Towards Objective Measures of Speech Perception

Jonathan Z. Simon

Department of Electrical & Computer Engineering

Department of Biology

Institute for Systems Research

University of Maryland

Objective Measures of Speech Perception

- What do I mean by ***objective measure***?
  - EEG/MEG measures of cortical activity
  - Stimulus: naturalistic, long-duration speech
  - Not addressed here:
    - subcortical activity
    - other non-invasive measures (fNIRS, fMRI)
    - other forms of speech

Objective Measures of Speech Perception

- What do I mean by ***speech perception***?
  - Beyond intelligibility
  - Allow for role of cognition
  - Role of attention
  - Importance of language in speech perception
  - Importance of speech meaning (semantics)
  - Processing effort? (not addressed here)

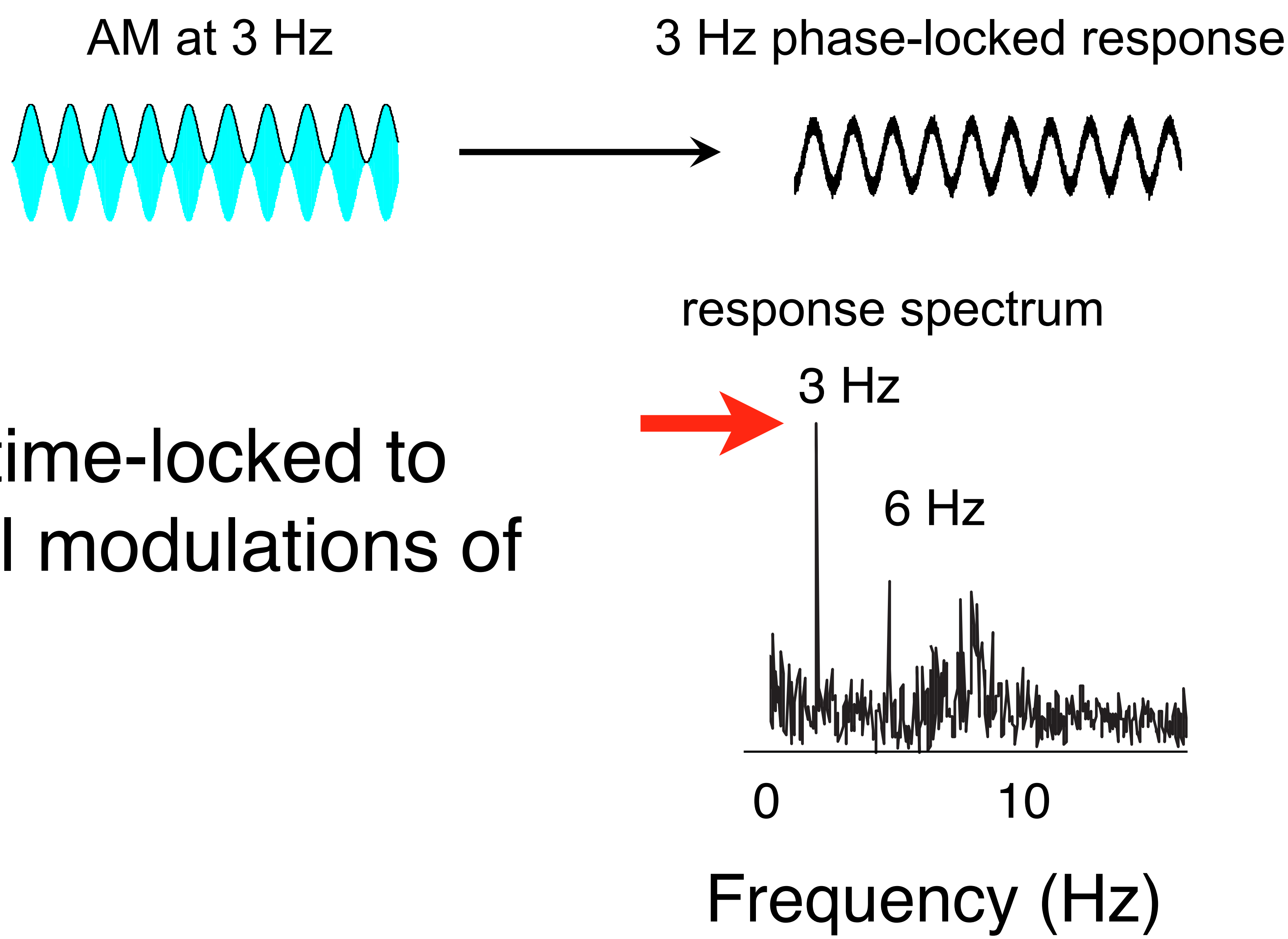
Outline

- Background & motivation
  - Neural responses in time
  - Response prediction from a stimulus via Temporal Response Function (TRF)
  - Stimulus reconstruction from responses
- Towards objective measures of
  - Speech intelligibility
  - Lexical processing of speech
  - Semantic processing of speech

Outline

- Background & motivation
  - Neural responses in time
  - Response prediction from a stimulus via Temporal Response Function (TRF)
  - Stimulus reconstruction from responses
- Towards objective measures of
  - Speech intelligibility
  - Lexical processing of speech
  - Semantic processing of speech

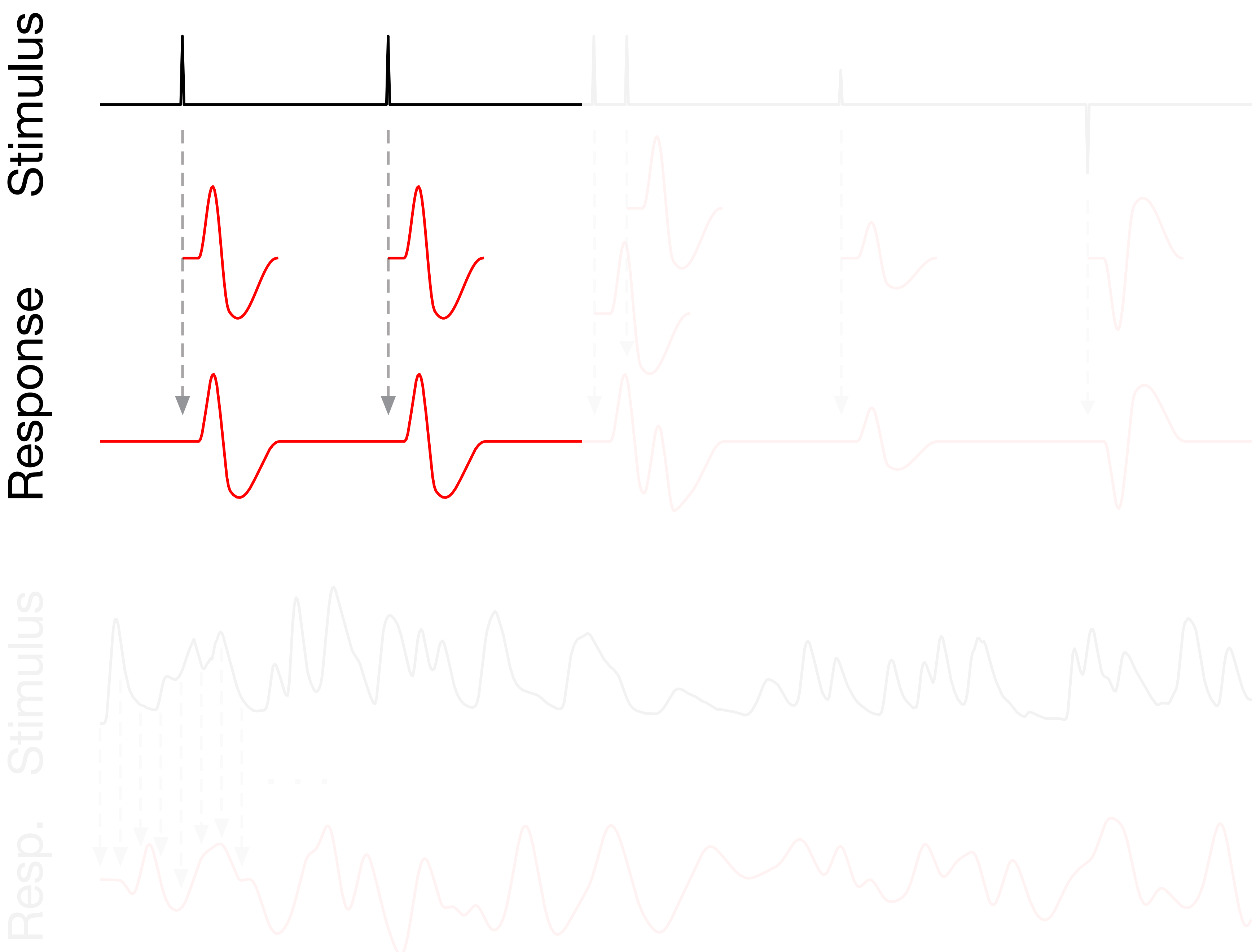
EEG & MEG Responses in Time



Activity time-locked to temporal modulations of sounds

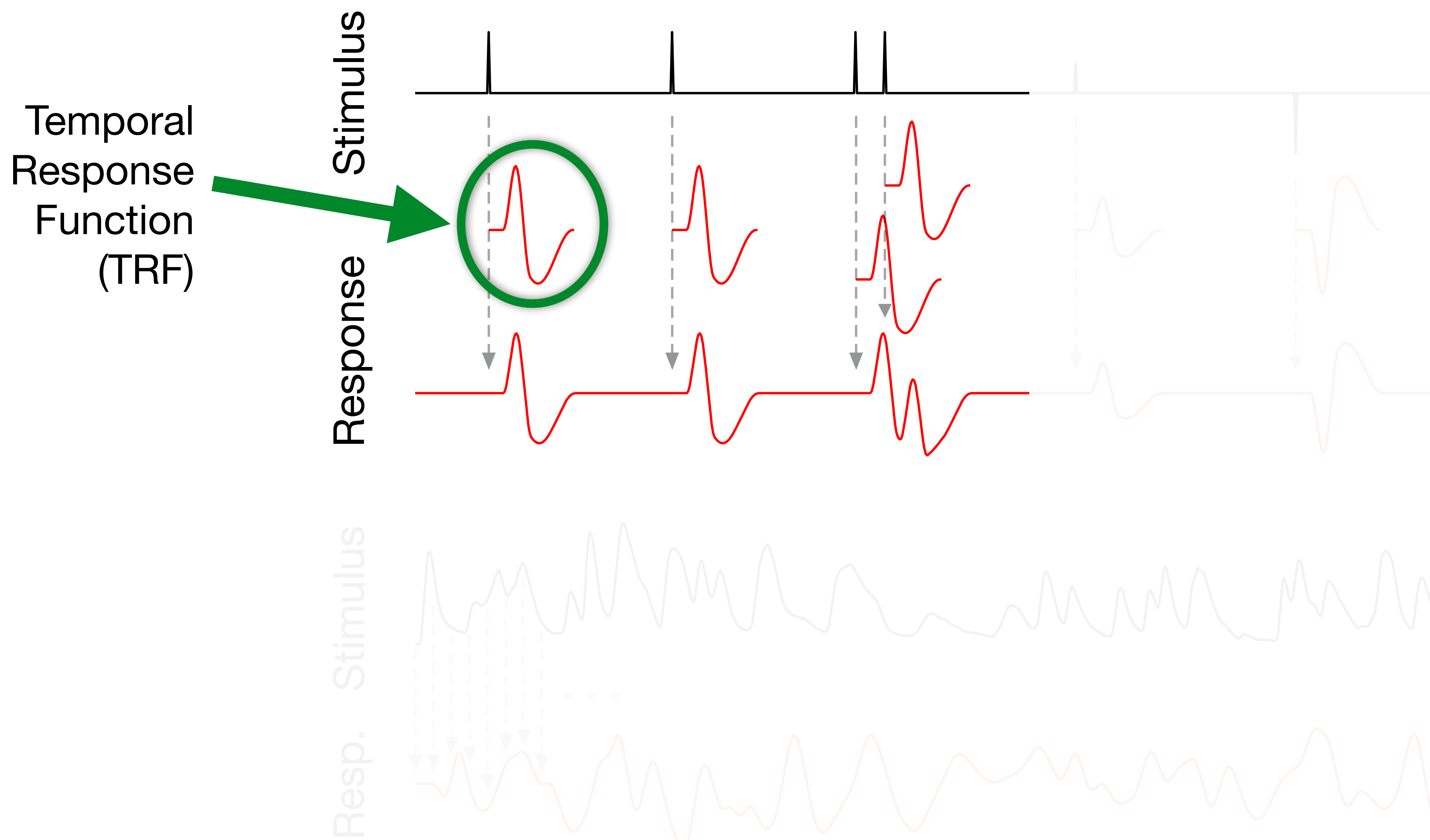
Ding & Simon, J Neurophysiol (2009)

