Continuous speech and its neural representations, through auditory cortex and beyond

Jonathan Z. Simon

University of Maryland, College Park

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

JHU, 15 Oct 2021

Acknowledgements

Current Lab Members & Affiliates

- Shohini Bhattasali
- Regina Calloway
- Vrishab Commuri
- Jason Dunlap
- Theo Dutcher
- Tejas Guha
- Sydney Hancock
- Kevin Hu
- *Neha Joshi*
- Dushyanthi Karunathilake
- ***Joshua Kulasingham***
- *Mohsen Rezaeizadeh*
- Behrad Soleimani
- Ciaran Stone
- Richard Williams

Current & Recent Collaborators

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

Recent Lab Members & Affiliates

- Sahar Akram
- Ross Baehr
- Anurupa Bhonsale
- ***Christian Brodbeck***
- ***Francisco Cervantes Constantino***
- Aura Cruz Heredia
- Proloy Das
- Lien Decruy
- Marisel Villafane Delgado
- *Nai Ding*
- *Marlies Gilles*
- Alex Jiao
- Natalia Lapinskaya
- Sina Miran
- David Nahmias
- Alex Presacco
- Krishna Puvvada
- *Jonas Vanthornhout*
- James Williams
- Peng Zan

Funding & Support

Computational Sensorimotor Systems Lab

Continuous speech and its neural representations, through auditory cortex and beyond

To Intelligibility, and Beyond!

University of Maryland 1856

Computational Sensorimotor Systems Lab

Continuous speech and its neural representations, through auditory cortex and beyond

Cortical Representations of Continuous Speech

Continuous speech

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

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

The Botany of Desire — Michael Pollan

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

A Child's History of England — Charles Dickens

In the bosom of one of those spacious coves which indent the eastern shore of the Hudson, at that broad expansion of the river denominated by the ancient Dutch navigators …

The Legend of Sleepy Hollow — Washington Irving

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

The Old Man and the Sea — Ernest Hemingway

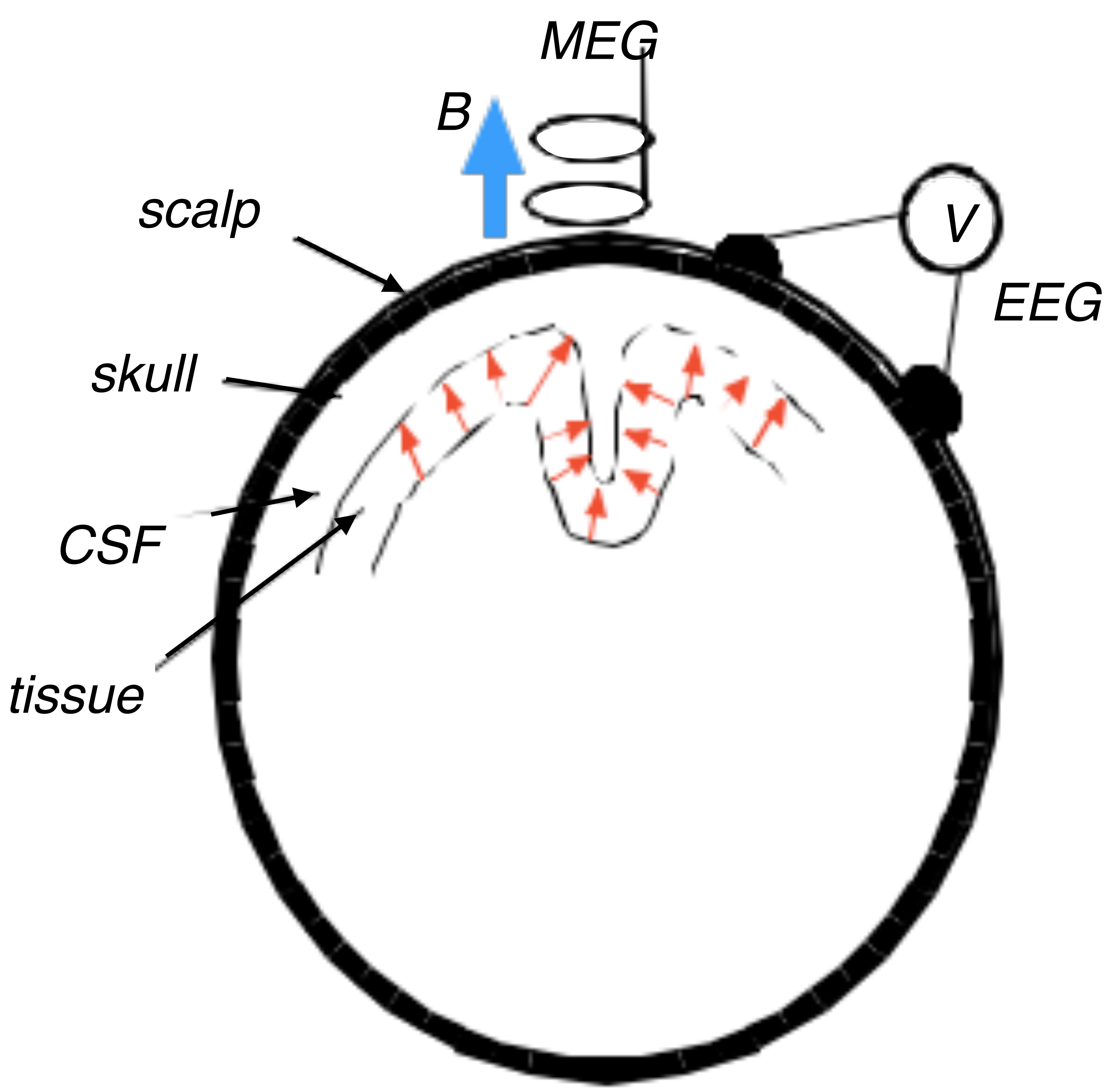
Cortical Representations of Continuous Speech

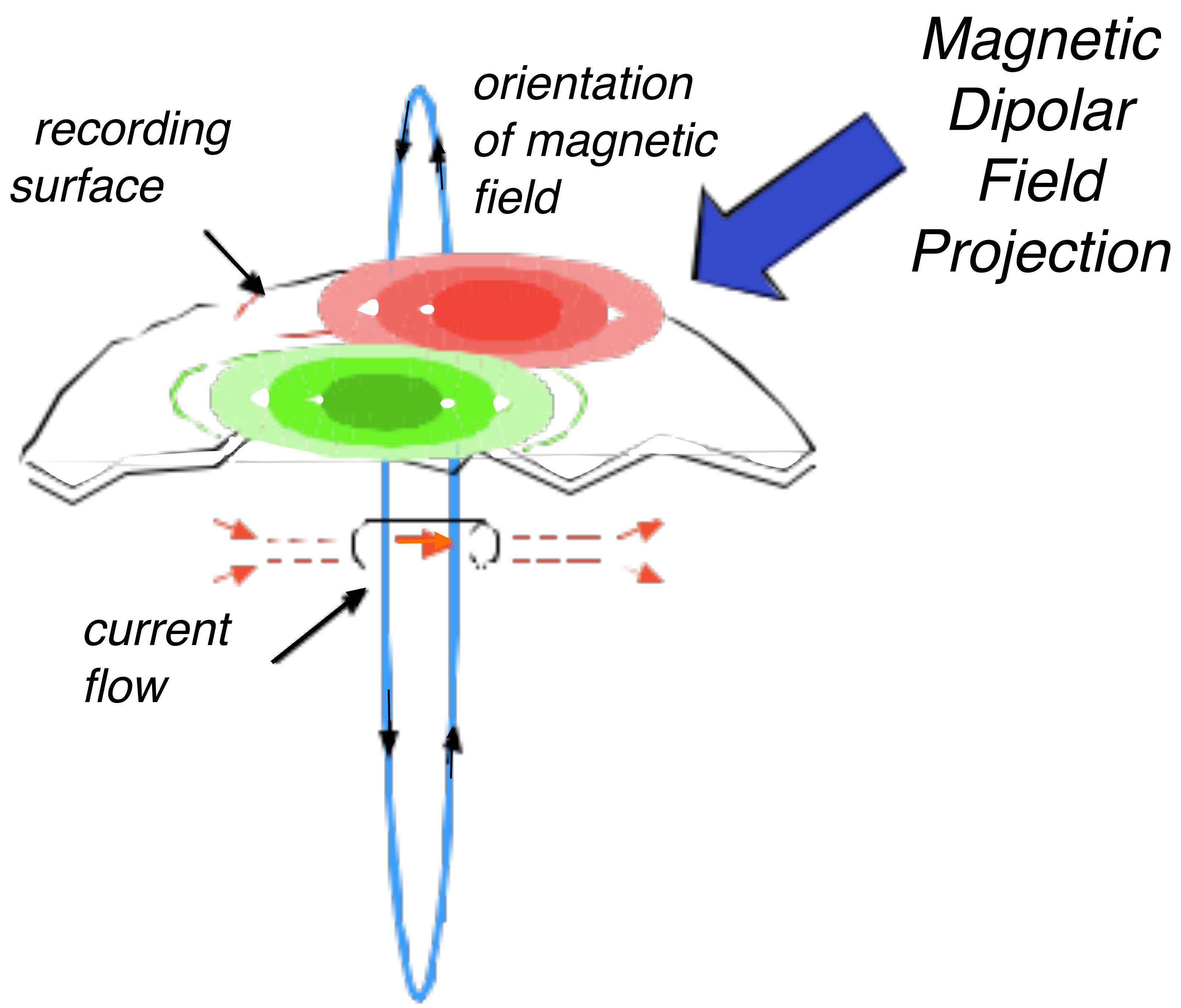
Temporal neural patterns ⇆ temporal patterns in speech

- 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*

Neural Representations & MEG

Photo by Fritz Goro

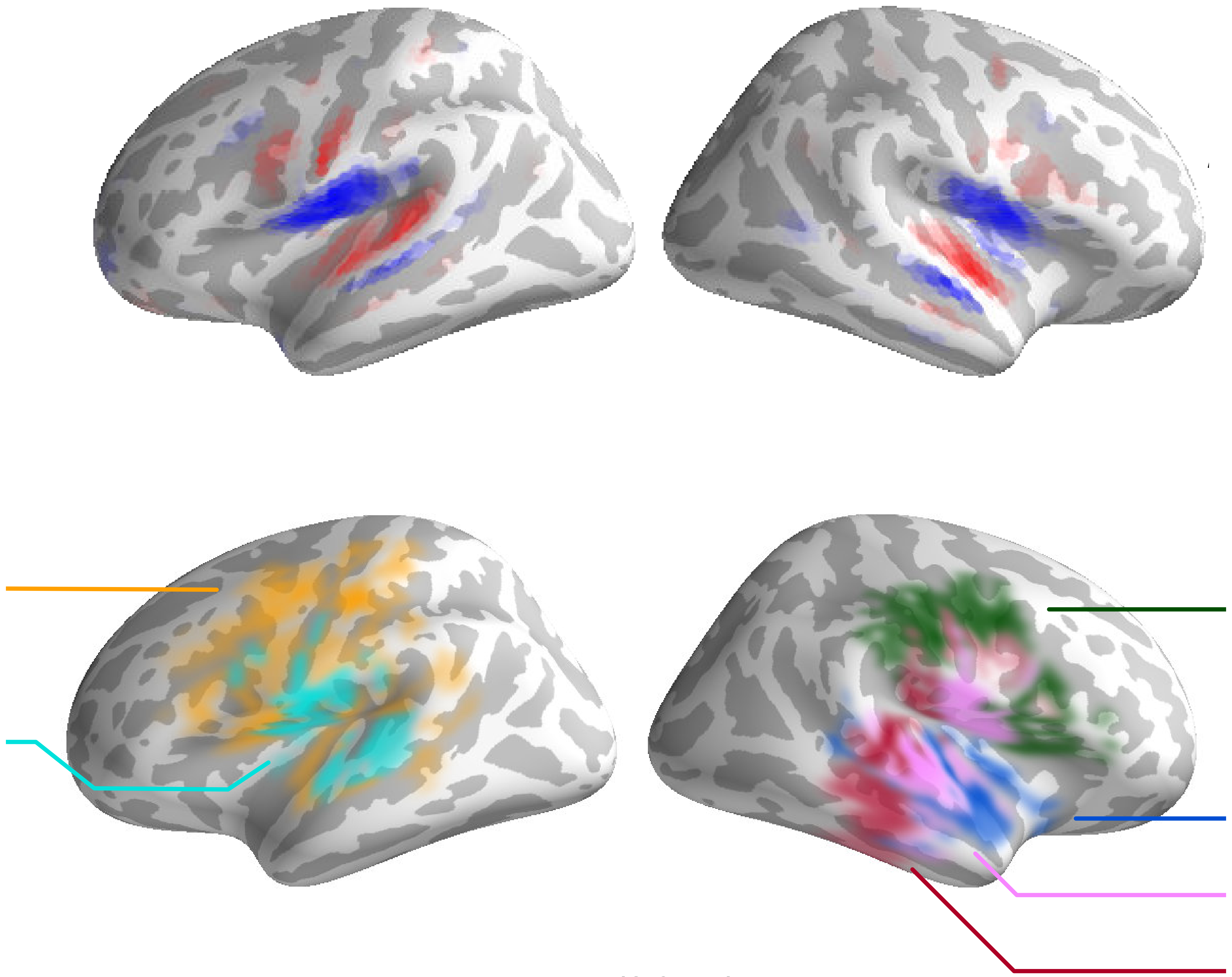




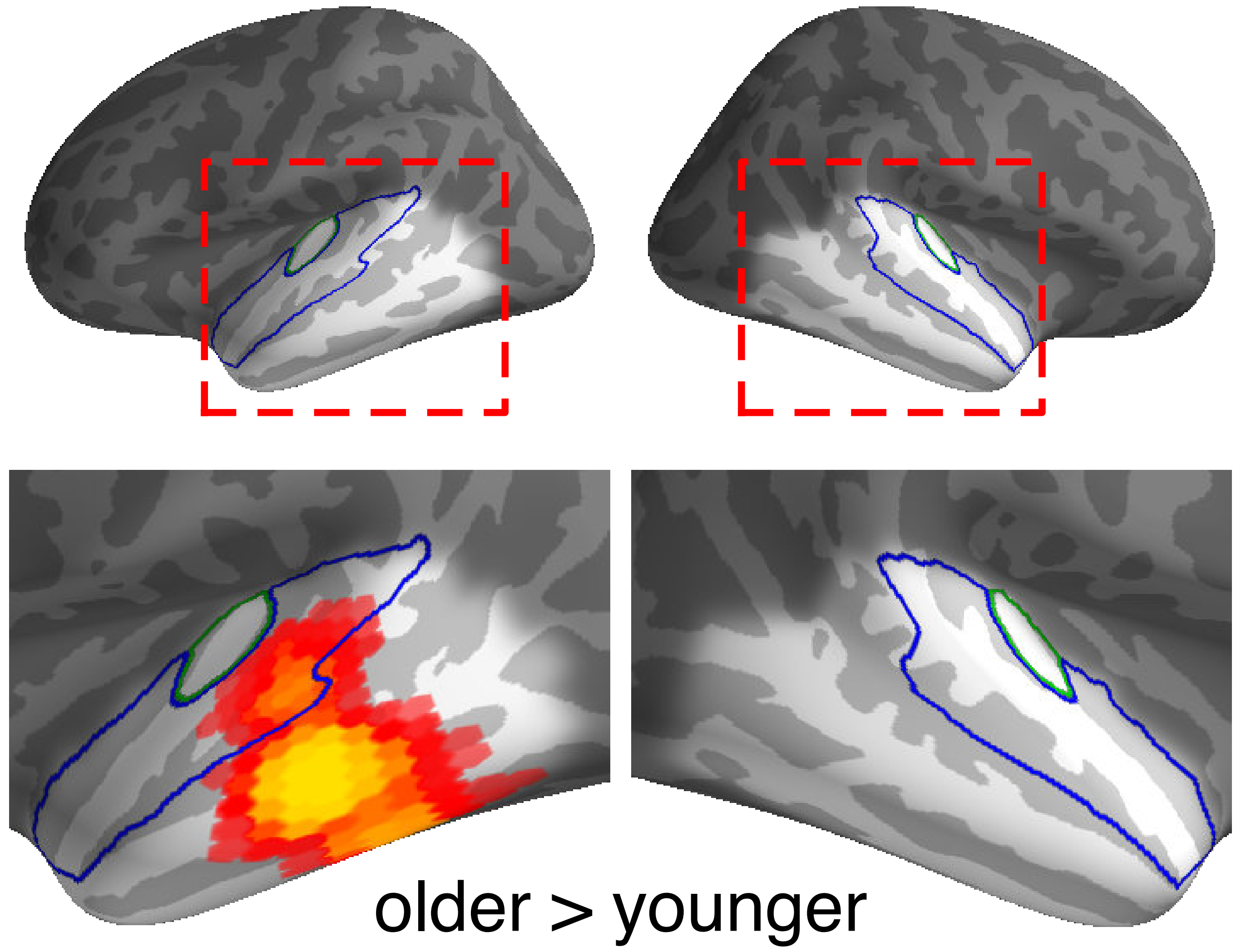
- Direct electrophysiological measurement
  - not hemodynamic
  - real-time
- No unique solution for distributed source

