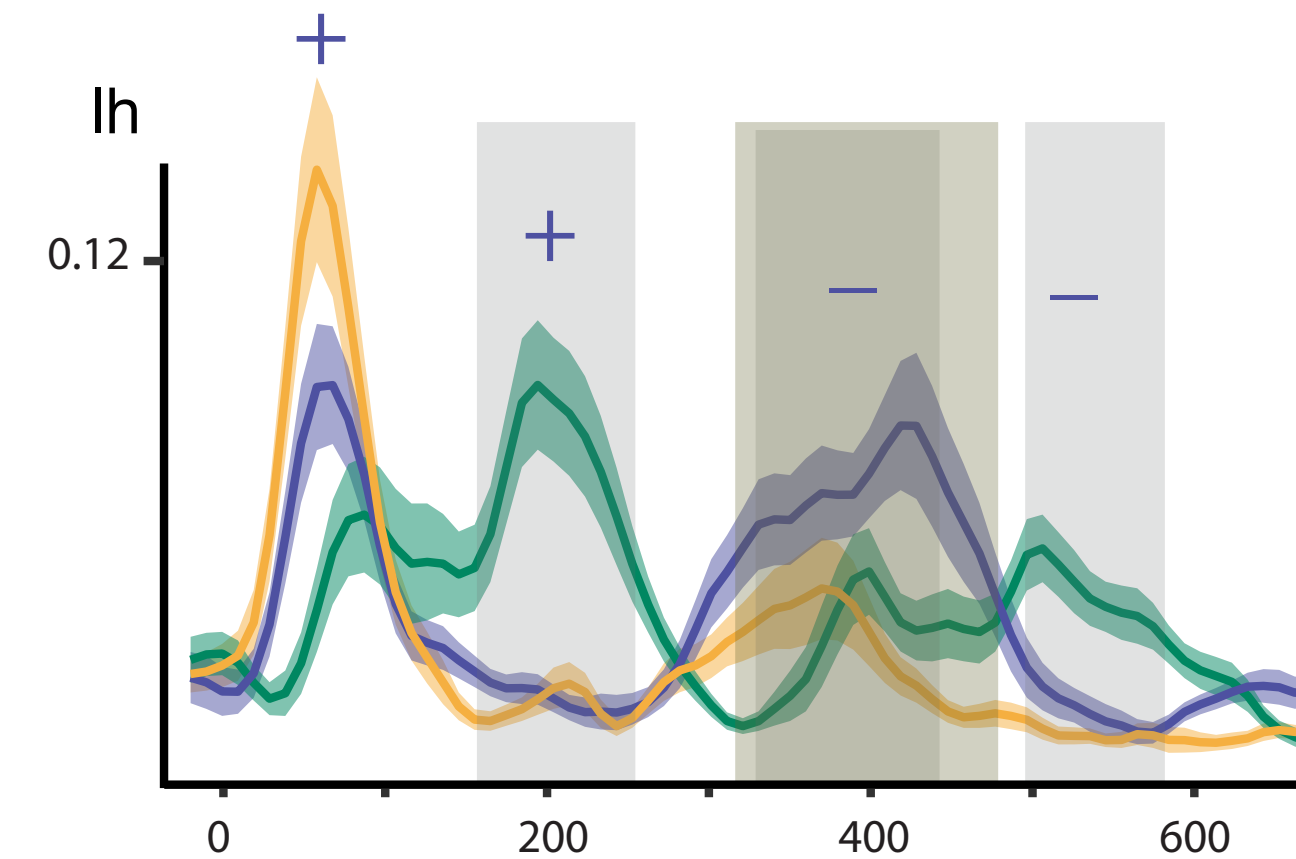Early phoneme processing (~85 ms)

- scrambled > narrative

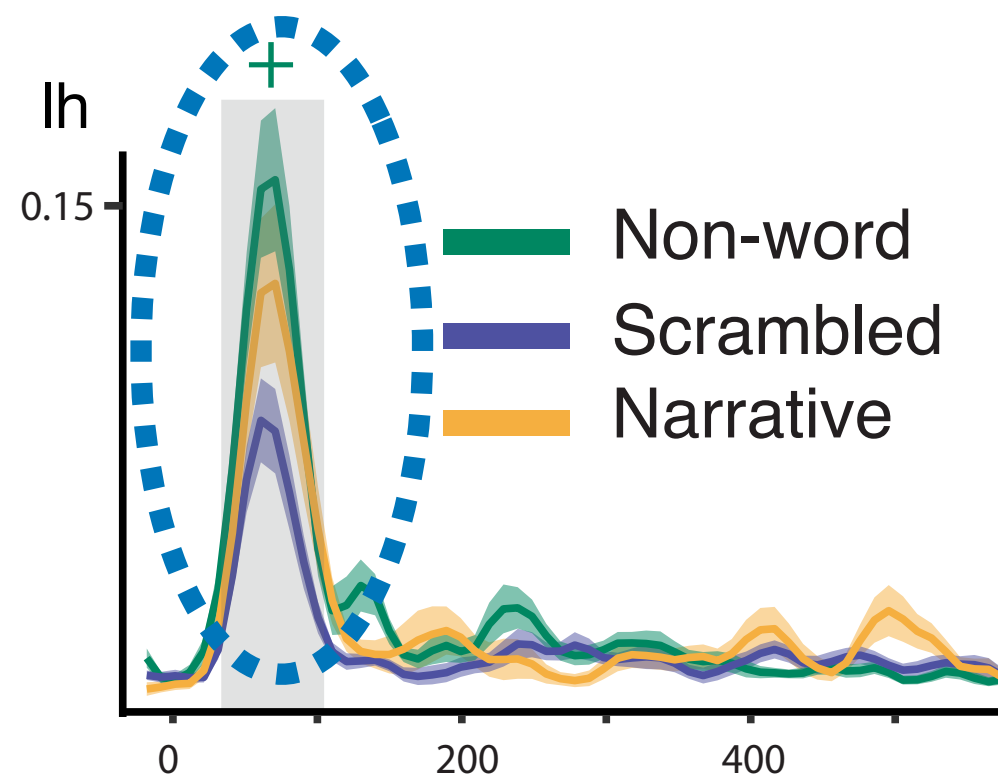Late phoneme processing (~350 ms)

- Words > non-words

Late context processing

- N400-like response (reduced for narrative)

- Additional/delayed peaks in non-words (difference in stimulus distributions)

