

Dynamic Processing of Background Speech at the Cocktail Party: Evidence for Early Active Cortical Stream Segregation

Christian Brodbeck¹, Alex Jiao¹, L. Elliot Hong²,
Jonathan Z. Simon¹

¹ *University of Maryland, College Park*

² *University of Maryland School of Medicine, Baltimore*

Outline

- Cocktail party listening
 - ▶ Speech segregation & cortical processing of ignored speech
 - ▶ MEG representations of speech
- Methods
- Results
- Summary

Outline

- Cocktail party listening
 - ▶ Speech segregation & cortical processing of ignored speech
 - ▶ MEG representations of speech
- Methods
- Results
- Summary

Listening to Speech at the Cocktail Party



Listening to Speech at the Cocktail Party



Listening to Speech at the Cocktail Party



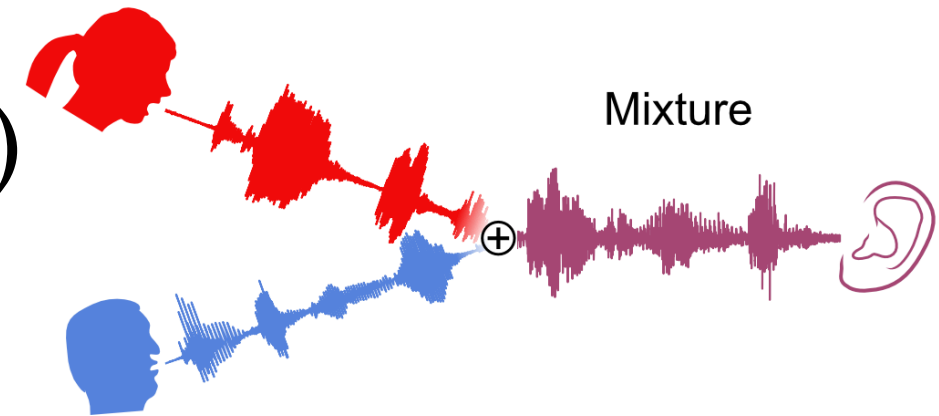
Listening to Speech at the Cocktail Party



Cocktail Party Problem

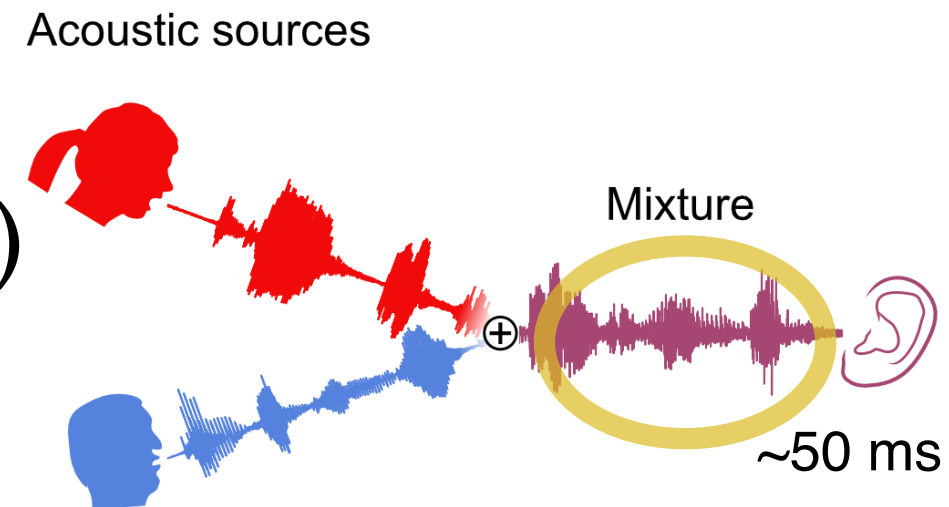
- Acoustic scene
 - Acoustic mixture (as in periphery)
 - Acoustic sources (talkers)

Acoustic sources



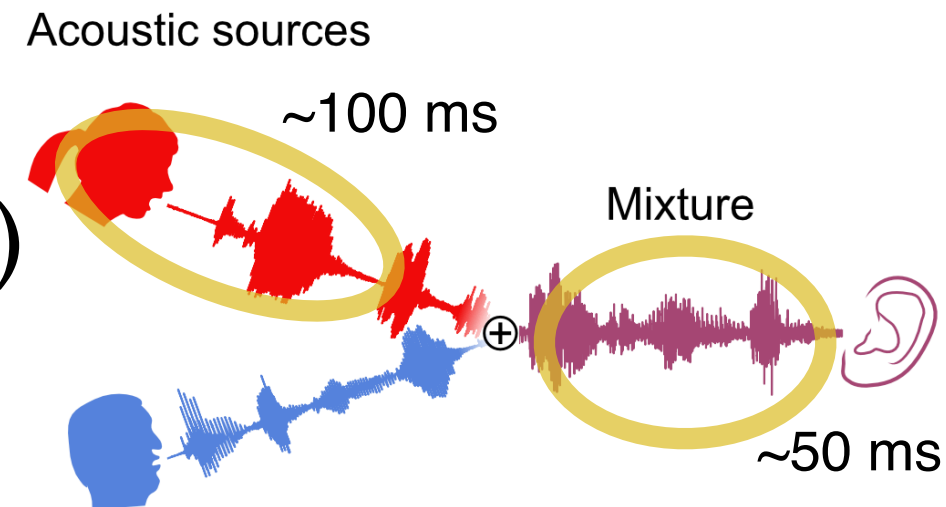
Cocktail Party Problem

- Acoustic scene
 - Acoustic mixture (as in periphery)
 - Acoustic sources (talkers)
- Cortical representations
 - Early (~50 ms): acoustic mixture (Puvvada & Simon, 2017)



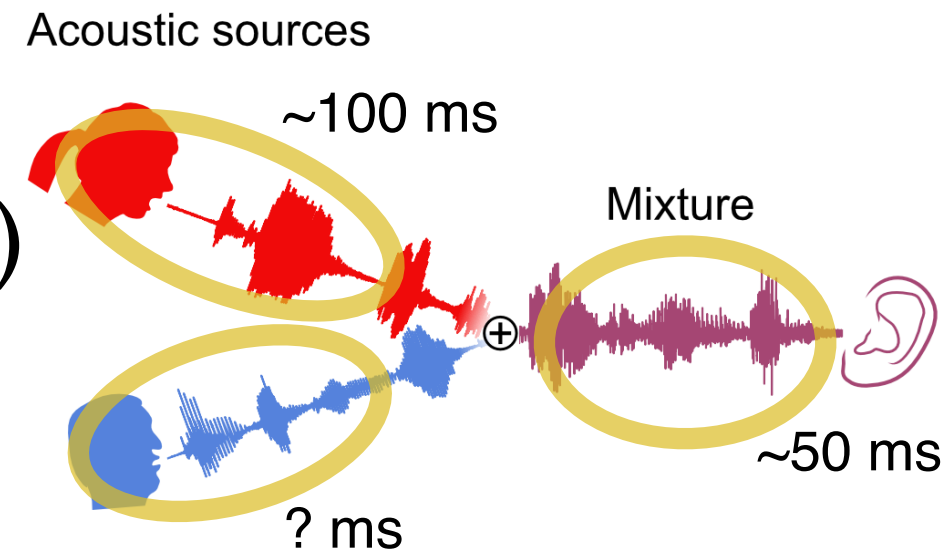
Cocktail Party Problem

- Acoustic scene
 - Acoustic mixture (as in periphery)
 - Acoustic sources (talkers)
- Cortical representations
 - Early (~50 ms): acoustic mixture (Puvvada & Simon, 2017)
 - Later (~100 ms): preferential for attended speech (Ding & Simon, 2012; O'Sullivan et al., 2019)



Cocktail Party Problem

- Acoustic scene
 - Acoustic mixture (as in periphery)
 - Acoustic sources (talkers)
- Cortical representations
 - Early (~50 ms): acoustic mixture (Puvvada & Simon, 2017)
 - Later (~100 ms): preferential for attended speech (Ding & Simon, 2012; O'Sullivan et al., 2019)
- How is ignored speech separated from the mixture in auditory cortex?
 - How is either speech source separated?
 - Passive mechanisms vs. active mechanisms?



Outline

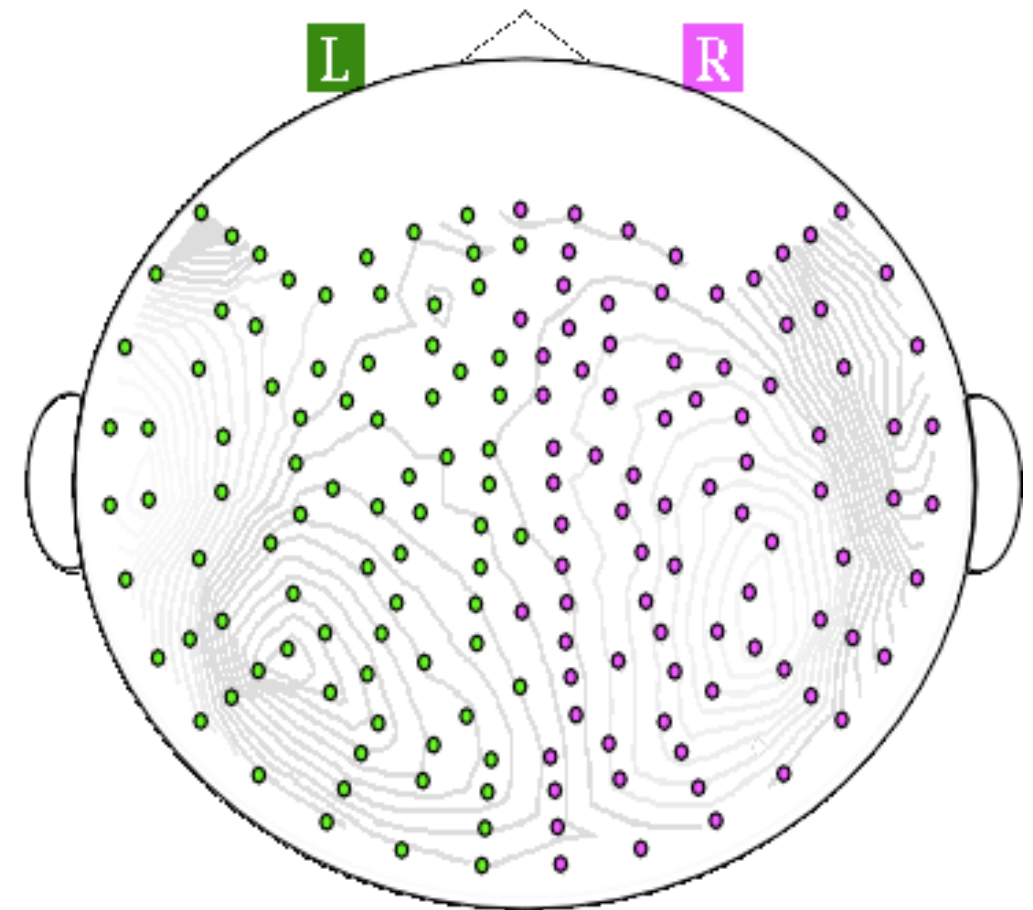
- Cocktail party listening
 - ▶ Speech segregation & cortical processing of ignored speech
 - ▶ MEG representations of speech
- Methods
- Results
- Summary