Predicting EEG/MEG Responses

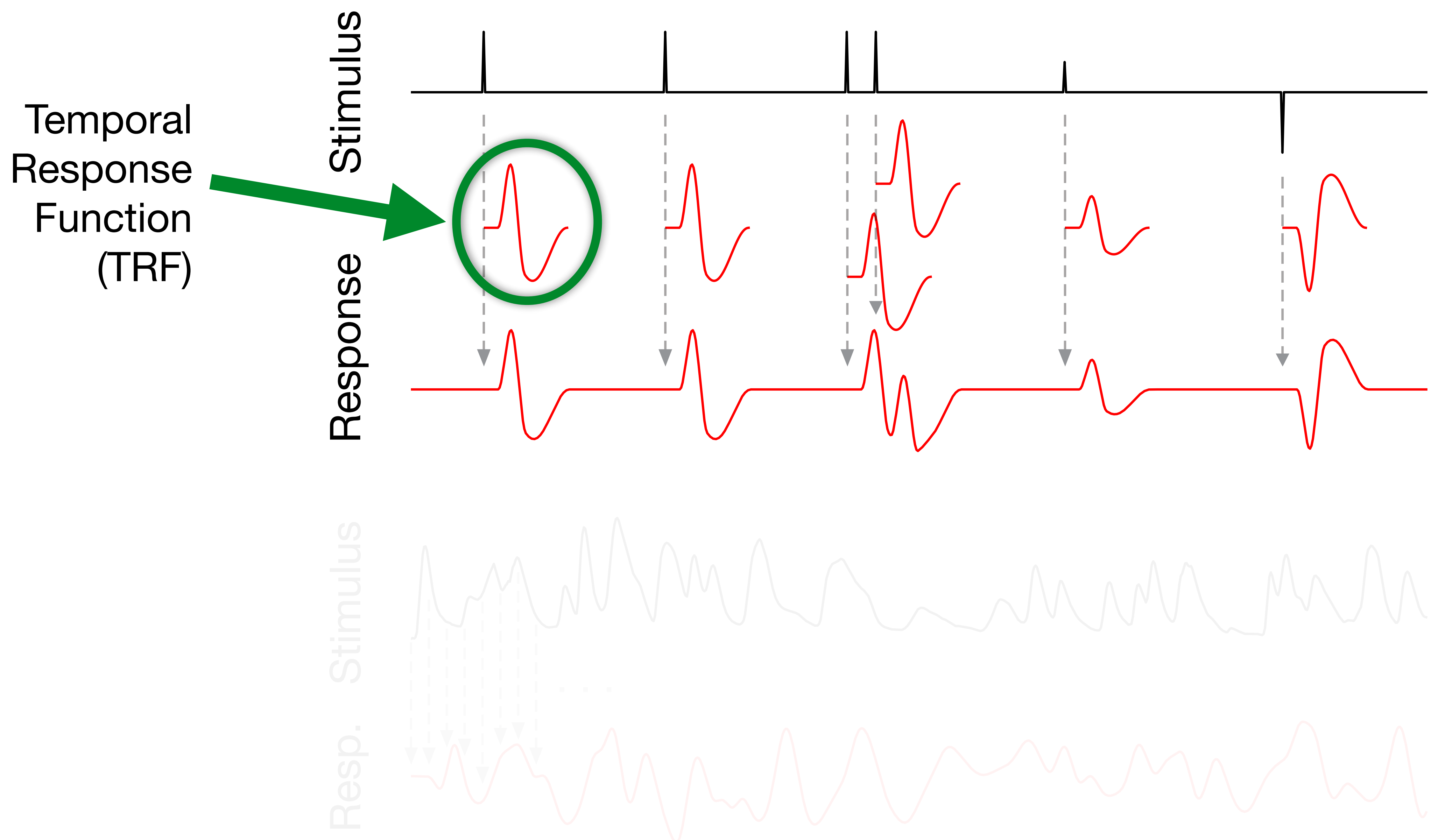


Predicting EEG/MEG Responses

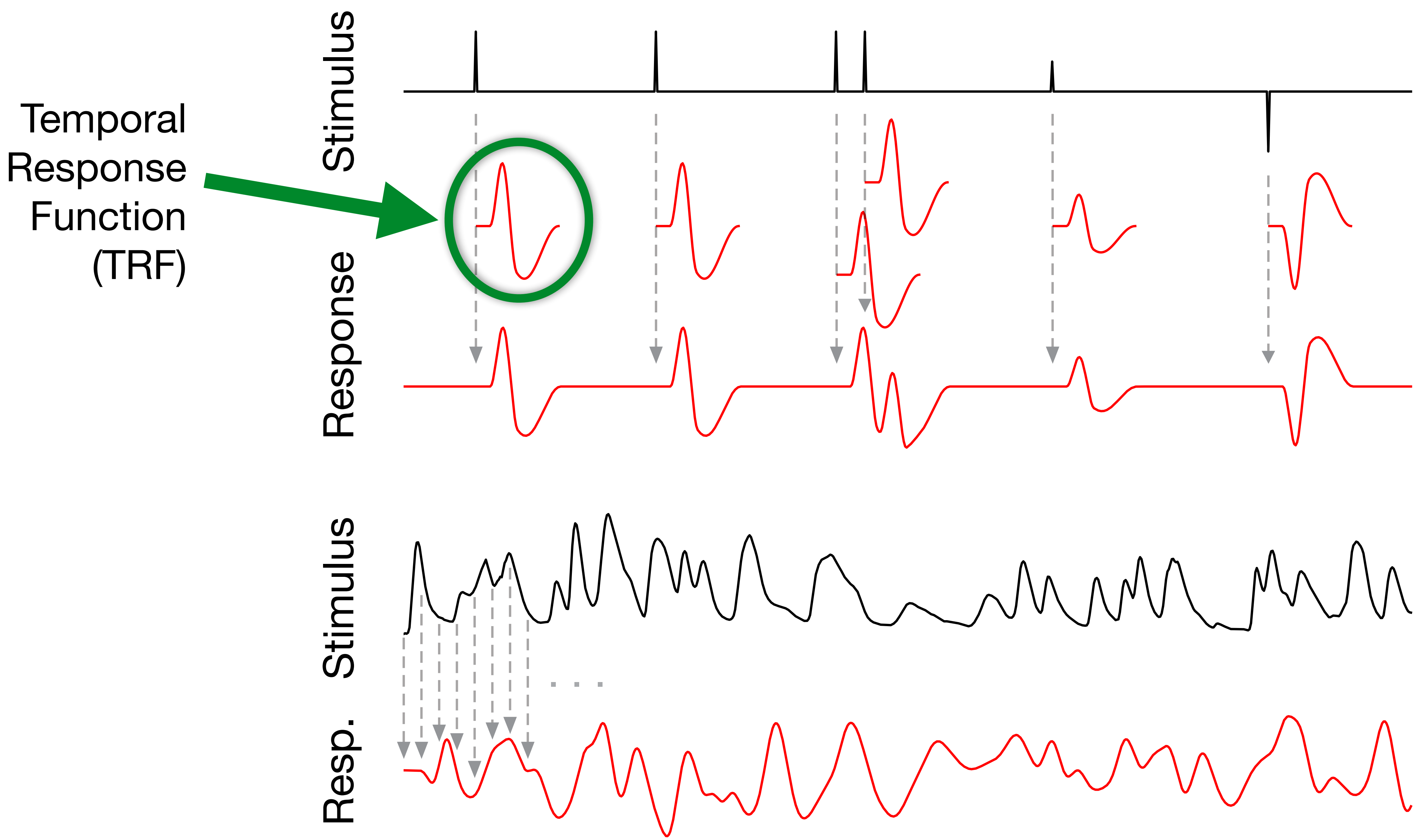
Predicting EEG/MEG Responses



Predicting EEG/MEG Responses



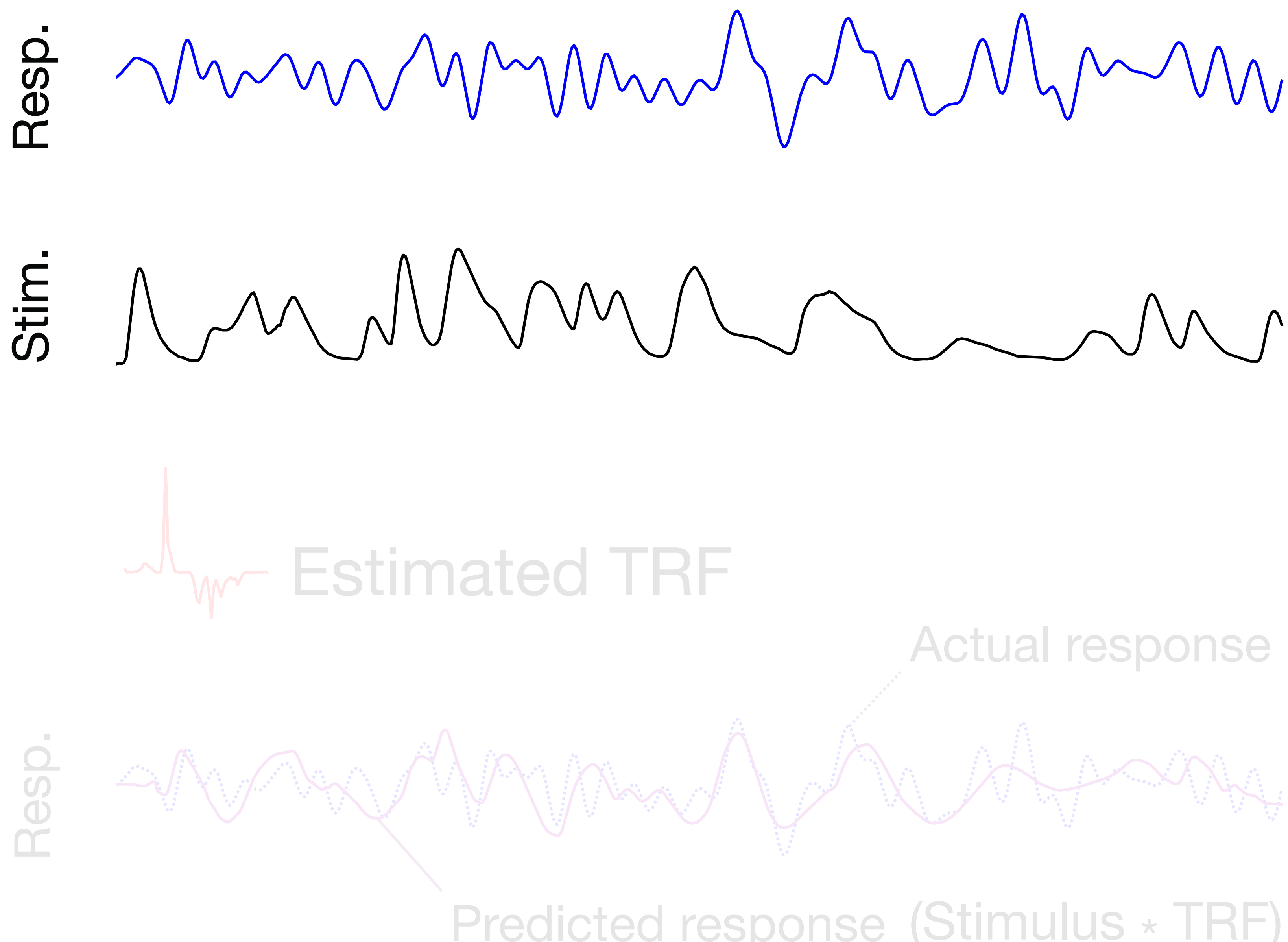
Predicting EEG/MEG <u>Responses</u>



Predicting EEG/MEG Responses

Temporal Response Function (TRF) estimation:

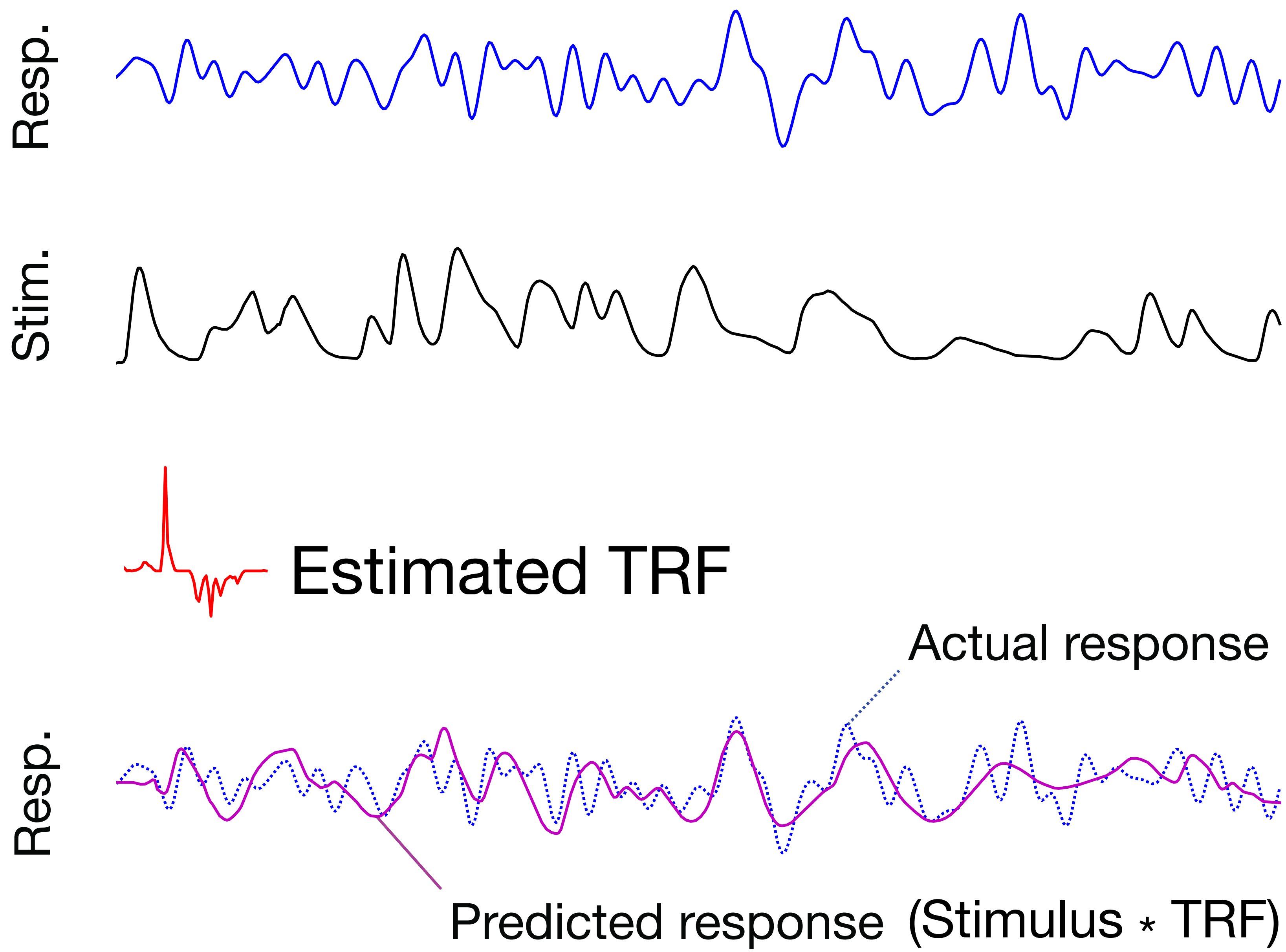
Stimulus and response are known; find the best TRF to produce the response from the stimulus:



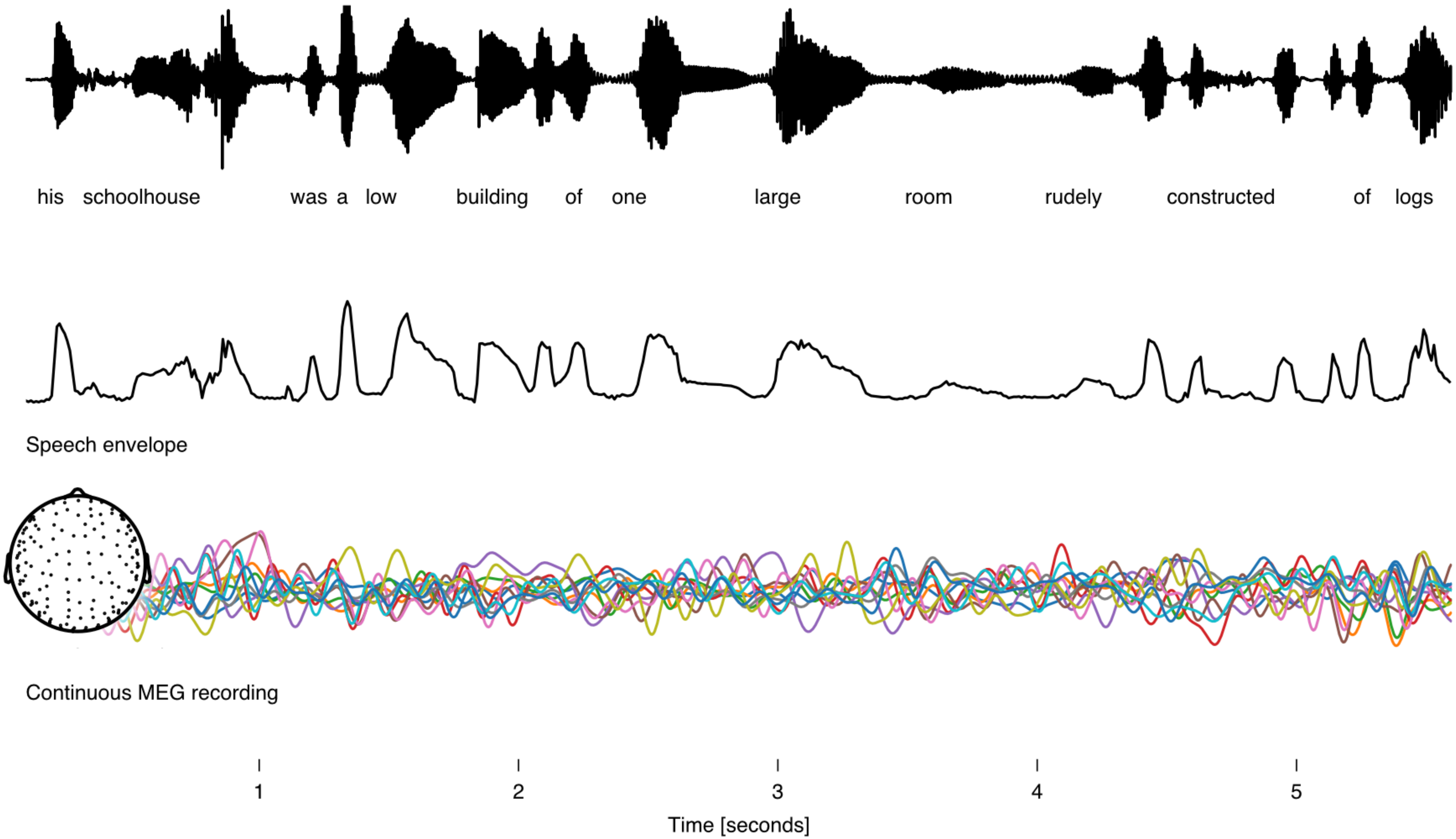
Predicting EEG/MEG Responses

Temporal Response Function (TRF) estimation:

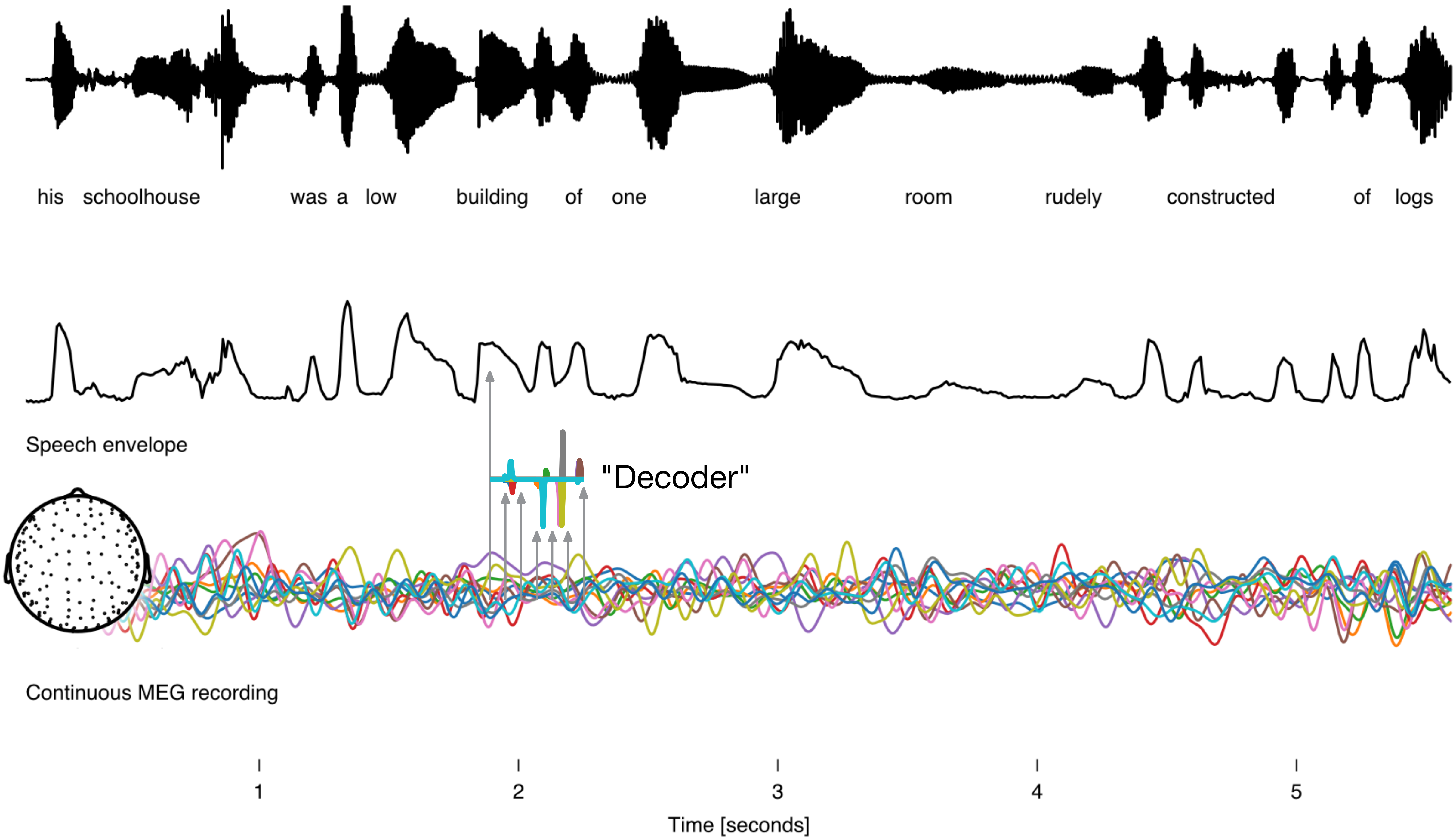
Stimulus and response are known; find the best TRF to produce the response from the stimulus:



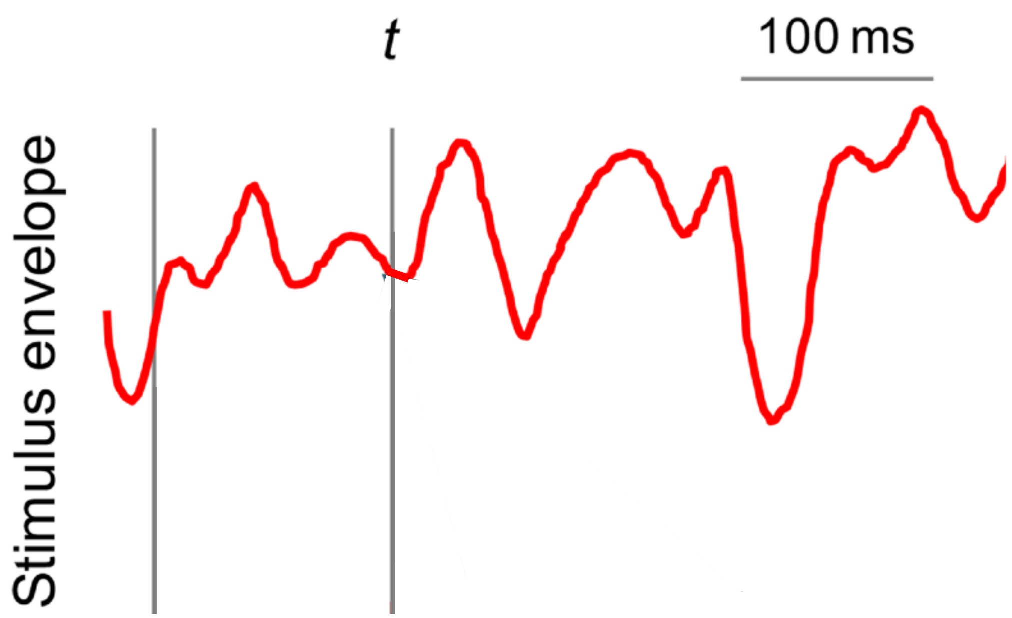
Stimulus Reconstruction in Time



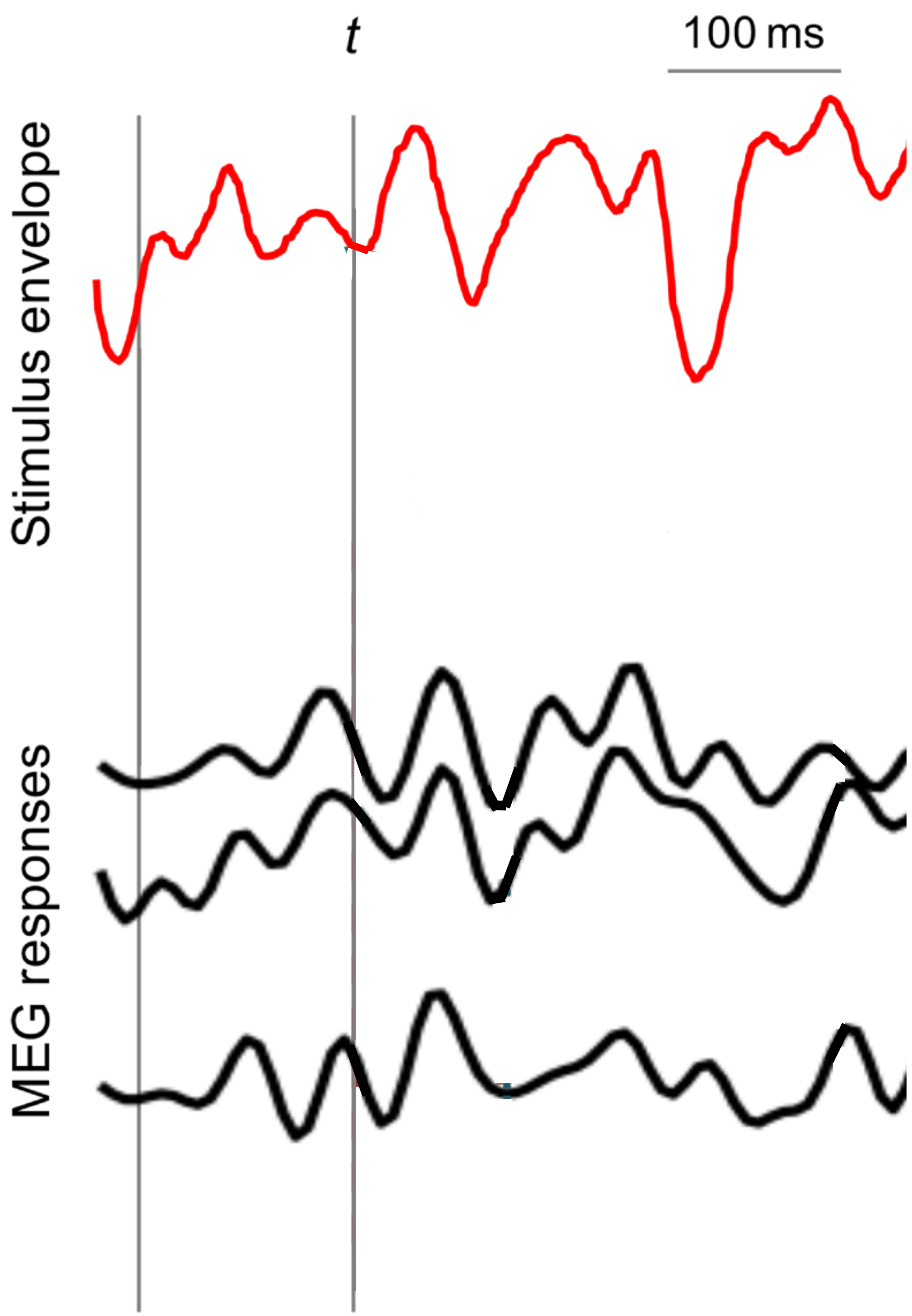
Stimulus Reconstruction in Time



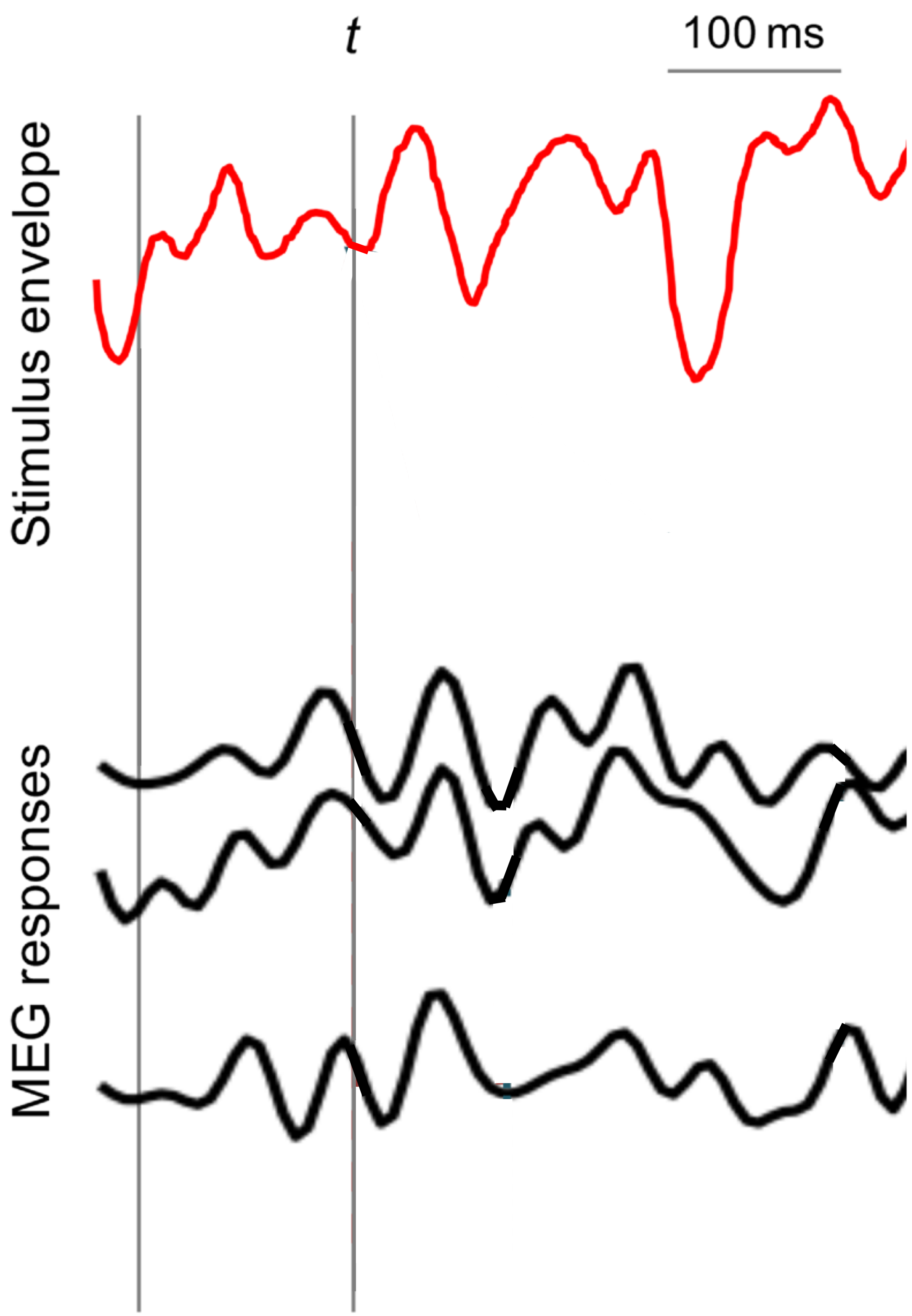
Stimulus Reconstruction in Time



<u>Stimulus</u> Reconstruction in Time



Stimulus Reconstruction in Time

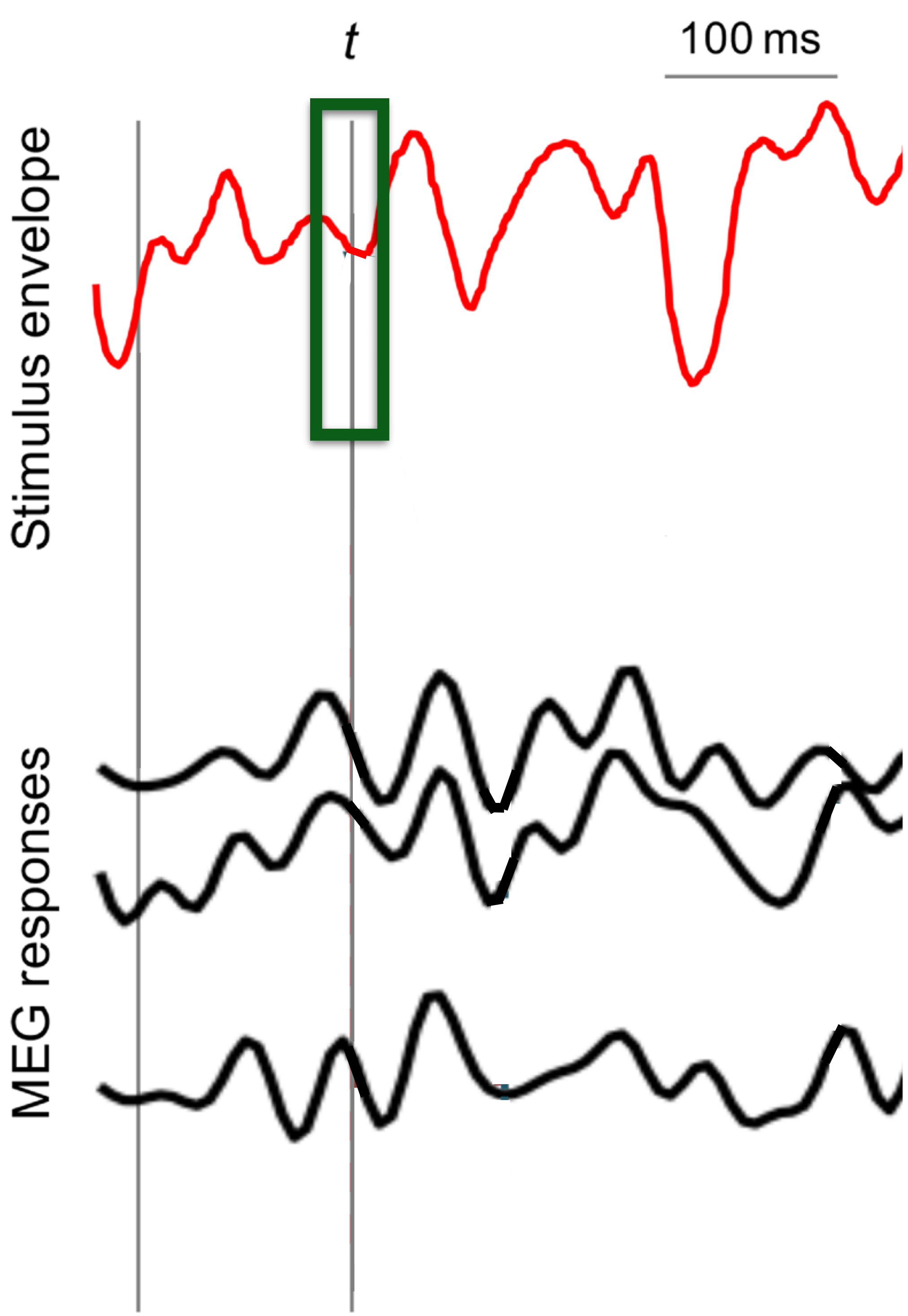


Stimulus Reconstruction:

Stimulus and response are known;

Find the best matrix in time ***and space*** to produce the **stimulus** from the **response**

Stimulus Reconstruction in Time

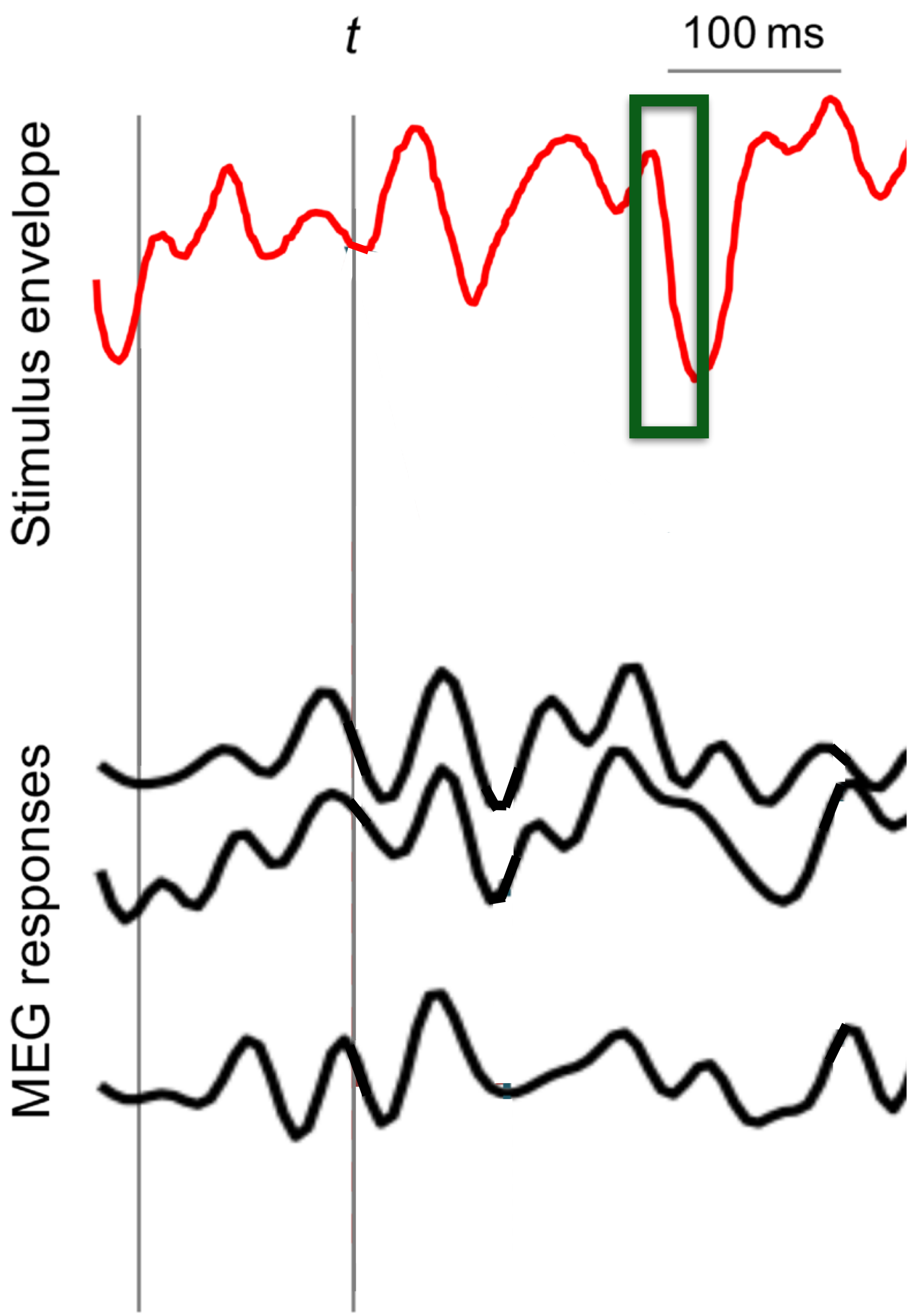


Stimulus Reconstruction:

Stimulus and response are known;

Find the best matrix in time ***and space*** to produce the **stimulus** from the **response**

<u>Stimulus</u> Reconstruction in Time

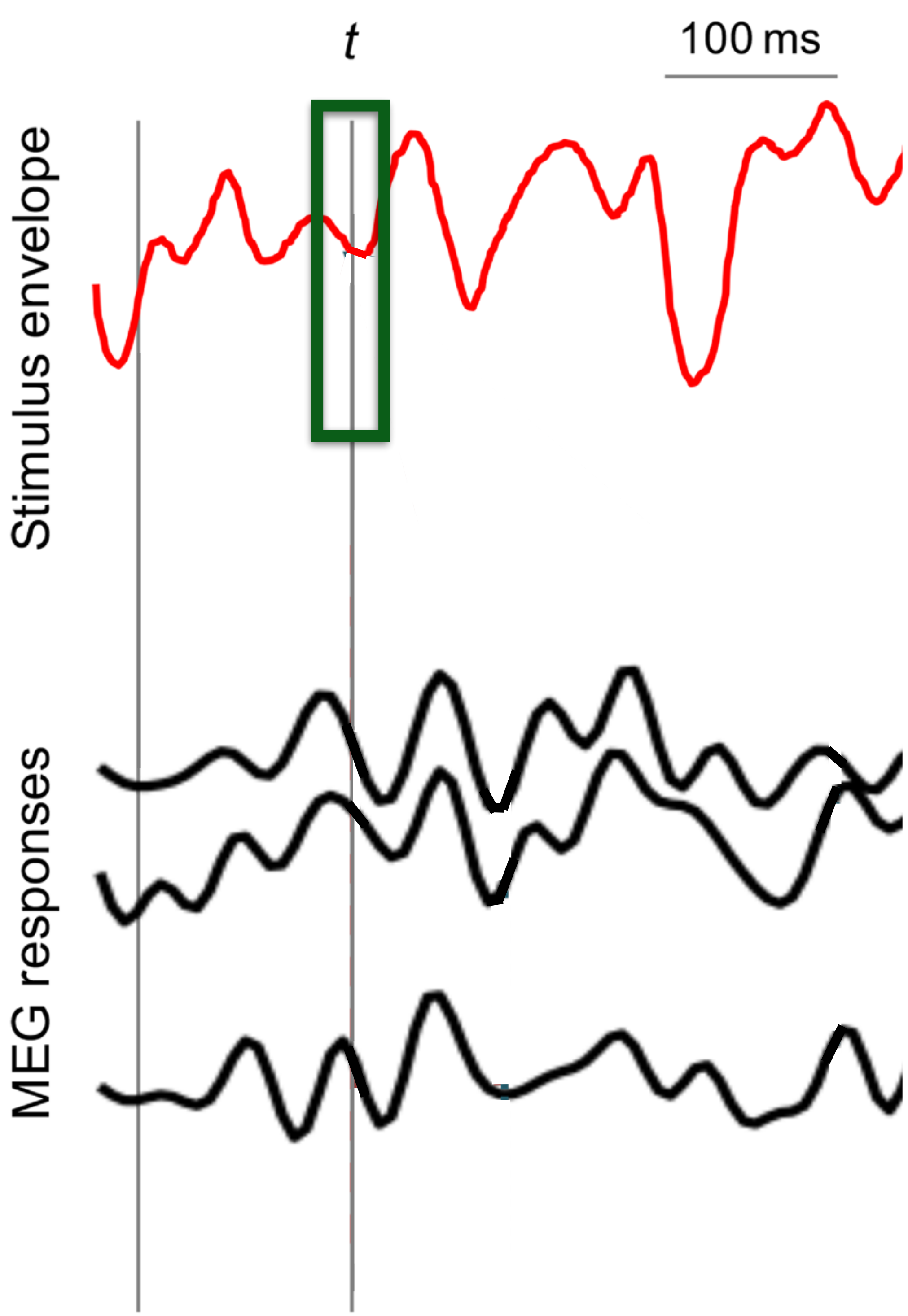


Stimulus Reconstruction:

Stimulus and response are known;

Find the best matrix in time ***and space*** to produce the **stimulus** from the **response**

<u>Stimulus</u> Reconstruction in Time

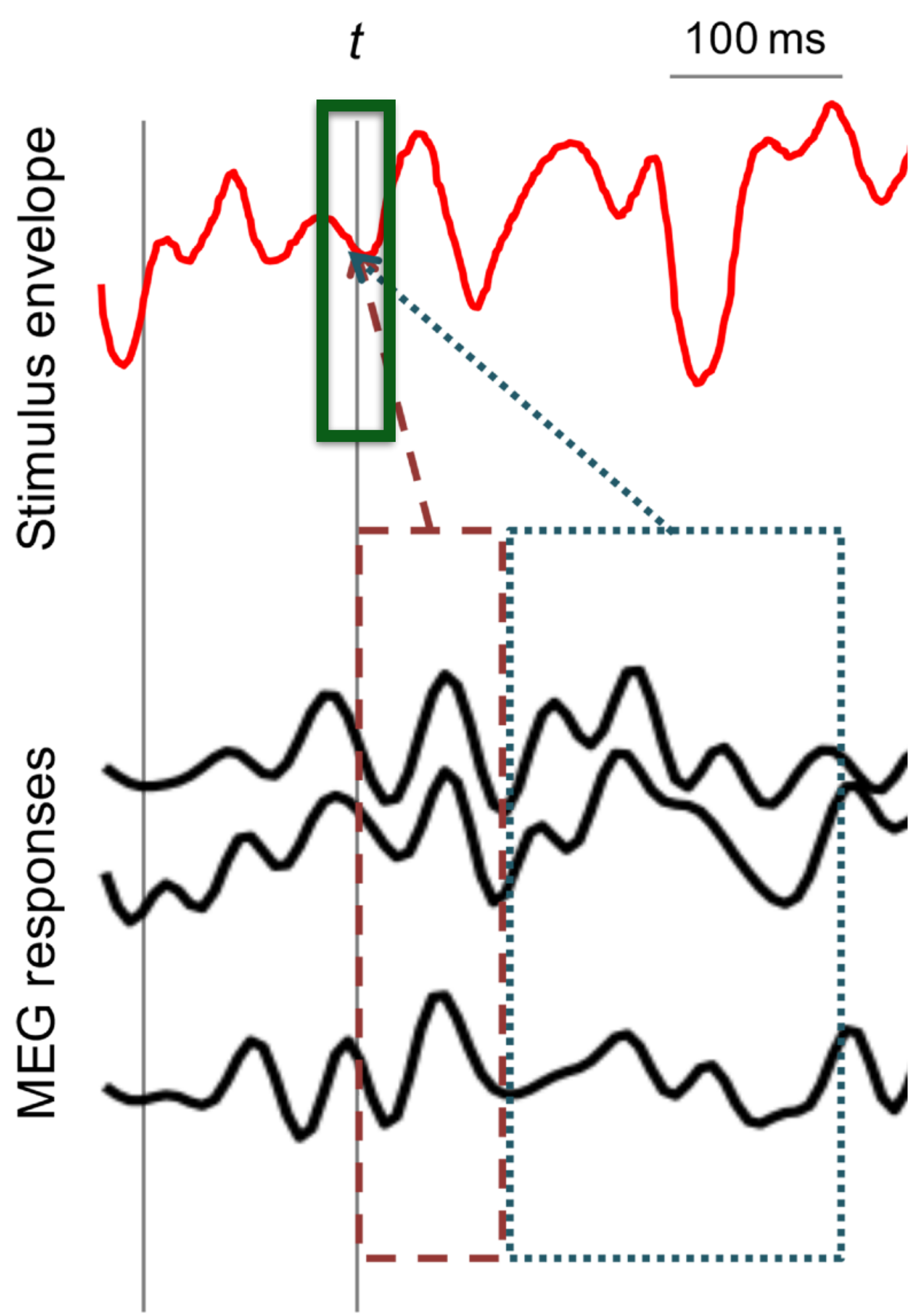


Stimulus Reconstruction:

Stimulus and response are known;

Find the best matrix in time ***and space*** to produce the **stimulus** from the **response**

Stimulus Reconstruction in Time



Stimulus Reconstruction:

Stimulus and response are known;

Find the best matrix in time ***and space*** to produce the **stimulus** from the **response**

Cortical Representations of Continuous Speech

- For long duration continuous speech
- Encoding & decoding (complementary)
- Linear model
- Acoustics: spectrotemporal **envelope**
- Envelope rates: ~ 1 - 10 Hz