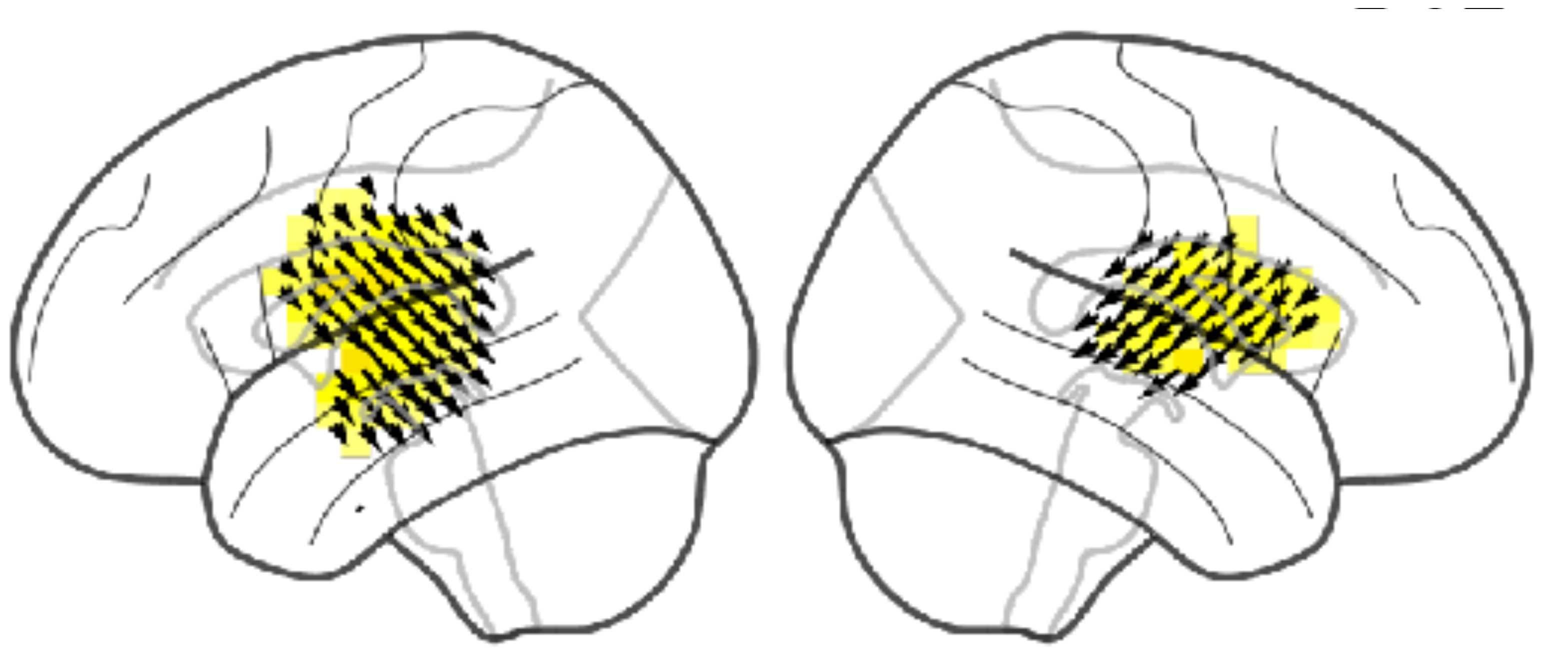
- Measures spatially synchronized cortical activity
- Fine temporal resolution (~ 1 ms)
- Moderate spatial resolution (~ 1 cm)

Spatial Distributions of MEG Neural Currents



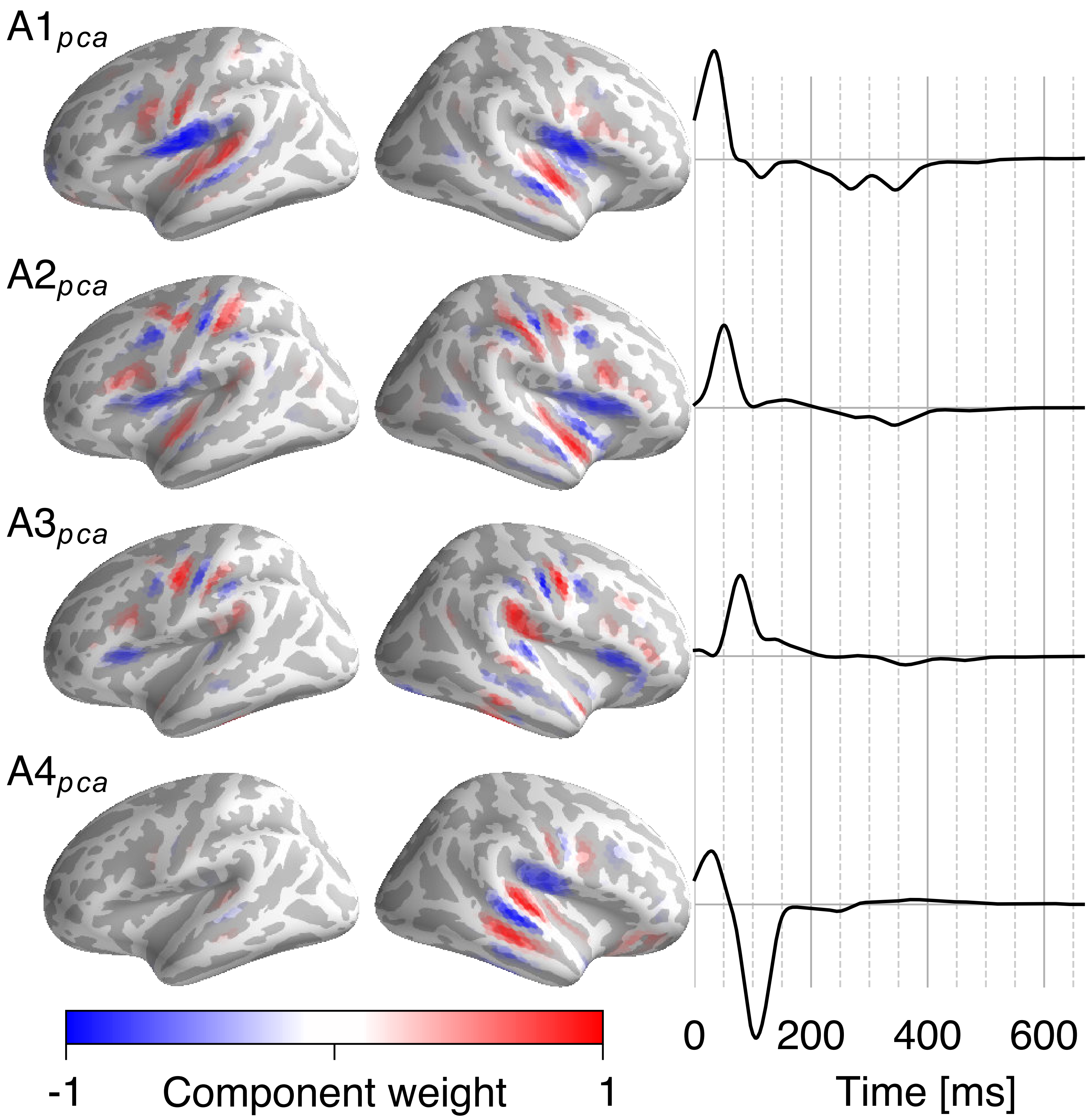
Brodbeck et al., NeuroImage (2017)
Brodbeck et al., Acta Acust united Ac (2018)

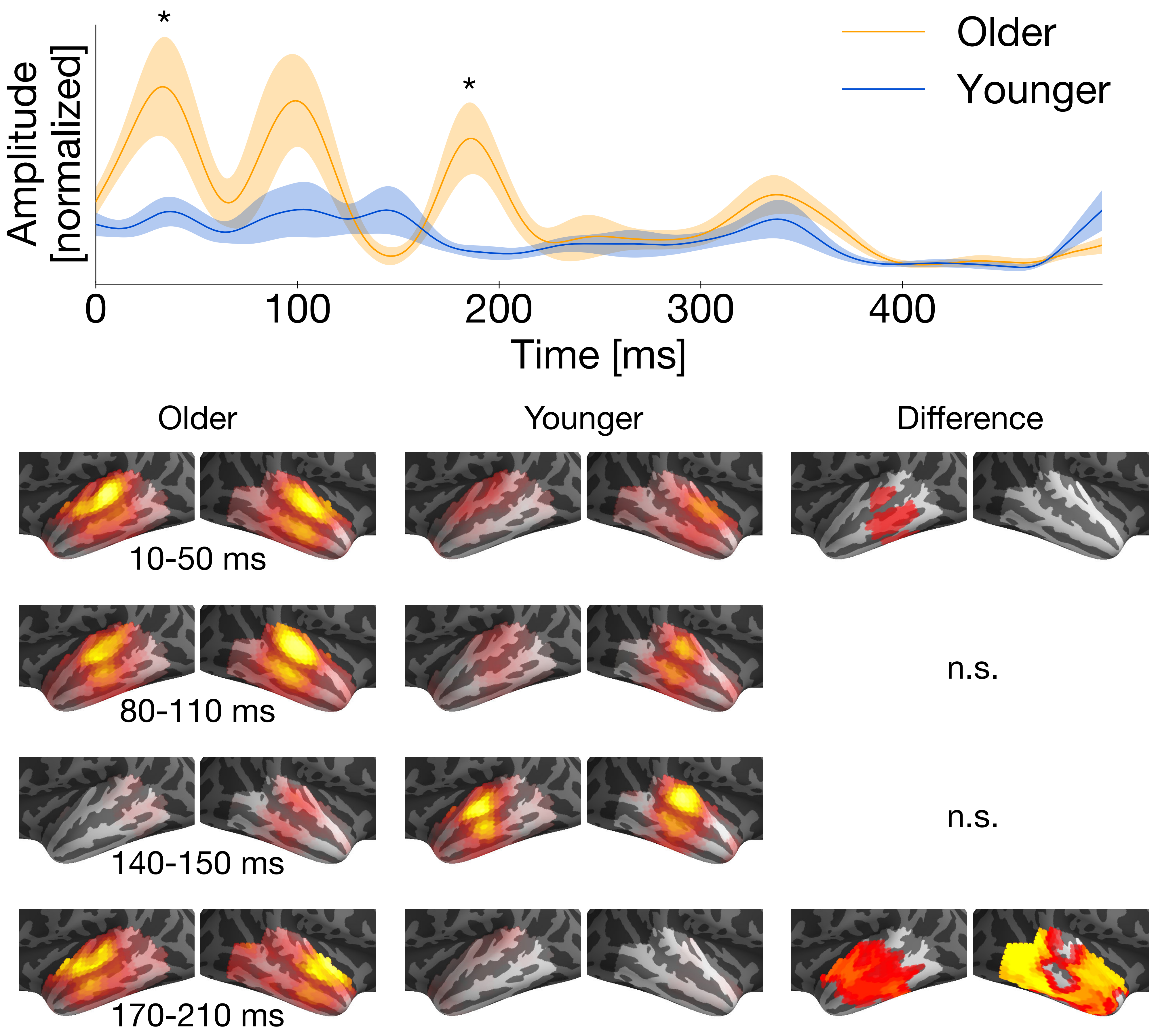




Das et al., NeuroImage (2020)

Spatiotemporal Distribution of Neural Currents





Brodbeck et al., NeuroImage (2017)
Brodbeck et al., Acta Acust united Ac (2018)

Cortical Representations of Continuous Speech

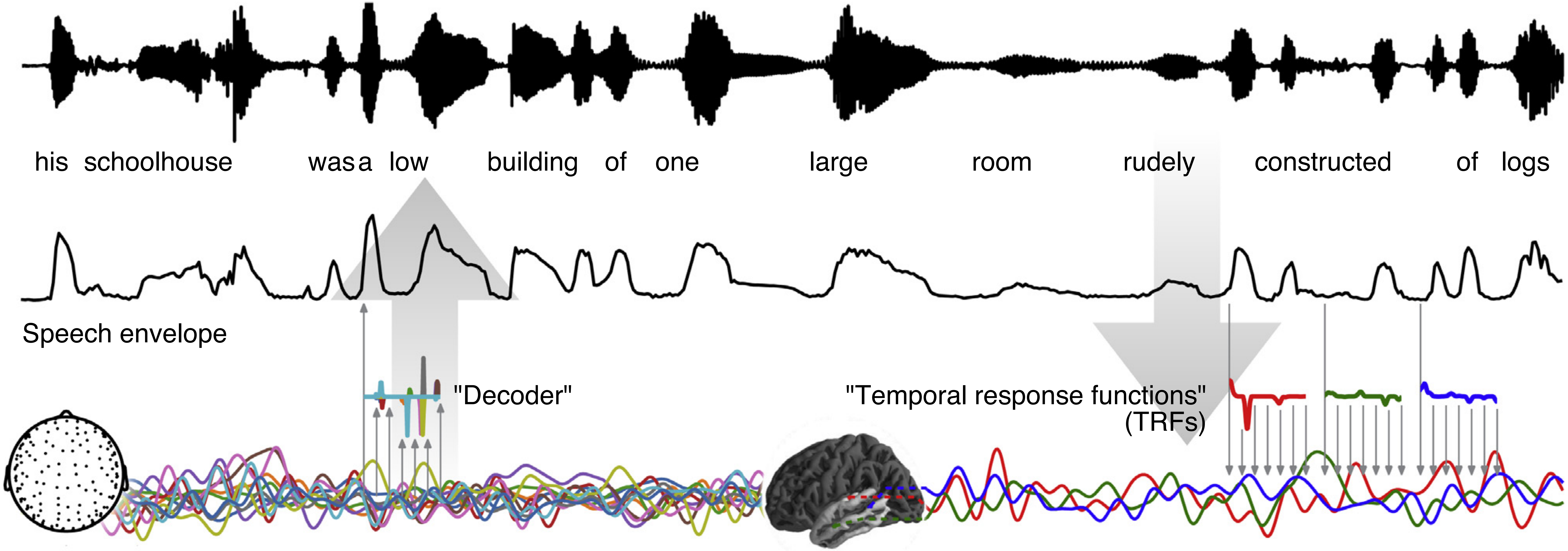
Neural Representations of Speech

- driven oscillations at pitch frequencies (mostly subcortical)
  - acoustic onset tracking
    - speech envelope rhythmic following
      - phoneme-based responses
        - phoneme-context-based responses
          - sentence-structure rhythm following
            - semantic structure tracking
- plus connections to **intelligibility/perception/behavior**

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

Cortical Representations of Speech

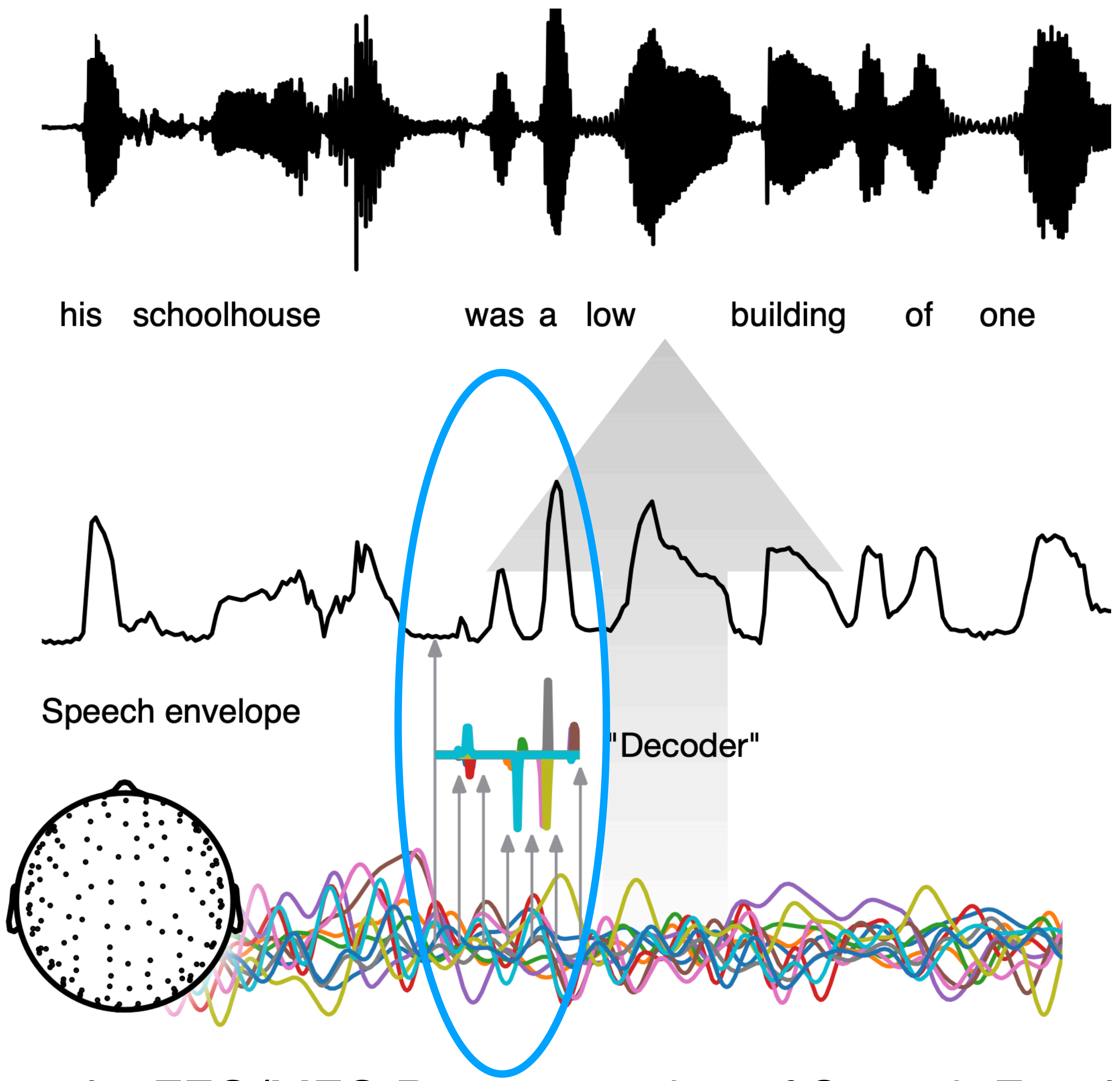
- Measure *time-locked* responses to temporal pattern of speech features (in humans)
- Any speech feature of interest: acoustic envelope, lexical, pitch, semantic, etc.
- Infer spatio-temporal neural origins of neural responses