Outline

- Cocktail party listening
 - ▶ Speech segregation & cortical processing of ignored speech
 - ▶ **MEG representations of speech**
- Methods
- Results
- Summary

Magnetoencephalography (MEG)

- Non-invasive, passive, silent neural recordings from cortex
- Simultaneous whole-head recording (~200 sensors)
- Sensitivity
 - high: ~ 100 fT (10^{-13} Tesla)
 - low: $\sim 10^4 - \sim 10^6$ neurons
- Temporal resolution: ~ 1 ms
- Spatial resolution
 - coarse: ~ 1 cm
 - ambiguous



Neural Signals & MEG

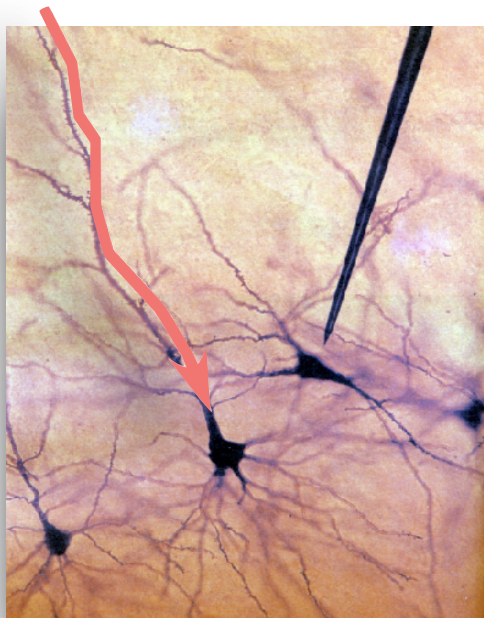
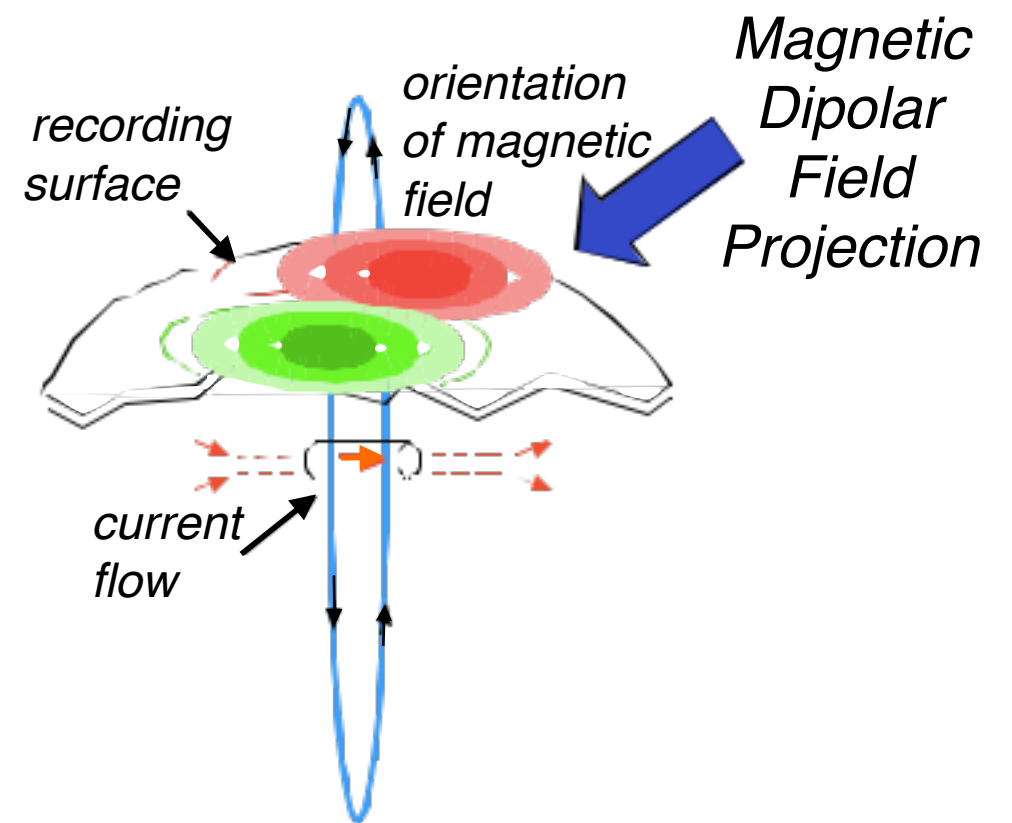
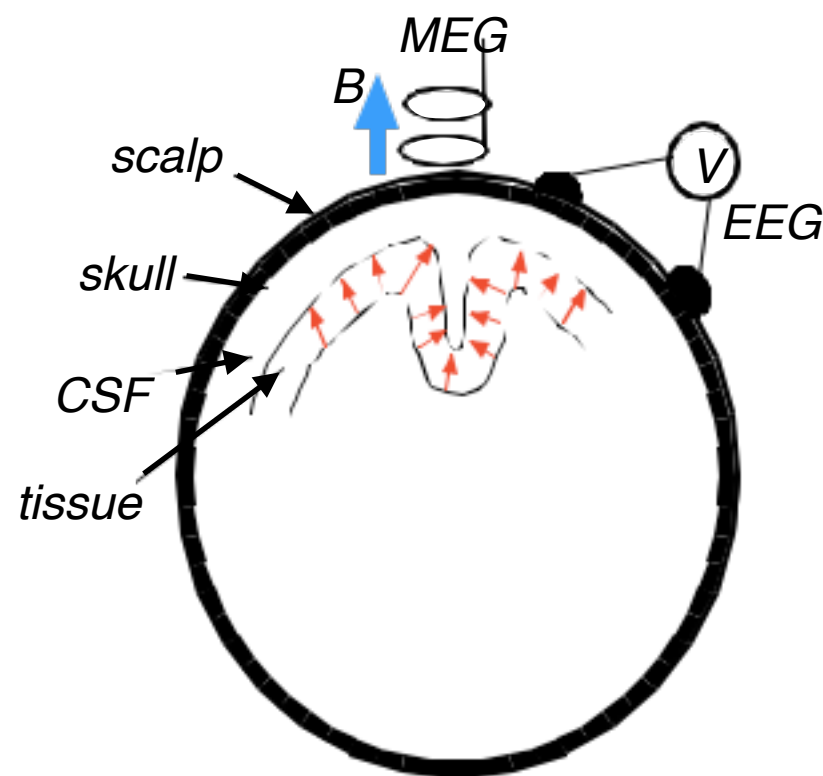


Photo by Fritz Goro

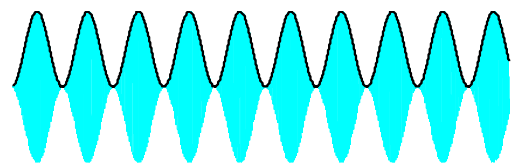


- Direct electrophysiological measurement
 - not hemodynamic
 - real-time
- No unique solution for distributed source