Only left hemisphere shown,
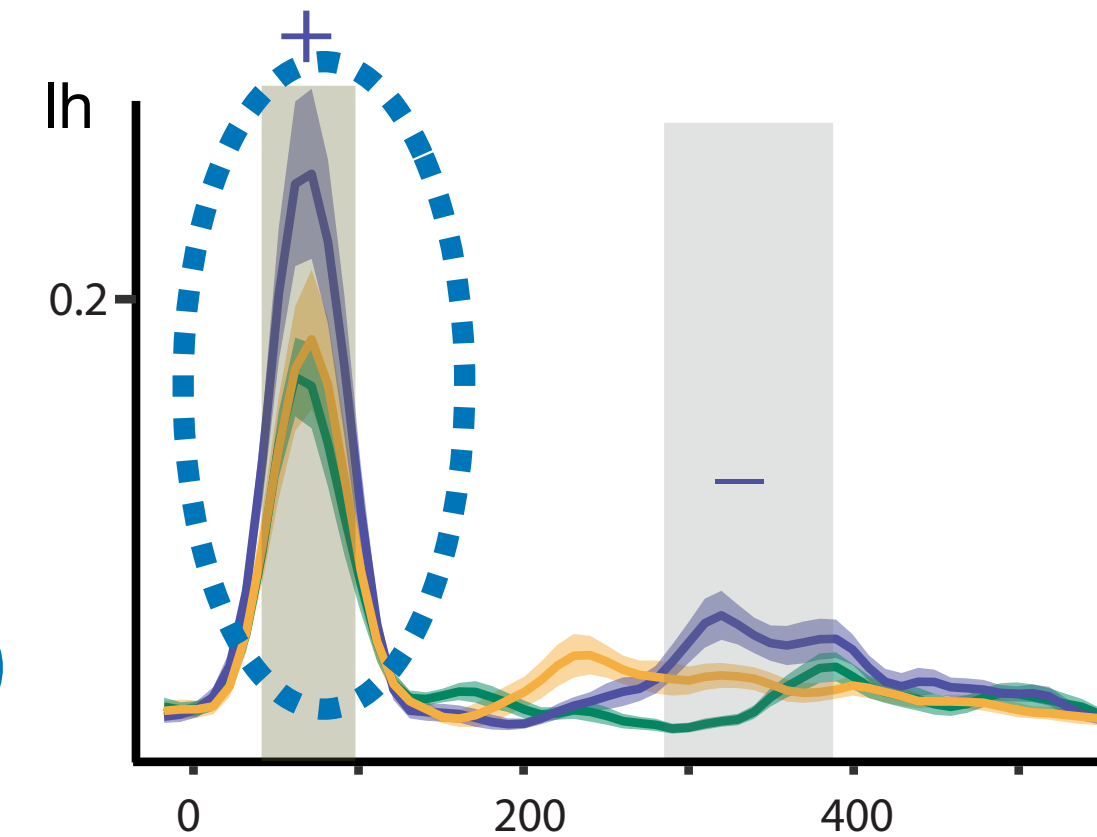condition based differences disappear in right
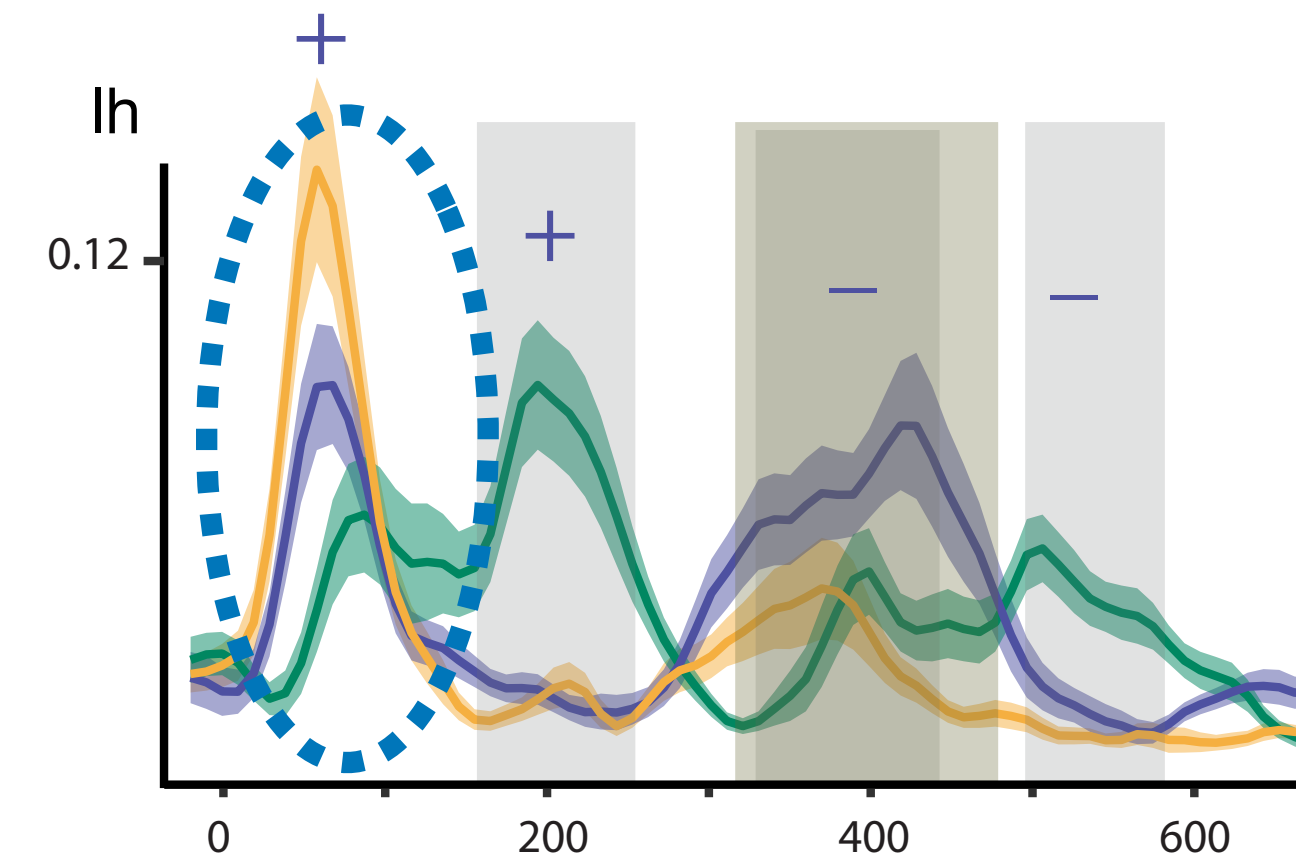
9

Phonemic TRF Results

Phoneme Onset

lh

0.15

- Non-word
- Scrambled
- Narrative

0   200   400

Top-down

Bottom-up

Sub-Lexical

Acoustic

Speech
Stimuli

120

80

60

0

time (ms)

time (ms)

- Non-words response largest
- No later processing

Phoneme Surprisal

lh

0.2

0   200   400

Early phoneme processing (~85 ms)