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

Cortical Representations: Decoding

- Reconstruct past stimulus features (from present neural responses)
  - how much information, regarding this class of stimulus features, is visible in the brain?
- Typically speech envelope (dynamic, ongoing)
  - other features possible but less common
- Moderate time resolution (10s to 100s of ms)
  - spatial resolution more iffy

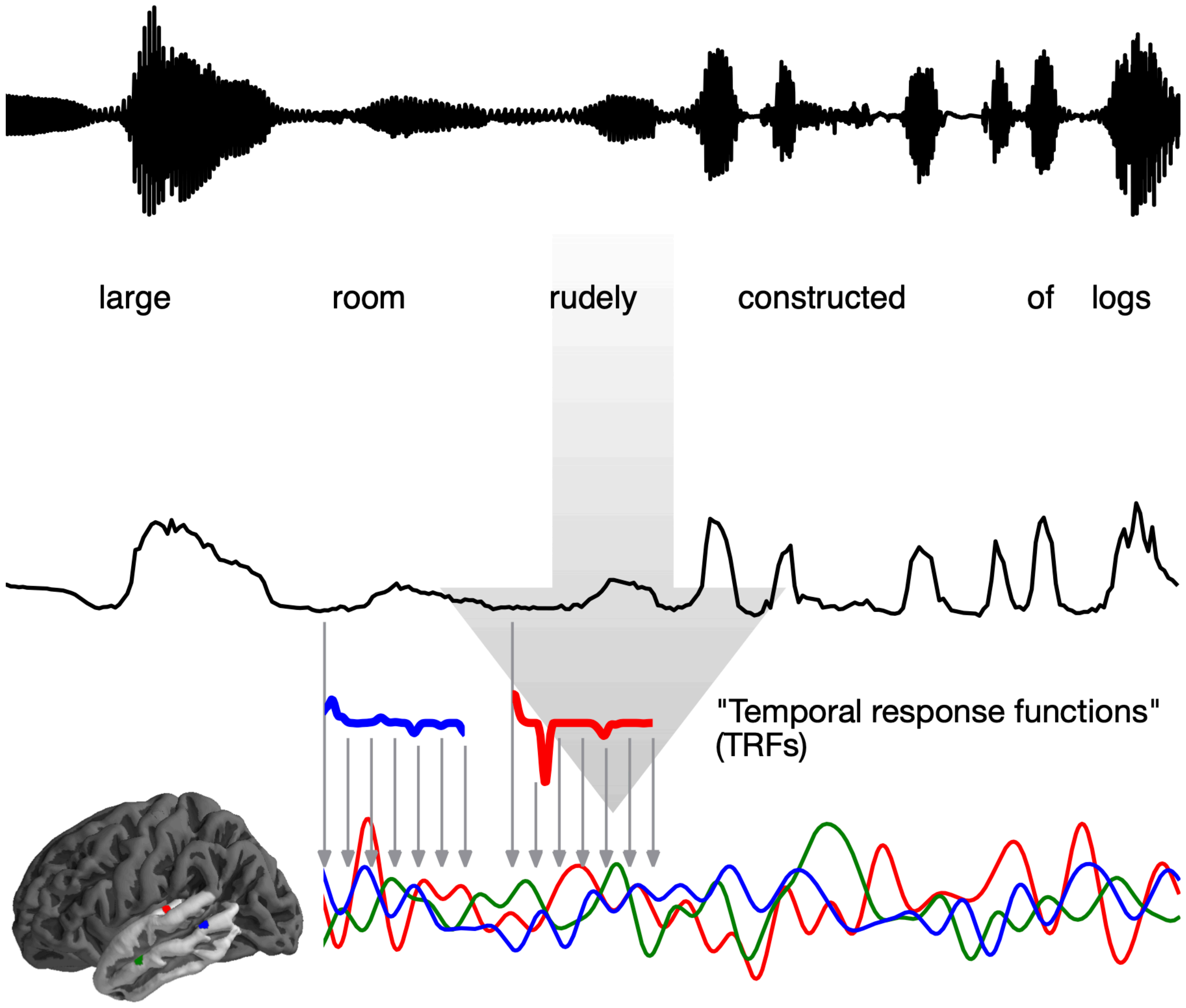


Example: EEG/MEG Reconstruction of Speech Envelope

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

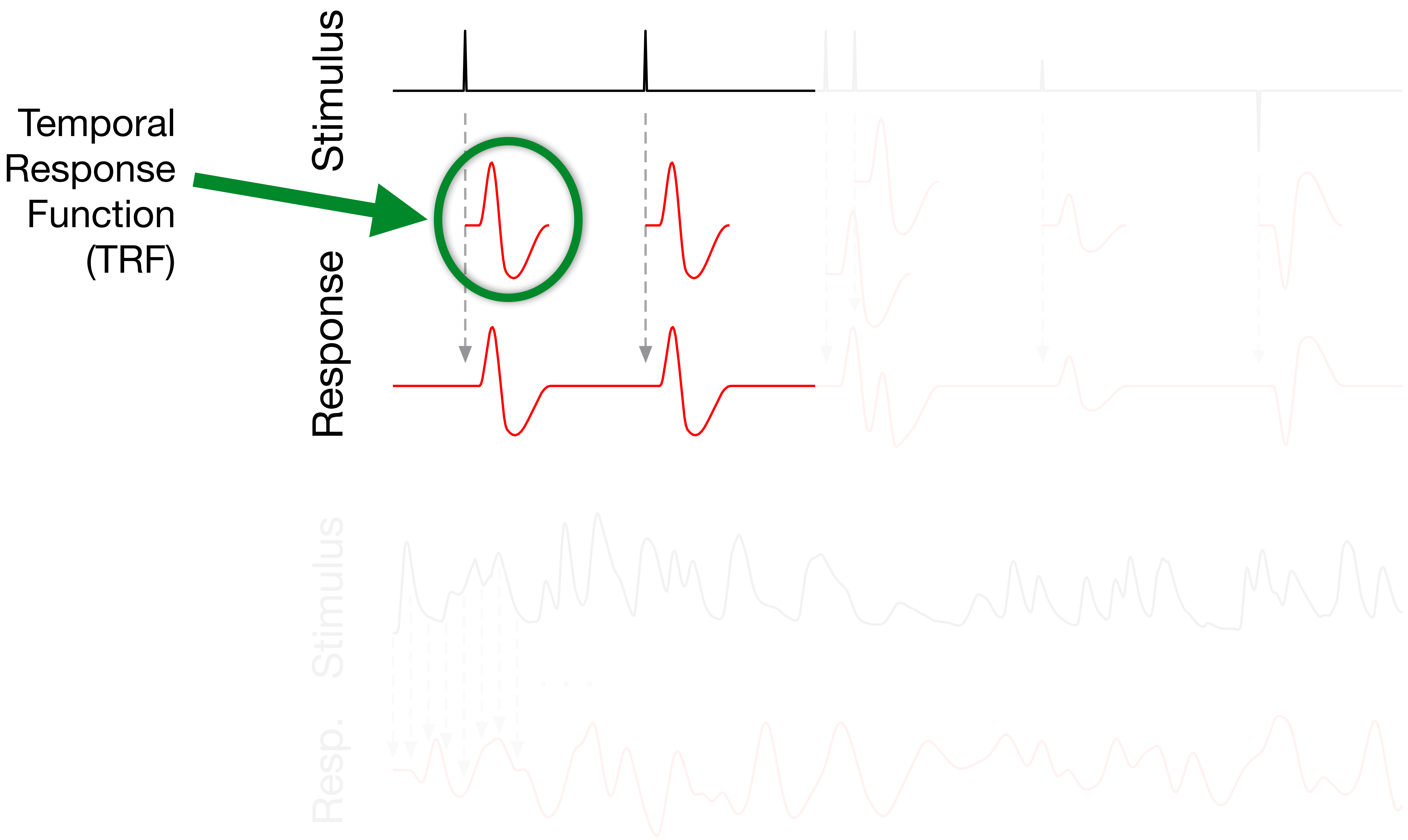
Cortical Representations: Encoding

- Predicting future neural responses from present stimulus features,
  - wide variety of stimulus features
  - via Temporal Response Function (TRF)
- Why look at encoding? It *often* tells us more about the brain
  - TRF analogous to evoked response
  - peak amplitude ≈ processing intensity
  - peak latency ≈ source location
  - multiple TRFs simultaneously



Example: MEG Prediction of Voxel Responses

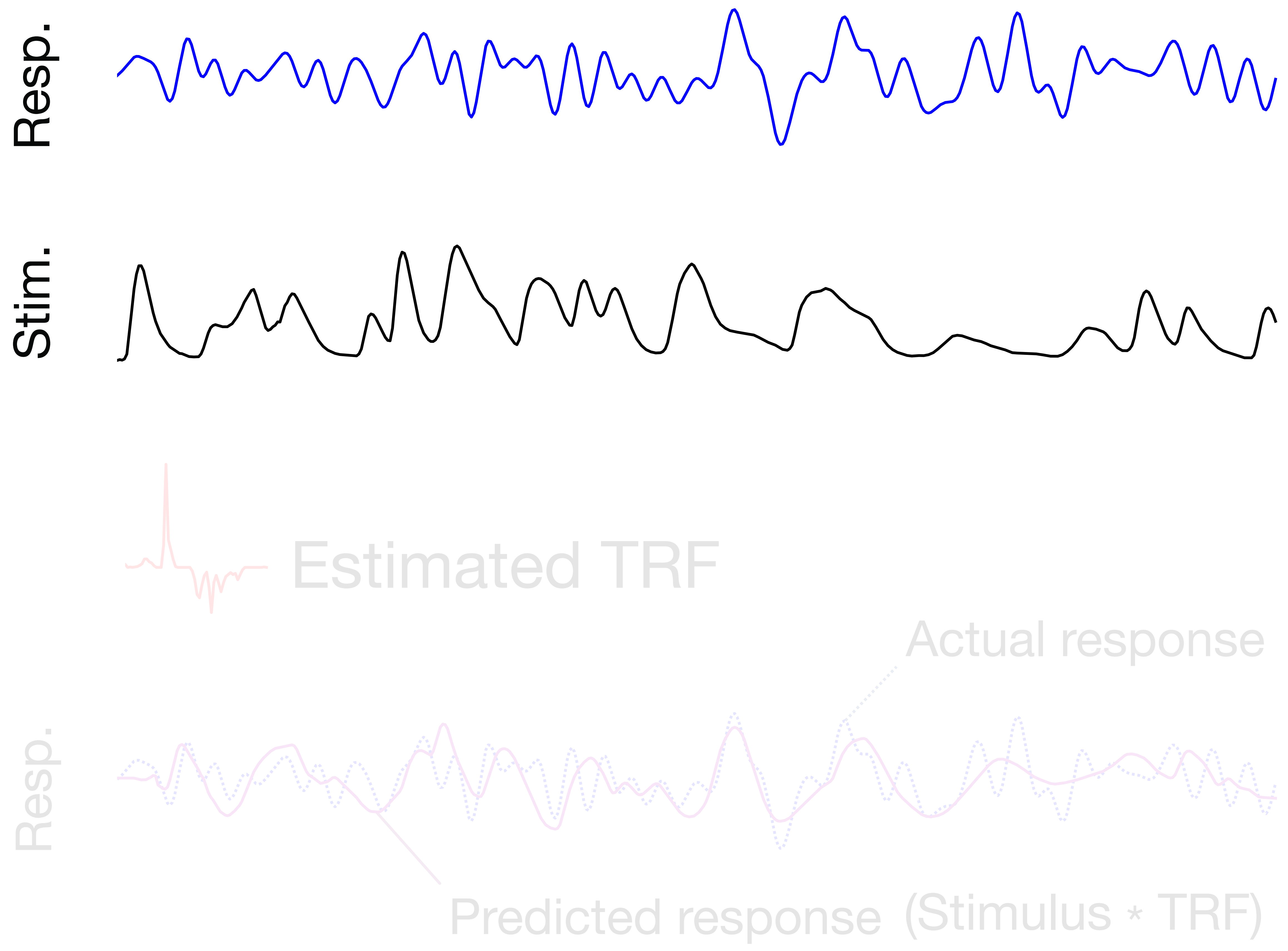
Temporal Response Functions



TRF Model Estimation & Fit

Temporal Response Function (TRF) estimation:

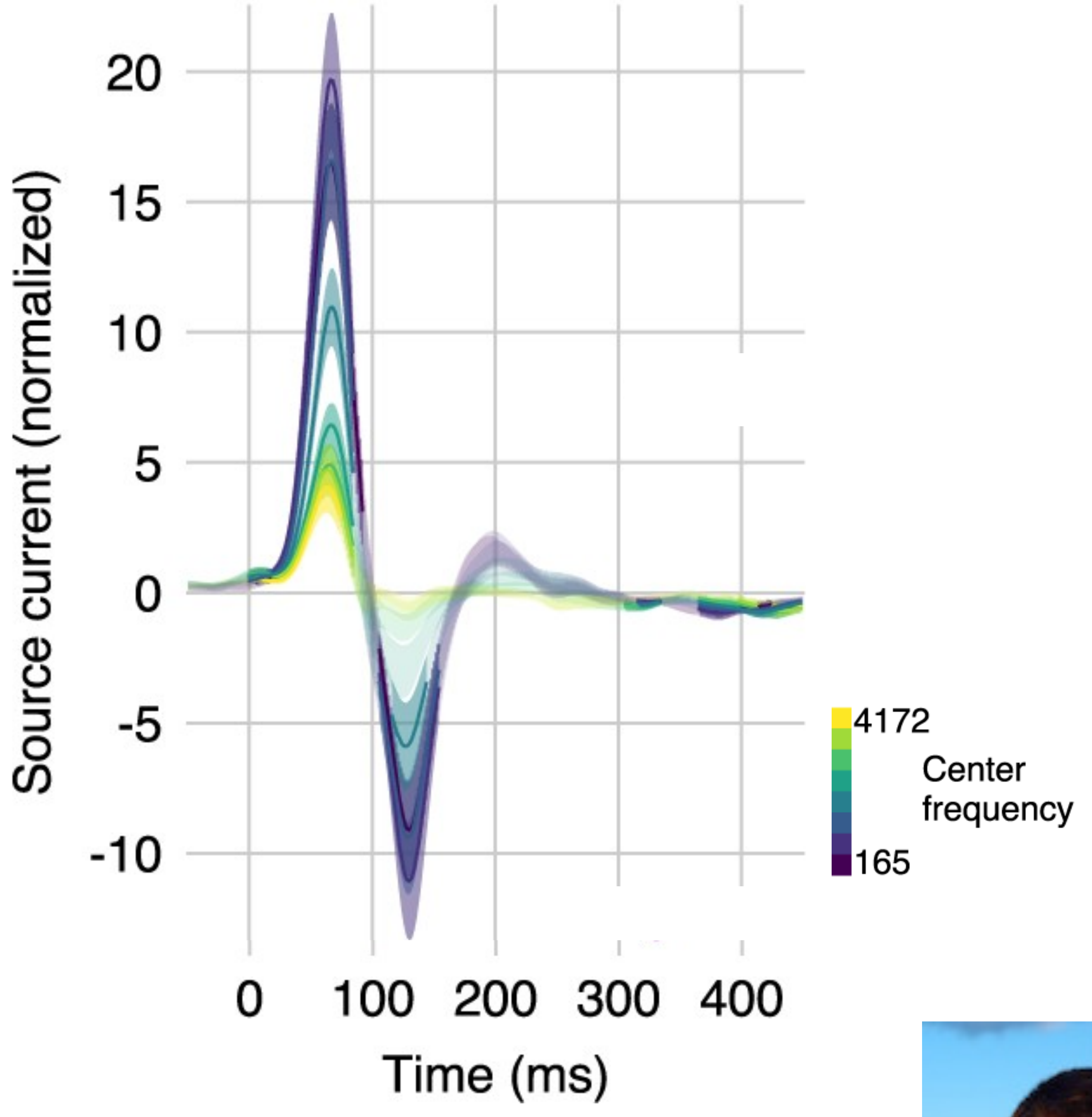
Stimulus and response are known; find the best TRF to produce the response from the stimulus:



Example: Representation of Speech Envelope

- TRF interpretable a la evoked response
  - Has M50 (~"P1") & M100 (~"N1") peaks, but from instantaneous speech envelope
  - early peak localizes to primary auditory areas (HG)
  - later peak localizes to associative areas (PT)
  - caveat: actually from envelope *onset*
- This is from a single talker, clean speech
  - simple but limiting
  - what about noise? other speakers? attention?
  - can the speech representation be cleaned?