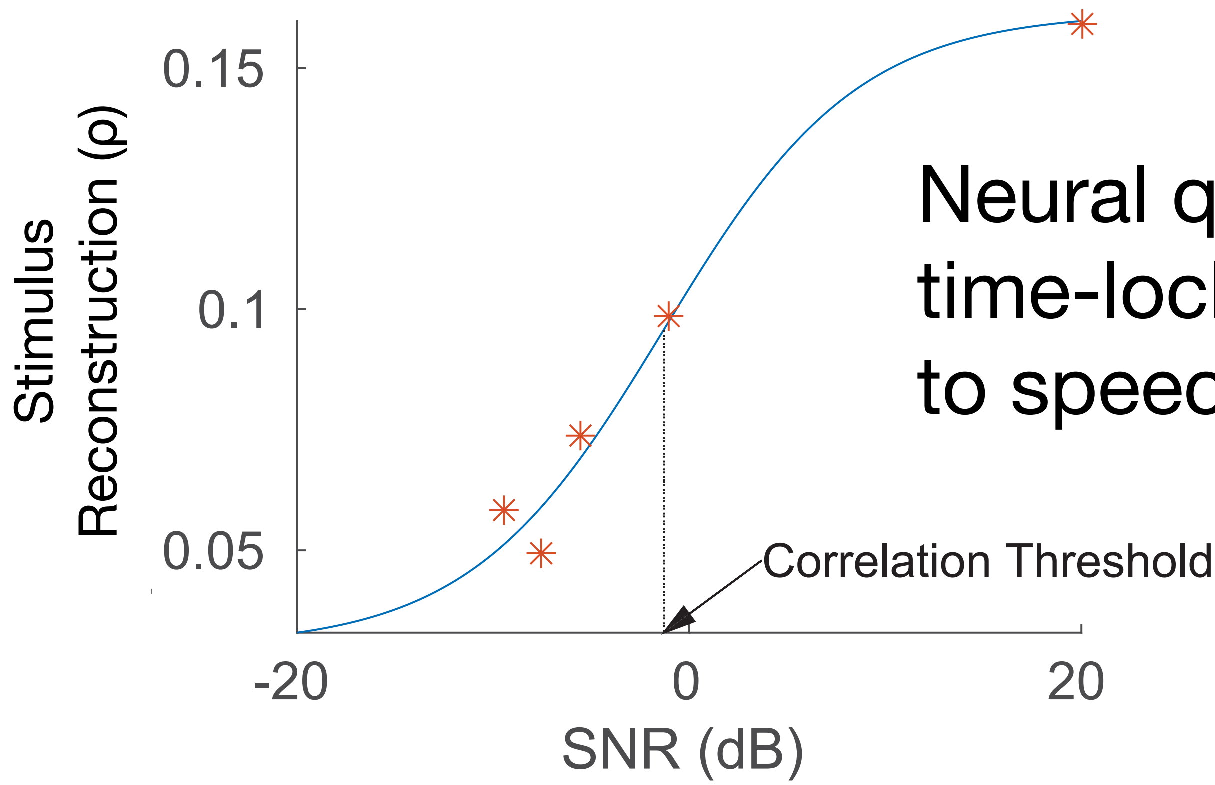
Outline

- Background & motivation
  - Neural responses in time
  - Response prediction from a stimulus via Temporal Response Function (TRF)
  - Stimulus reconstruction from responses
- Towards objective measures of
  - Speech intelligibility
  - Lexical processing of speech
  - Semantic processing of speech

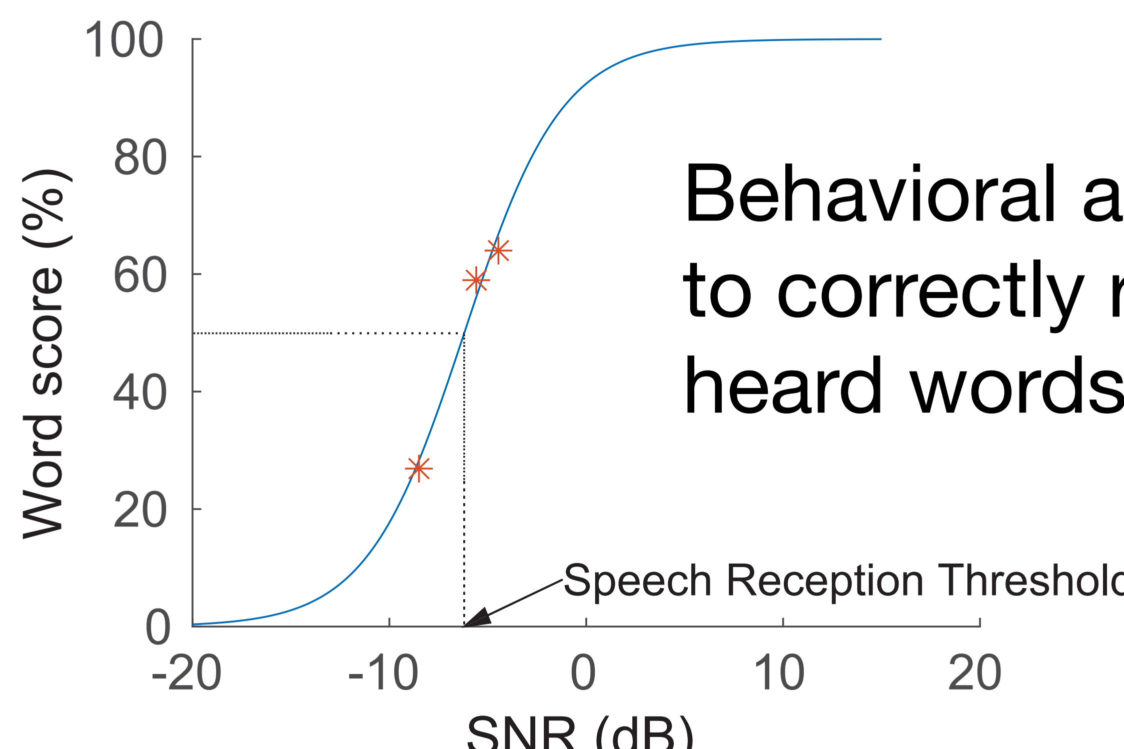
Outline

- Background & motivation
  - Neural responses in time
  - Response prediction from a stimulus via Temporal Response Function (TRF)
  - Stimulus reconstruction from responses
- Towards objective measures of
  - Speech intelligibility
  - Lexical processing of speech
  - Semantic processing of speech

Stimulus Reconstruction ➜ Intelligibility



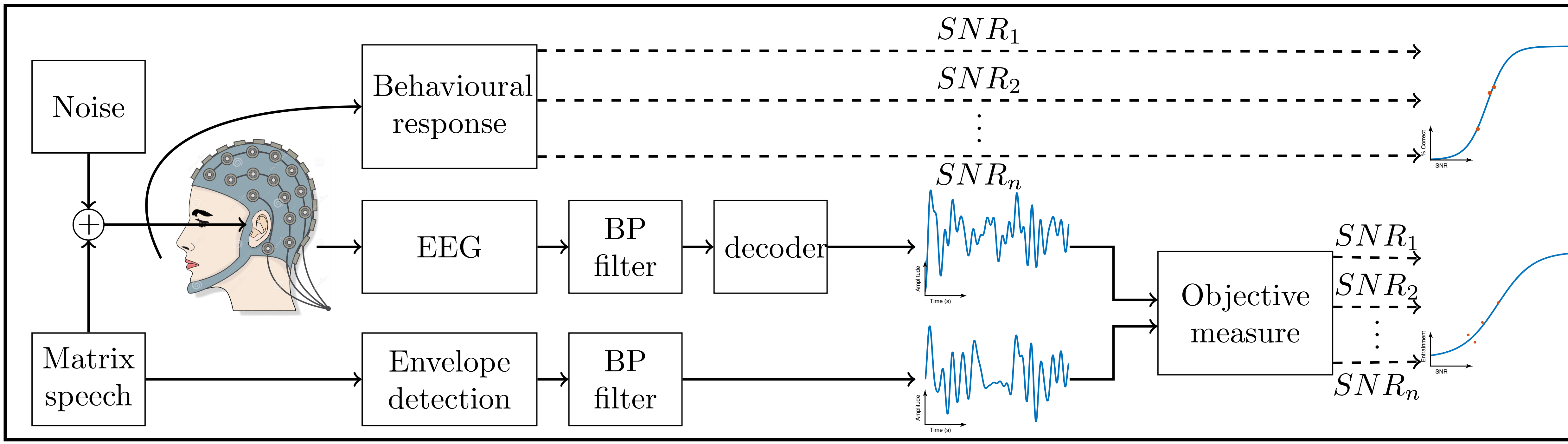
Neural quality of time-locked response to speech acoustics



Behavioral ability to correctly report heard words

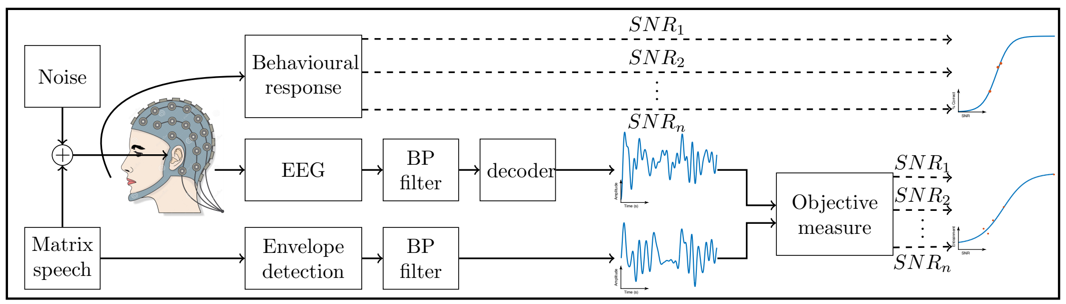
Vanthornhout et al., JARO (2018)

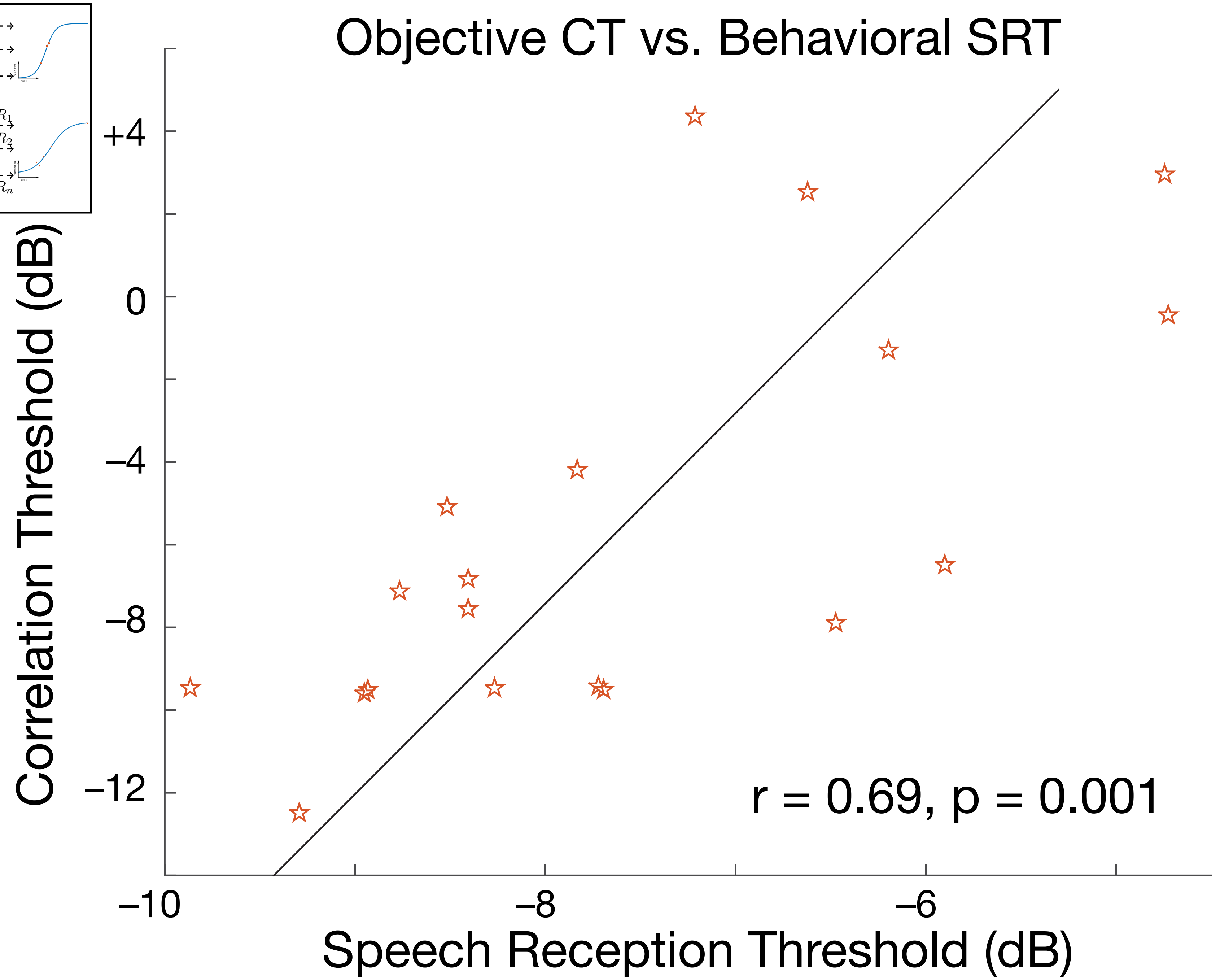
Stimulus Reconstruction ➜ Intelligibility



Vanthornhout et al., JARO (2018)

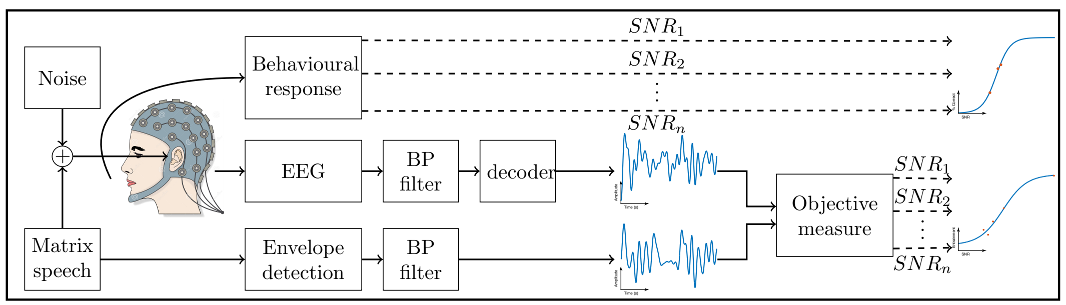
Stimulus Reconstruction ➜ Intelligibility





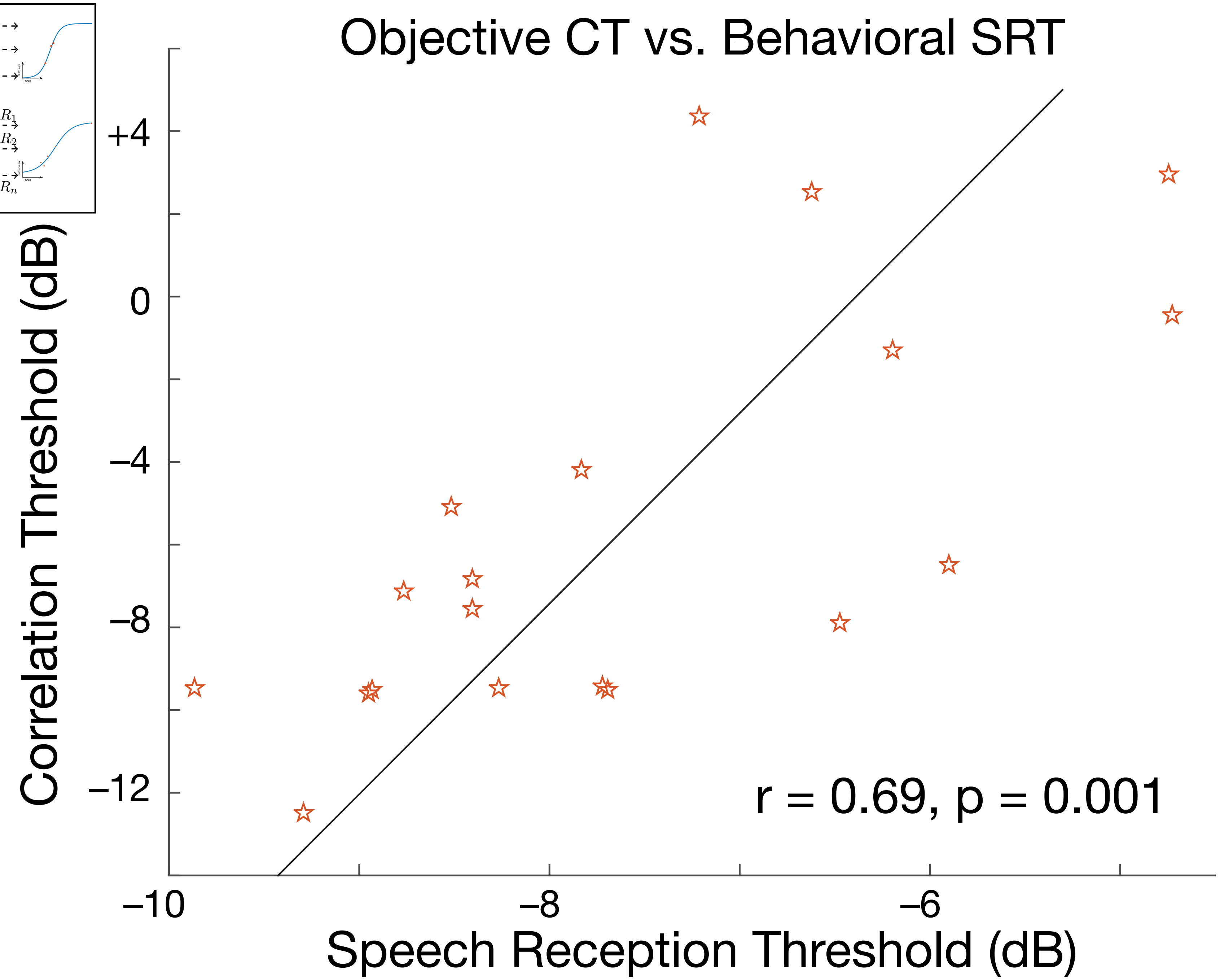
Vanthornhout et al., JARO (2018)

Stimulus Reconstruction ➜ Intelligibility



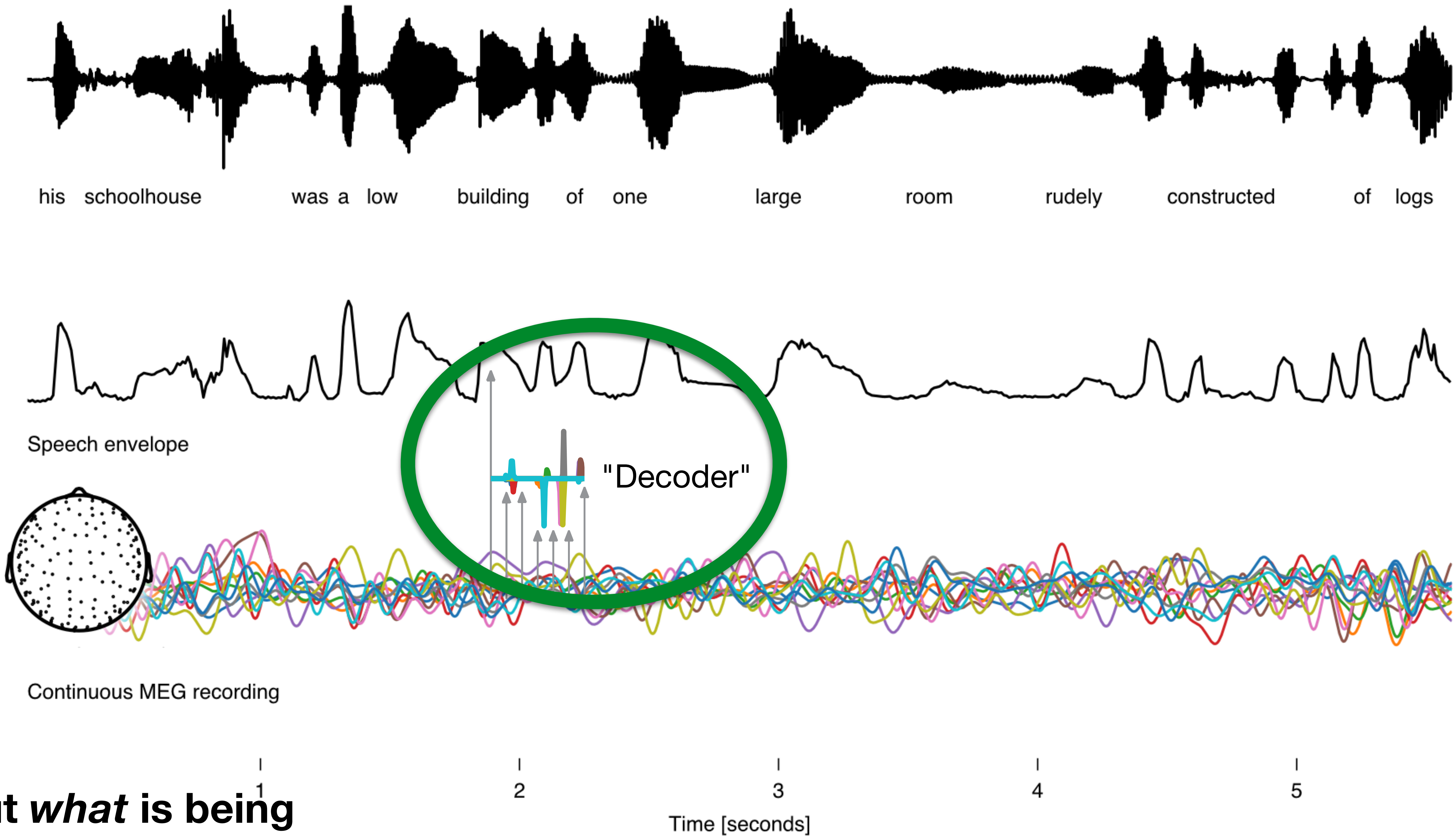
But *what* is being measured neurally?