- scrambled > narrative

Late phoneme processing (~350 ms)

- Words > non-words

Cohort Entropy

lh

0.12

0   200   400   600

Late context processing

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

Phonemic TRF Results

Phoneme Onset

- Non-words response largest

- No later processing

Top-down

Bottom-up

High level meaning

350

120

0

Sub-Lexical

Acoustic

80

60

0

Speech Stimuli
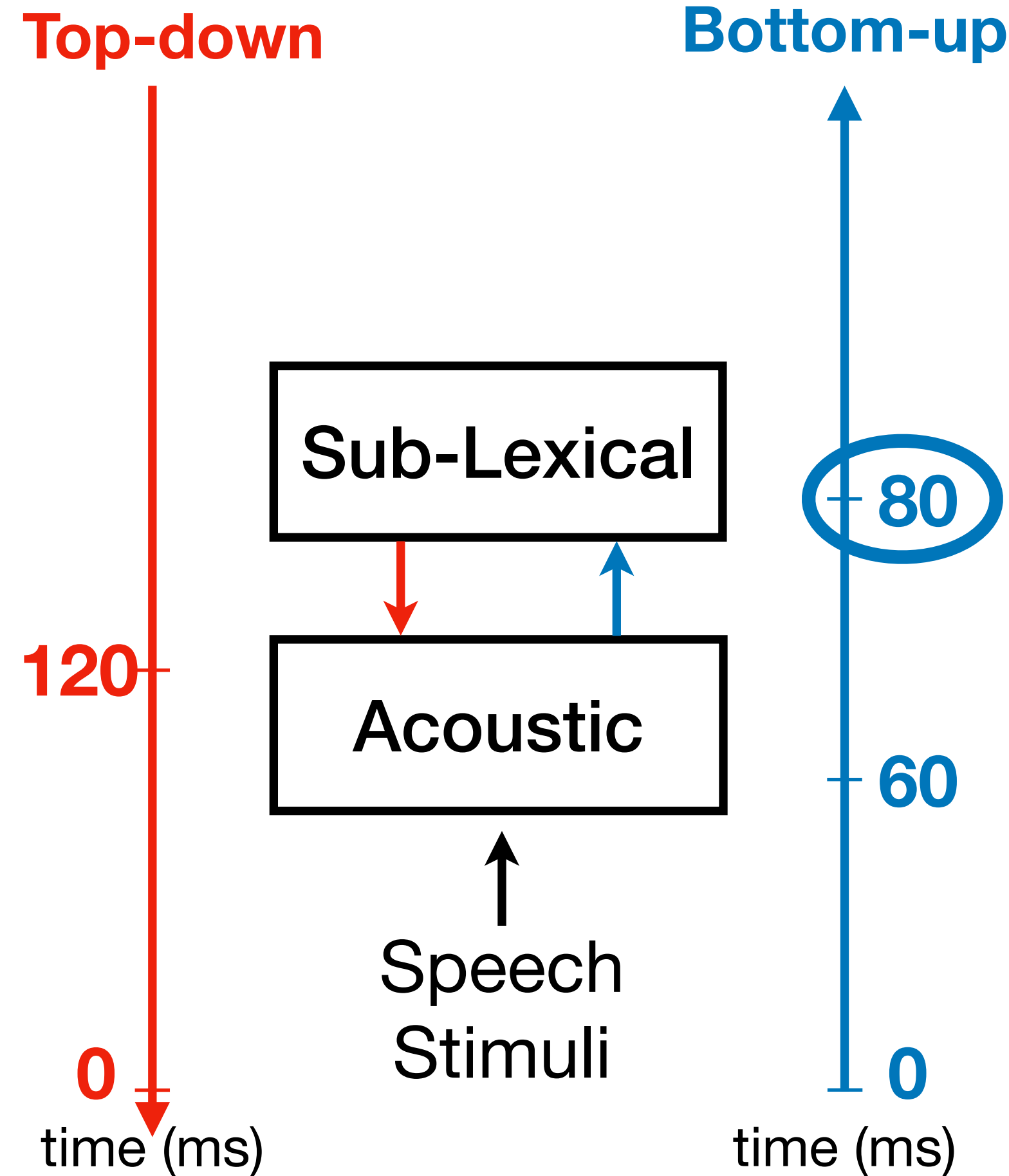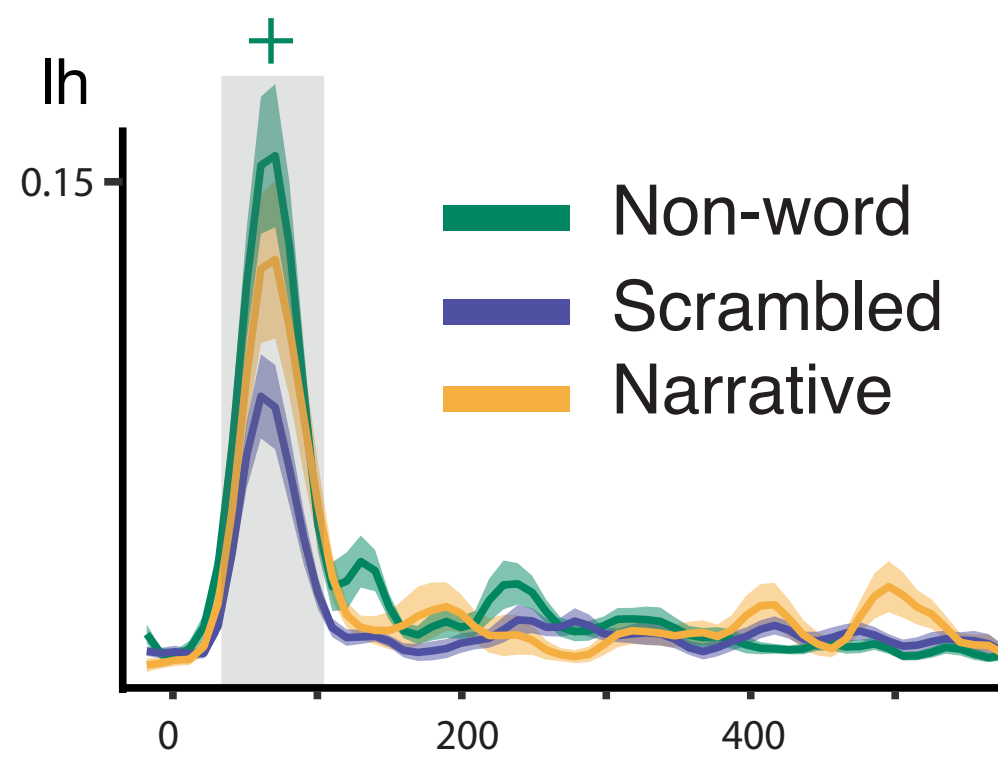
time (ms)   time (ms)