Temporal Response Fields

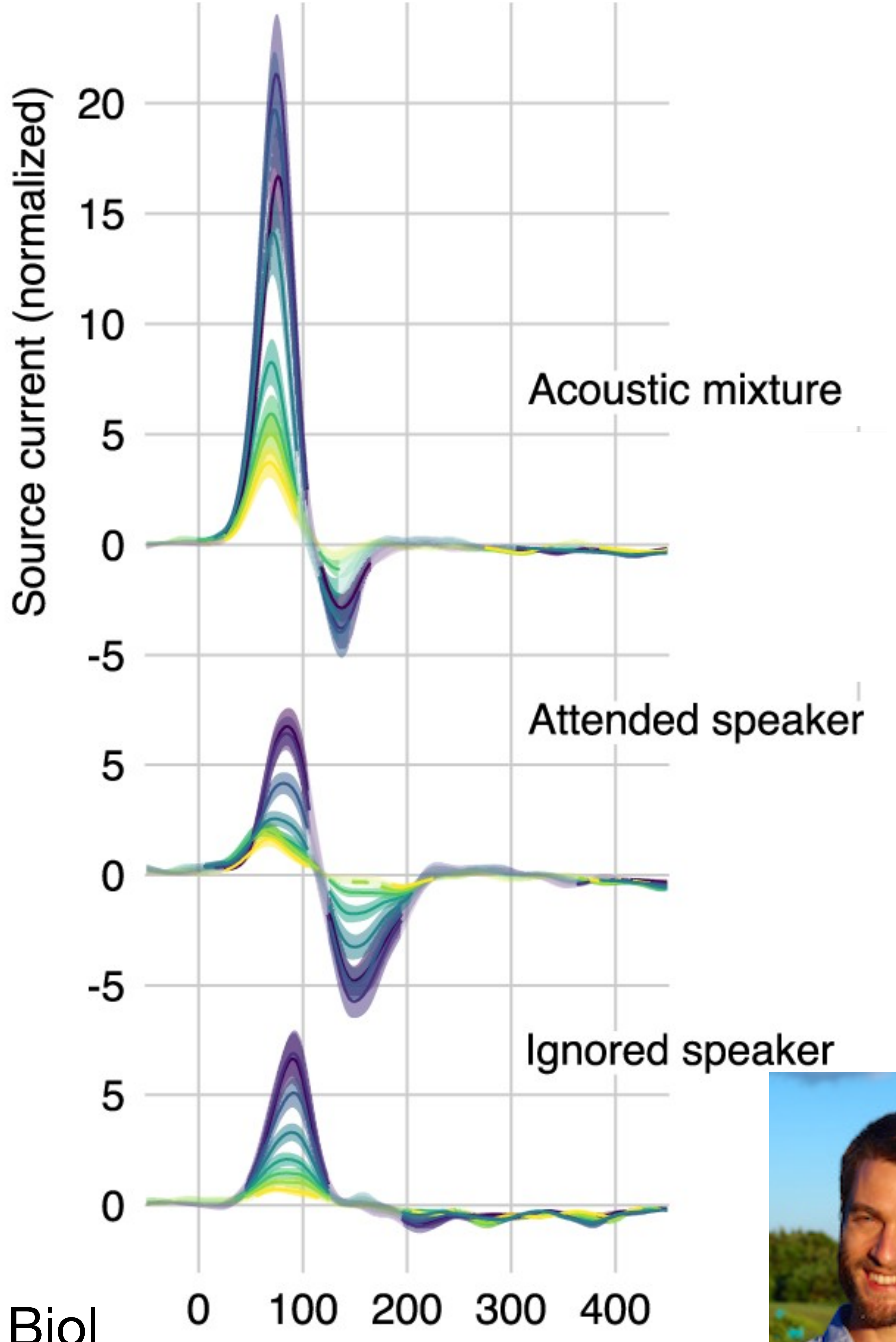


Brodbeck et al. (2020) *Neural Speech Restoration at the Cocktail Party …*, PLoS Biol

Cortical Representations: Selective Attention

Two competing speakers, selectively attend to one

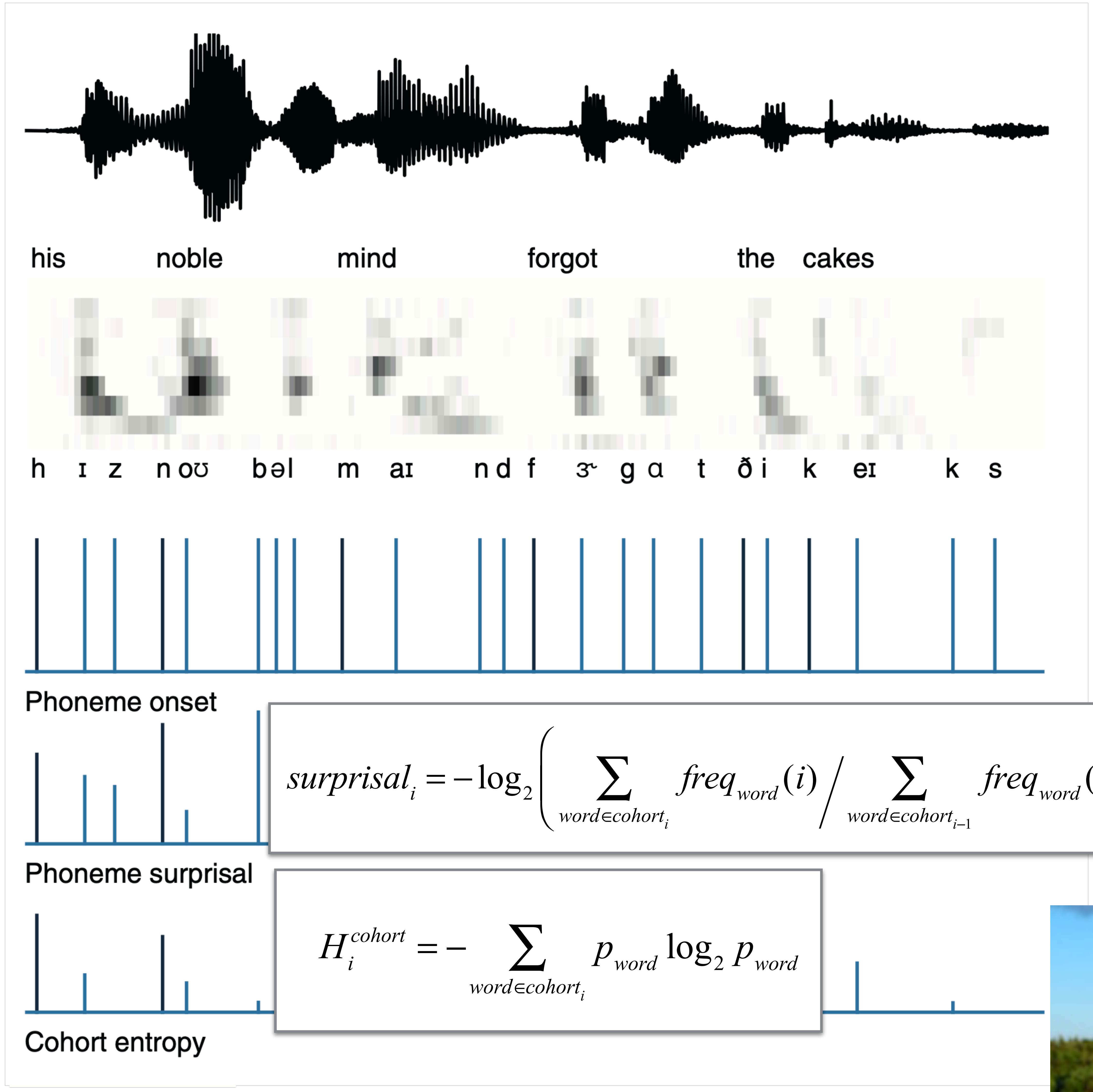
- more illuminating since more complex auditory scene
- need more care re: “stimulus” responsible for responses
  - acoustic mixture entering ears
  - foreground speech
  - background speech
- estimate all TRFs simultaneously
  - compete to explain variance



Brodbeck et al. (2020) *Neural Speech Restoration at the Cocktail Party …*, PLoS Biol

Cortical Representations: Language Features

- Language-based speech features
  - phonemes
  - words & word boundaries
  - phoneme context
- All TRFs estimated simultaneously
  - compete to explain variance



Brodbeck et al. (2018) *Rapid Transformation from Auditory to Linguistic Representations …*, Curr Biol

Phoneme Surprisal

Number of times a word that starts with this sequence occurs in SUBTLEX

K EY …
52908
(90 words)

Number of words that start with this sequence

Surprisal

K EY M …	23875 (45%)	(4 words)	“came”, “Cambridge”, …
K EY S …	16048 (30%)	(13 words)	“case”, “cases”, “caseworker”, “casein”, …
K EY K …	2598 (5%)	(3 words)	“cake”, “caked”, “cakes”
K EY N …	1337 (3%)	(13 words)	“cane”, “canine”, “Canaan”, “Kane”, “Keynesian”, …

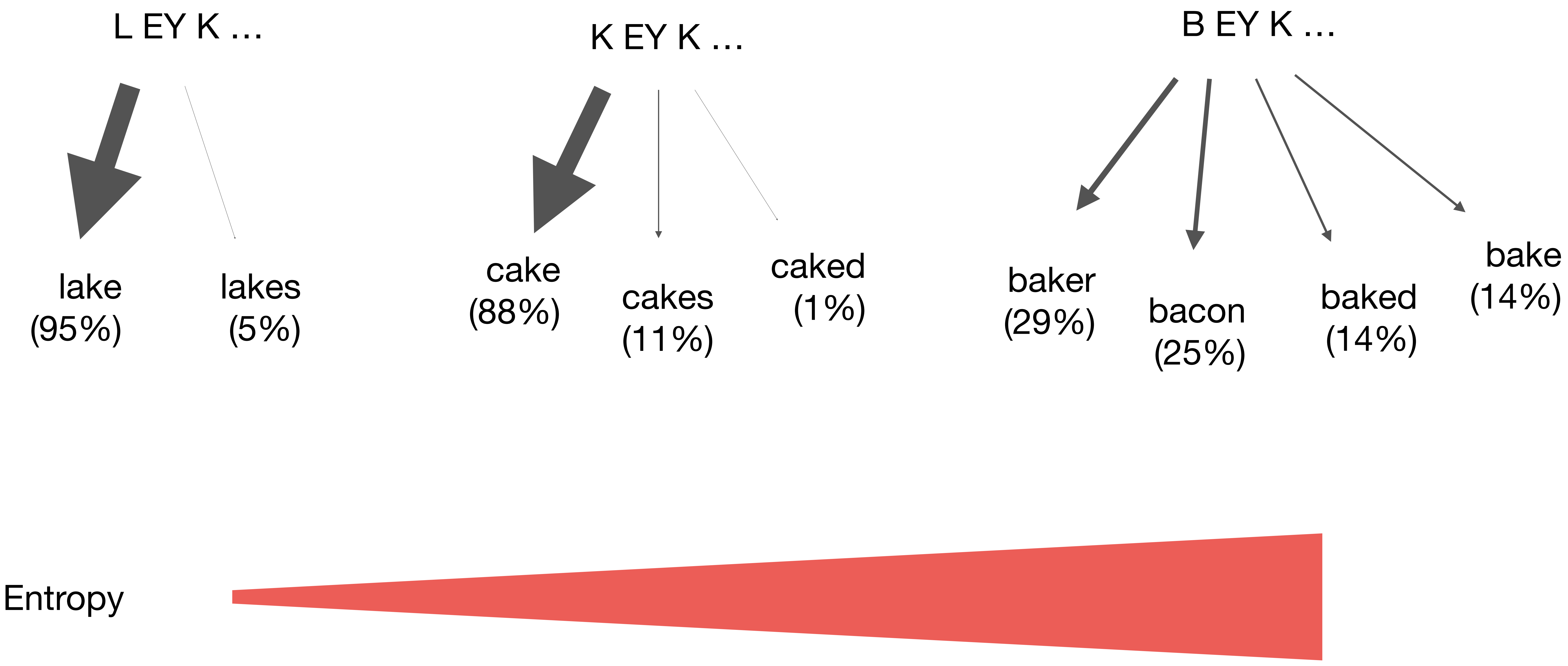
⋮

SUBTLEX:
- 51 million words
- movie subtitle database

Cohort Entropy

Cohort entropy

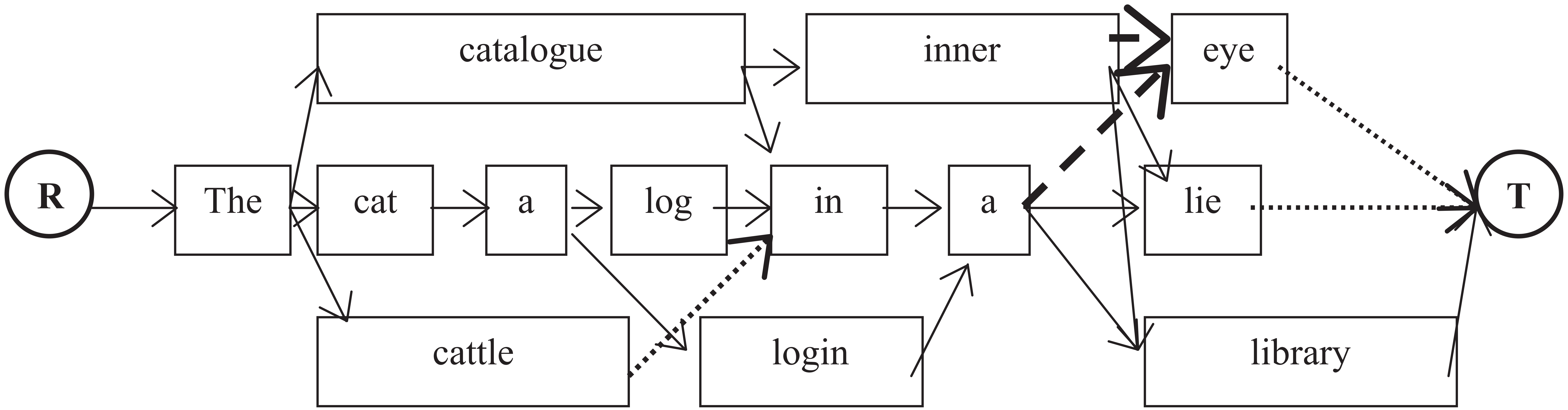
- How unpredictable is the current word?



Word Onsets

Do we…

- Anticipate word boundaries based on context?
- Infer them later based on consistency?

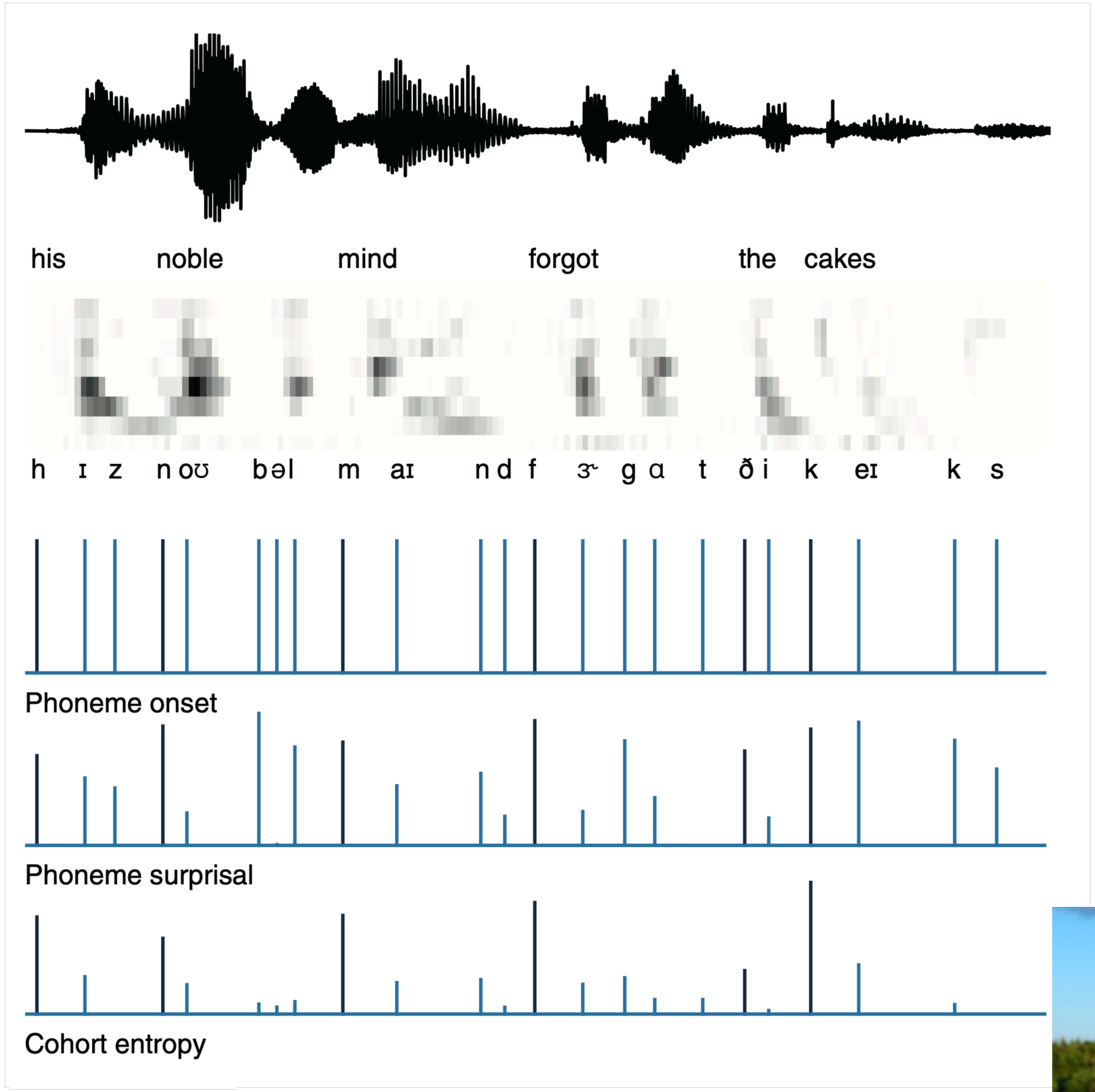


"The catalogue in a library"

(Norris & McQueen, 2008)