Cortical responses, but where? (or when?)



Vanthornhout et al., JARO (2018)

Stimulus Reconstruction in Time



But *what* is being measured neurally?

Cortical responses, but where? (or when?)

Stimulus Reconstruction ➜ Intelligibility

Integration window span indicates latencies of interest

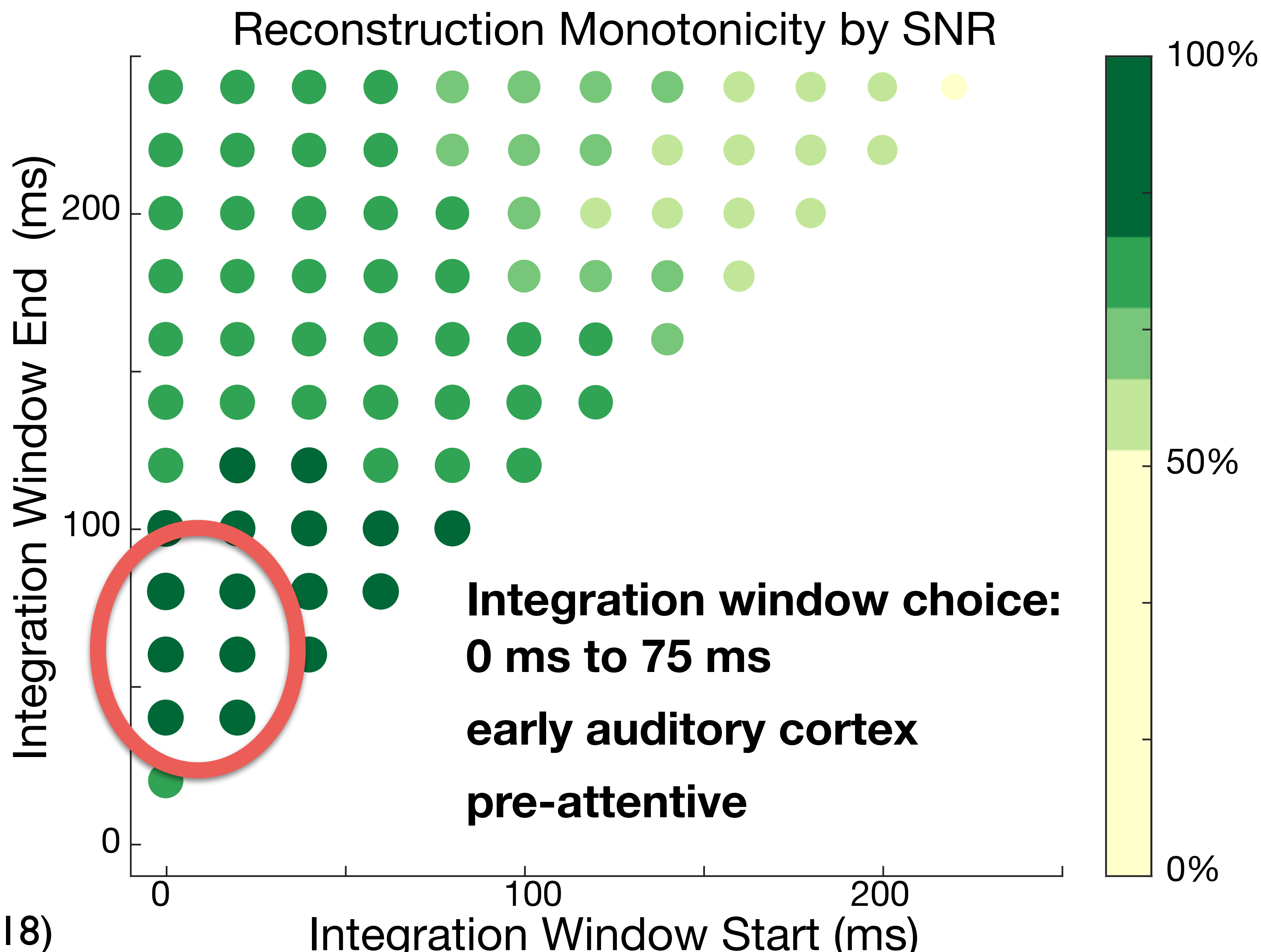
- choose window for reconstruction
- not based on highest correlation (of reconstructed stimulus)
- based on reconstruction **monotonicity** as a function of SNR.

Vanthornhout et al., JARO (2018)

Stimulus Reconstruction ➜ Intelligibility

Integration window span indicates latencies of interest

- choose window for reconstruction
- not based on highest correlation (of reconstructed stimulus)
- based on reconstruction **monotonicity** as a function of SNR.



Vanthornhout et al., JARO (2018)

Stimulus Reconstruction ➜ Intelligibility

- Continuous speech envelope reconstruction (neurometric) threshold predicts behavioral speech reception threshold (SRT).
- Uses long duration continuous speech
- Based on robust *acoustic* speech representation
- Early auditory cortex most critical (pre-attentive)

Stimulus Reconstruction ➜ Intelligibility

- UPDATES from the Francart Lab
  - Response prediction (~~stimulus reconstruction~~)
  - Theta band
  - Speech Envelope ➜ Spectrogram
  - Added new representation: phonetic features*

*Role of phonetic features vs. spectrogram onsets?

Lesenfants et al., Hear Res (2019)

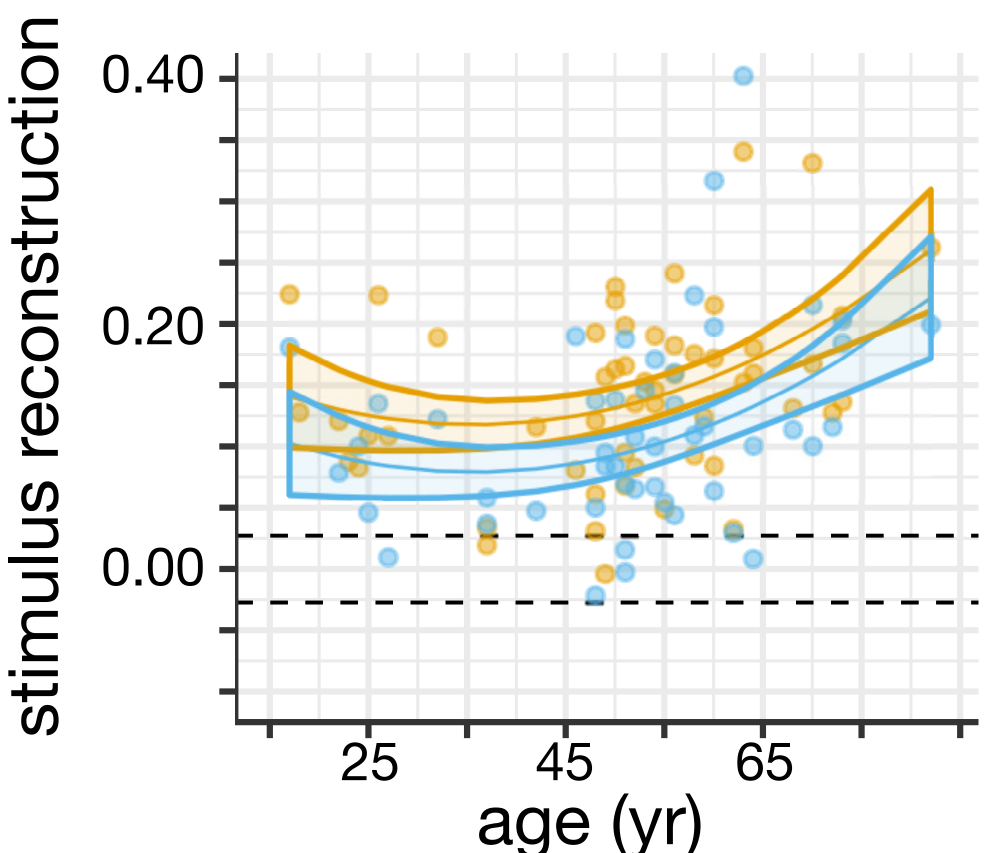
Phonetic Features vs. Spectrogram Onsets

- + ‘phonetic features’ representation increases EEG response prediction: Di Liberto et al. (2015).
- Adding only *acoustic* spectrogram *onsets* gives same predictive benefits as phonetic features for MEG responses: Daube et al. (2019).
- Also seen in Simon lab: Brodbeck et al. (2018).

➡ Phonetic features too correlated with acoustic onsets, in natural speech, to isolate them

Stimulus Reconstruction ➜ Intelligibility

- UPDATES from the Francart Lab
  - Age really matters: Decruy et al. (2019)



Not just linear but quadratic uptick

Cognitive decline also matters

In agreement with Presacco et al. (2016a, 2016b).

Outline

- Background & motivation
  - Neural responses in time
  - Response prediction from a stimulus via Temporal Response Function (TRF)
  - Stimulus reconstruction from responses
- Towards objective measures of
  - Speech intelligibility
  - Lexical processing of speech
  - Semantic processing of speech

Outline

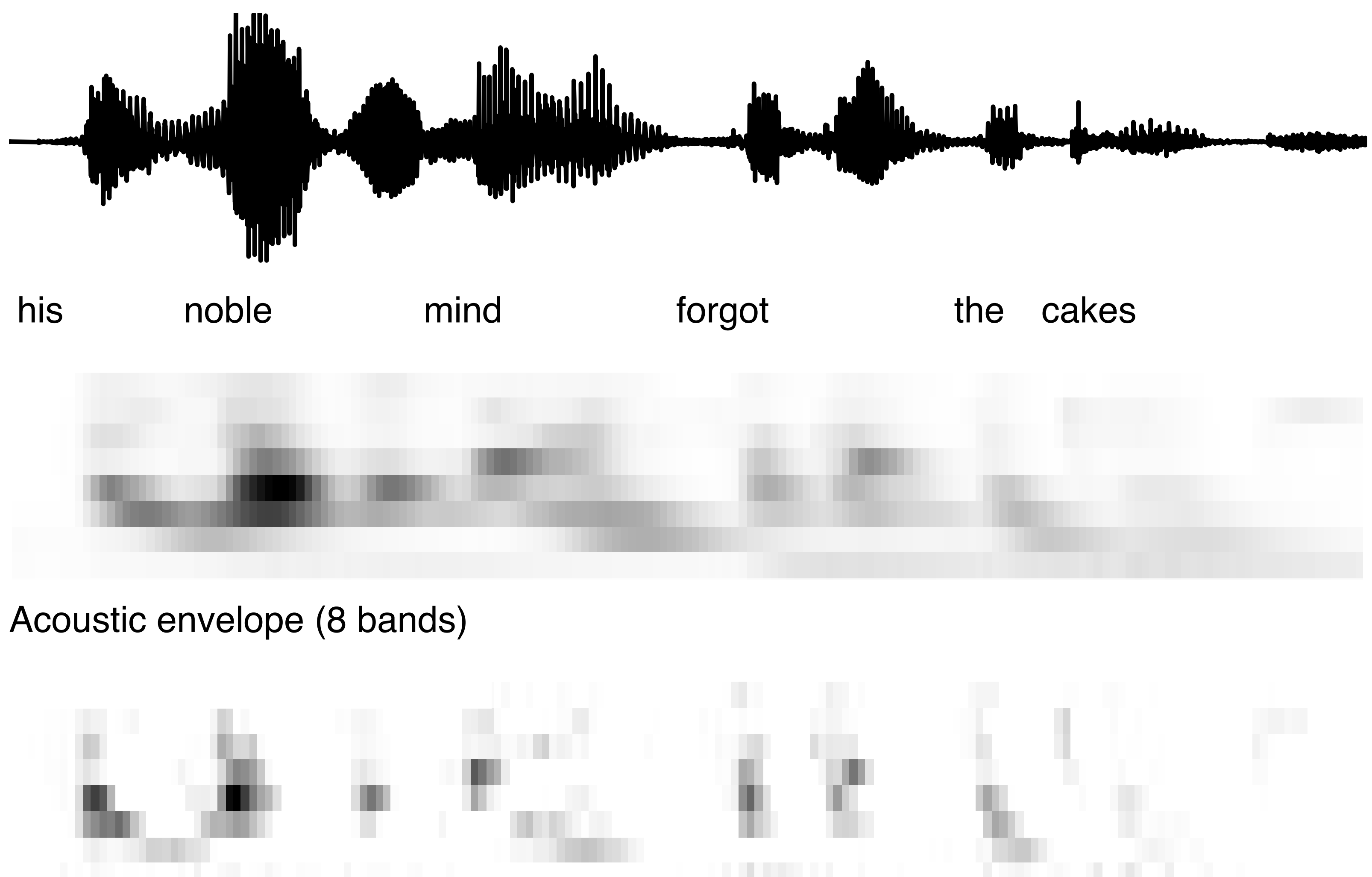
- Background & motivation
  - Neural responses in time
  - Response prediction from a stimulus via Temporal Response Function (TRF)
  - Stimulus reconstruction from responses
- Towards objective measures of
  - Speech intelligibility
  - Lexical processing of speech
  - Semantic processing of speech

Lexical Processing

- Processing by early auditory cortex critical
- Using more than global speech envelope helps
- Another level of speech perception:
  - Transforming speech sounds into words
  - “Lexical processing”
    - Language-based but not via word meaning

Brodbeck et al., Curr Biol (2018)

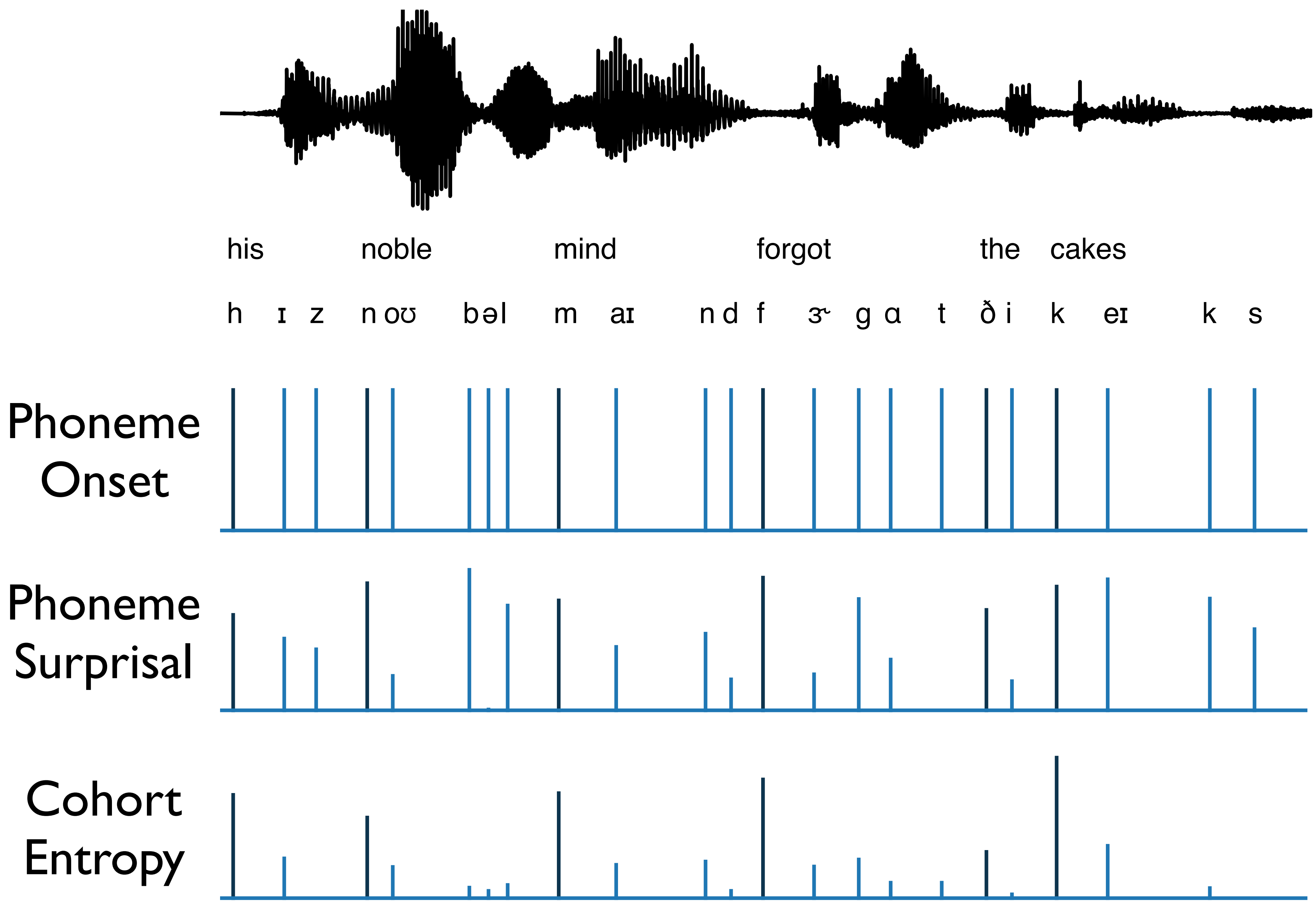
Acoustic to Lexical Speech Processing



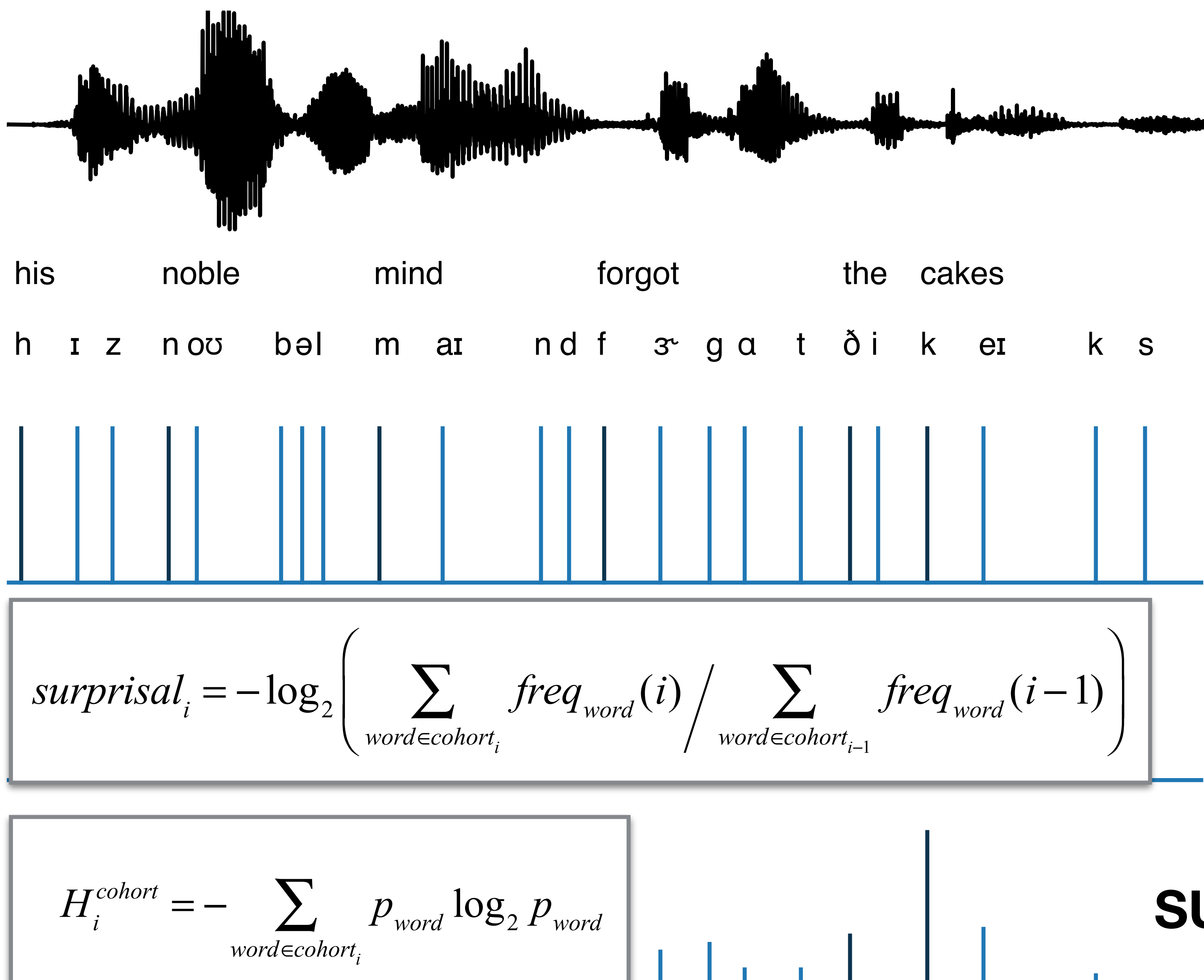
Acoustic Envelope (8 bands)