Phoneme Surprisal

Early phoneme processing (~85 ms)

- scrambled > narrative

Late phoneme processing (~350 ms)

- Words > non-words

Cohort Entropy

Late context processing

- N400-like response (reduced for narrative)

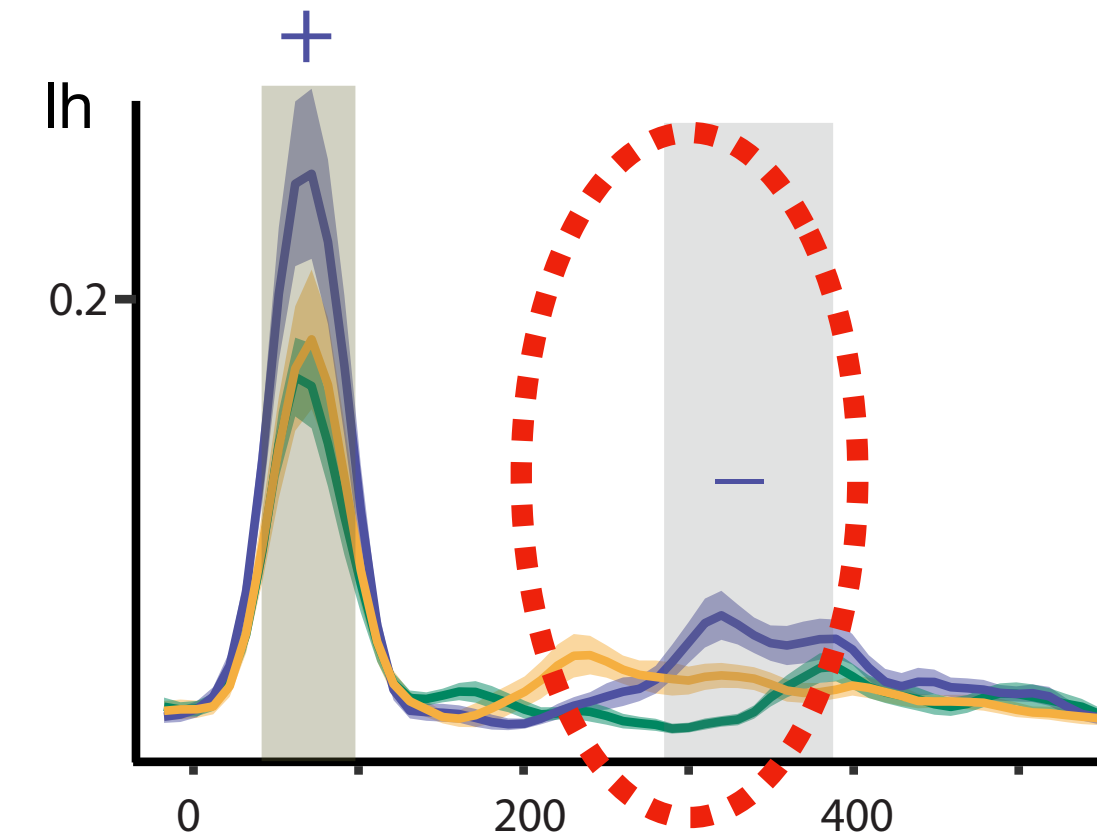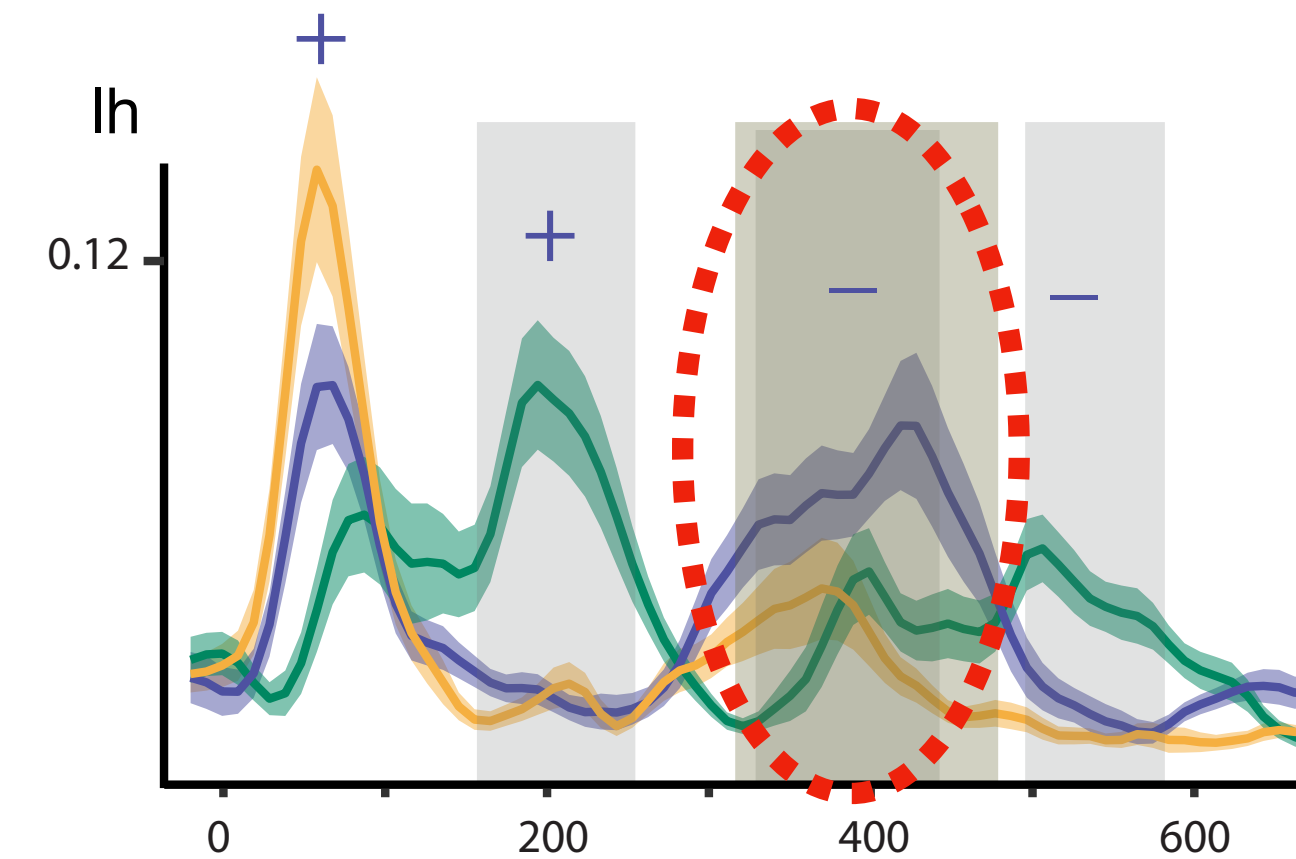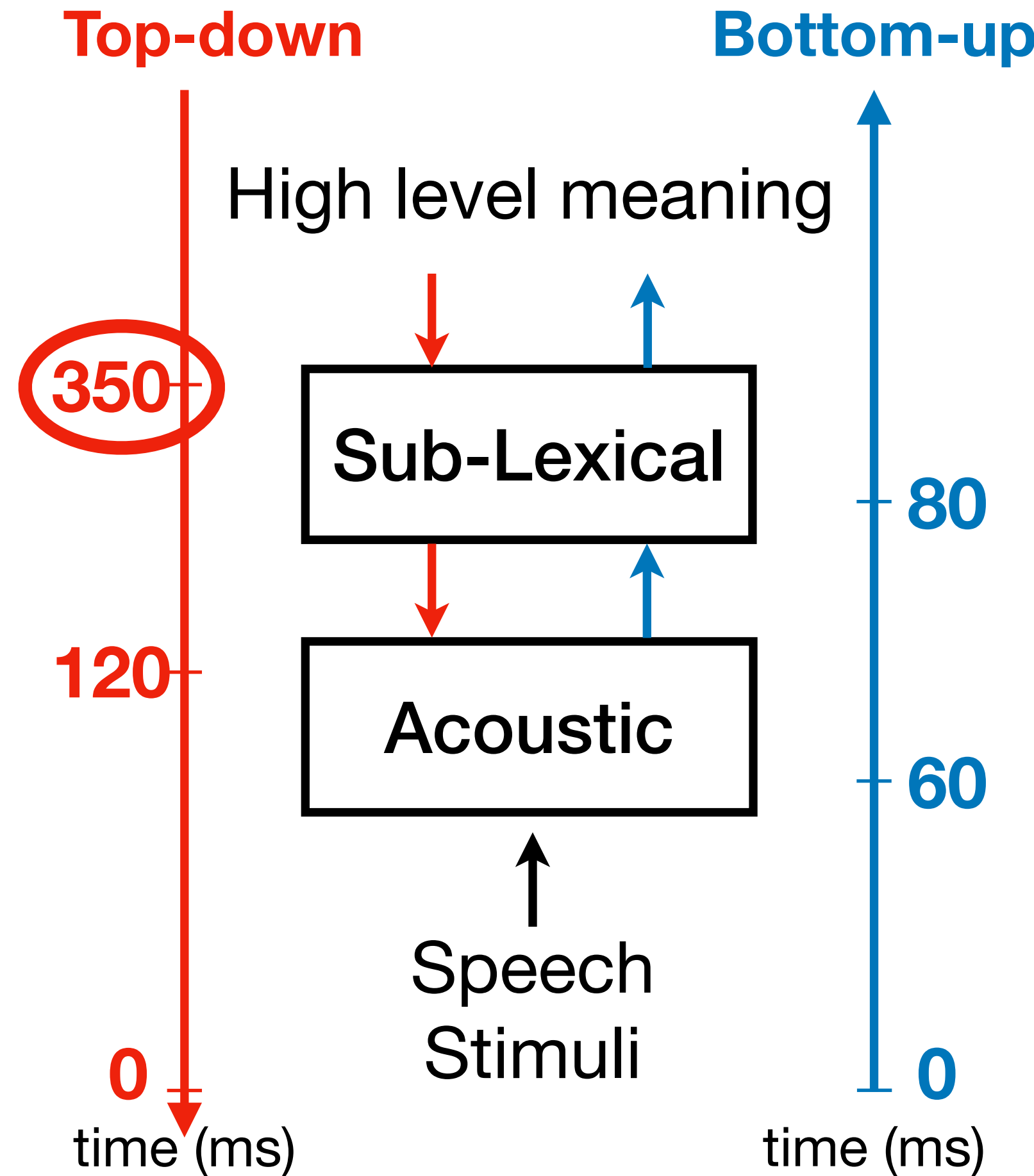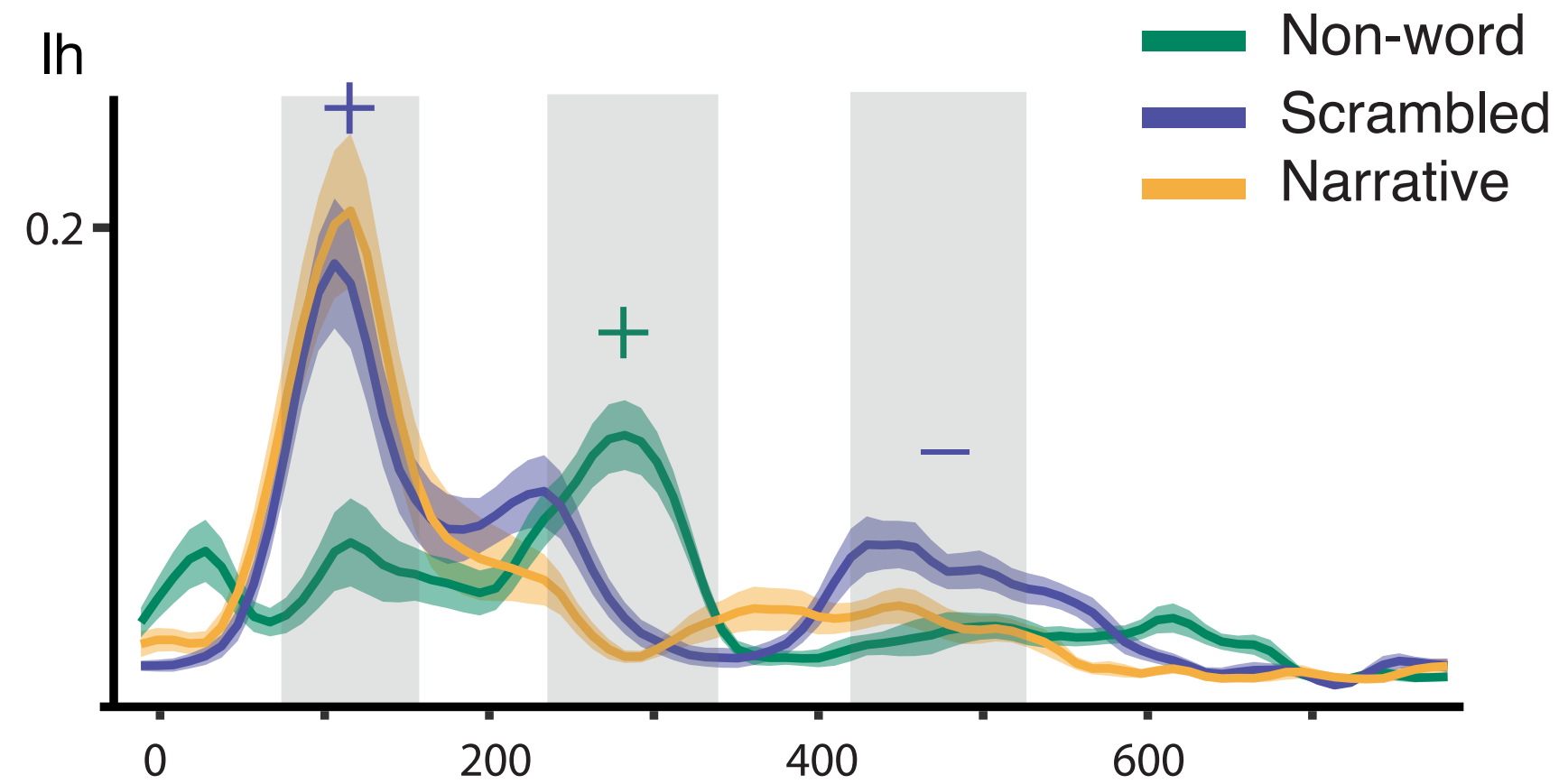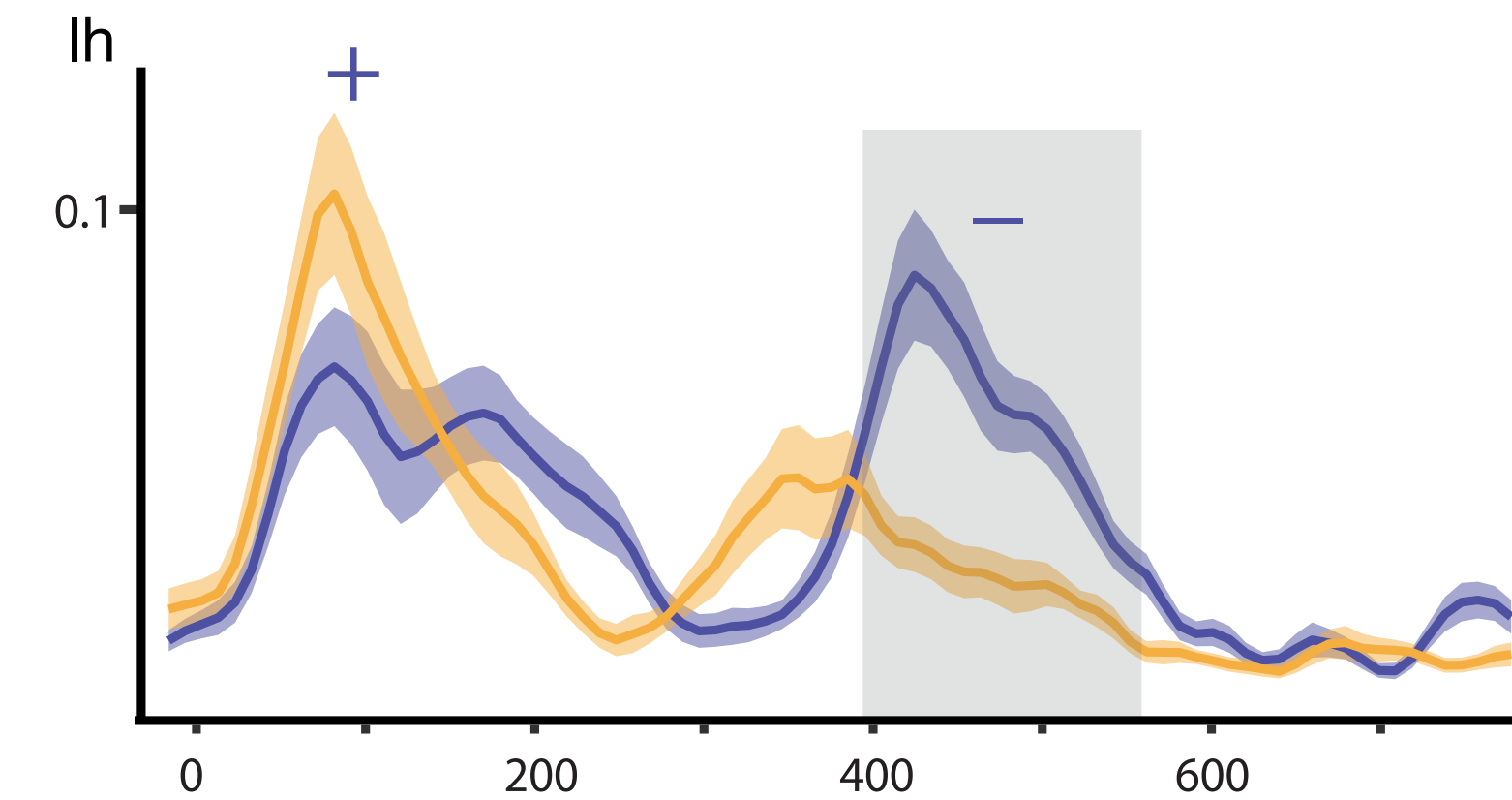- Additional/delayed peaks in non-words (difference in stimulus distributions)