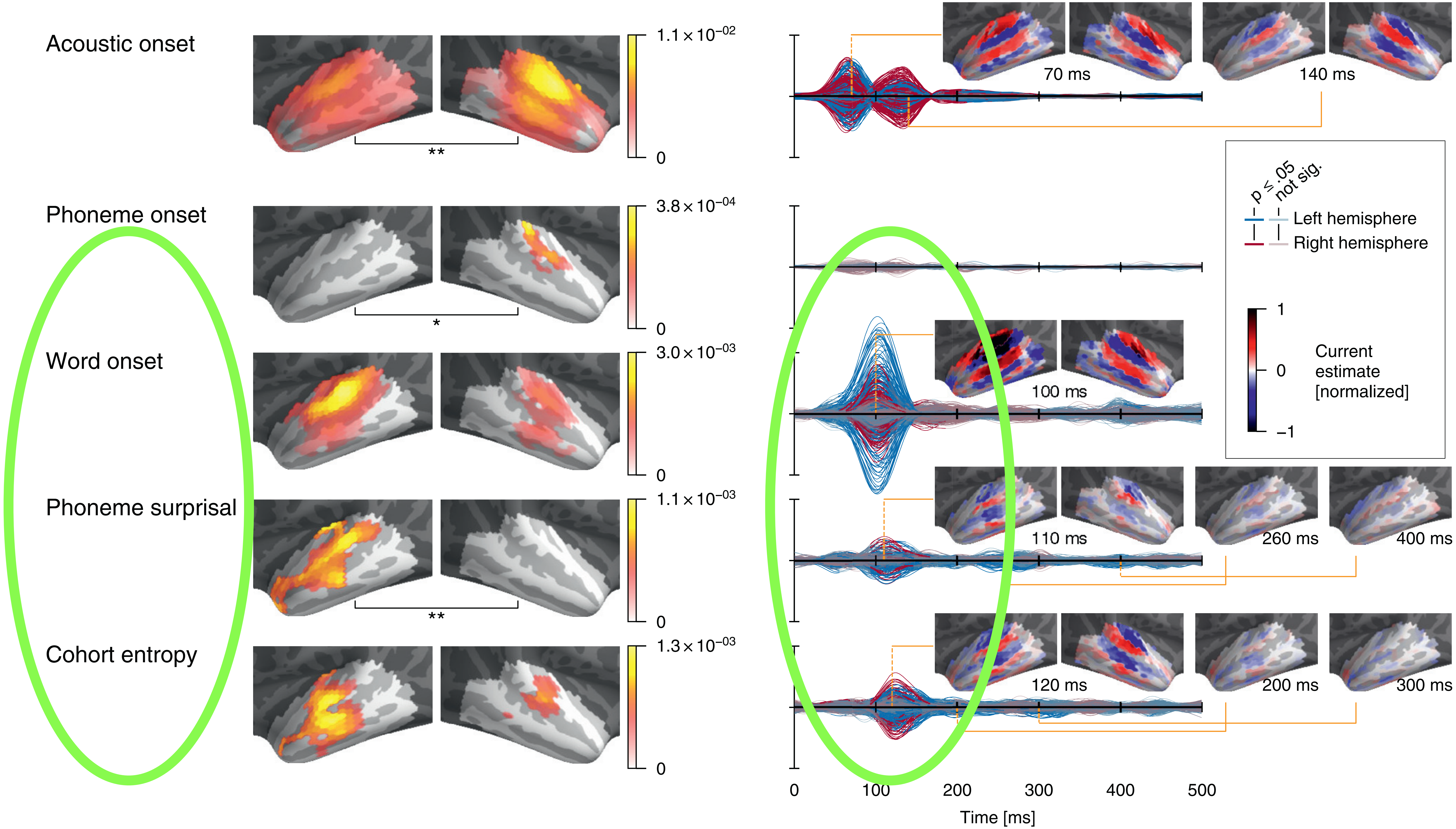
Cortical Representations: Language Features

- Language-based speech features
  - phonemes
  - words & word boundaries
  - phoneme context
- All TRFs estimated simultaneously
  - compete to explain variance



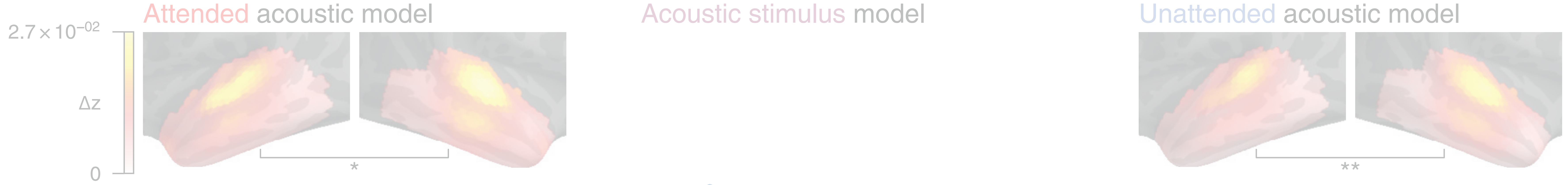
Brodbeck et al. (2018) *Rapid Transformation from Auditory to Linguistic Representations …*, Curr Biol

Language-feature based TRFs

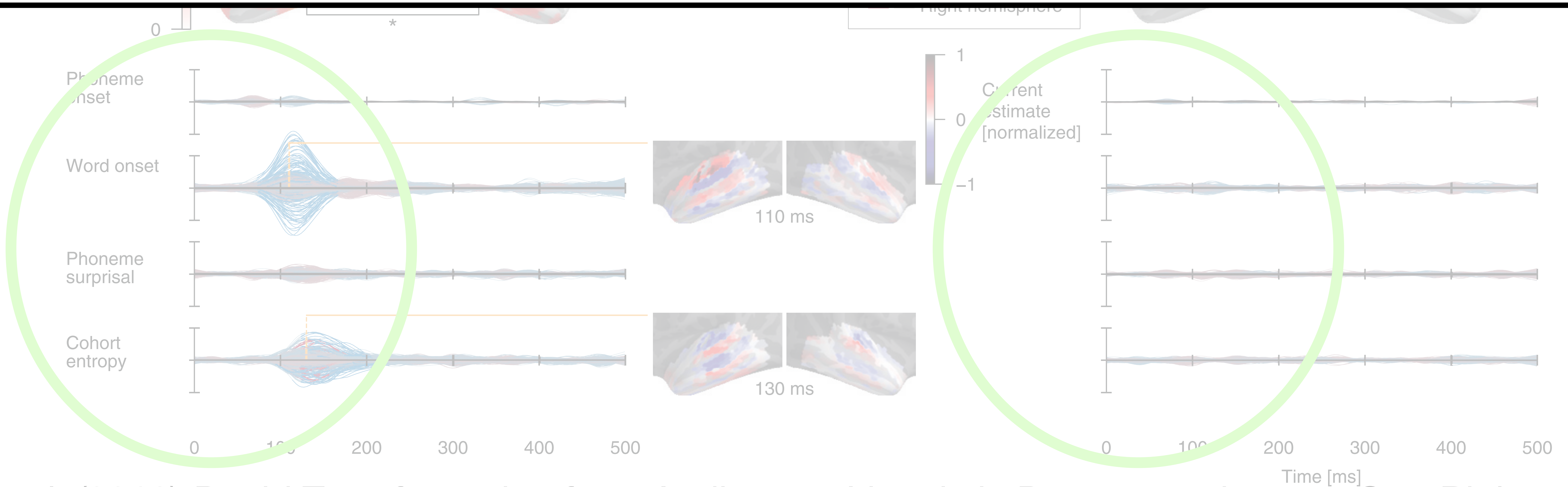


Brodbeck et al. (2018) *Rapid Transformation from Auditory to Linguistic Representations …*, Curr Biol

Attention + Language-feature based TRFs

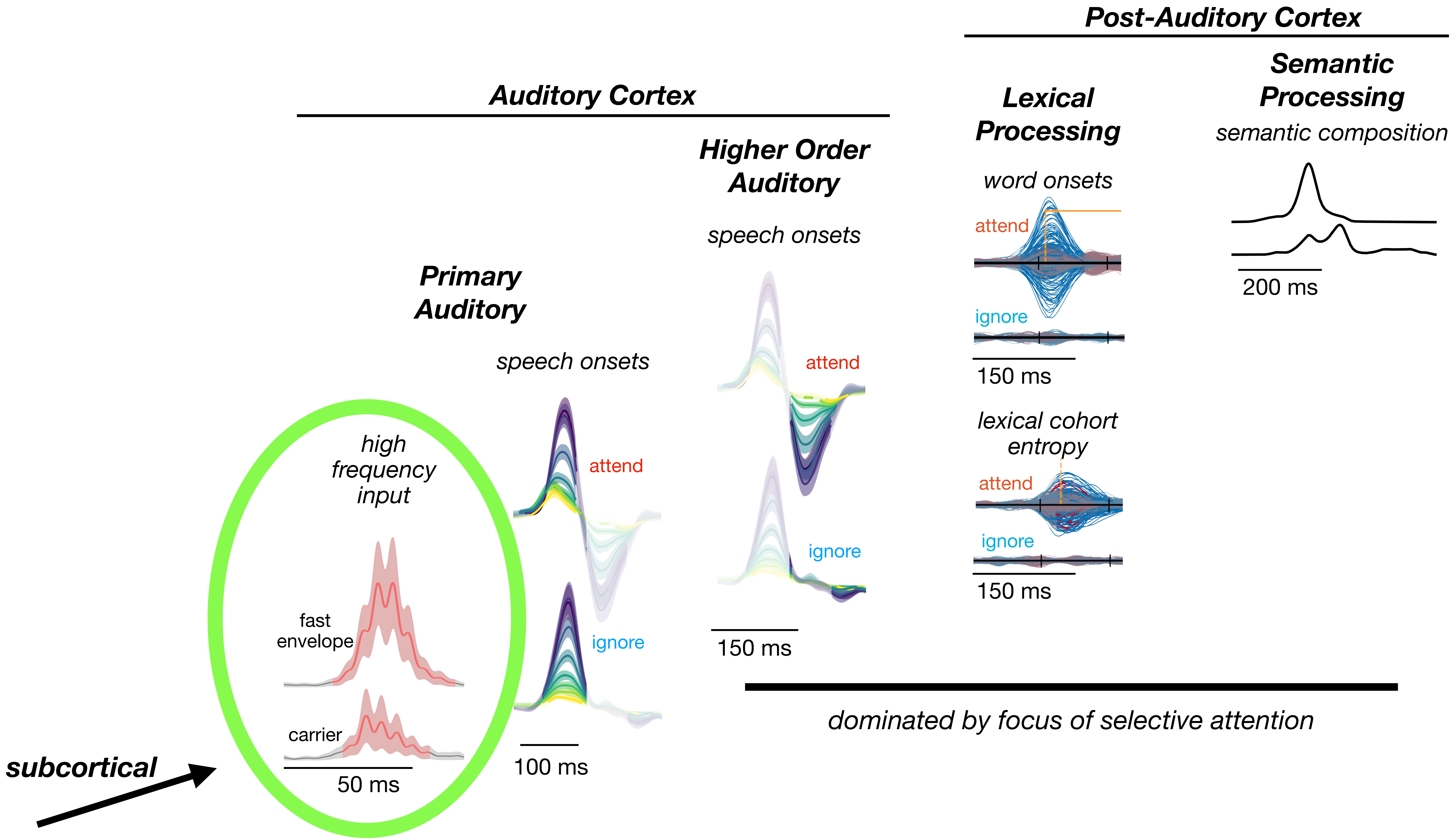


See also:
Gillis et al., (2021) bioRxiv
Neural Markers of Speech Comprehension: Measuring EEG Tracking of Linguistic Speech Representations, Controlling the Speech Acoustics

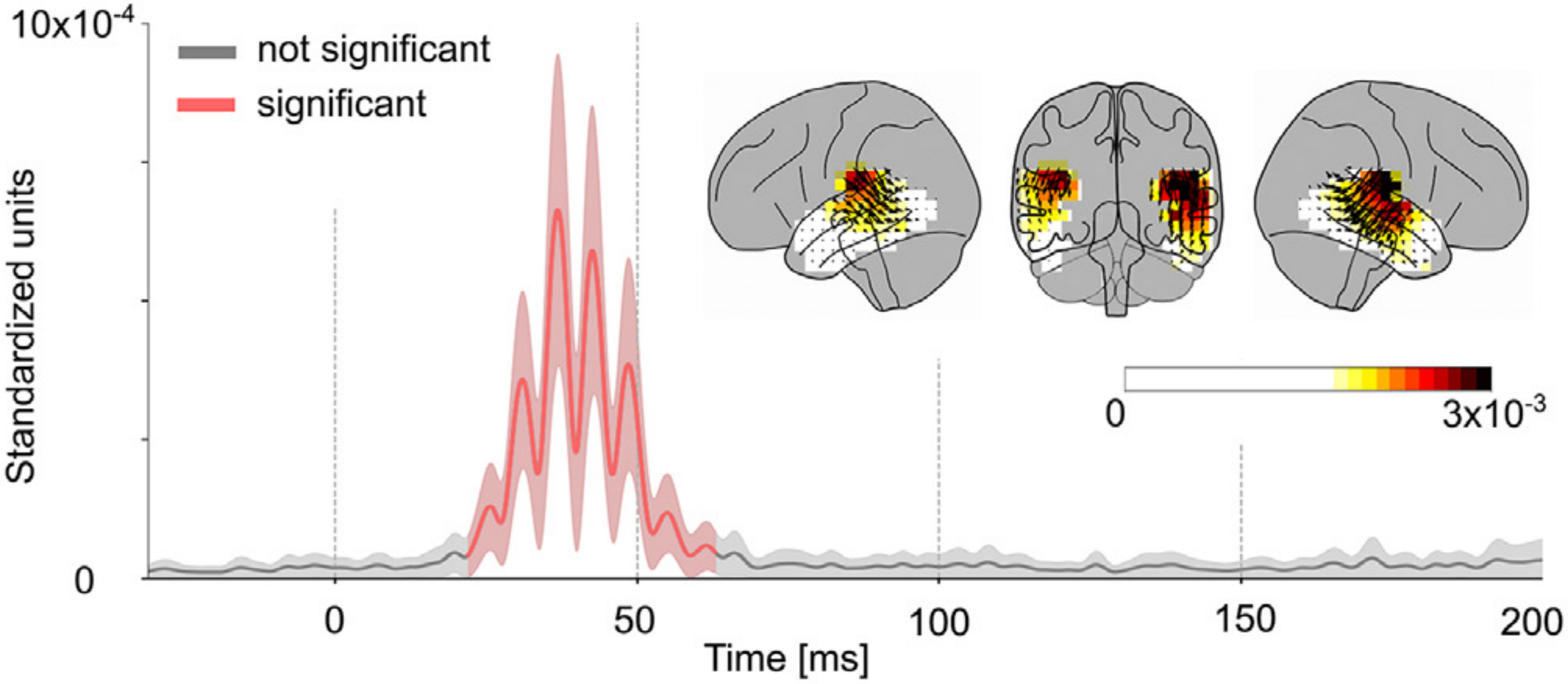


Brodbeck et al. (2018) *Rapid Transformation from Auditory to Linguistic Representations …*, Curr Biol

Cortical Representations Across Cortex

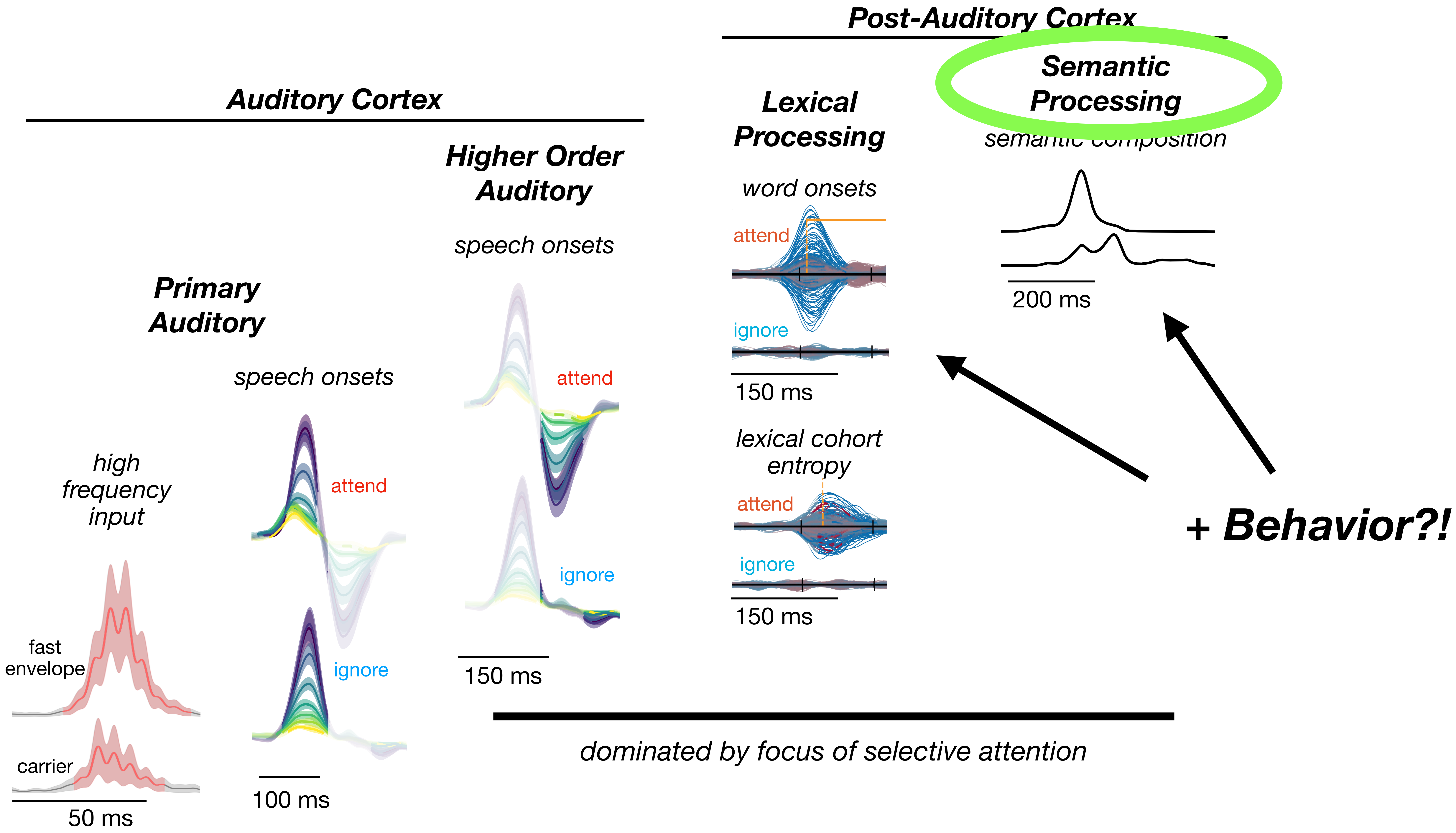


Fast & Early Cortical Representations



Kulasingham et al. (2020) *High Gamma Cortical Processing of Continuous Speech …*, NeuroImage

Cortical Representations Across Cortex



To Intelligibility, and Beyond?