Acoustic Onset (8 bands)

Acoustic to Lexical Speech Processing



Acoustic to Lexical Speech Processing



Phoneme Onset

Phoneme Surprisal

$$surprisal_i = -\log_2\left(\sum_{word \in cohort_i} freq_{word}(i) \Big/ \sum_{word \in cohort_{i-1}} freq_{word}(i-1) \right)$$

Cohort Entropy

$$H_i^{cohort} = -\sum_{word \in cohort_i} p_{word} \log_2 p_{word}$$

SUBTLEX:

51 million words

movie subtitle database

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)

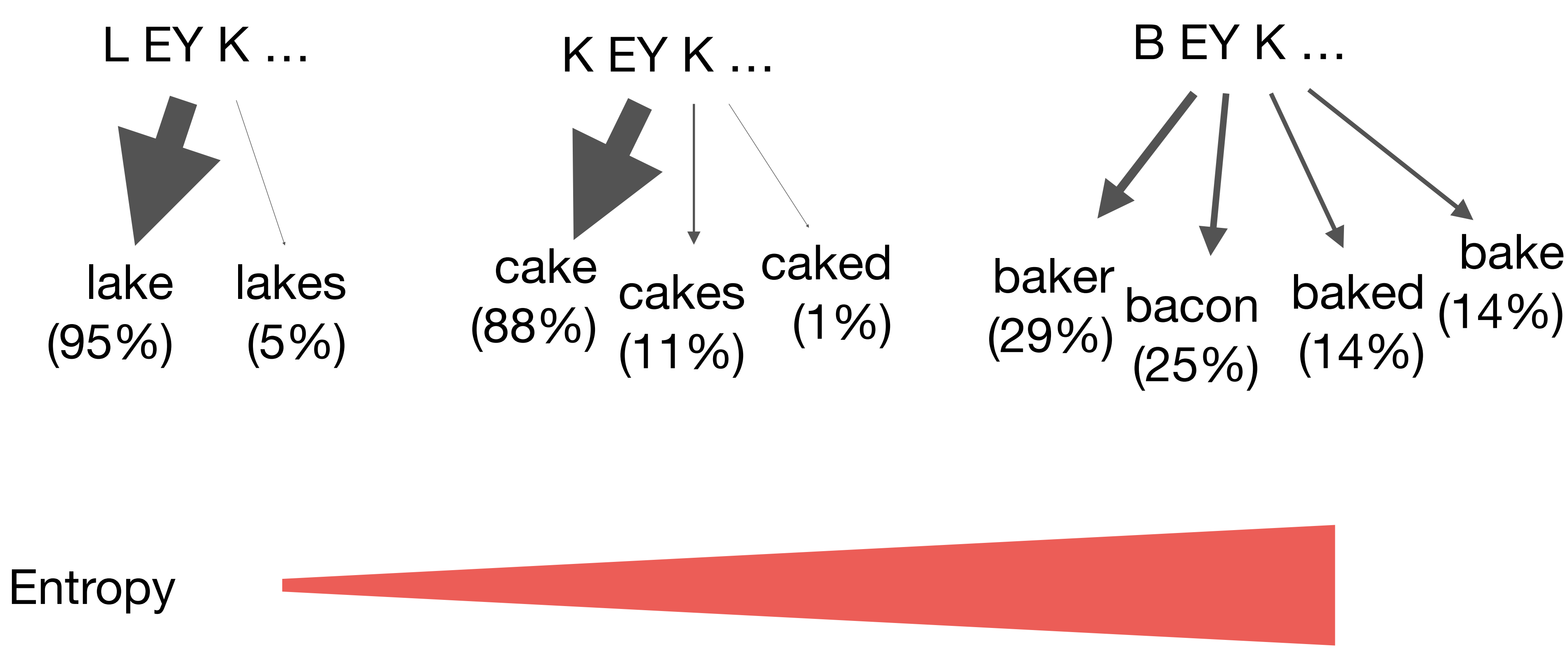
"cane", "canine", "Canaan", "Kane", "Keynesian", …

⋮

Entropy

Cohort entropy

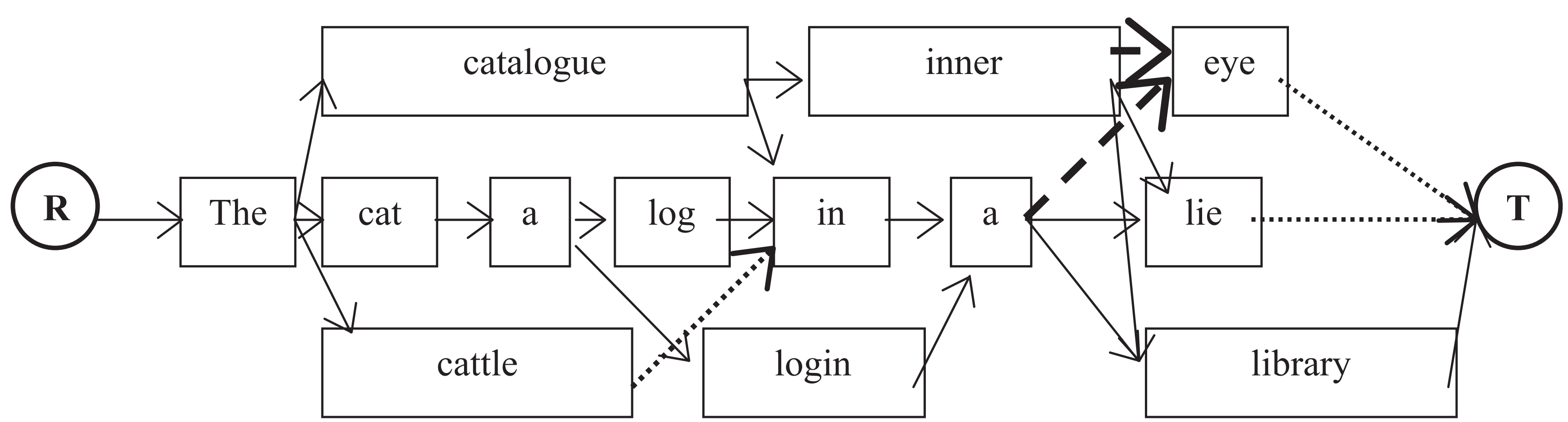
- How unpredictable is the current word?



Word onsets

Do we…

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

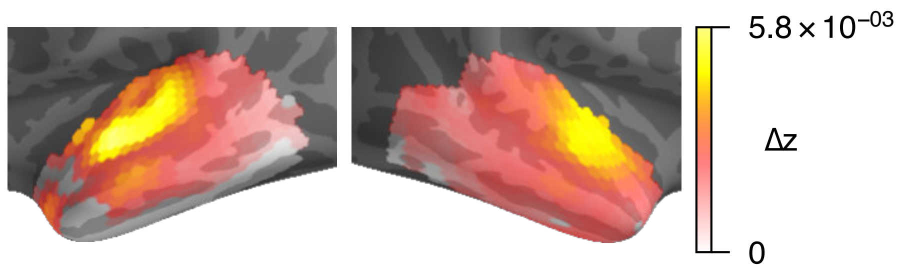


“The catalogue in a library”

(Norris & McQueen, 2008)

Acoustic Results

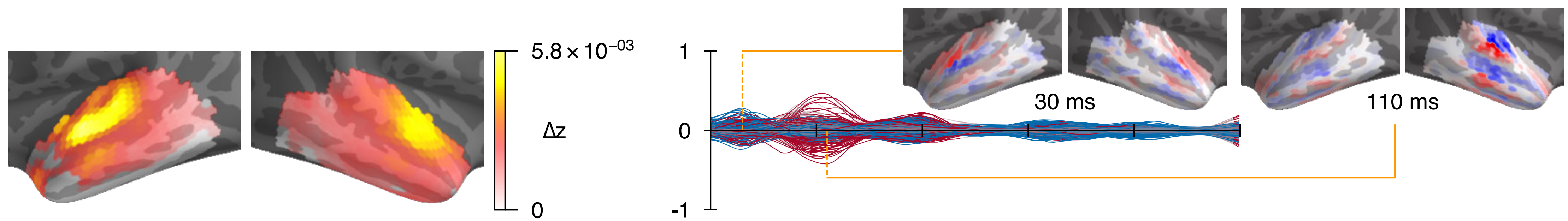
Acoustic Envelope



Brodbeck et al., Curr Biol (2018)

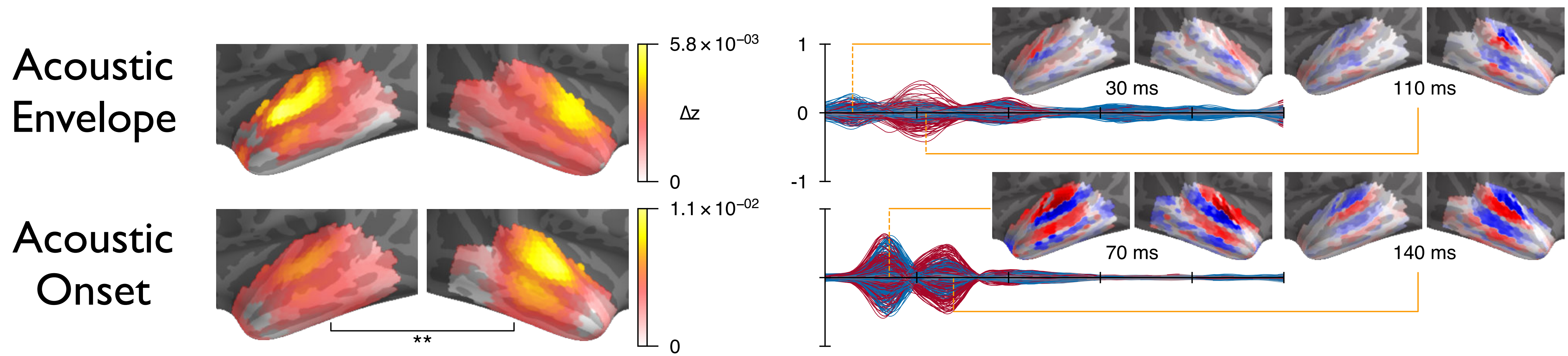
Acoustic Results

Acoustic
Envelope



Brodbeck et al., Curr Biol (2018)

Acoustic Results

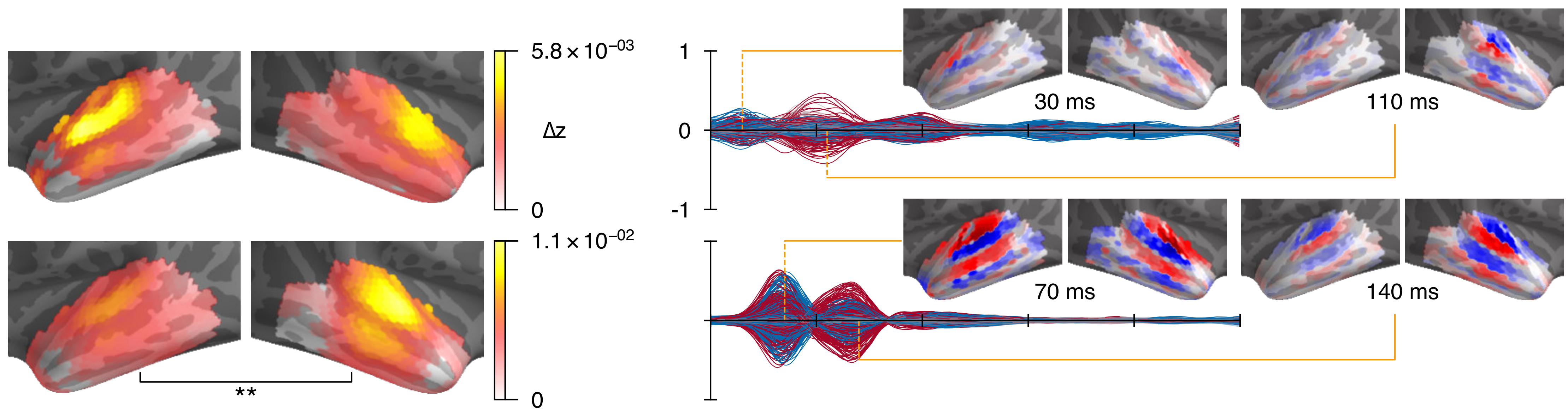


Brodbeck et al., Curr Biol (2018)

Acoustic Results

Acoustic Envelope

Acoustic Onset

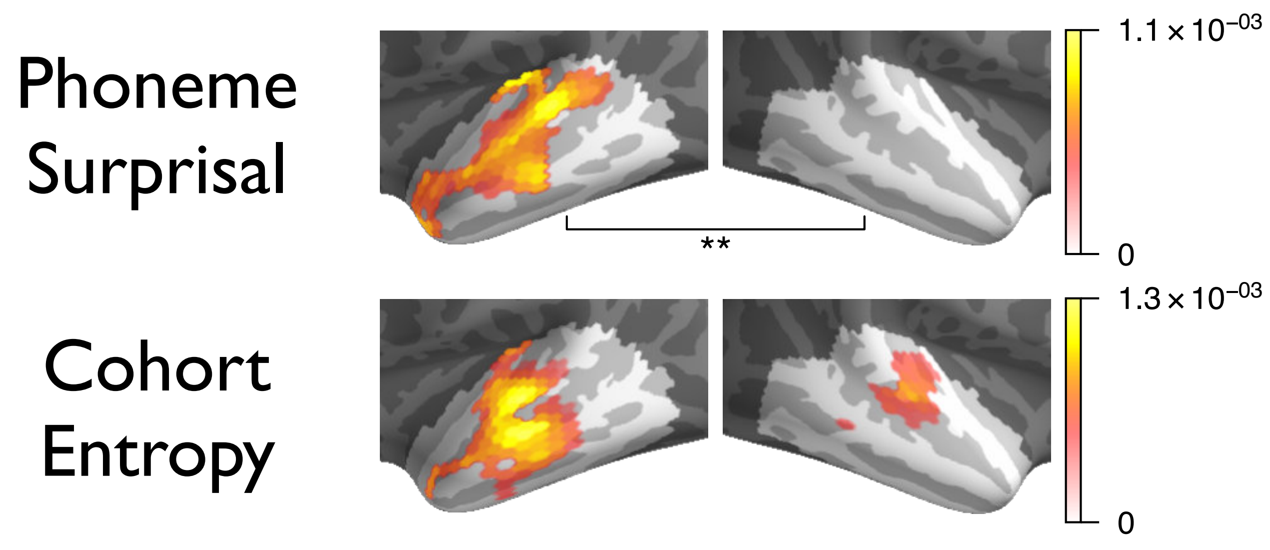


cf. Daube et al., Curr Biol (2019)

- Onset explains more variance
- Latency(ies) as expected
- Strongly bilateral
- Onset stronger in right hemisphere

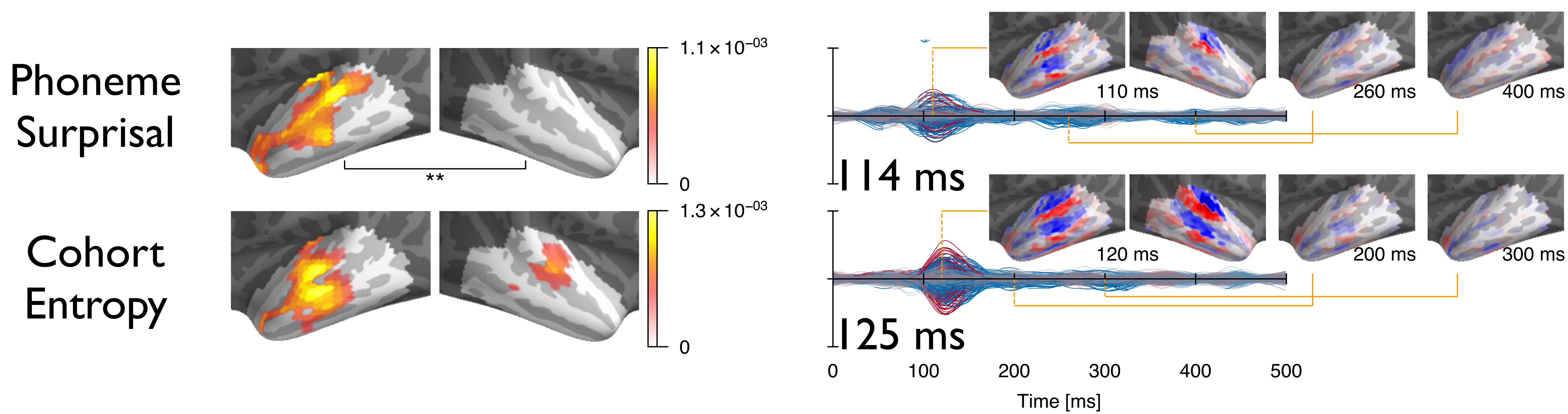
Brodbeck et al., Curr Biol (2018)

Neural Lexical Processing



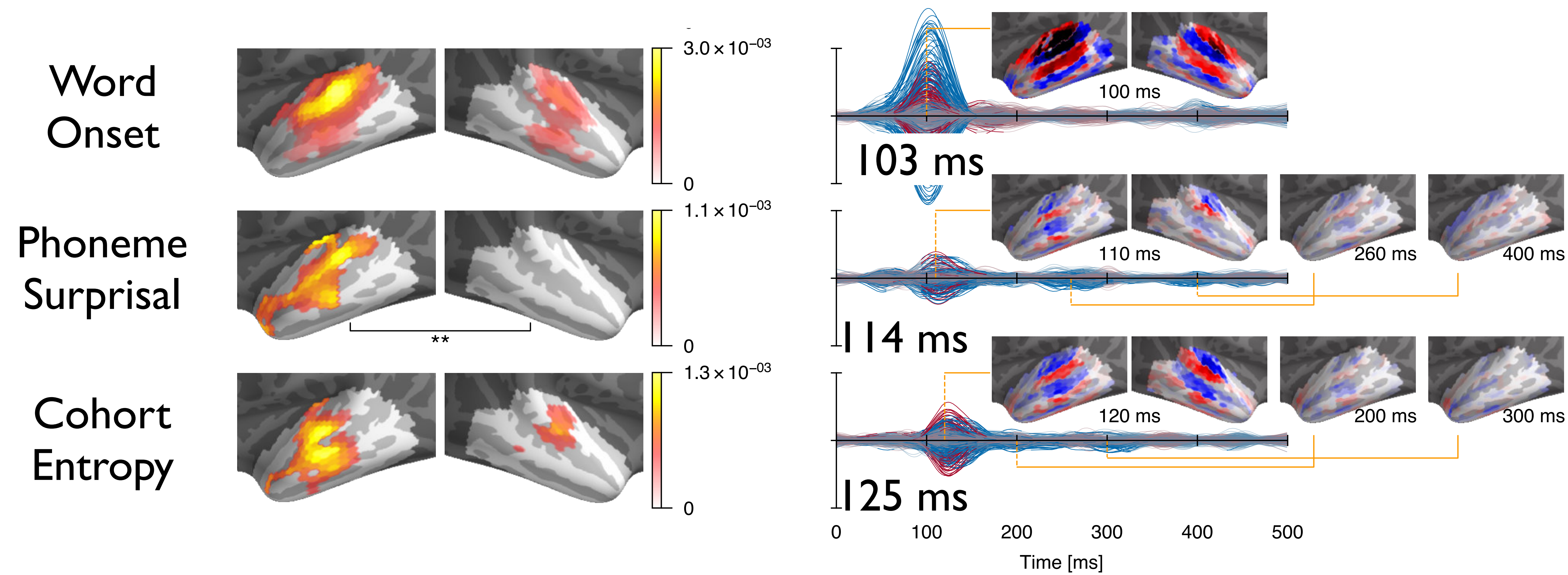
Brodbeck et al., Curr Biol (2018)

Neural Lexical Processing



Brodbeck et al., Curr Biol (2018)