Word-based TRF Results

Word Onset

Non-word
Scrambled
Narrative

Unigram surprisal

- Scrambled words response > narrative response 450 ms

- Different neural mechanisms for non-words (searching for mental lexicon?)

- words: Left hemisphere > Right

- non-words: Left hemisphere ≈ Right

- N400-like response

- Reduction in surprisal when more context

- Left hemisphere > Right

- no condition based difference in Right hemisphere

Only left hemisphere shown

Word-based TRF Results

Word Onset

Legend:
- Non-word (green)
- Scrambled (purple)
- Narrative (orange)

Top-down (red axis, time (ms)): 350, 120, 0

Bottom-up (blue axis, time (ms)): 100, 80, 60, 0

Boxes: Lexical → Sub-Lexical → Acoustic ← Speech Stimuli

Unigram surprisal

- Scrambled words response > narrative response 450 ms

- Different neural mechanisms for non-words (searching for mental lexicon?)

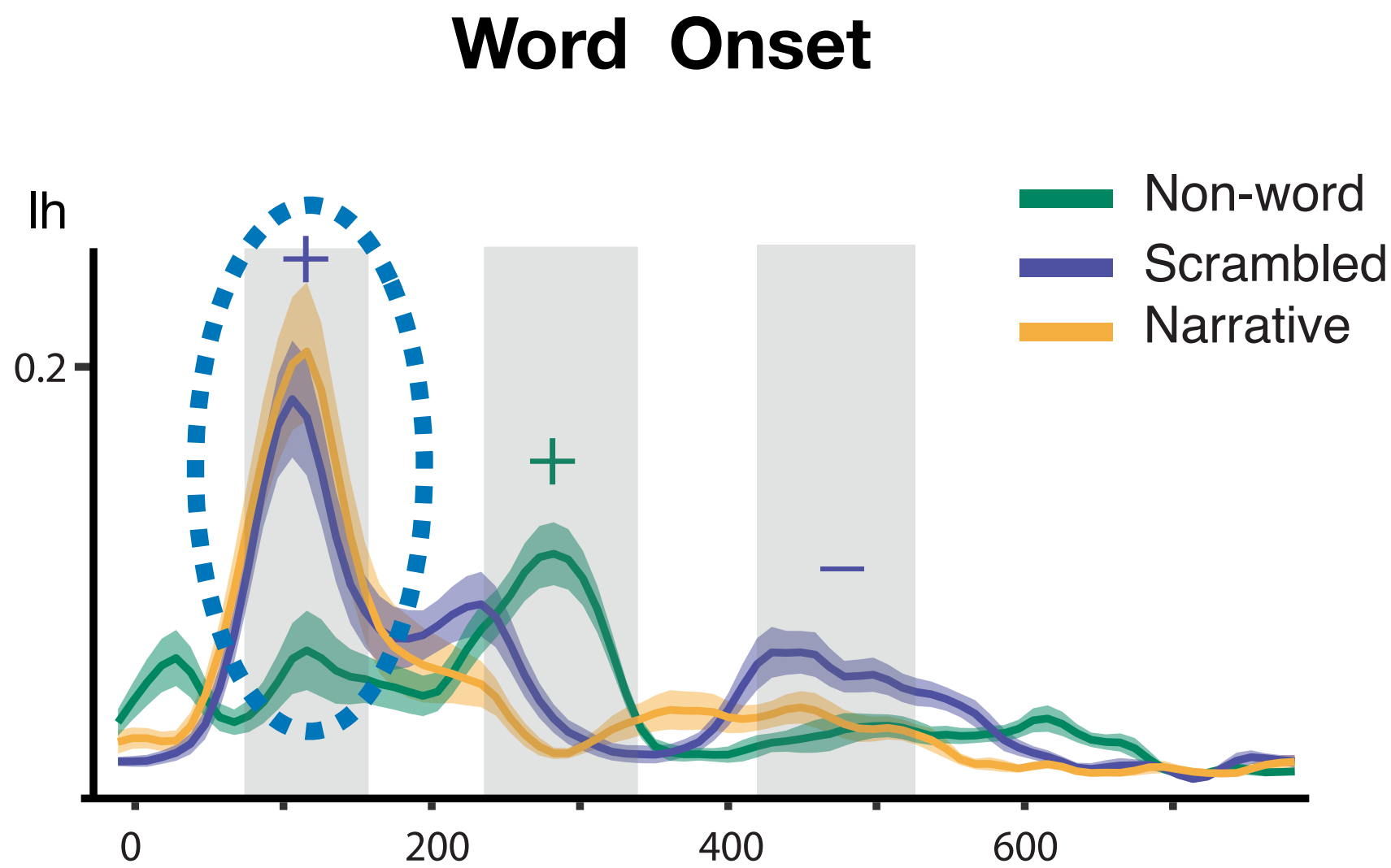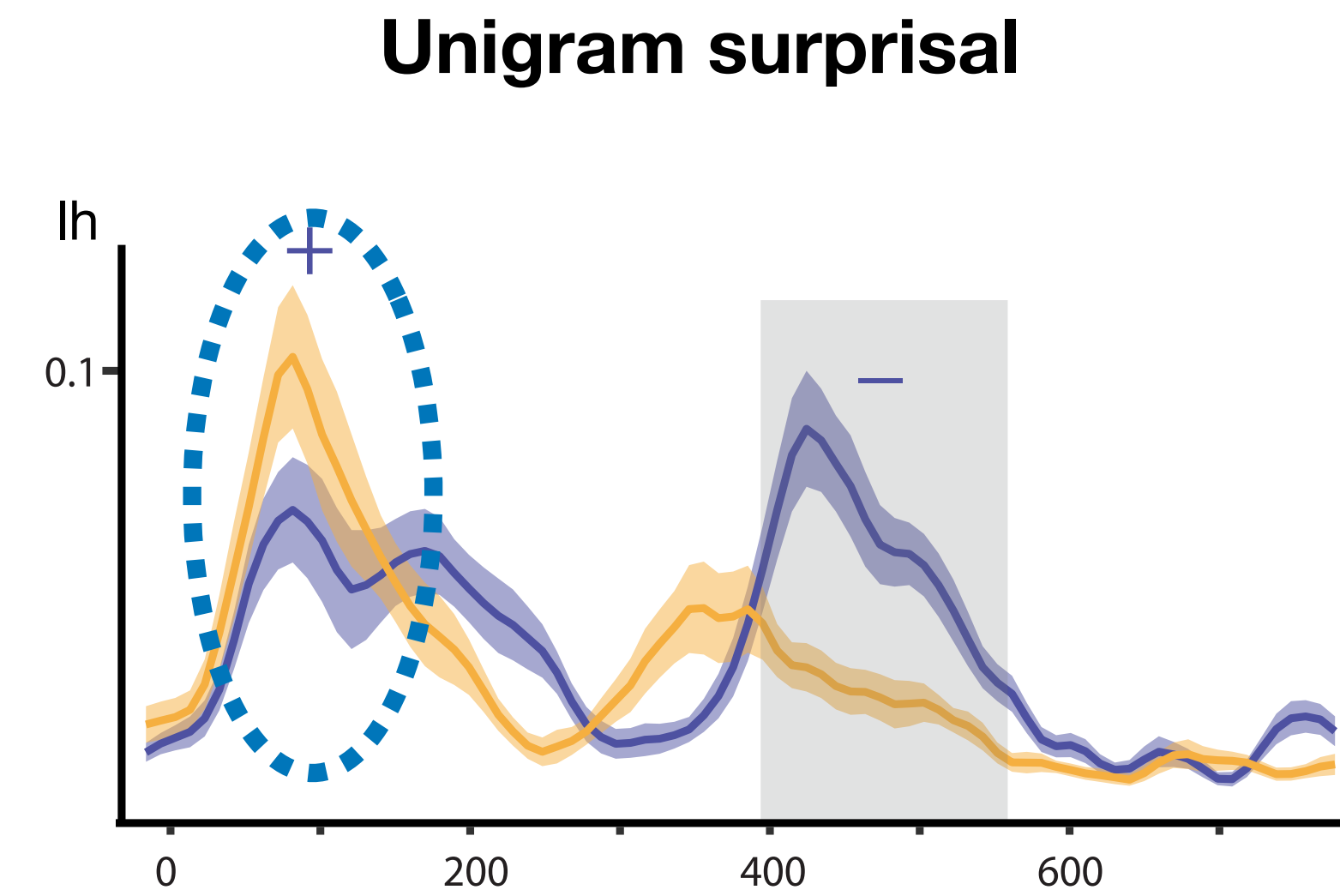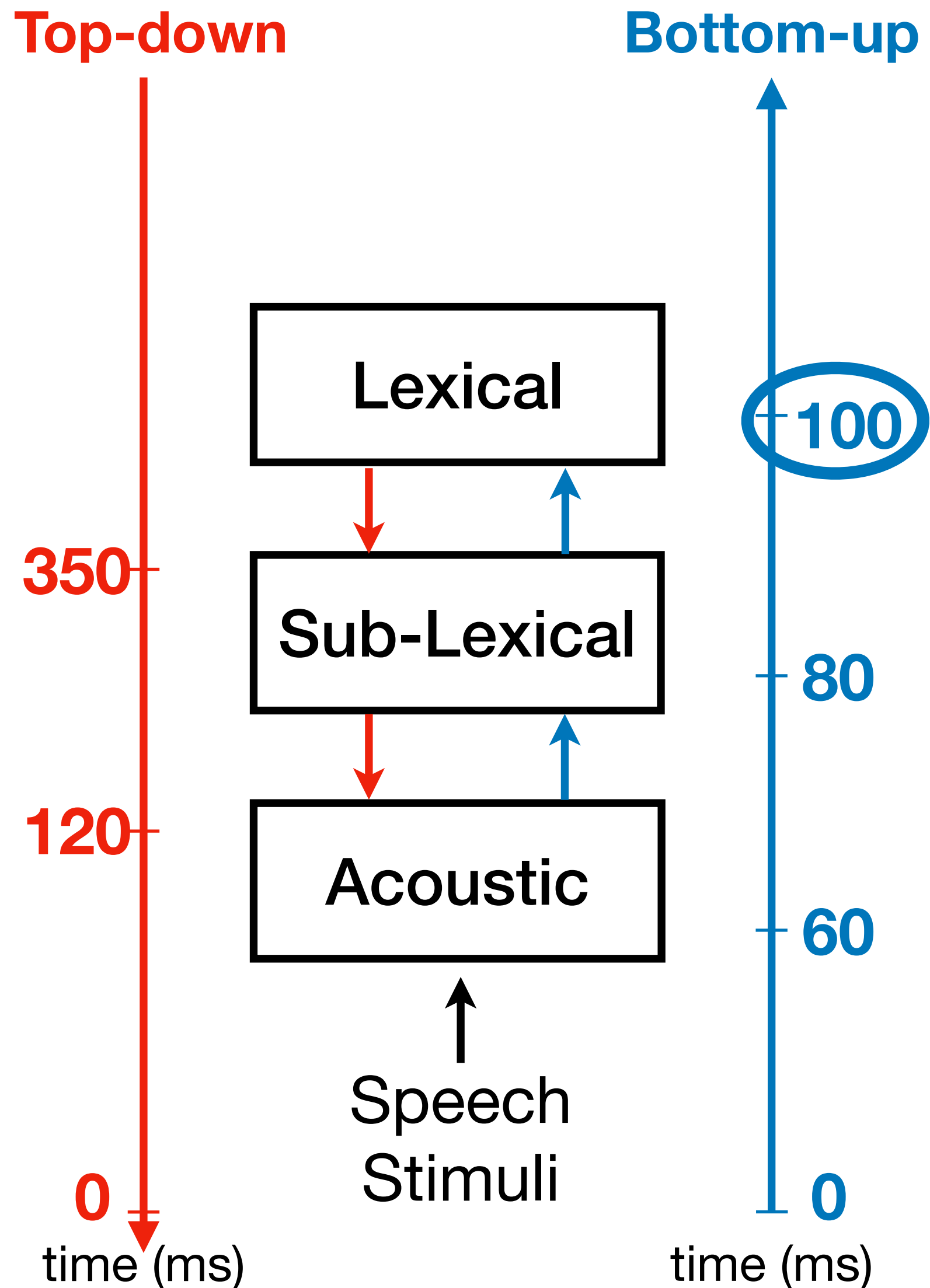- words: Left hemisphere > Right