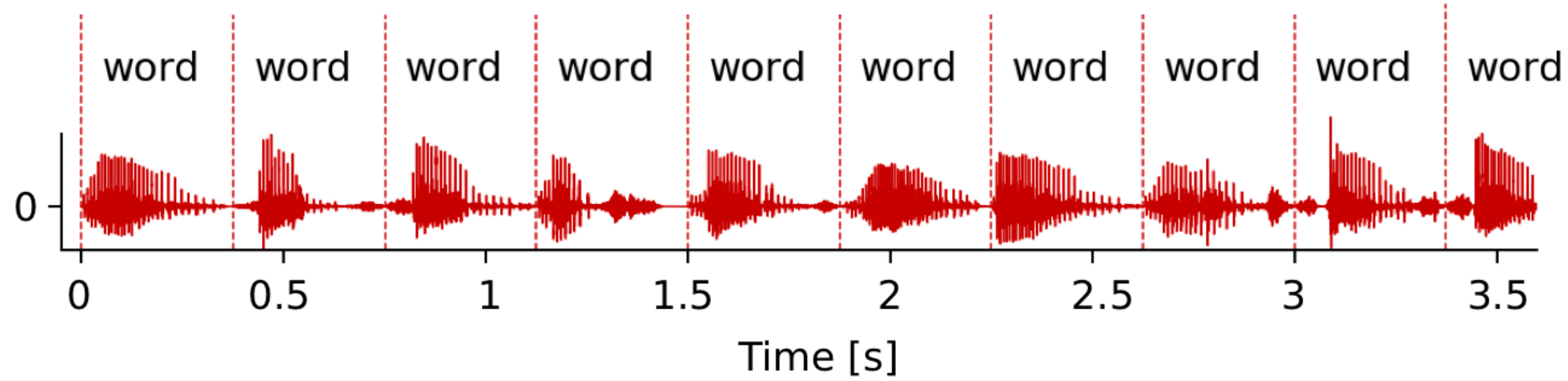
Cortical Representations of Speech Understanding

- Behavioral correlates of speech understanding
  - implies language comprehension
  - higher order comprehension (?)
    - sentence structure
    - other structures, e.g. poetic, logical
- Neural correlates of speech understanding
  - rhythms of higher order structures, even if *totally absent in the acoustics*
    - sentence structures

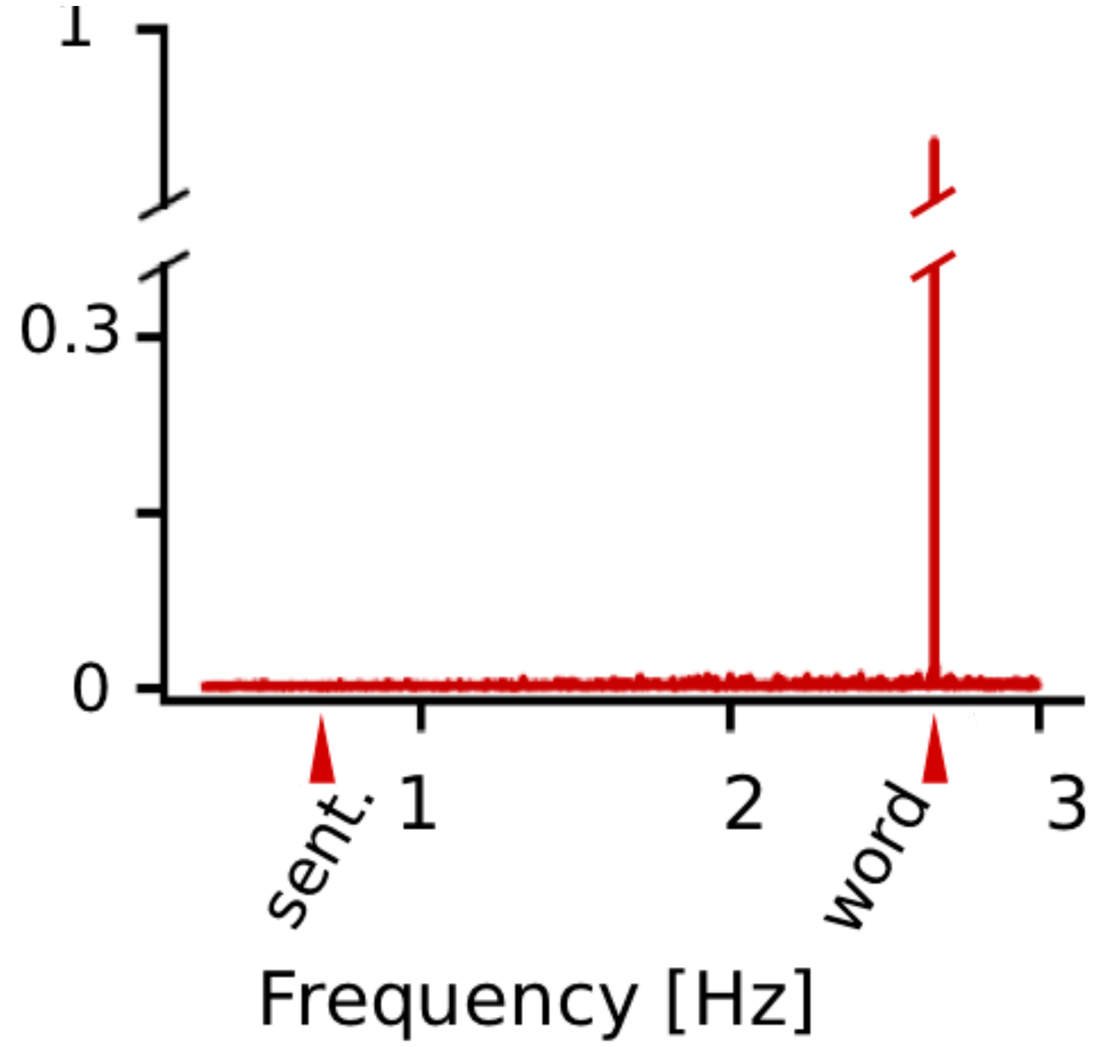
Ding et al., Nat Neurosci 2016

Isochronous Speech

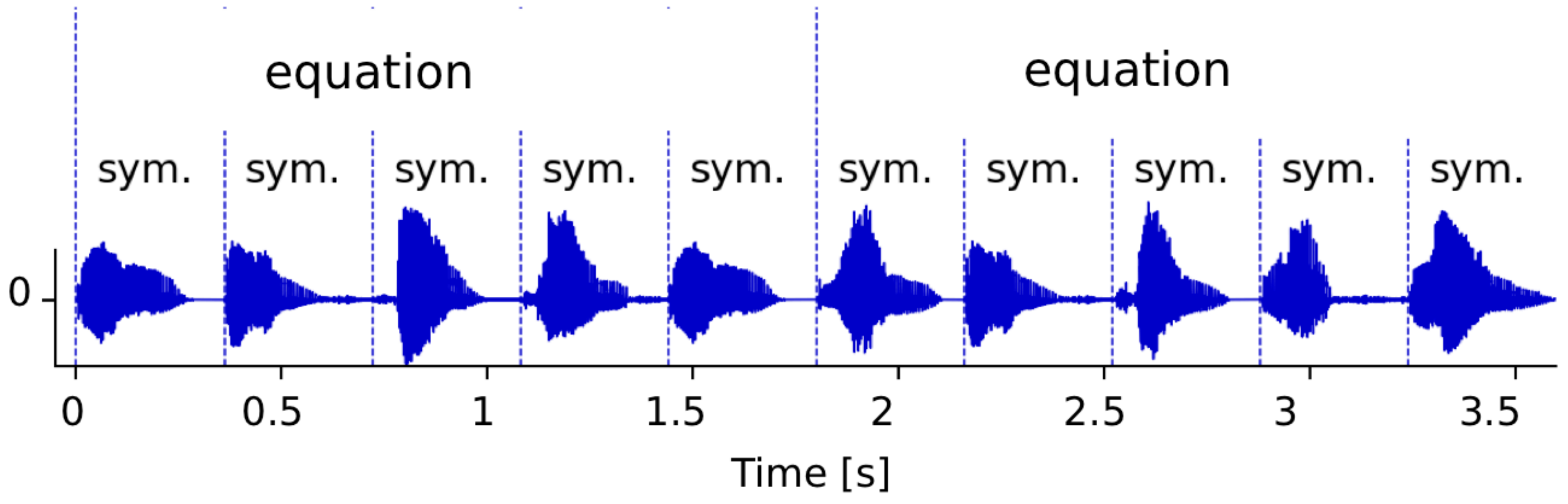
Acoustics



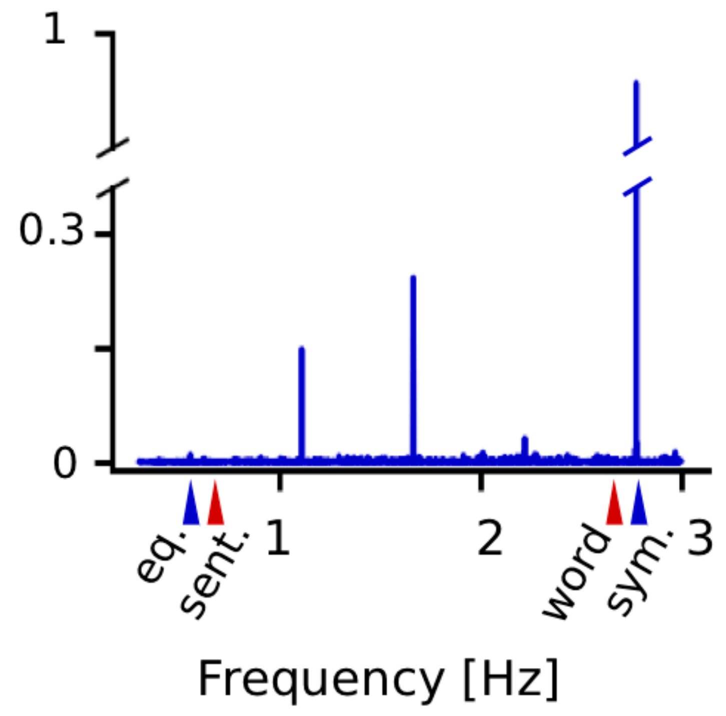
Acoustical Spectrum (envelope)