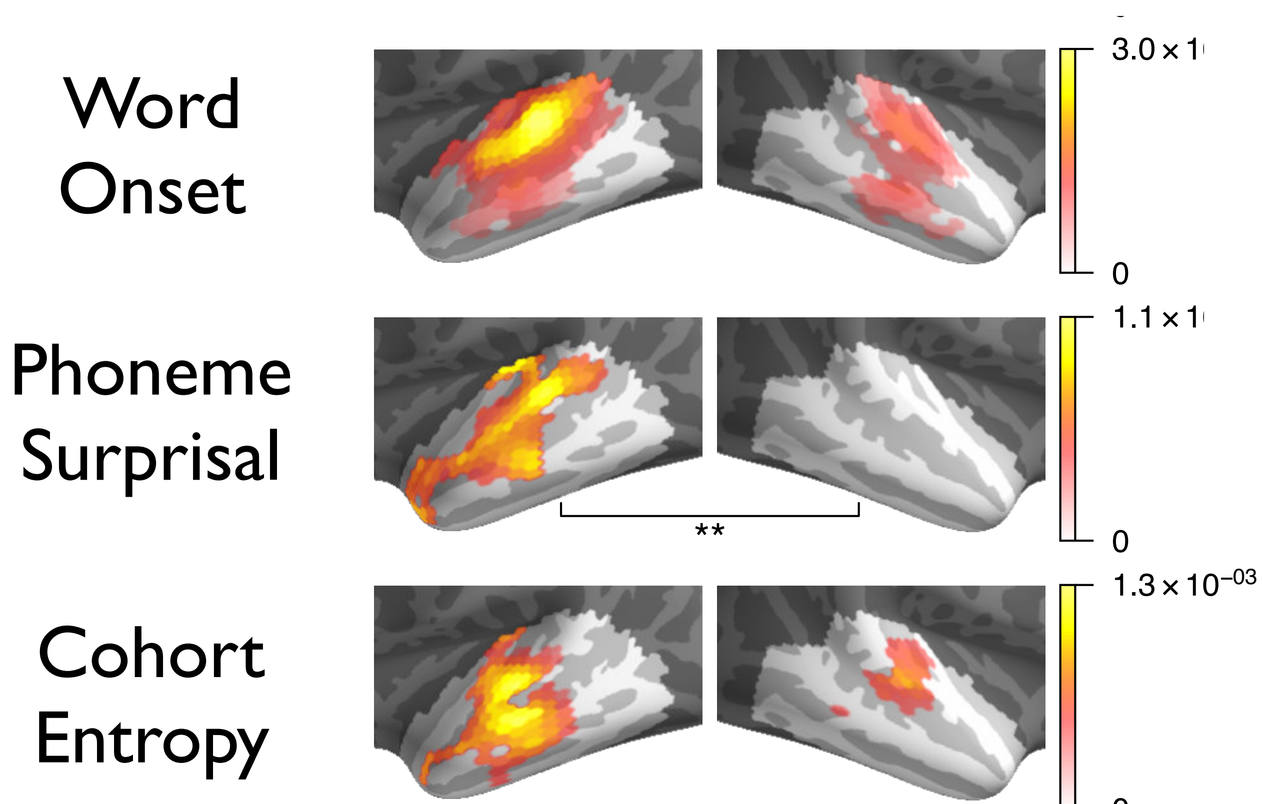
Neural Lexical Processing

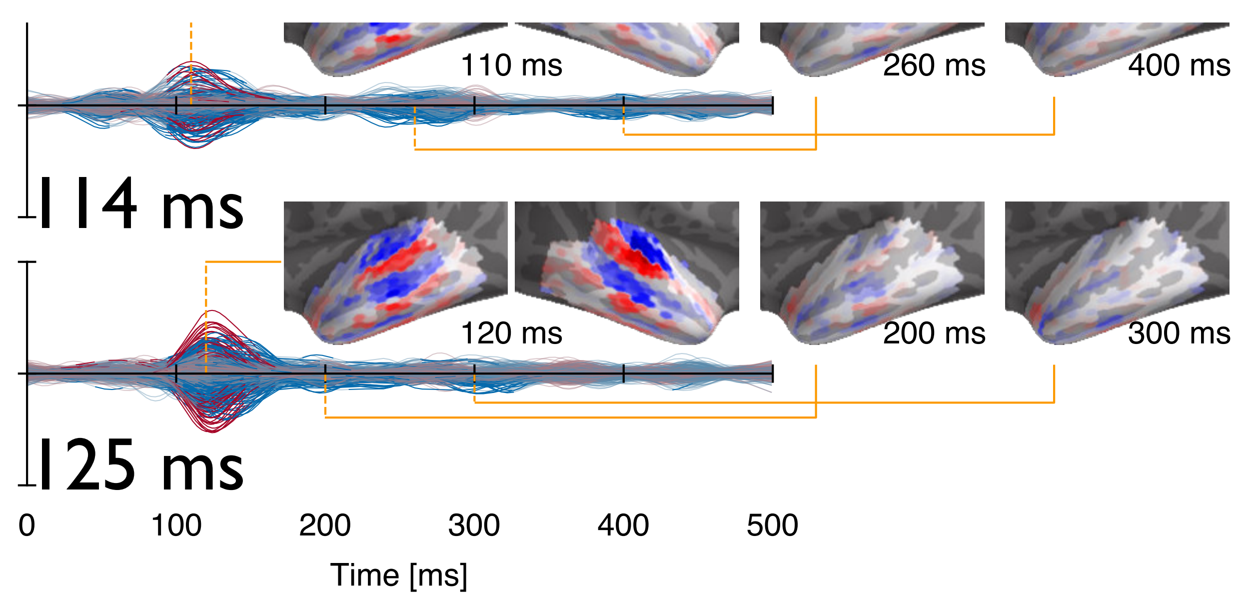


Brodbeck et al., Curr Biol (2018)

Neural Lexical Processing

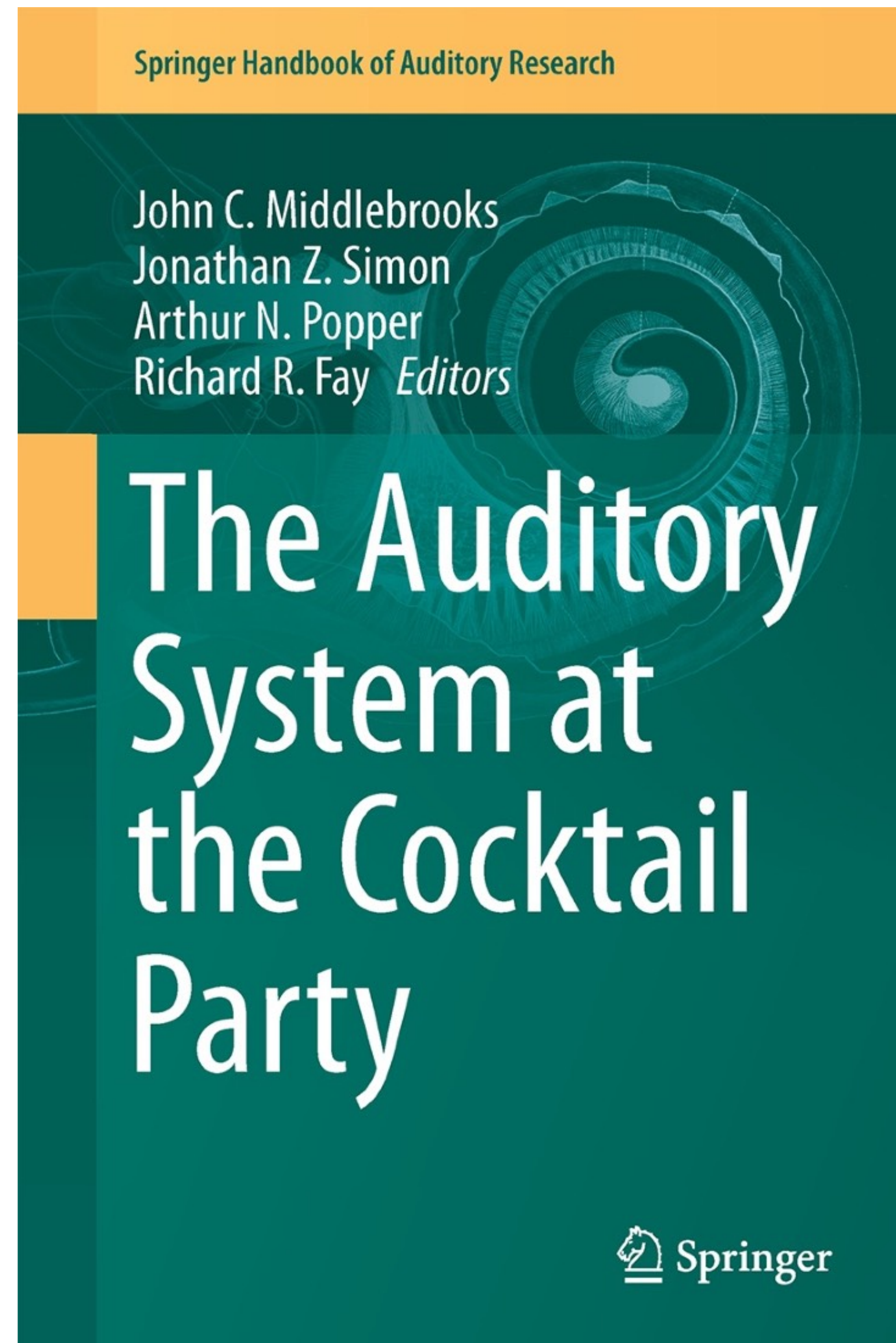
- Rapid transformation to lexical
- Word boundaries identified
- Surprisal = local measure of phoneme prediction error (predictive coding?)
- Cohort entropy = global measure of lexical competition across cohort
- Strongly left hemisphere dominant





Brodbeck et al., Curr Biol (2018)

Listening at the Cocktail Party



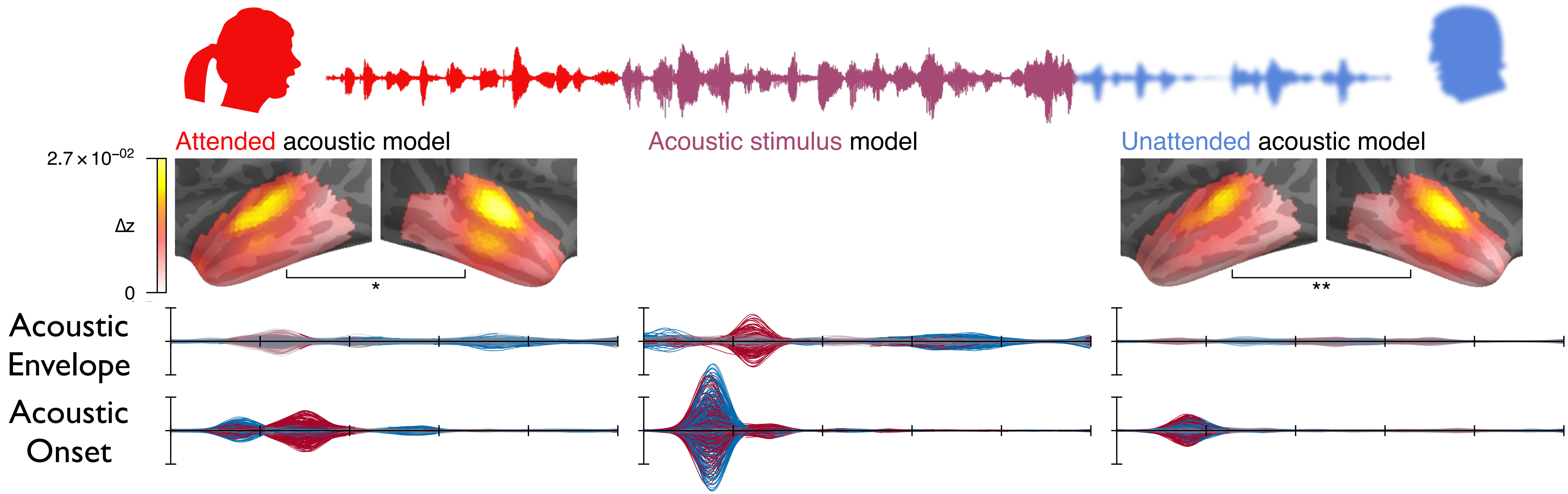
Acoustic Attention

2 competing speakers, equal loudness, attend to one

Brodbeck et al., Curr Biol (2018)

Acoustic Attention

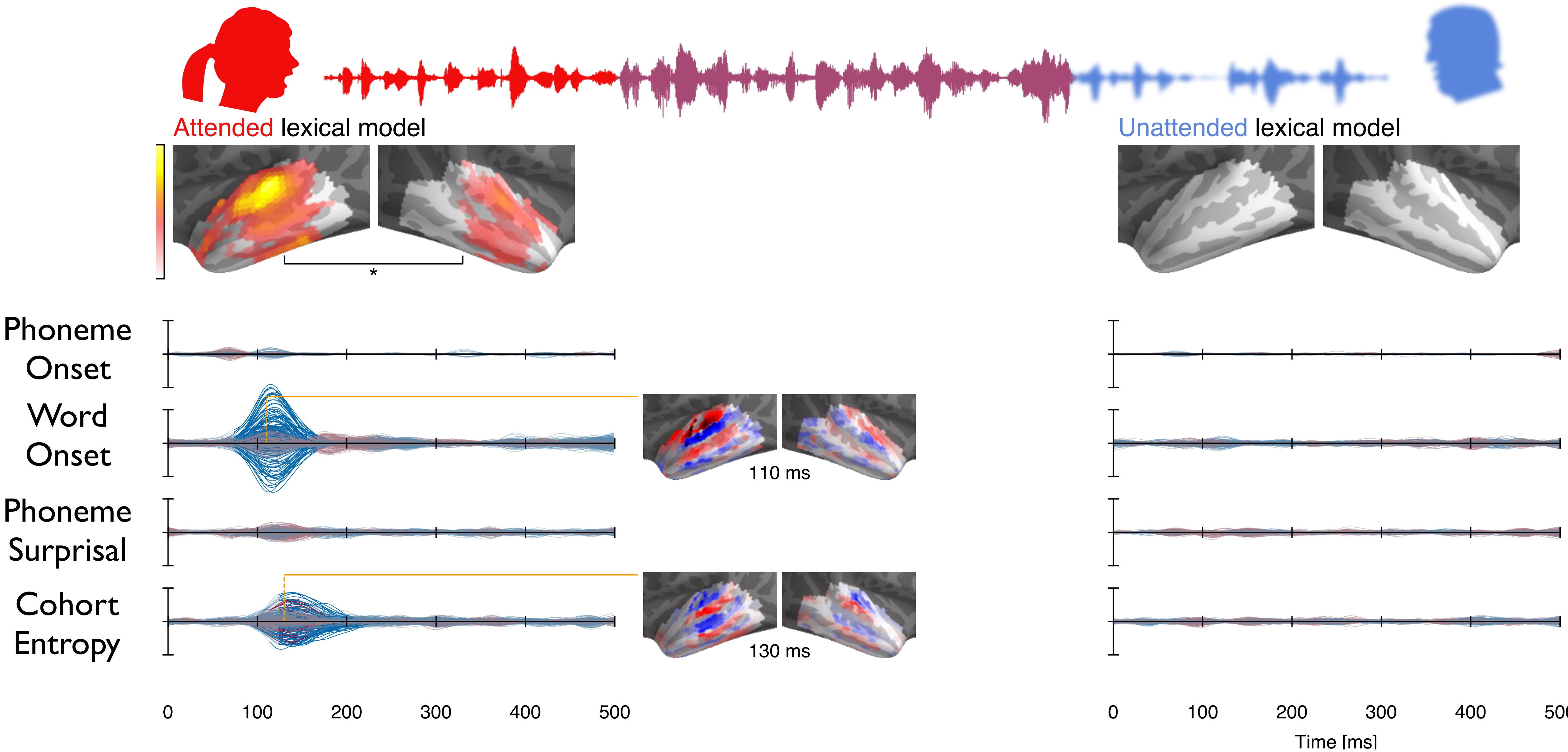
2 competing speakers, equal loudness, attend to one



- Onset Representation Dominates
- Attended Dominates Later

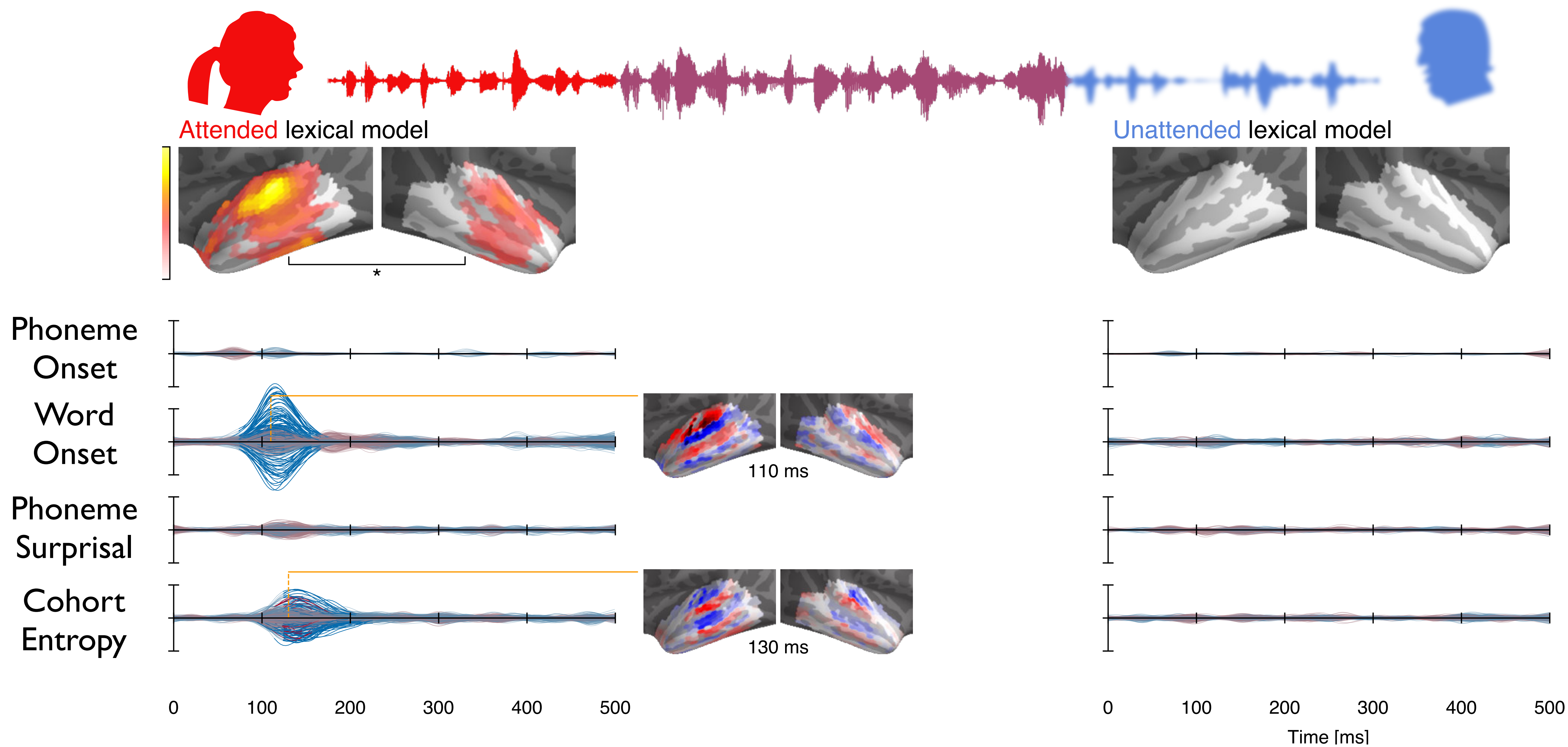
Brodbeck et al., Curr Biol (2018)

Lexical Attention



Brodbeck et al., Curr Biol (2018)

Lexical Attention



- Only attended speech processed lexically
- Lexical processing slowed by ~15 ms

Brodbeck et al., Curr Biol (2018)

Lexical Processing

- Speech perception at level of transforming speech sounds into words
- "Post-acoustic" phoneme processing
- Word-based
- Attention required (?)
- Surprisingly early

Brodbeck et al., Curr Biol (2018)

Outline

- Background & motivation
  - Neural responses in time
  - Response prediction from a stimulus via Temporal Response Function (TRF)
  - Stimulus reconstruction from responses
- Towards objective measures of
  - Speech intelligibility
  - Lexical processing of speech
  - Semantic processing of speech

Outline

- Background & motivation
  - Neural responses in time
  - Response prediction from a stimulus via Temporal Response Function (TRF)
  - Stimulus reconstruction from responses
- Towards objective measures of
  - Speech intelligibility
  - Lexical processing of speech
  - Semantic processing of speech

Semantic Processing

- Speech perception includes perceiving the meaning of the speech
- Computational language models give several semantic measures: *semantic dissimilarity*
- Analysis of *Semantic-dissimilarity*-based TRF
  - potential basis of objective measure of perception of speech meaning

Broderick et al., Curr Biol (2018)