- non-words: Left hemisphere ≈ Right

- N400-like response

- Reduction in surprisal when more context

- Left hemisphere > Right

- no condition based difference in Right hemisphere

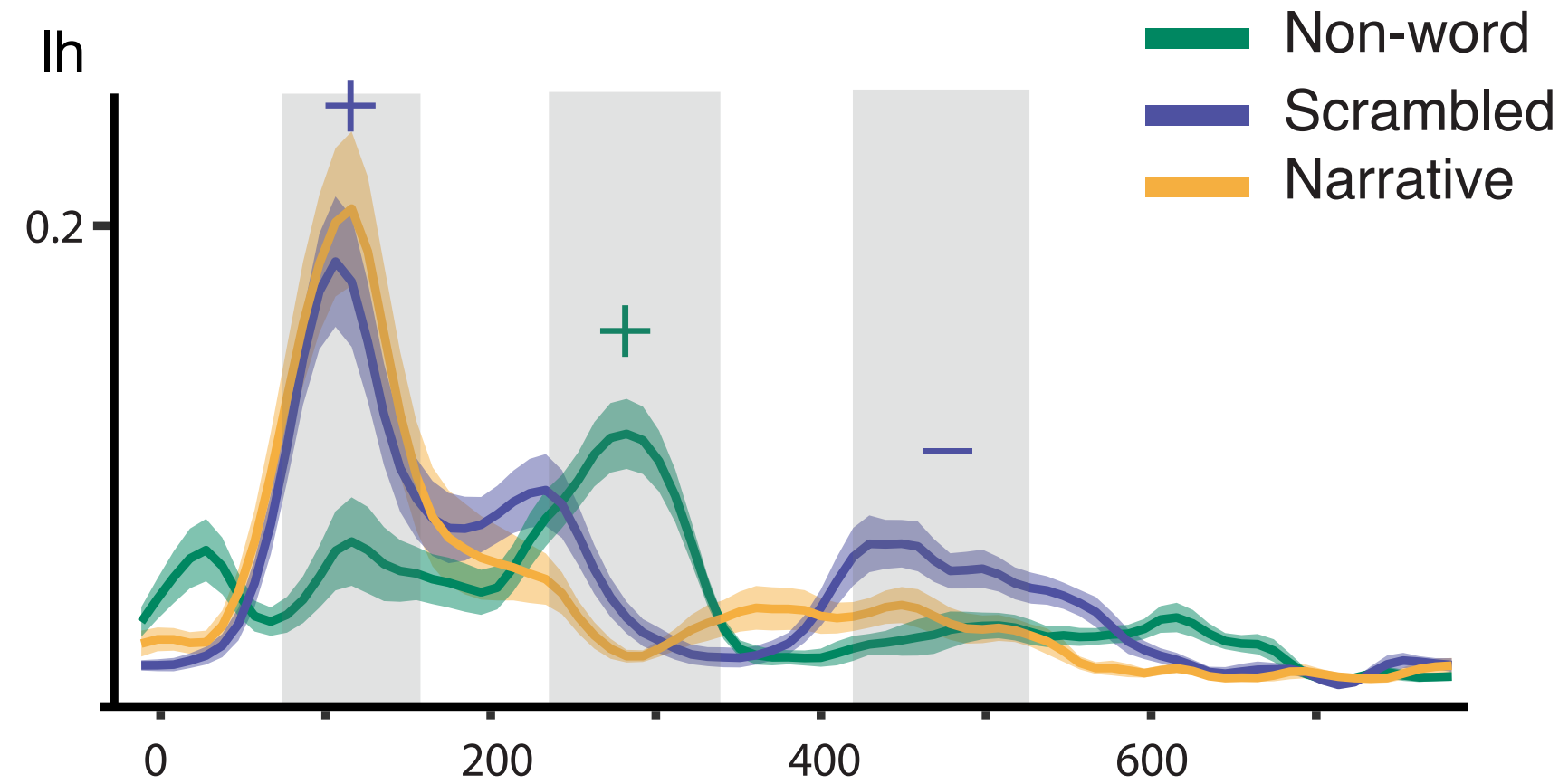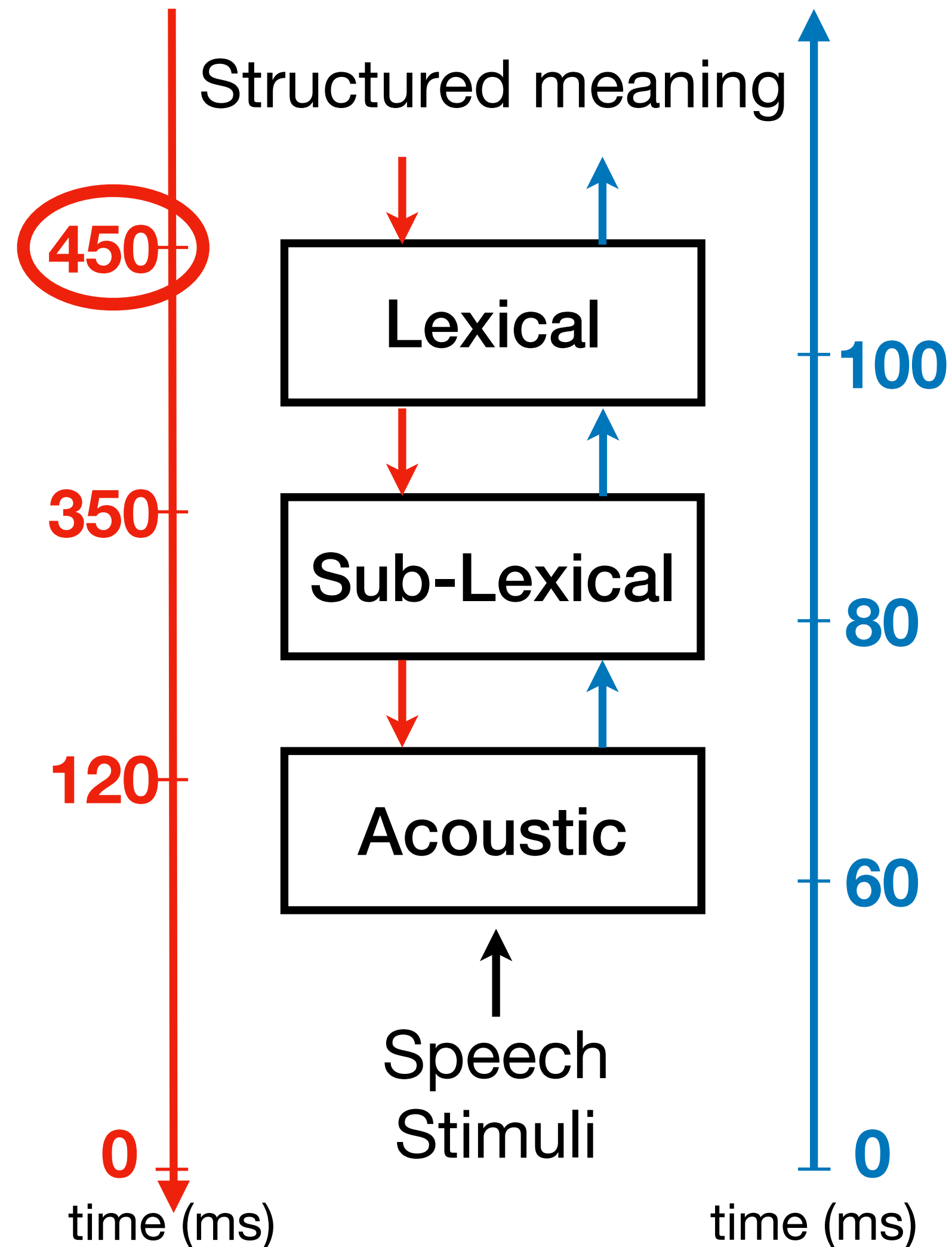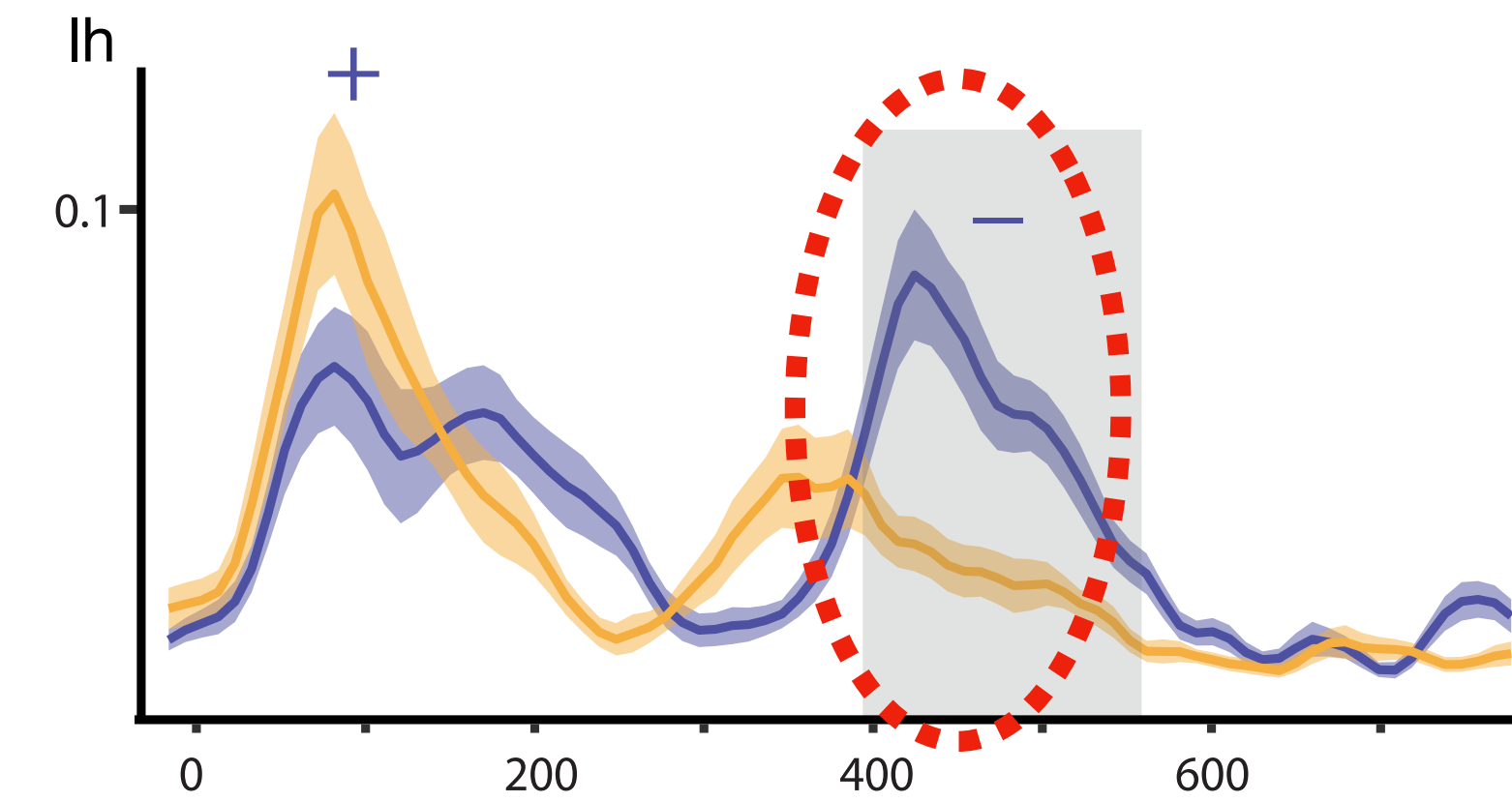# Word-based TRF Results

Word Onset

- Scrambled words response > narrative response 450 ms

- Different neural mechanisms for non-words (searching for mental lexicon?)