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

Time Course of MEG Responses

AM at 3 Hz

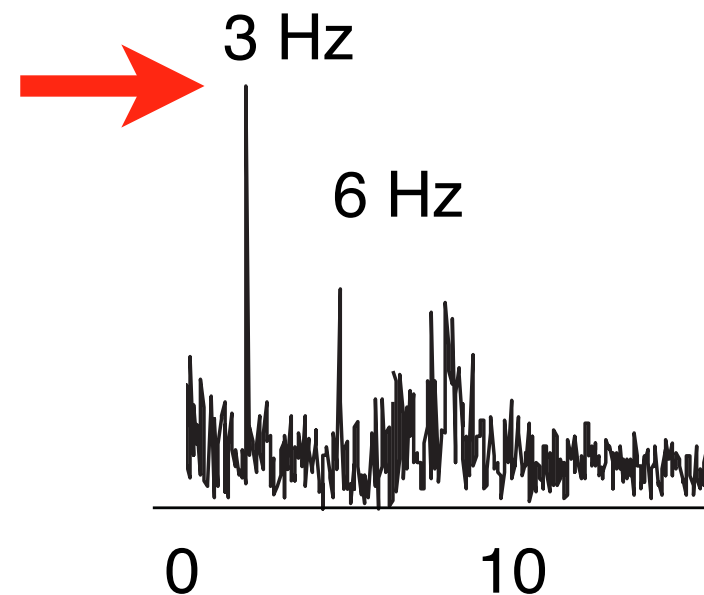


3 Hz phase-locked response



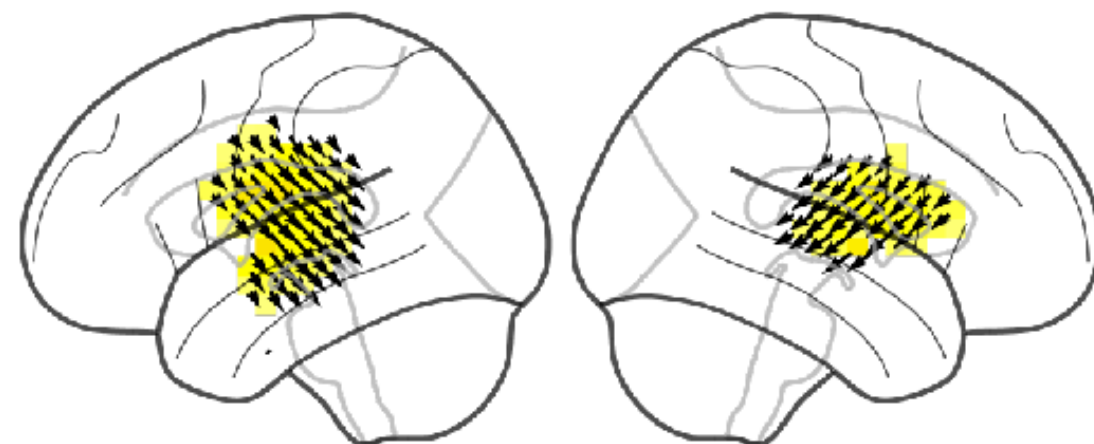
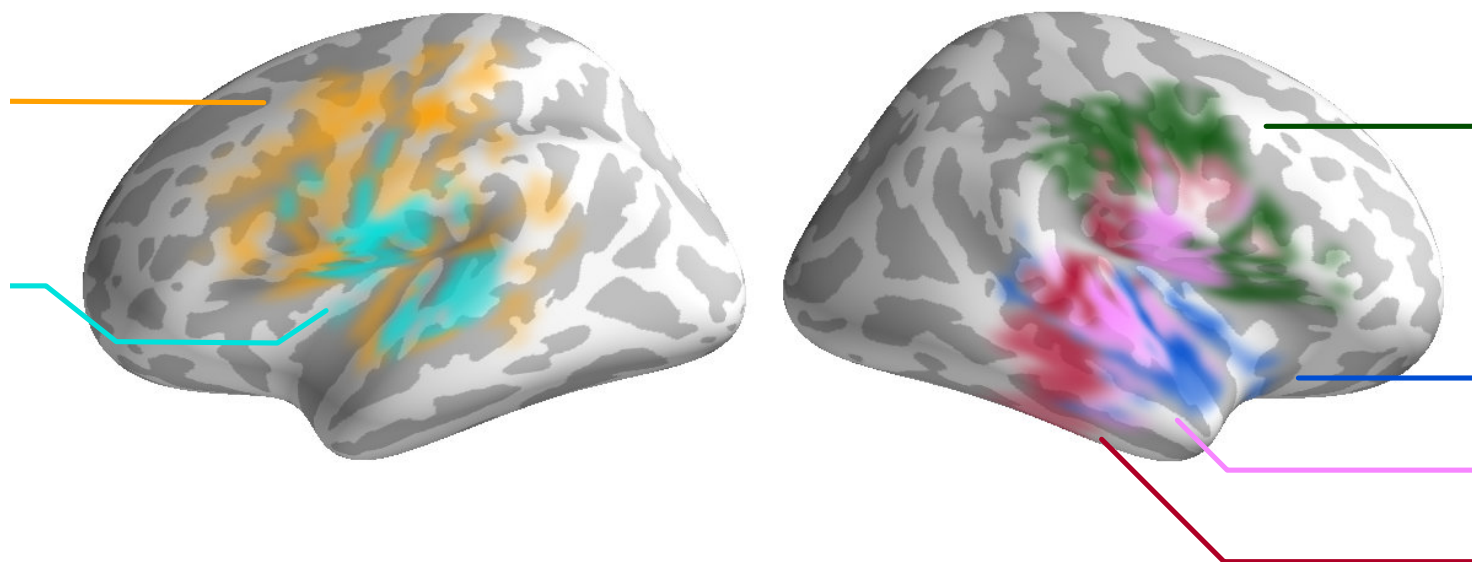
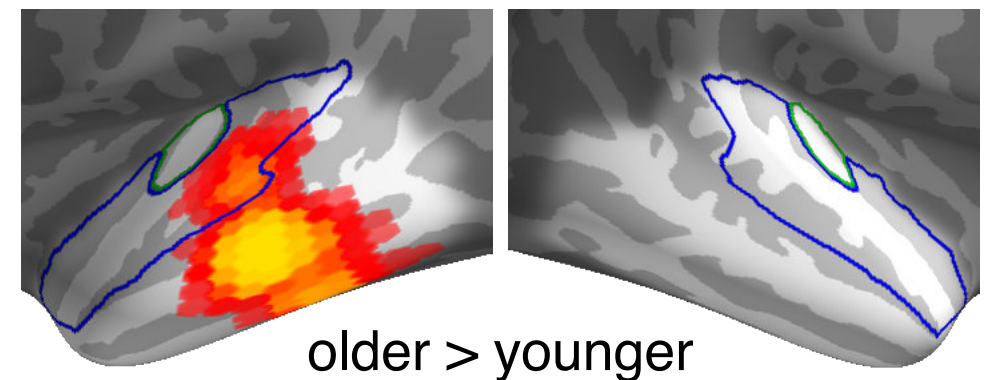
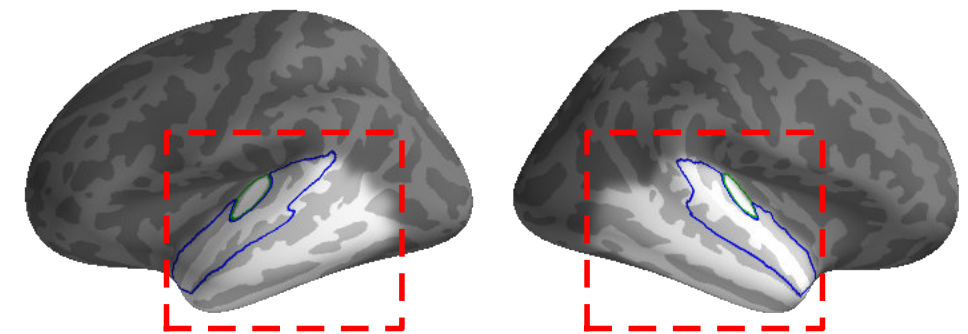
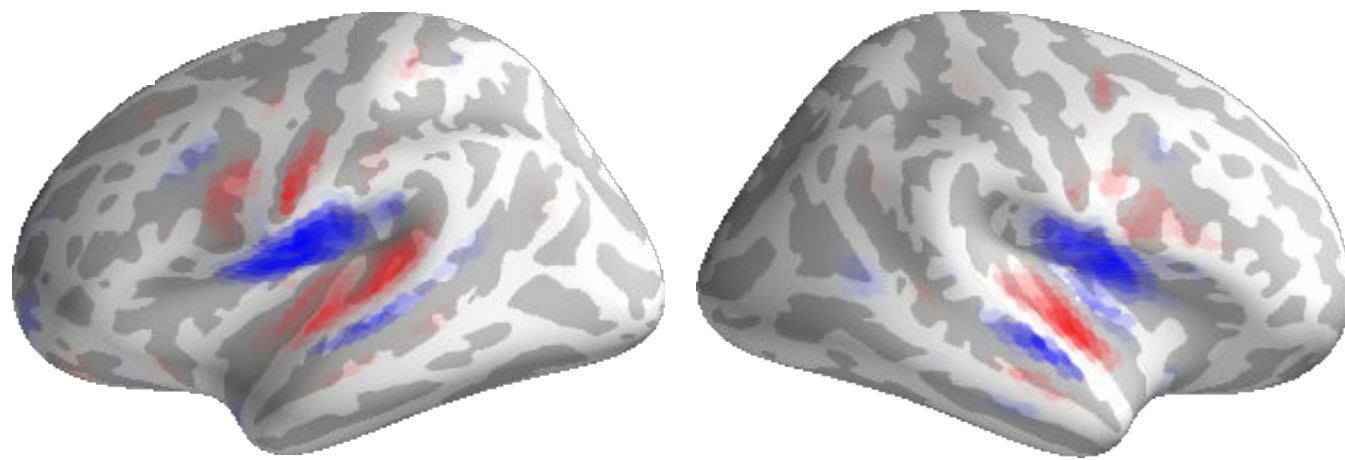
MEG activity time-locks
to temporal modulations
of sound

response spectrum (*subject R0747*)



Frequency (Hz)