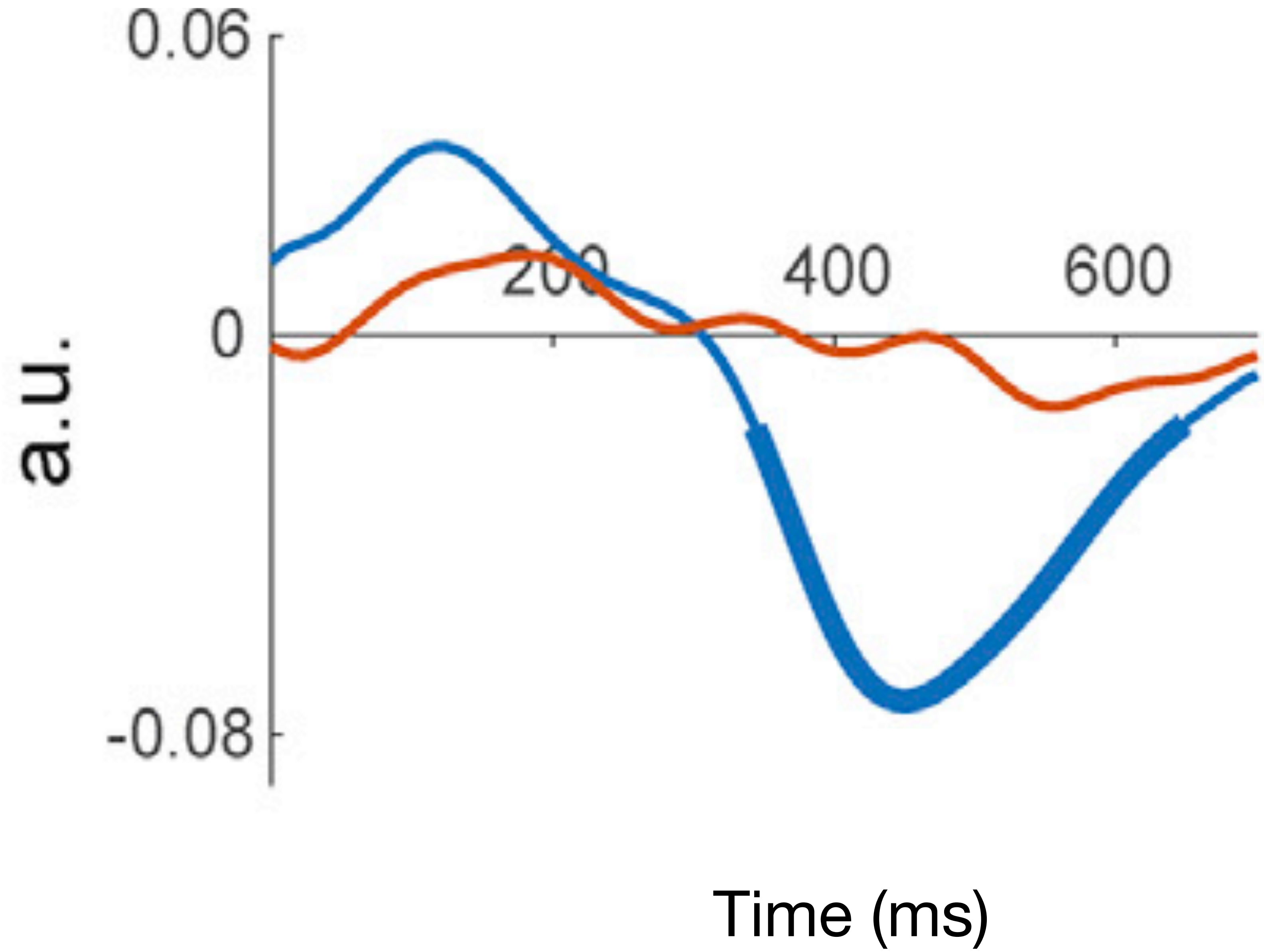
Semantic Processing

- Speech perception includes perceiving the meaning of the speech
- Computational language models give several semantic measures: *semantic dissimilarity*
- Analysis of *Semantic-dissimilarity*-based TRF
  - potential basis of objective measure of perception of speech meaning

Broderick et al., Curr Biol (2018)

Semantic Processing

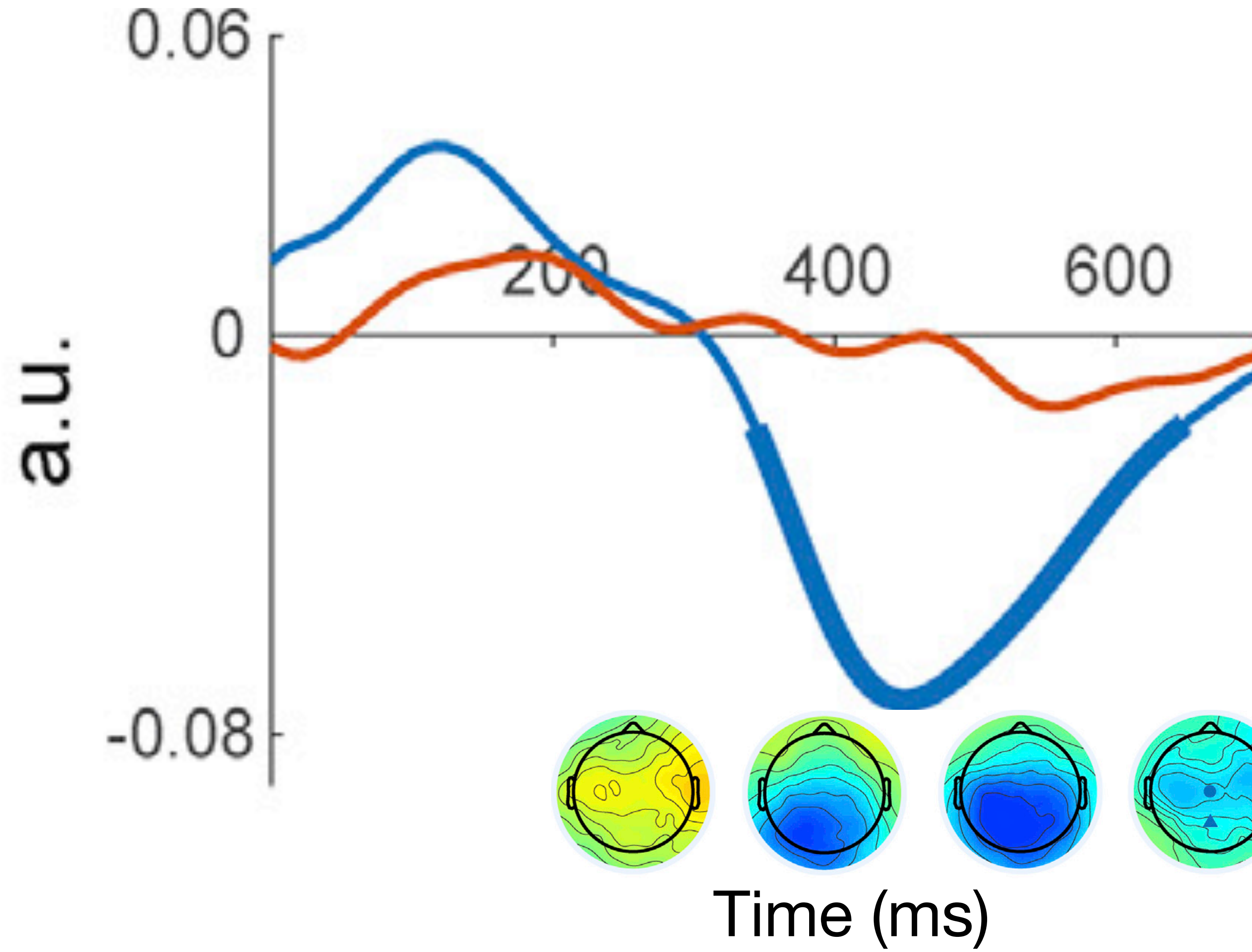
Semantic-dissimilarity TRF



Broderick et al., Curr Biol (2018)

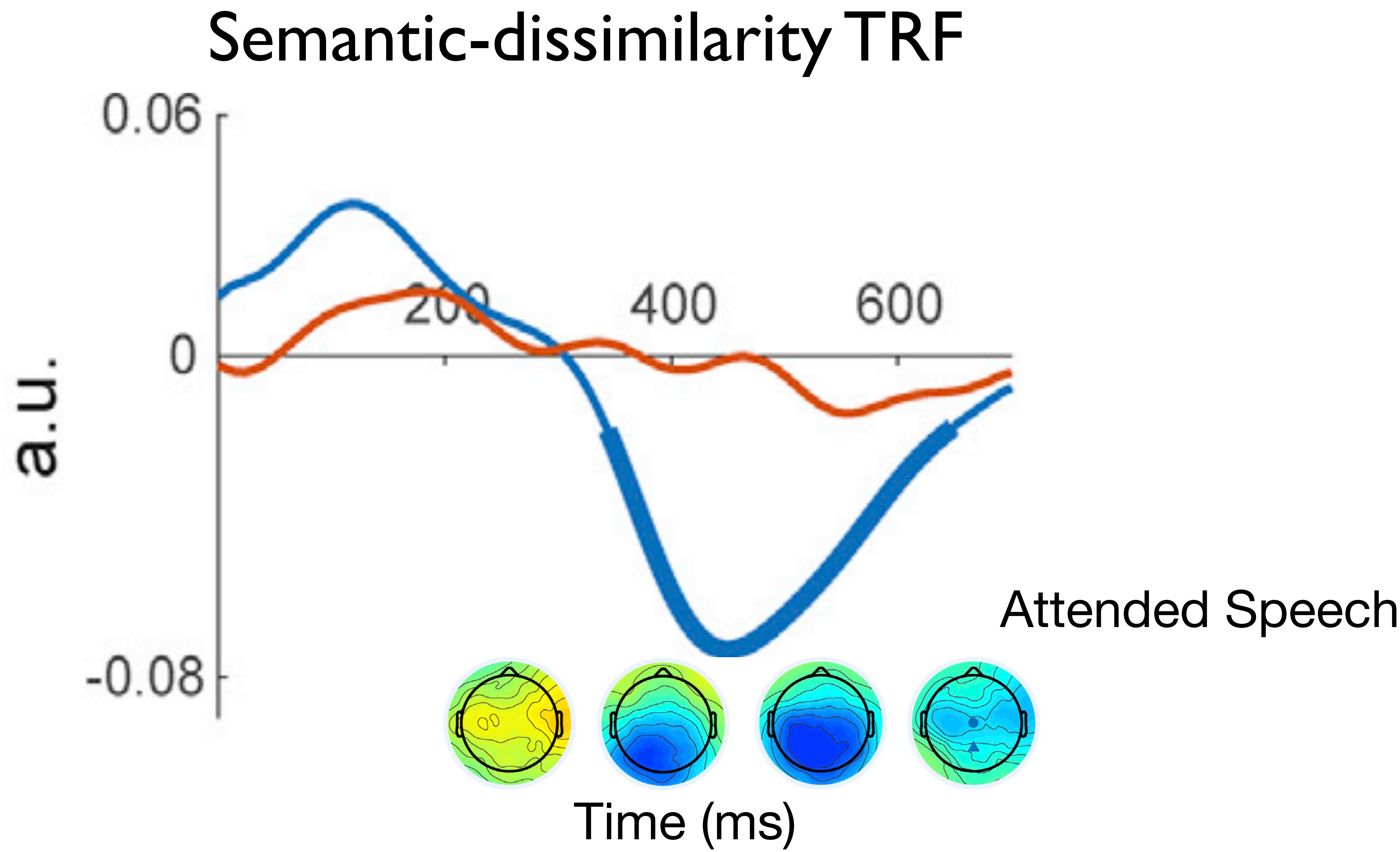
Semantic Processing

Semantic-dissimilarity TRF



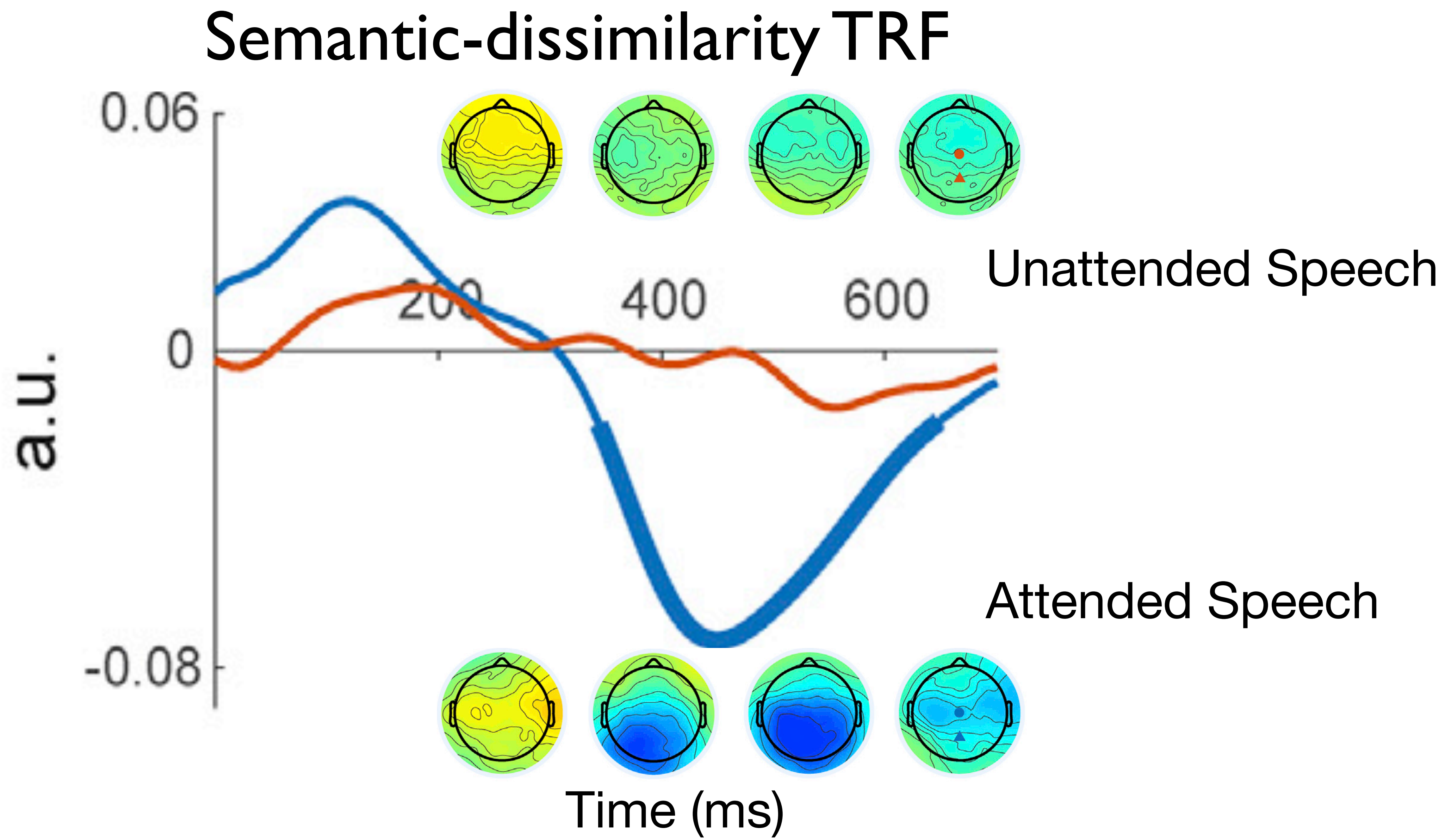
Broderick et al., Curr Biol (2018)

Semantic Processing



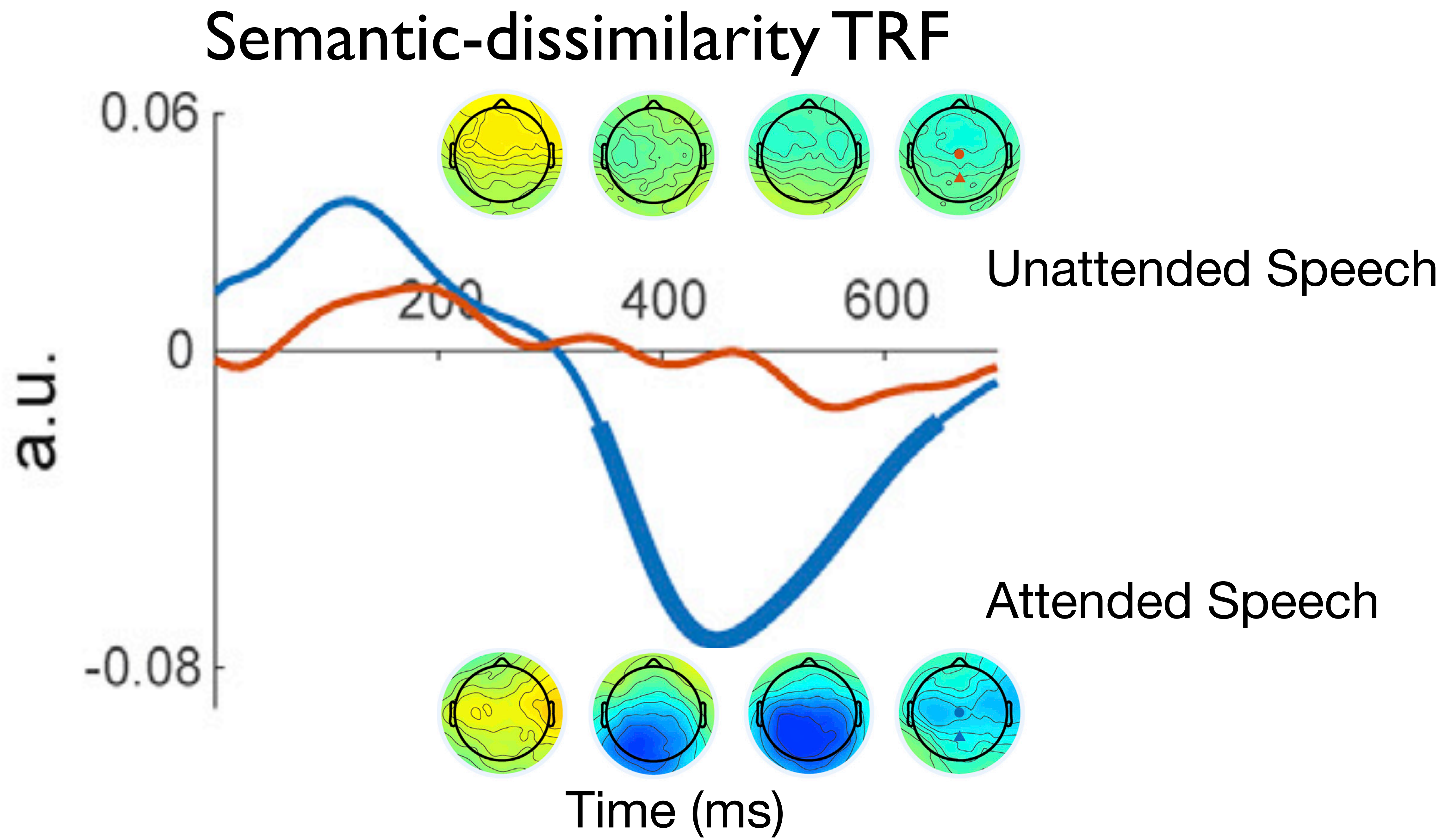
Broderick et al., Curr Biol (2018)

Semantic Processing



Broderick et al., Curr Biol (2018)

Semantic Processing



- This TRF reflects processing of semantics
- This semantic processing depends on attention

Broderick et al., Curr Biol (2018)

Summary

- Speech perception takes many forms
- Cortical processing of speech takes many forms
- Many potential ways to link the two
  - Faithful representation of speech acoustics
  - Processing speech sounds into words (lexical)
  - Semantic level processing
  - Cognitive aspects of perception allowed
- Cortical (temporal) processing of continuous speech processing: both encoding & decoding

Thank You

Acknowledgements

Current Lab Members & Affiliates

Christian Brodbeck
Alex Presacco
Proloy Das
Jason Dunlap
Theo Dutcher
Alex Jiao
Dushyanthi Karunathilake
Joshua Kulasingham
Natalia Lapinskaya
Sina Miran
David Nahmias
Peng Zan
Julian Jenkins
Pirazh Khorramshahi
Huan Luo
Mahshid Najafi
Krishna Puvvada
Jonas Vanthornhout
Ben Walsh
Yadong Wang
Juanjuan Xiang
Jiachen Zhuo

Past Lab Members & Affiliates

Nayef Ahmar
Sahar Akram
Murat Aytekin
Francisco Cervantes Constantino
Maria Chait
Marisel Villafane Delgado
Kim Drnec
Nai Ding
Victor Grau-Serrat

Collaborators

Pamela Abshire
Samira Anderson
Behtash Babadi
Catherine Carr
Monita Chatterjee
Alain de Cheveigné
Stephen David
Didier Depireux
Mounya Elhilali
Tom Francart
Jonathan Fritz
Michael Fu
Stefanie Kuchinsky
Steven Marcus
Cindy Moss
David Poeppel
Shihab Shamma

Past Undergraduate Students

Nicholas Asendorf
Ross Baehr
Anurupa Bhonsale
Sonja Bohr
Elizabeth Camenga
Katya Dombrowski
Kevin Hogan
Andrea Shome
James Williams

Funding NIH (*NIDCD*, *NIA*, *NIBIB*); NSF; DARPA; UMD; USDA