- words: Left hemisphere > Right

- non-words: Left hemisphere ≈ Right

Top-down   **Bottom-up**

Structured meaning

450

Lexical

100

350

Sub-Lexical

80

120

Acoustic

60

Speech
Stimuli

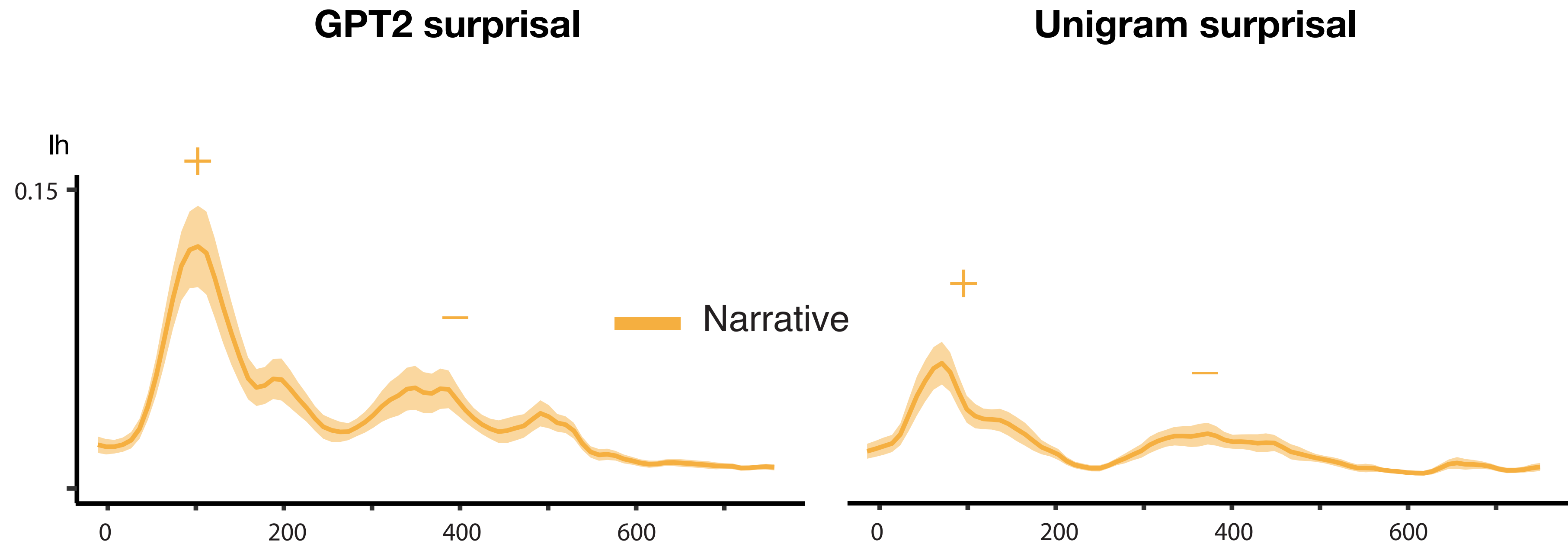0   0

time (ms)   time (ms)

Unigram surprisal

- N400-like response

- Reduction in surprisal when more context

- Left hemisphere > Right

- no condition based difference in Right hemisphere
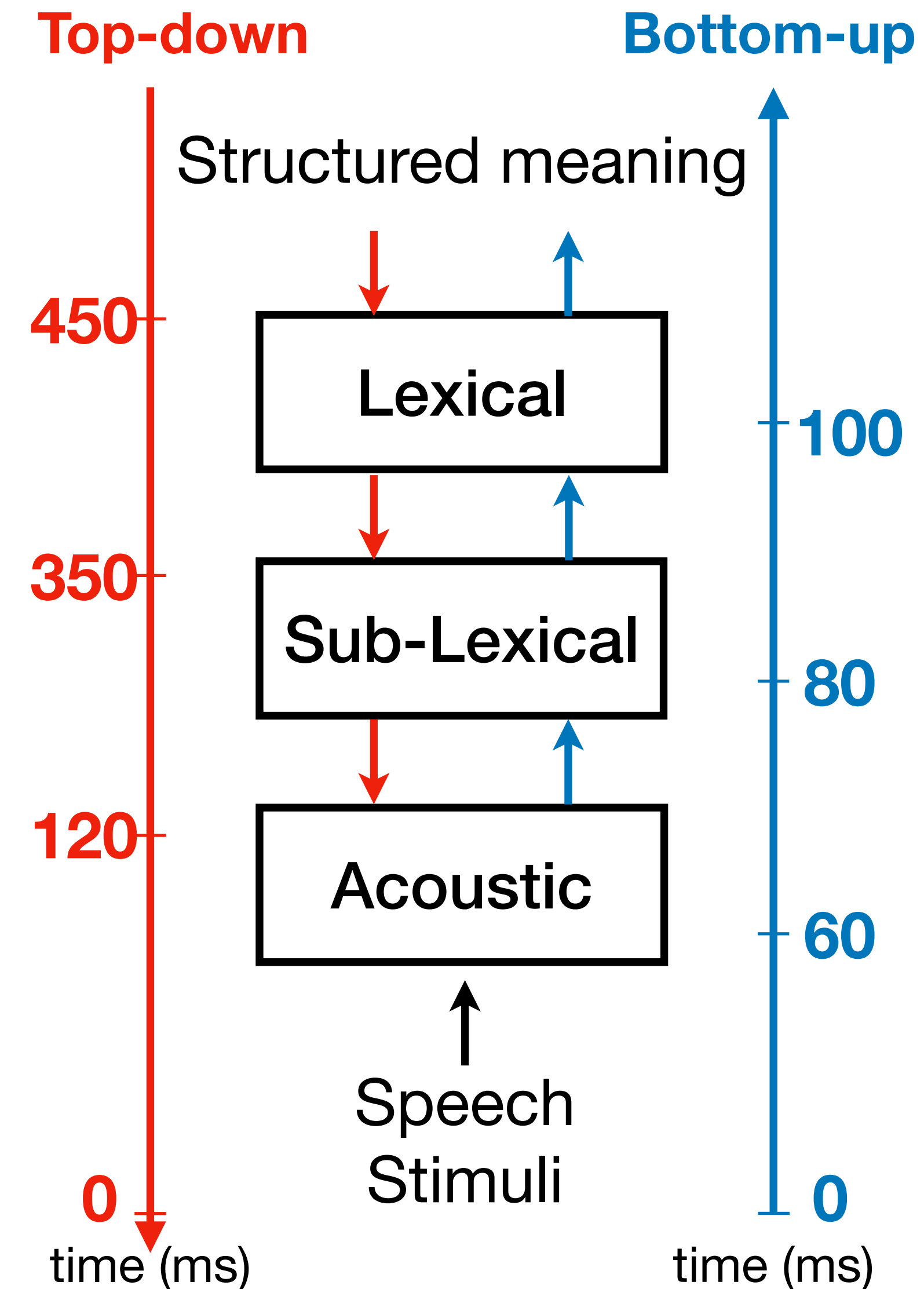
Contextual Word Surprisal Results

GPT2 surprisal

Unigram surprisal

- N400-like response in both predictors

- When context present, context based surprisal is better tracked compared to unigram surprisal

- Left hemisphere > Right

Summary

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

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

- Linguistic features are processed only when the linguistic boundaries are intelligible

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

Top-down   **Bottom-up**

Structured meaning

450

Lexical

100

350

Sub-Lexical

80

120

Acoustic

60

Speech Stimuli

0   0

time (ms)   time (ms)

Acknowledgements

CSSL LAB
- Mike Johns
- Karl Lerud
- Vrishab Commuri
- Charlie Fisher
- Morgan Belcher
- Ciaran Stone
- Kevin Hu

Participants

Funding

Thank you!!!