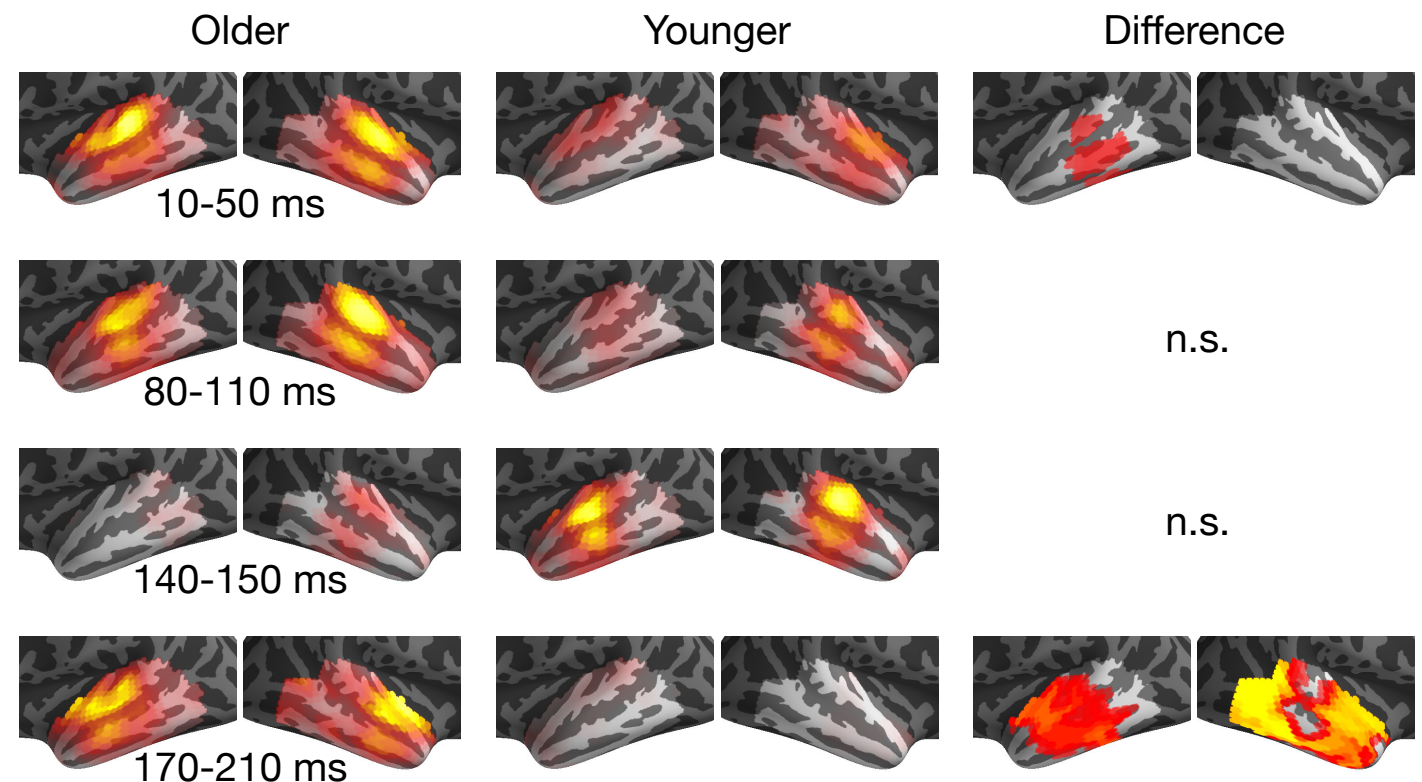
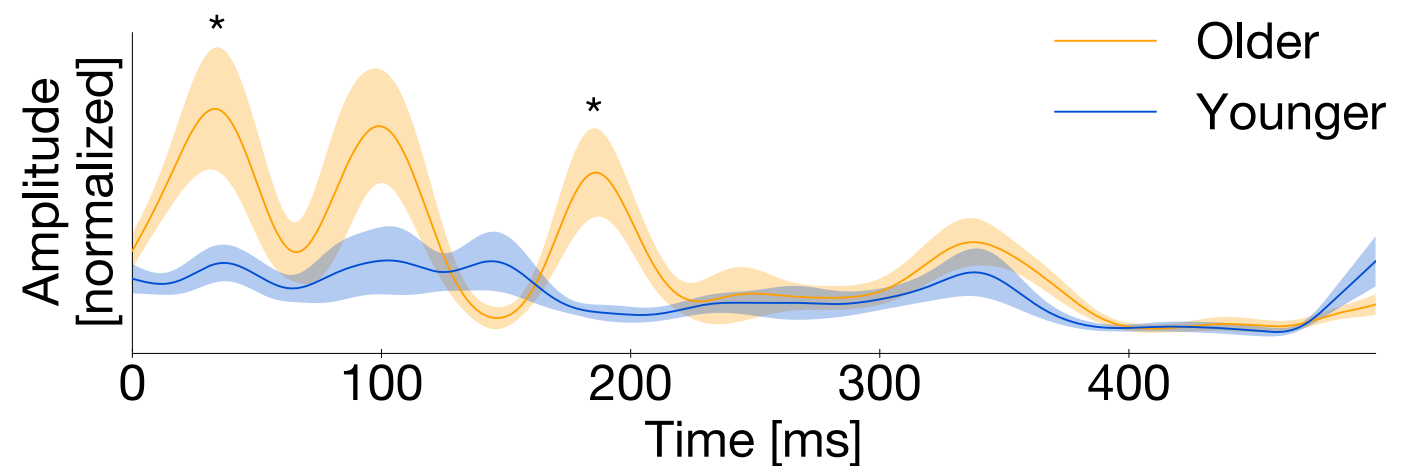
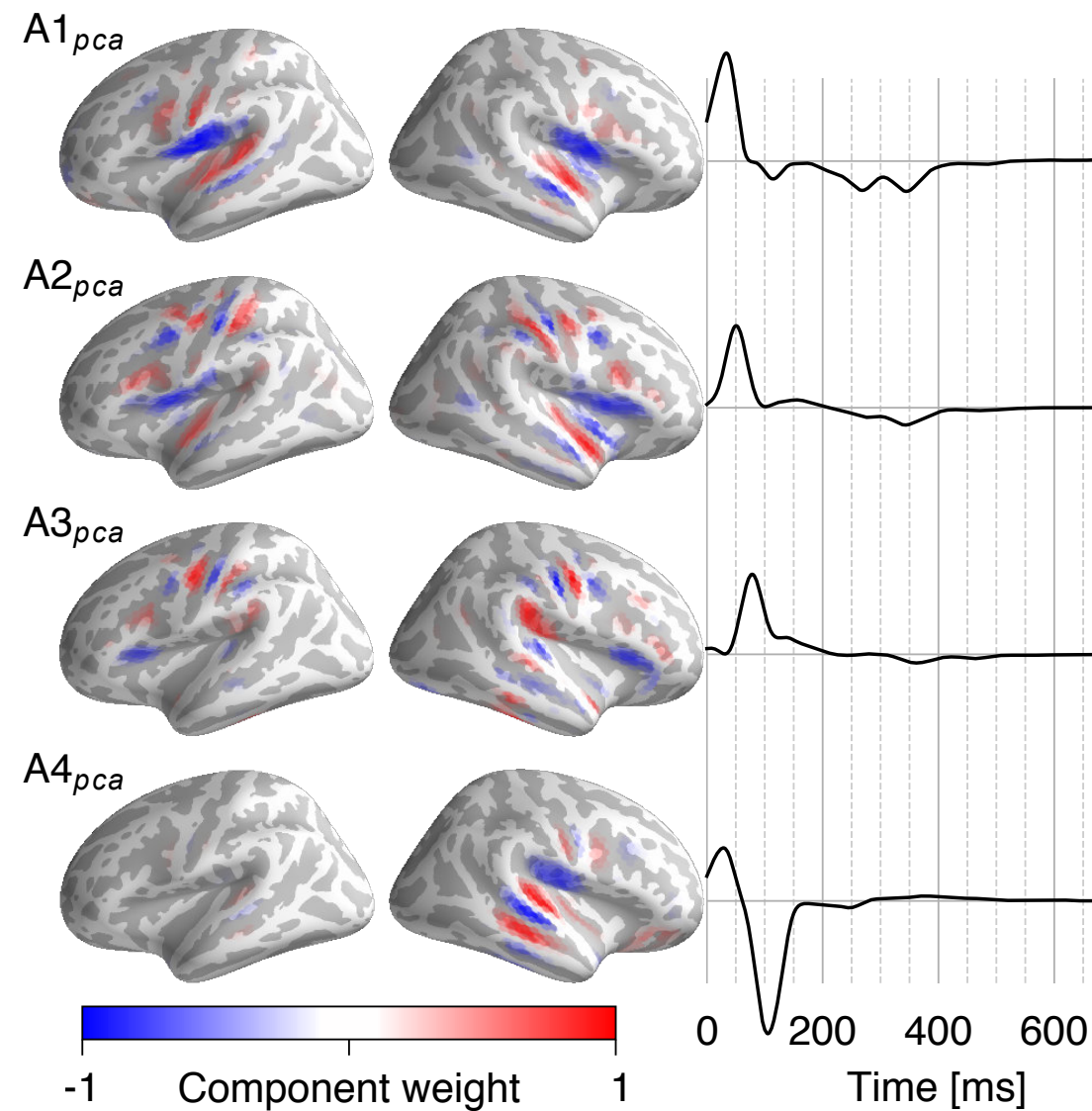
Spatial Distributions of MEG Neural Currents