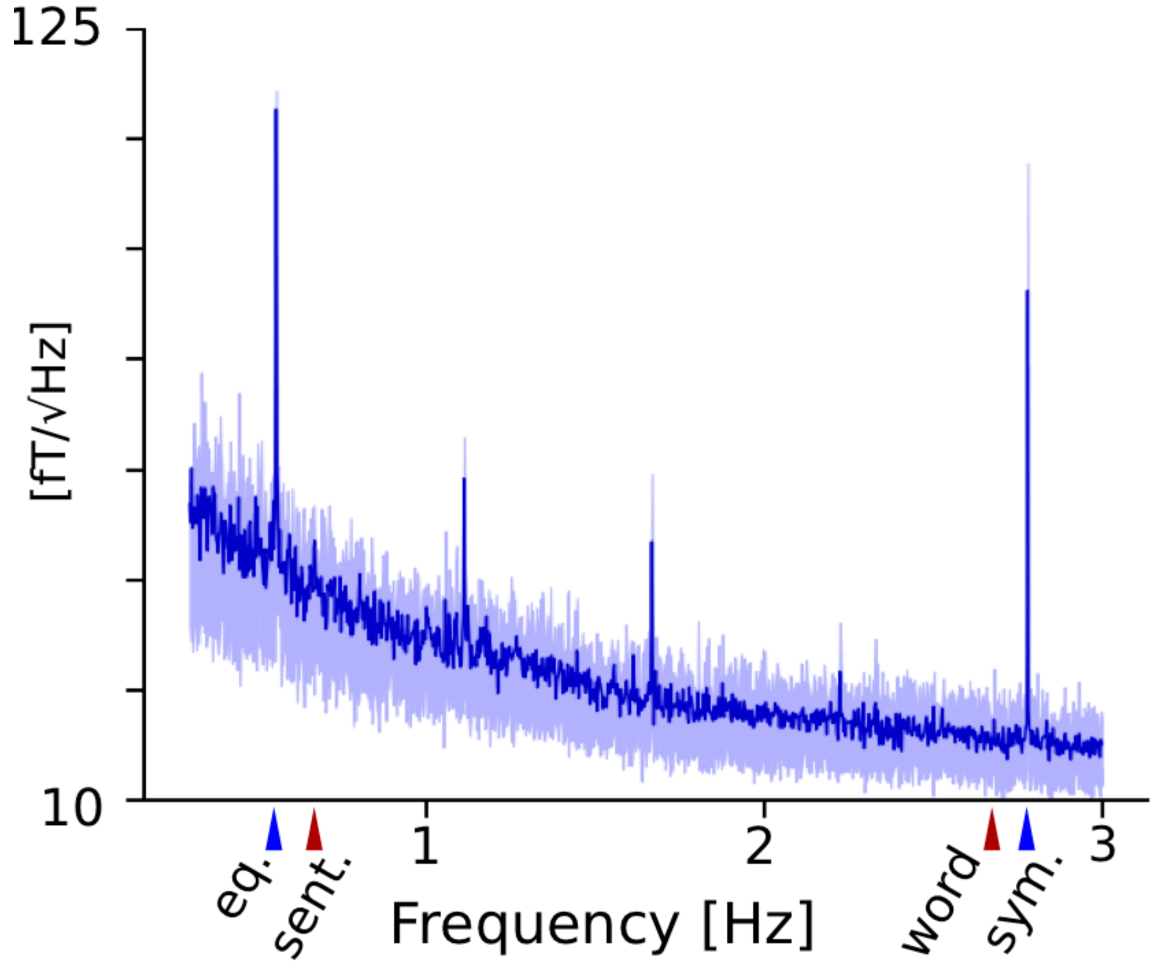
Isochronous Arithmetic



Acoustics



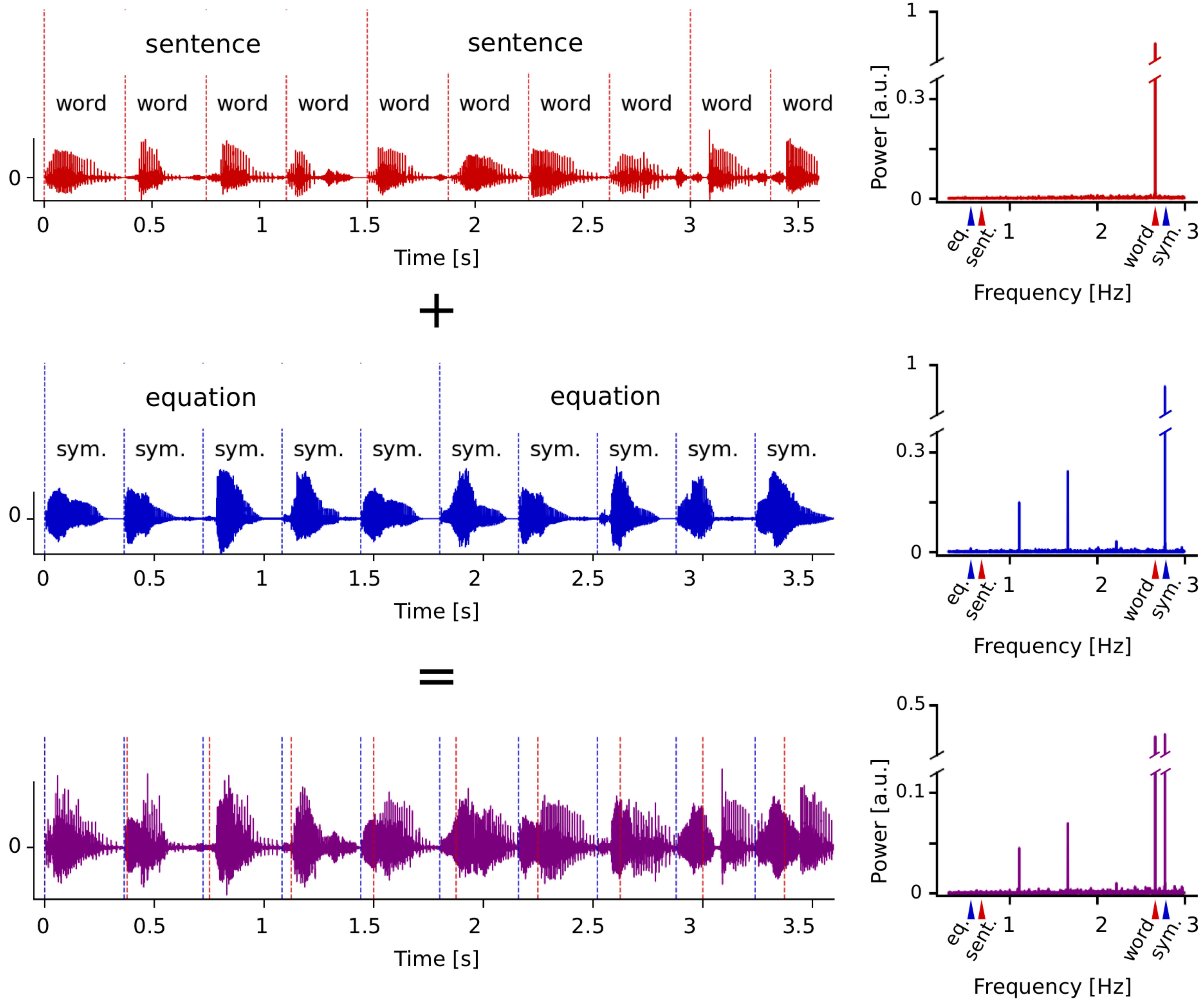
Acoustical Spectrum

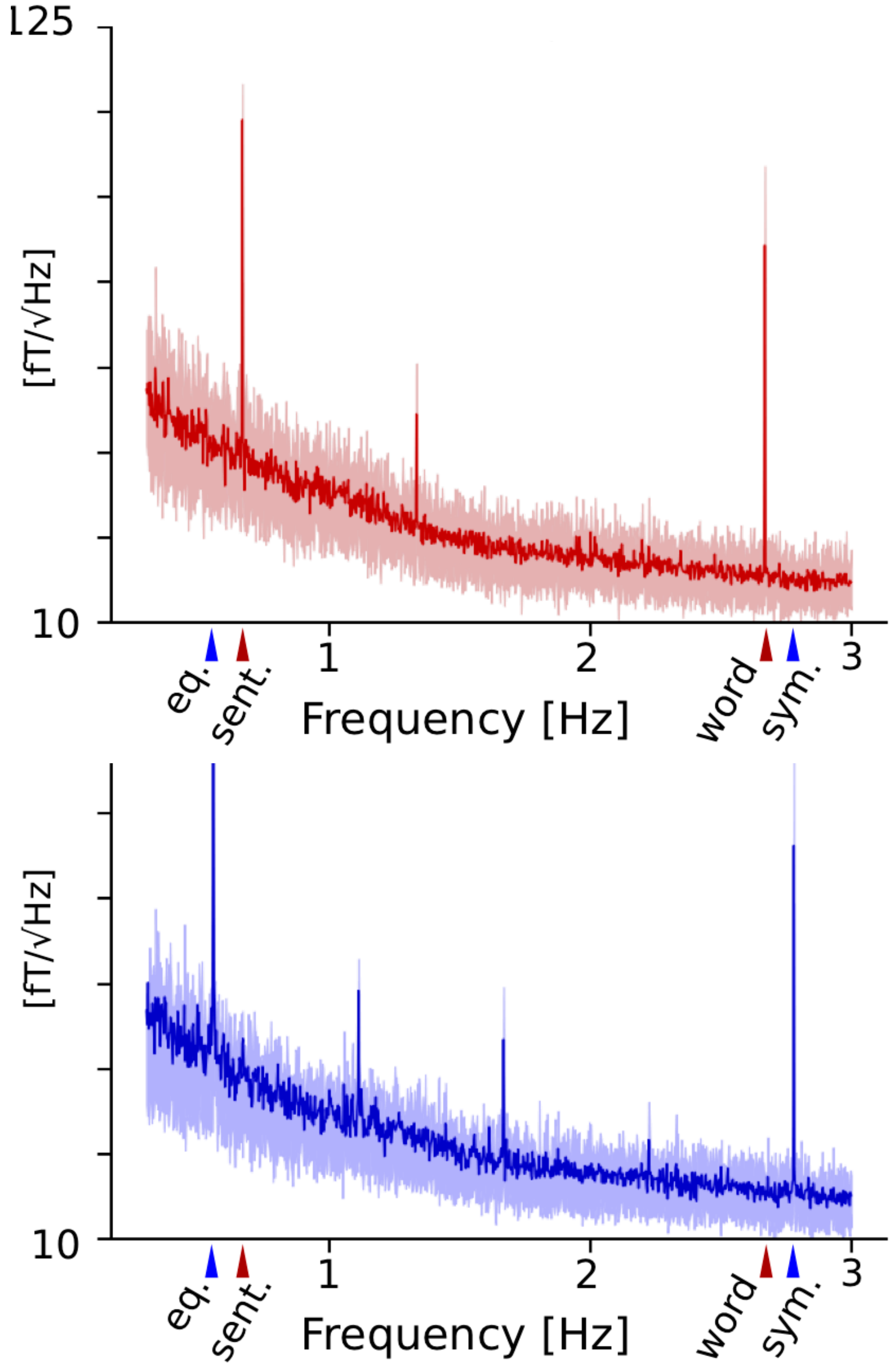


Neural Spectrum

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

Isochronous Cocktail Party





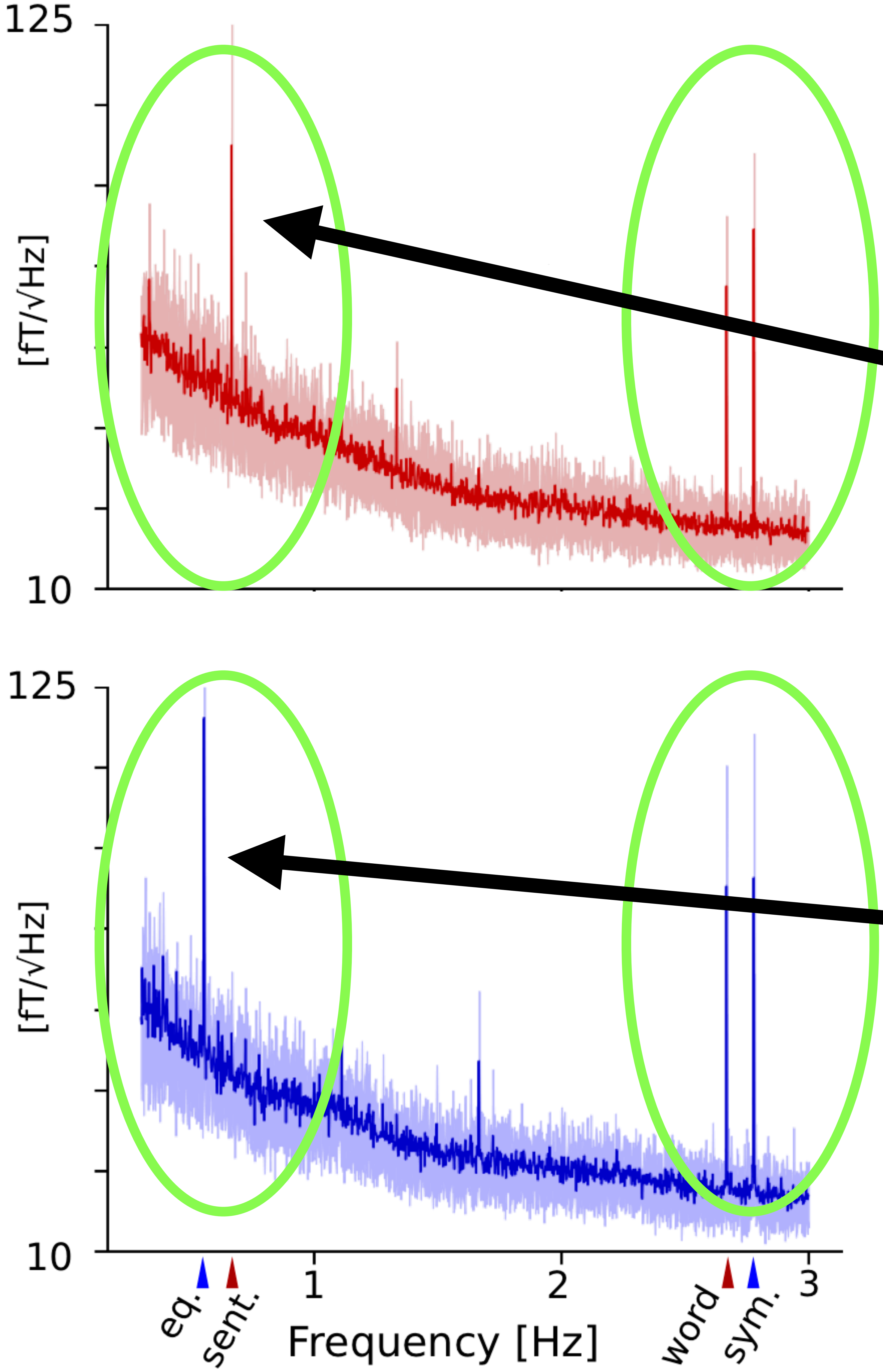
?

Isochronous Cocktail Party

Representations of Understanding

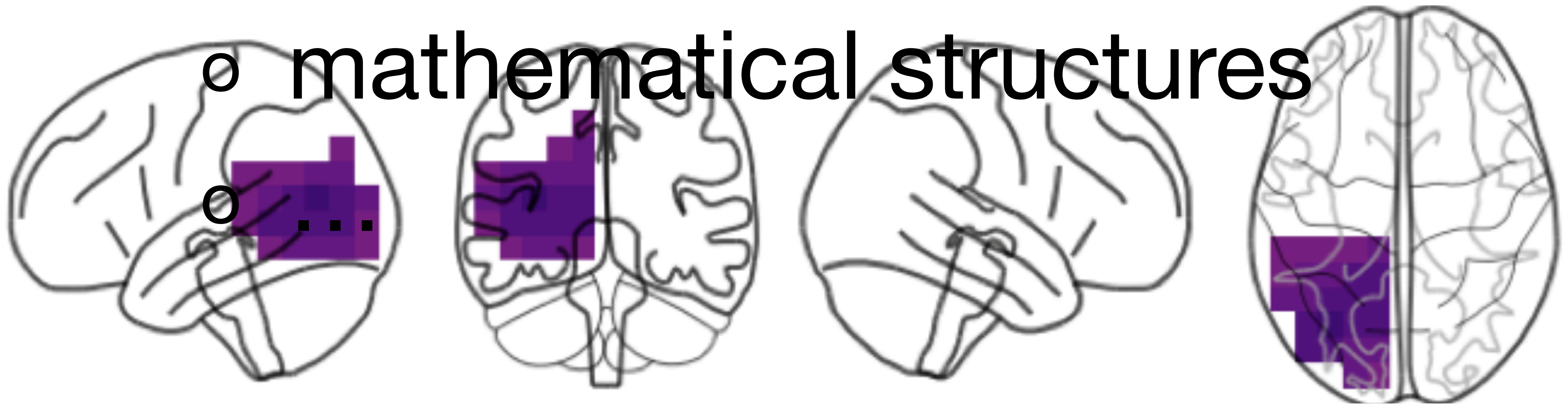
Attend to Sentences

Attend to Equations

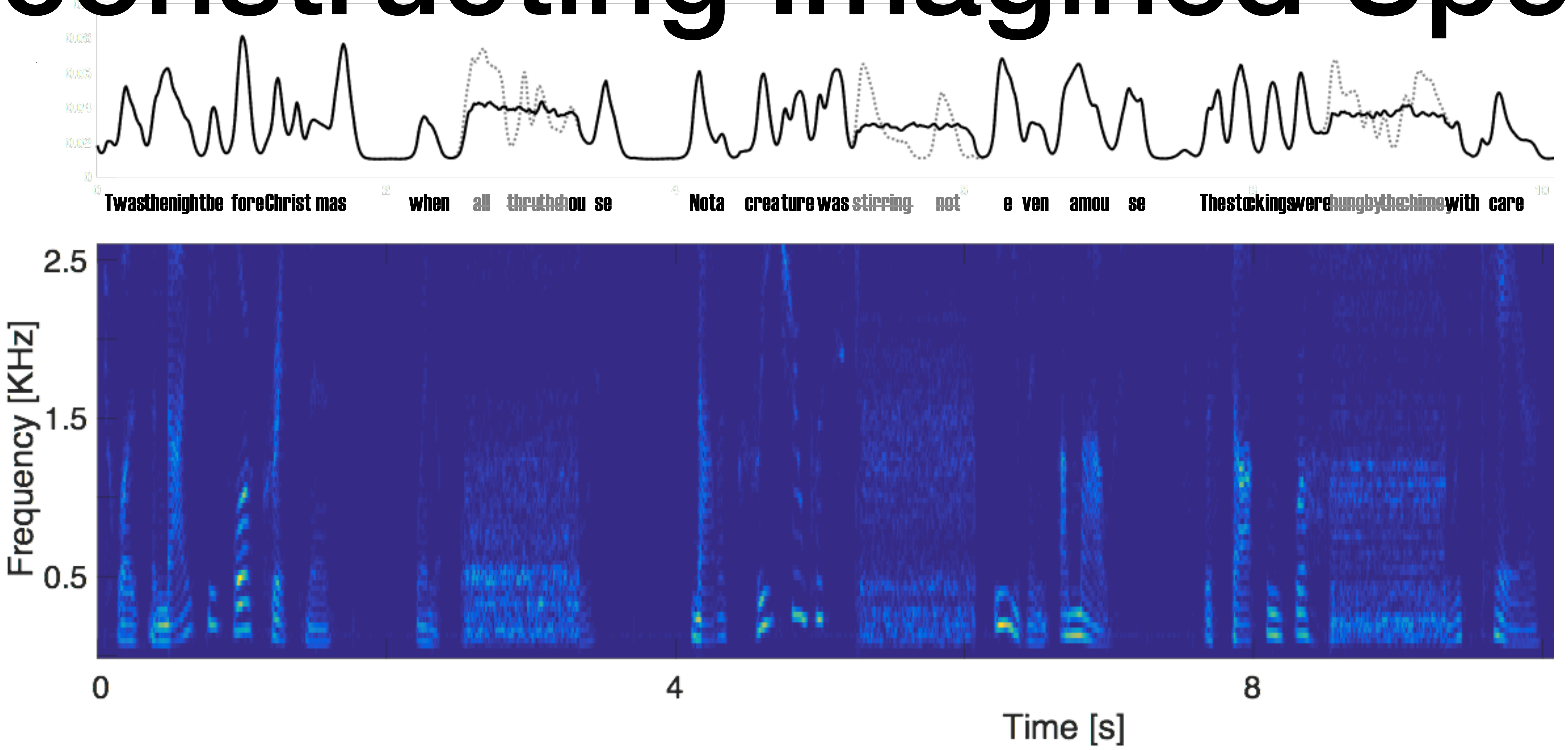


Neural Correlation with Behavior

- Neural correlates of understanding
  - rhythms of higher order structures
    - sentence structures
    - poetic structures
    - mathematical structures
    - …



Reconstructing Imagined Speech

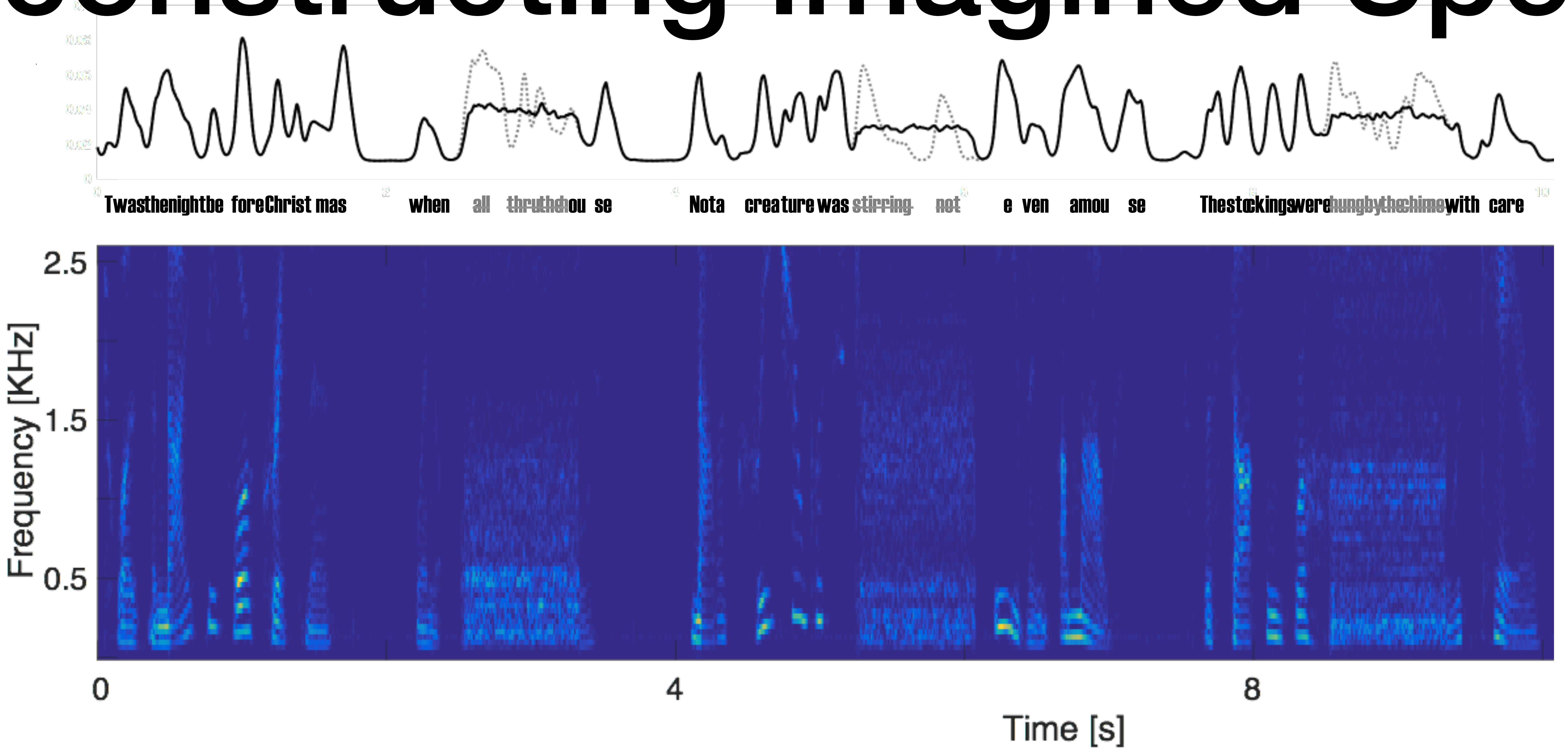


Can non-stationary, imagined speech be decoded?