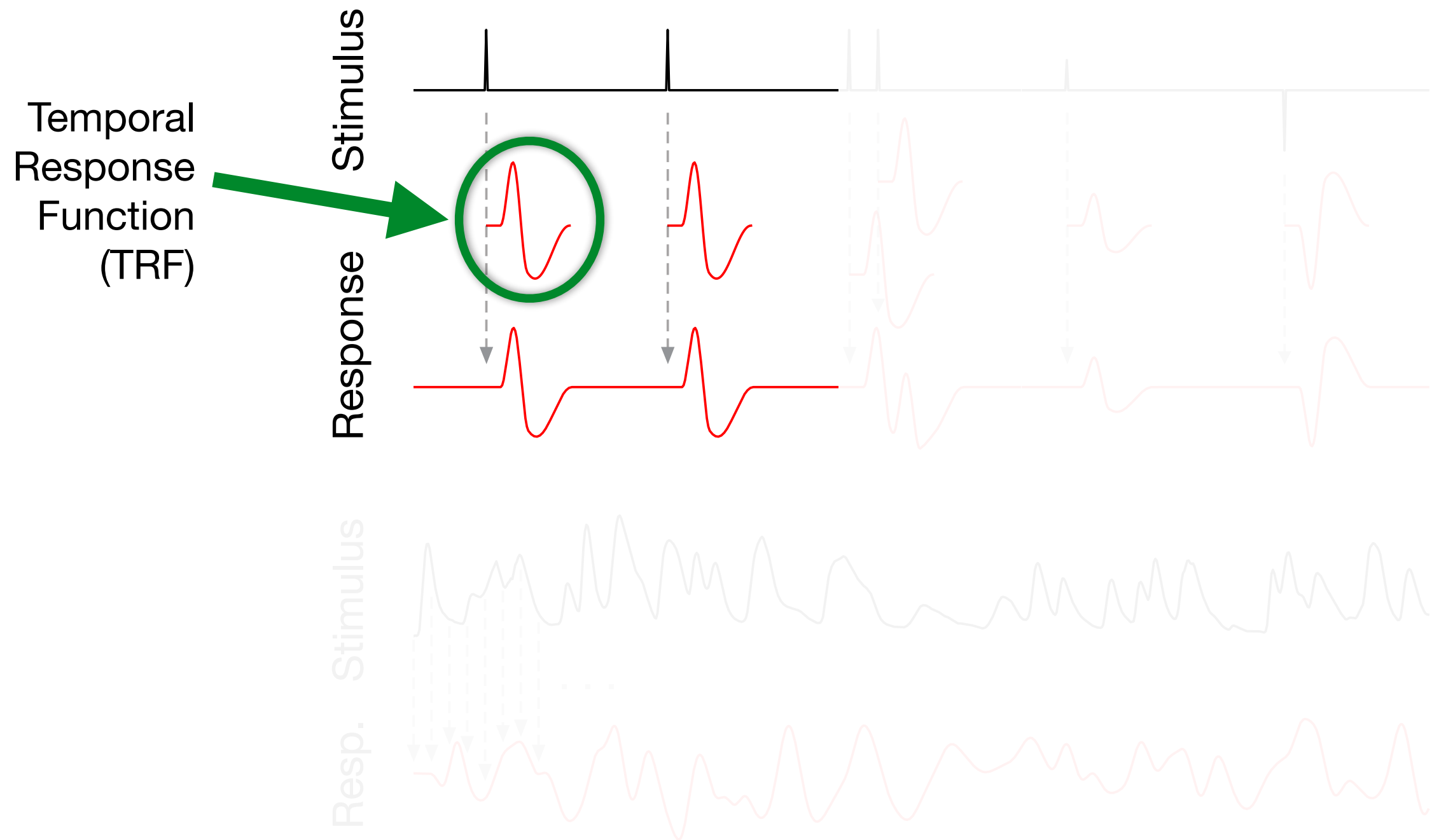
Brodbeck et al., Neurolmage (2017)
Brodbeck et al., Acta Acust united Ac (2018)

Das et al., Neurolmage (2020)

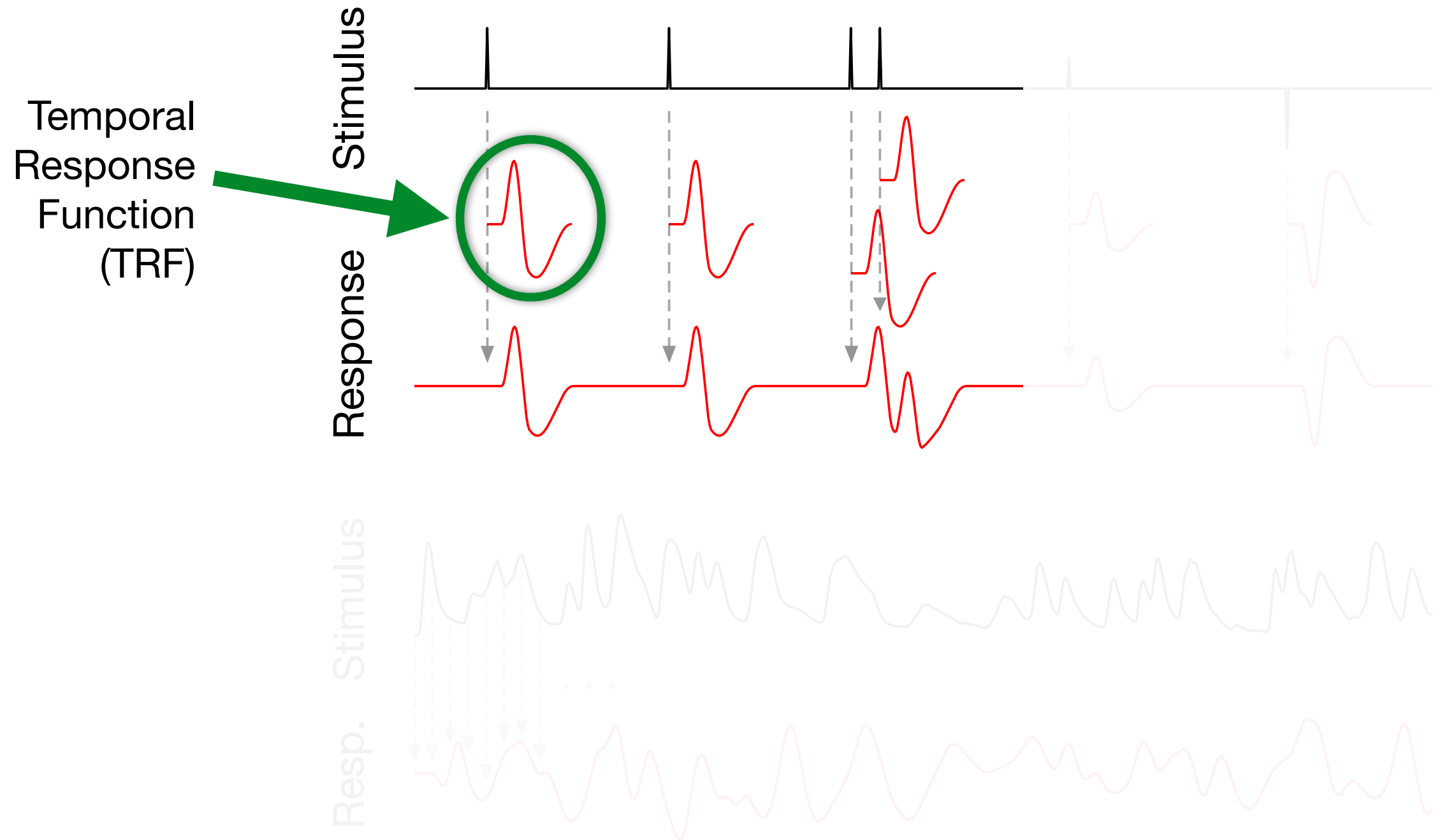
Spatiotemporal Distribution of Neural Currents



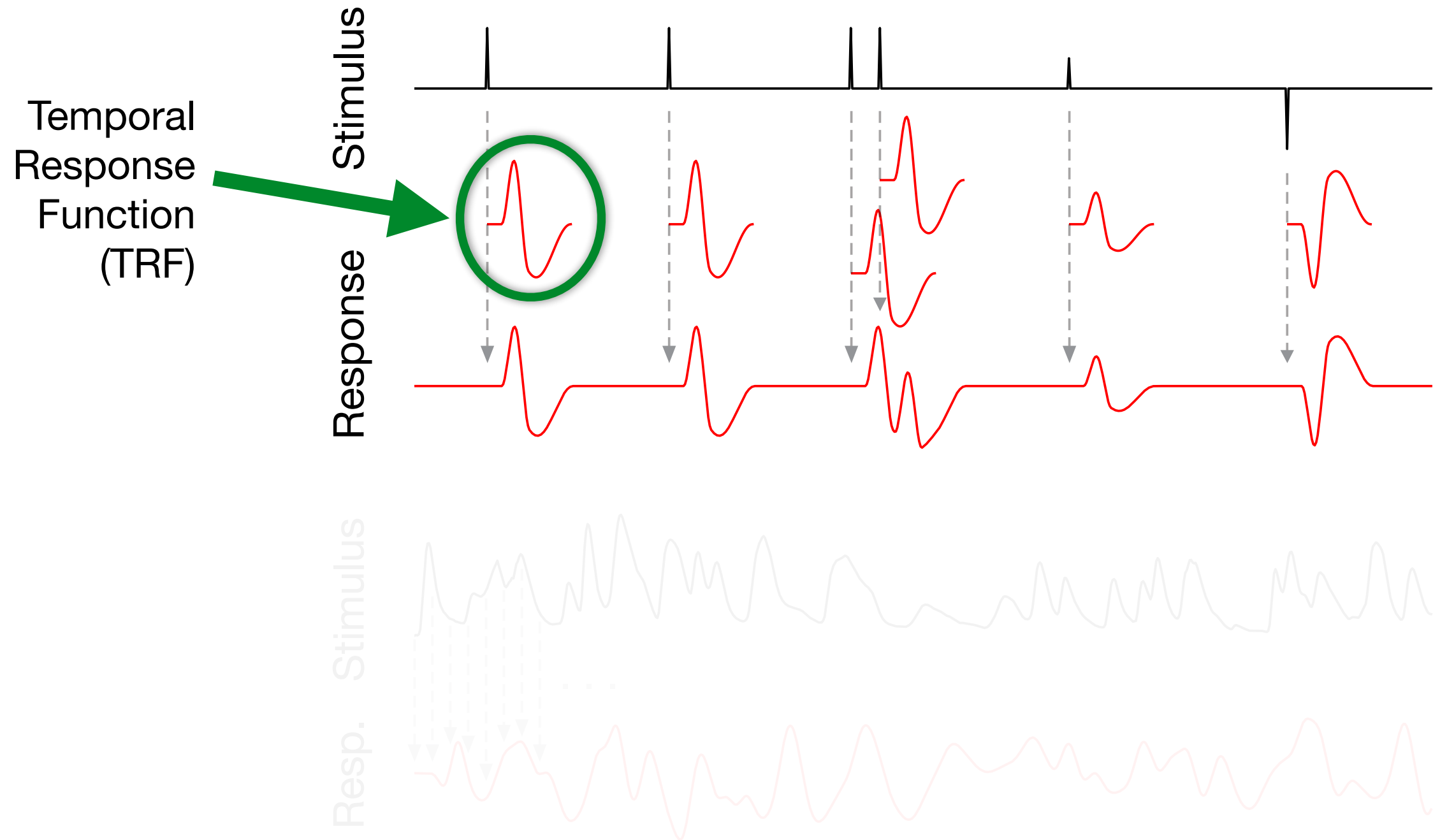
Temporal Response Functions



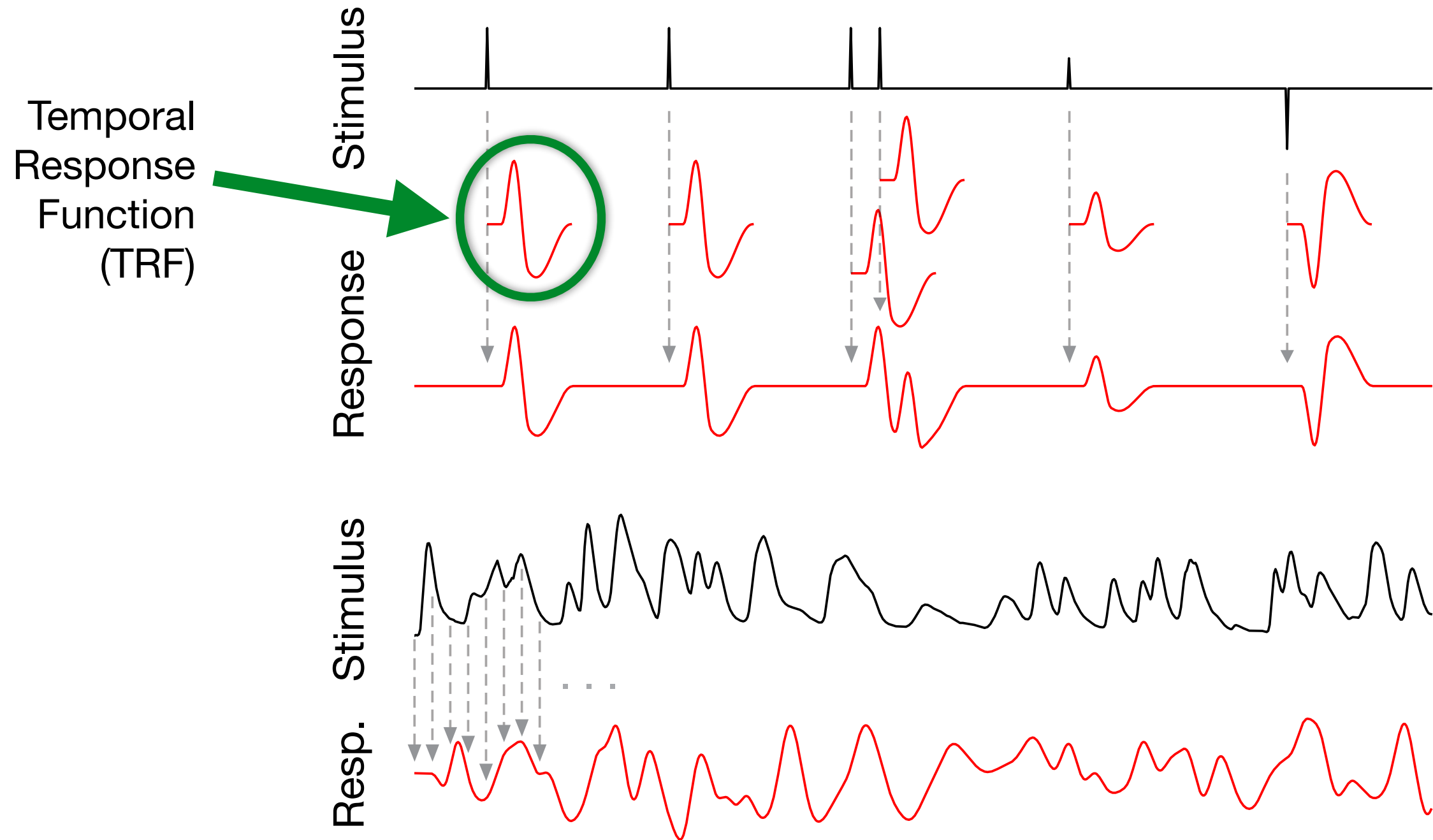
Temporal Response Functions



Temporal Response Functions



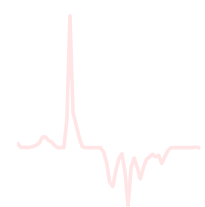
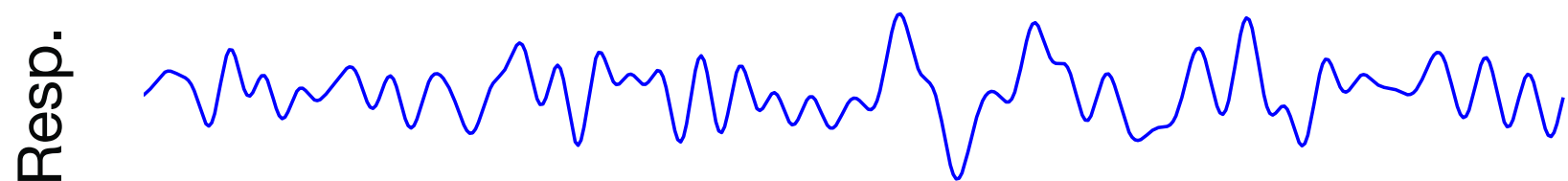
Temporal Response Functions



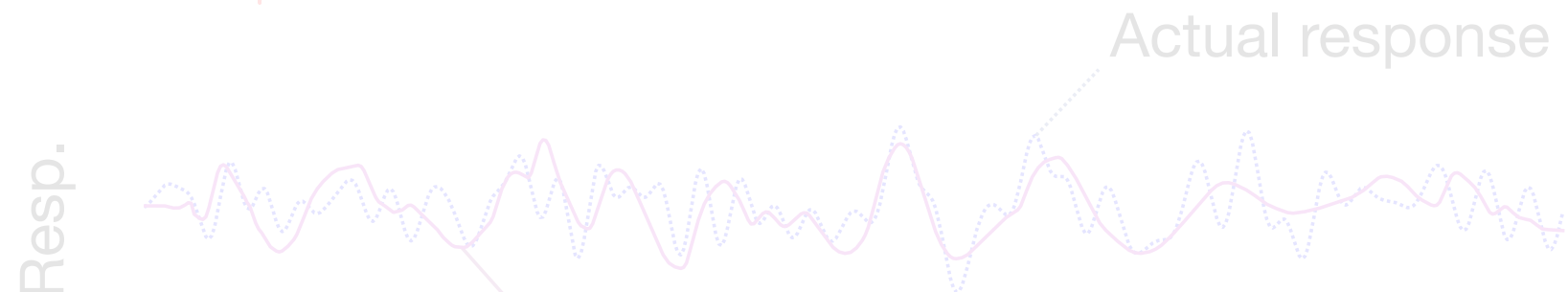
Temporal Response Functions

Temporal Response Function (TRF) estimation:

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



Estimated TRF



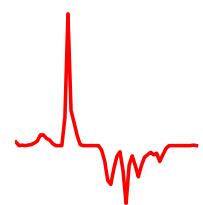
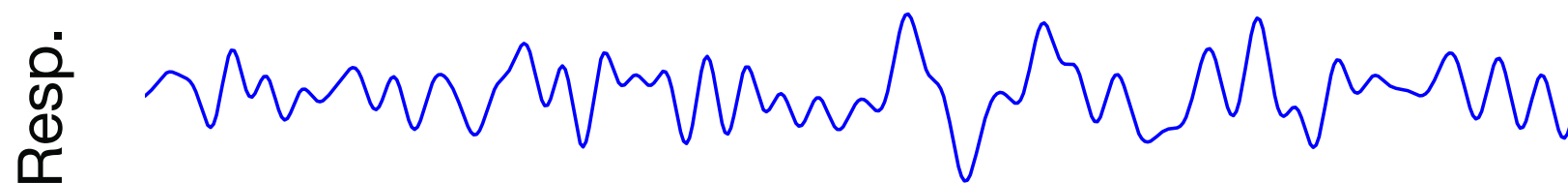
Actual response

Predicted response (Stimulus * TRF)

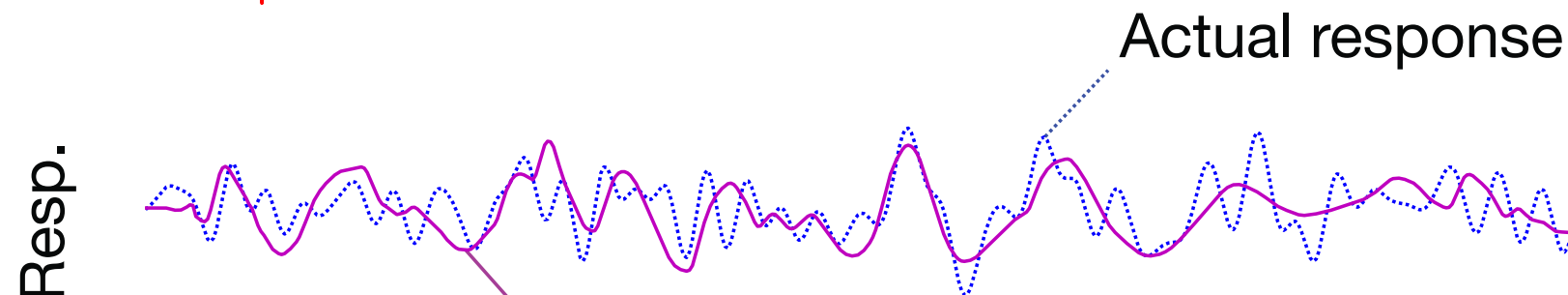
Temporal Response Functions

Temporal Response Function (TRF) estimation:

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



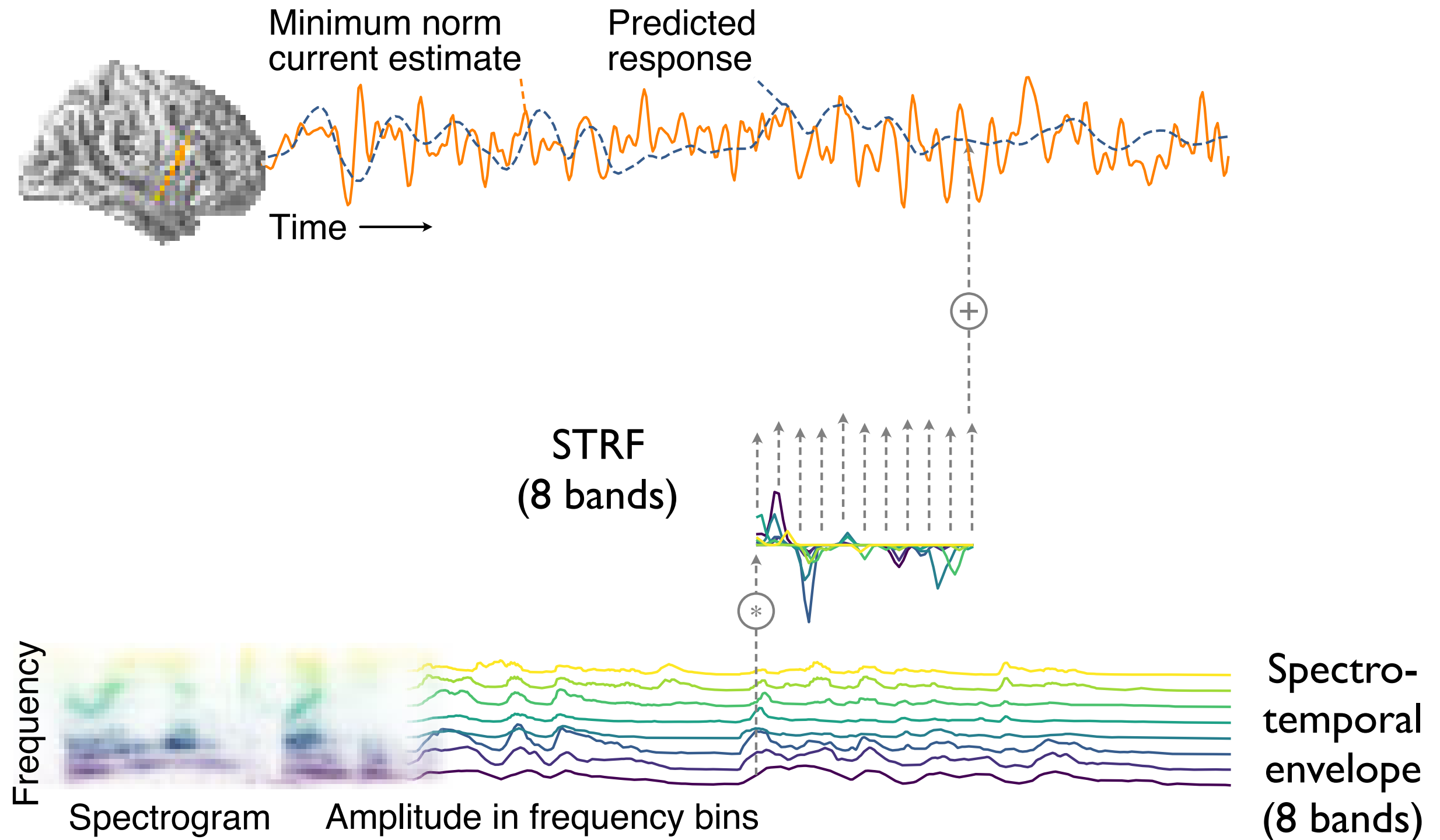
Estimated TRF



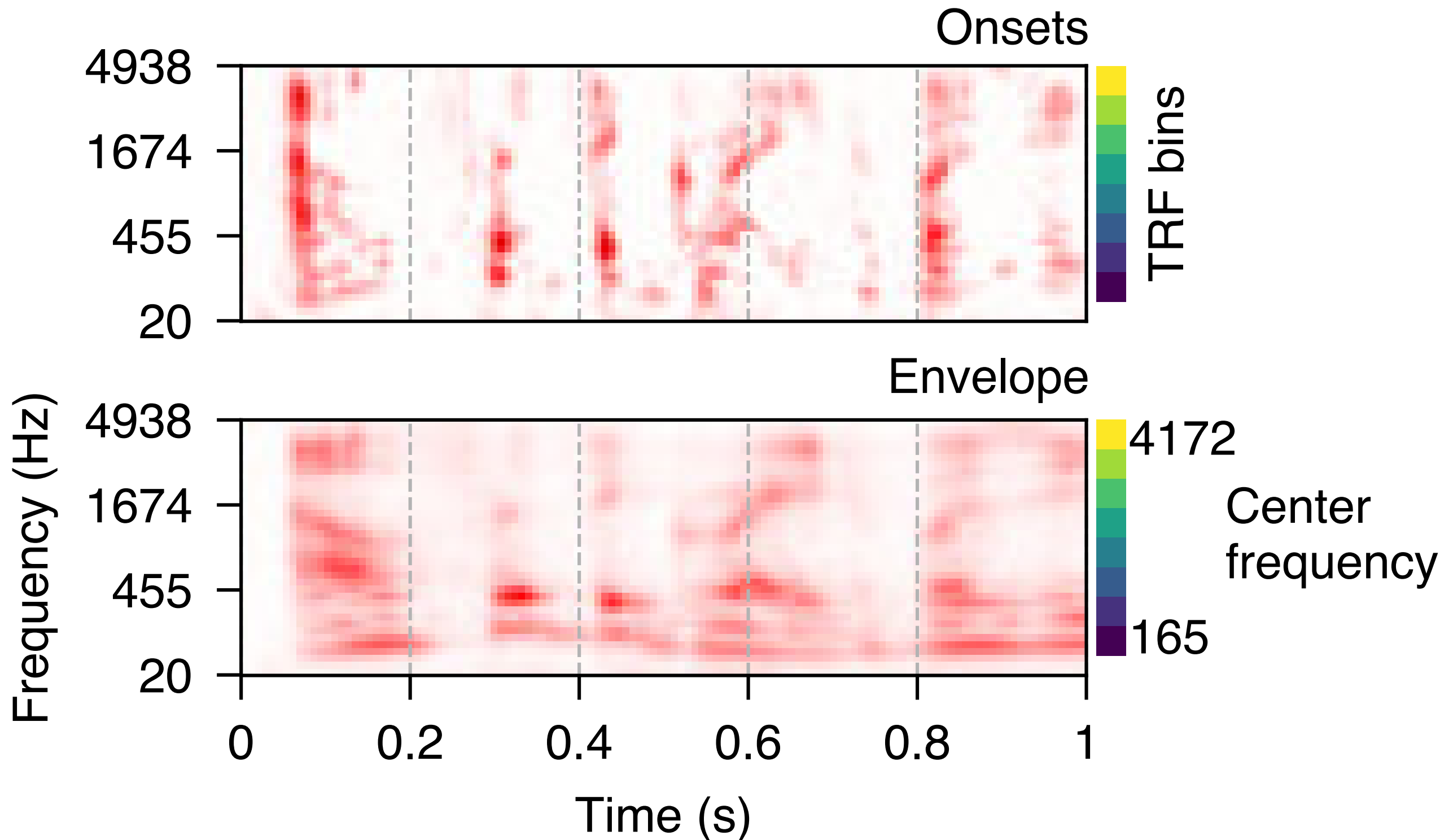
Actual response

Predicted response (Stimulus * TRF)