- might be aided by contextual knowledge/familiarity
- might be aided by strong rhythmicity

Cervantes Constantino & Simon (2018) *Restoration … Neural Processing of Continuous Speech …*, Front Syst Neurosci

Reconstructing Imagined Speech



Can non-stationary, imagined speech be decoded?

- might be aided by contextual knowledge/familiarity
- might be aided by strong rhythmicity

Cervantes Constantino & Simon (2018) *Restoration … Neural Processing of Continuous Speech …*, Front Syst Neurosci

Neural "Reconstruction" & Familiarity

Twas the night before Christmas, when all through the house
not a creature was stirring, not even a mouse.
The stockings were hung by the chimney with care,
in hopes that St. Nicholas soon would be there.

The children were nestled all snug in their beds,
while visions of sugar plums danced in their heads.
And Mama in her 'kerchief, and I in my cap,
had just settled our brains for a long winter's nap.

When out on the lawn there arose such a clatter,
I sprang from my bed to see what was the matter.
Away to the window I flew like a flash,
tore open the shutter, and threw up the sash.

The moon on the breast of the new-fallen snow
gave the lustre of midday to objects below,
when, what to my wondering eyes should appear,
but a miniature sleigh and eight tiny reindeer.

With a little old driver, so lively and quick,
I knew in a moment it must be St. Nick.
More rapid than eagles, his coursers they came,
and he whistled and shouted and called them by name.

"Now Dasher! Now Dancer! Now, Prancer and Vixen!
On, Comet! On, Cupid! On, Donner and Blitzen!
To the top of the porch! To the top of the wall!
Now dash away! Dash away! Dash away all!"

As dry leaves that before the wild hurricane fly,
when they meet with an obstacle, mount to the sky
so up to the house-top the coursers they flew,
with the sleigh full of toys, and St. Nicholas too.

And then, in a twinkling, I heard on the roof
the prancing and pawing of each little hoof.
As I drew in my head and was turning around,
down the chimney St. Nicholas came with a bound.

He was dressed all in fur, from his head to his foot,
and his clothes were all tarnished with ashes and soot.
A bundle of toys he had flung on his back,
and he looked like a peddler just opening his pack.

His eyes--how they twinkled! His dimples, how merry!
His cheeks were like roses, his nose like a cherry!
His droll little mouth was drawn up like a bow,
and the beard on his chin was as white as the snow.

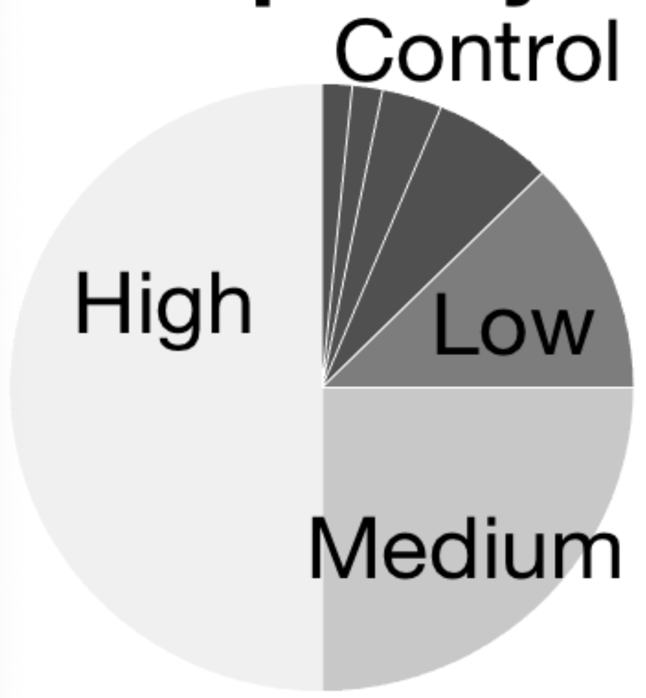
The stump of a pipe he held tight in his teeth,
and the smoke it encircled his head like a wreath.
He had a broad face and a little round belly,
that shook when he laughed, like a bowl full of jelly.

He was chubby and plump, a right jolly old elf,
and I laughed when I saw him, in spite of myself.
A wink of his eye and a twist of his head
soon gave me to know I had nothing to dread.

He spoke not a word, but went straight to his work,
and filled all the stockings, then turned with a jerk.
And laying his finger aside of his nose,
and giving a nod, up the chimney he rose.

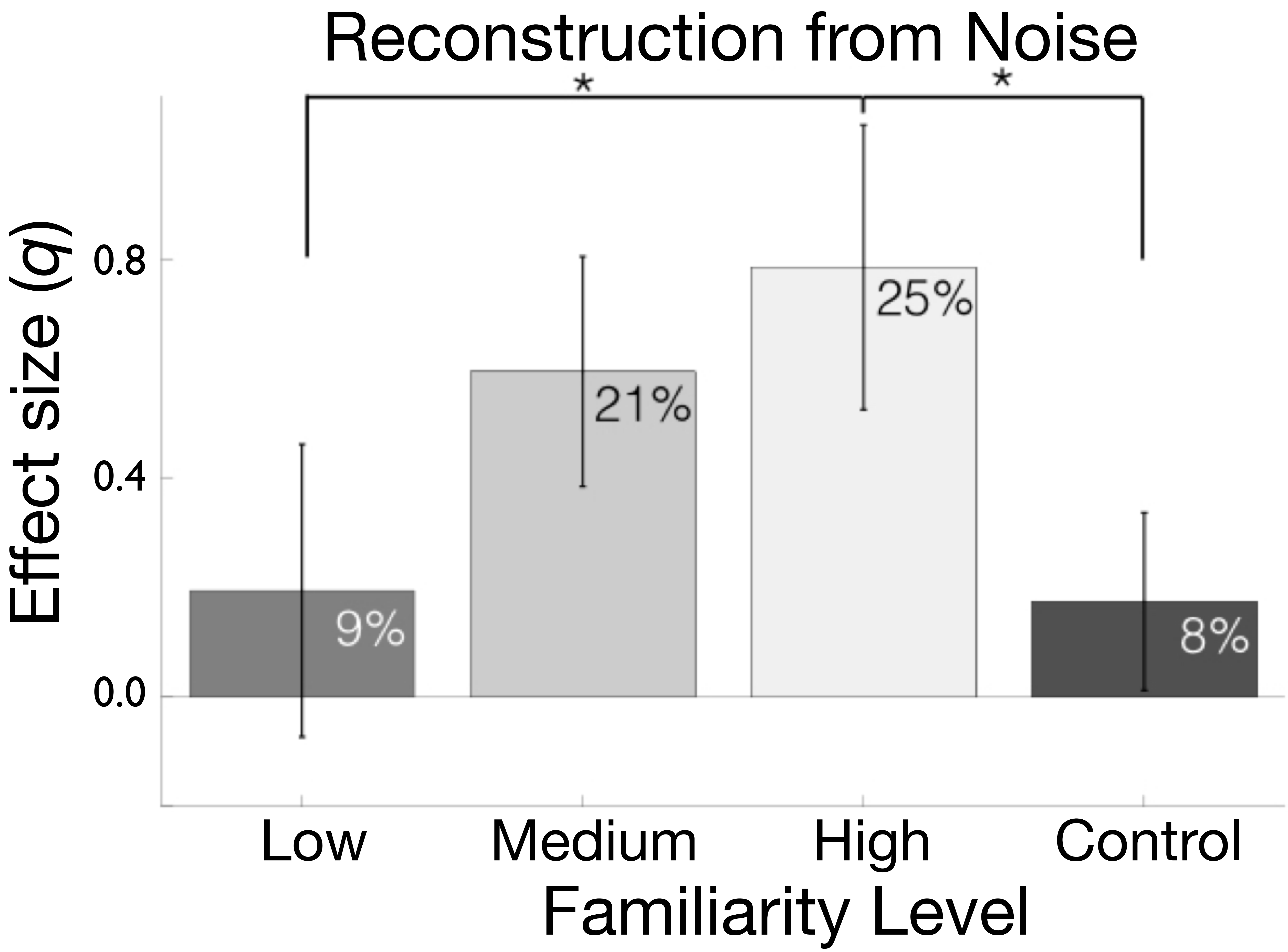
He sprang to his sleigh, to his team gave a whistle,
And away they all flew like the down of a thistle.
But I heard him exclaim, 'ere he drove out of sight,
"Happy Christmas to all, and to all a good night!"

Replay frequency



Controlling contextual knowledge of missing speech by exposure to the speech

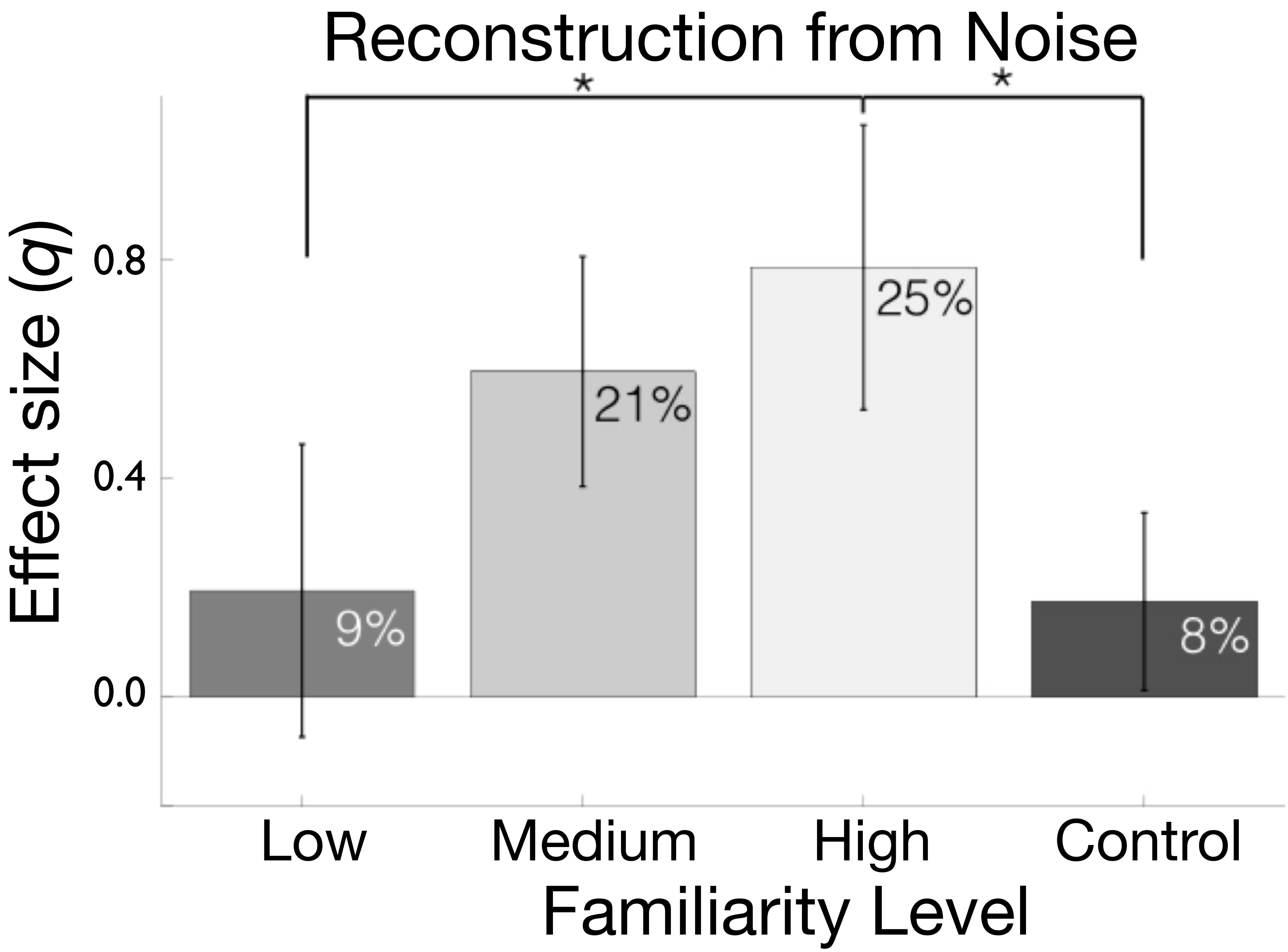
Imagined Speech "Reconstruction"



- Decoding of the ***missing*** speech token improves with prior experience
- Performance is a considerable fraction of that for clean speech

Cervantes Constantino & Simon (2018) *Restoration … Neural Processing of Continuous Speech …*, Front Syst Neurosci

Imagined Speech "Restoration"?

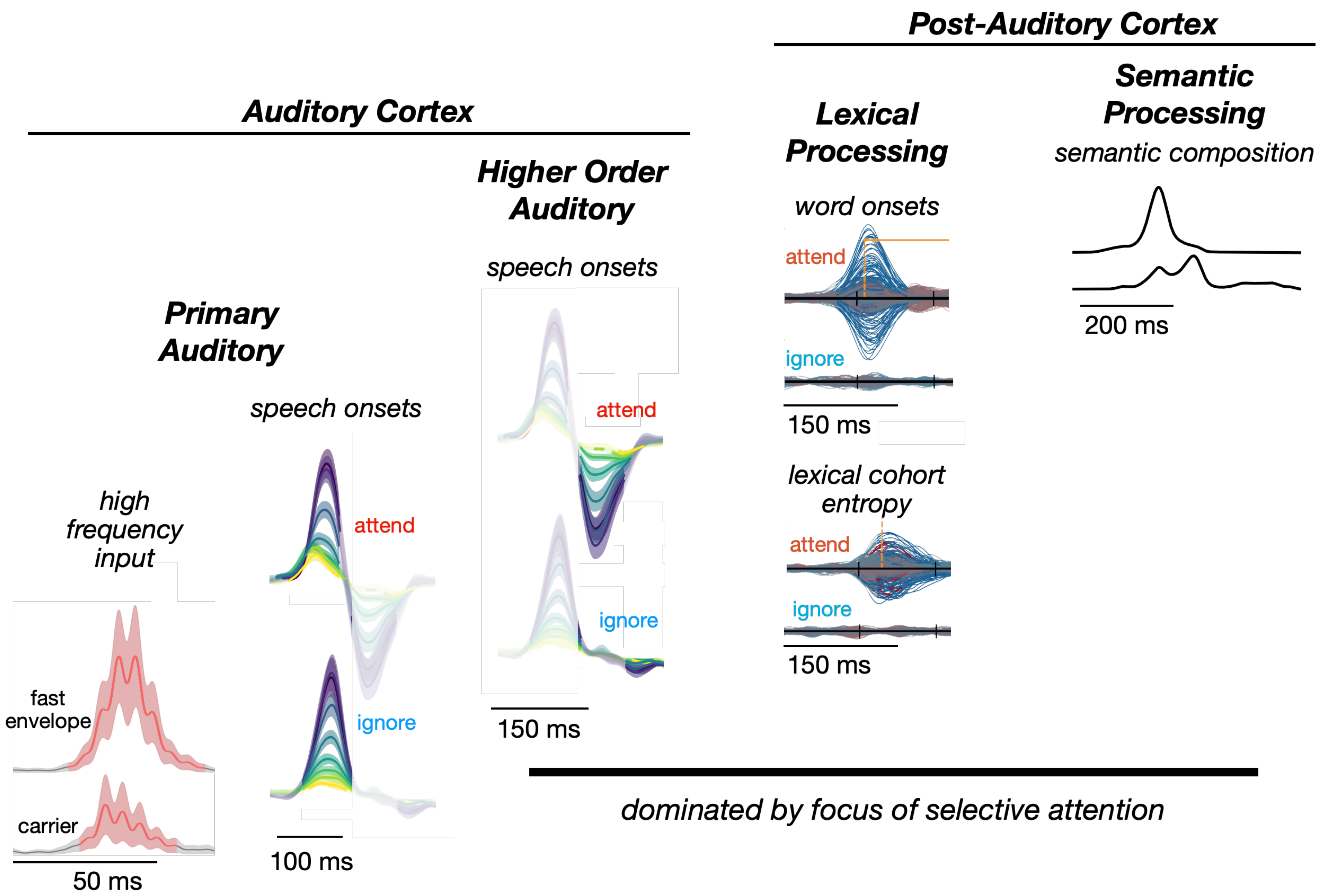


Cervantes Constantino & Simon (2018) *Restoration … Neural Processing of Continuous Speech …*, Front Syst Neurosci

Summary

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

- *Continuous speech* allows acquiring entire hierarchy from same stimulus
- Using *simultaneous TRFs* allows segregation of neural processes
- How is each process linked to intelligibility/understanding?
- Which links are predictive/causal?



thank you

These slides
available at:
ter.ps/simonpubs

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