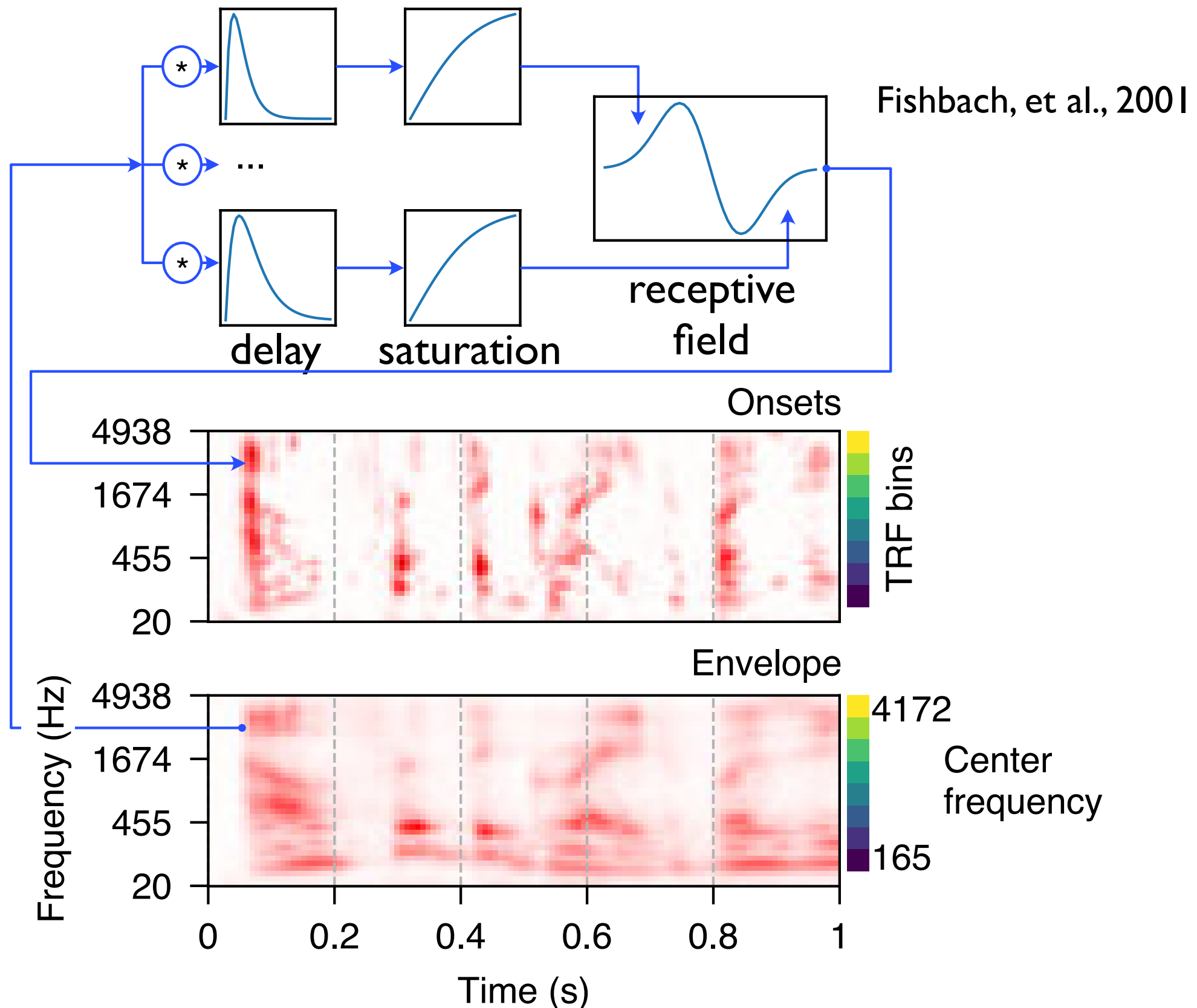
Spectro-Temporal Response Functions



Spectro-Temporal Onset vs Envelope



Spectro-Temporal Onset vs Envelope



Spectro-Temporal Onset vs Envelope

Onset Properties:

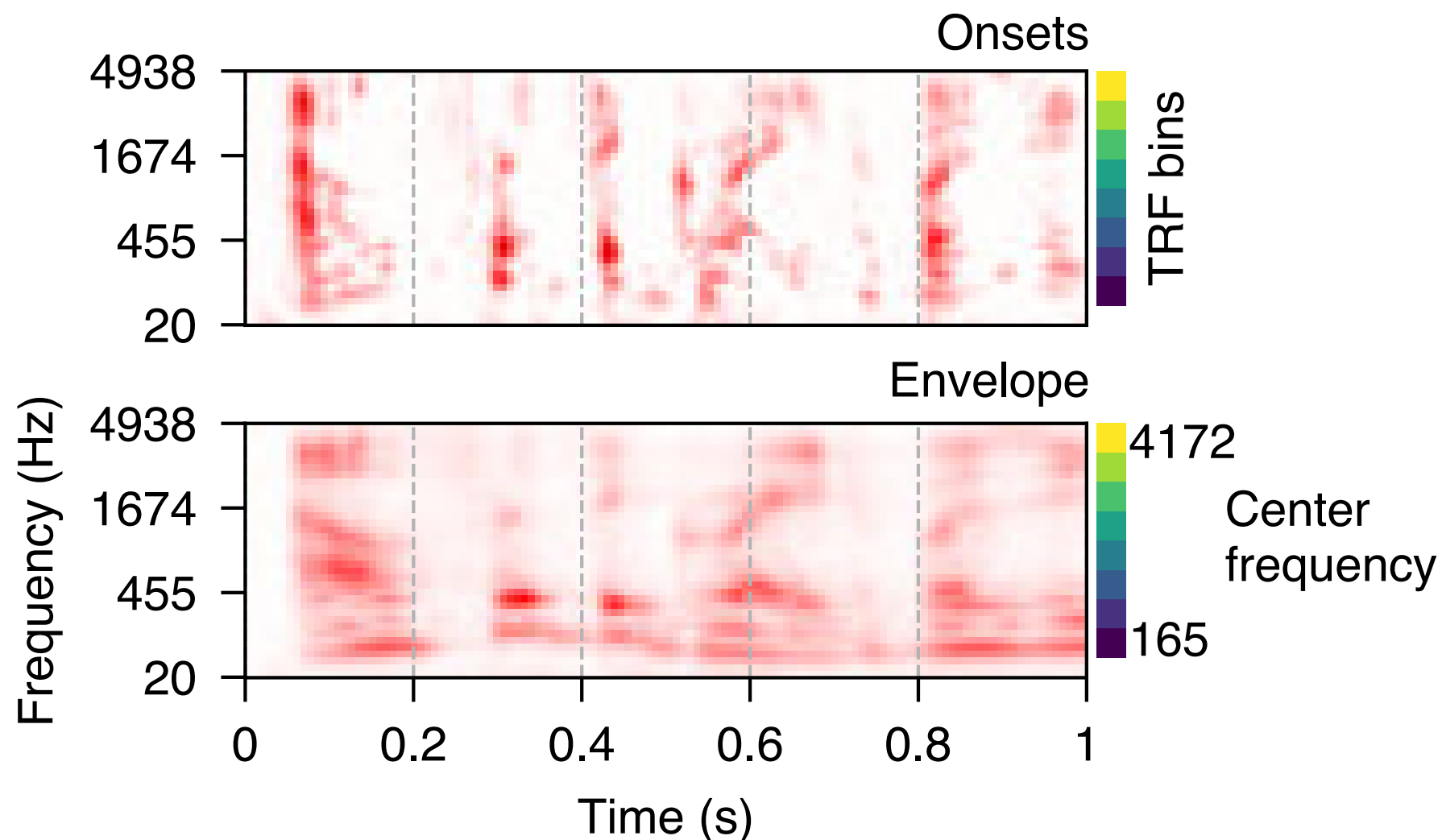
- Local increase in acoustic energy
- Prominent responses in auditory cortex
- Promote perceptual grouping
- Promote auditory object perception
- Can better distinguish between mixture and individual sources

Cervantes Constantino et al., 2017

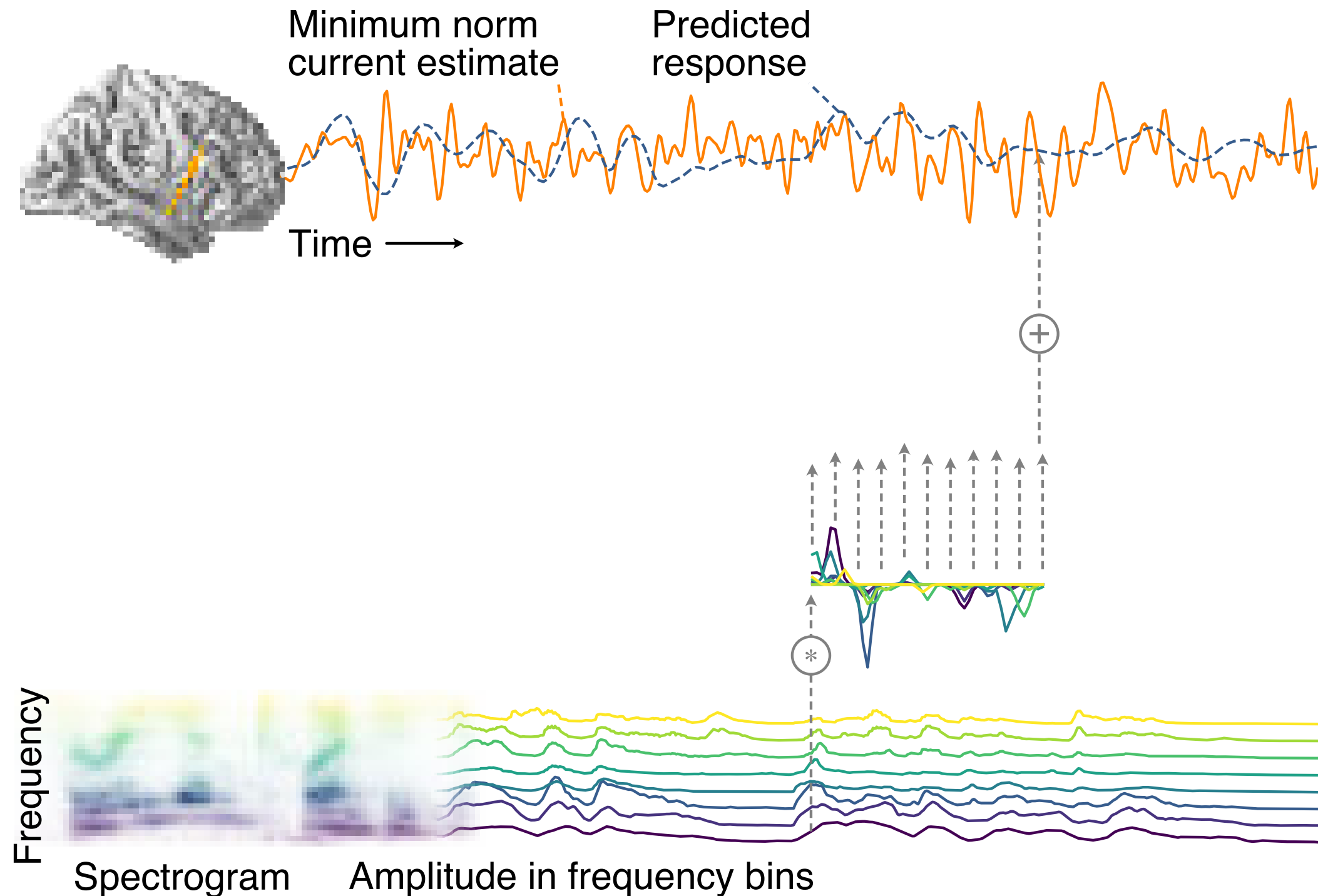
Hamilton et al., 2018

Daube et al., 2019

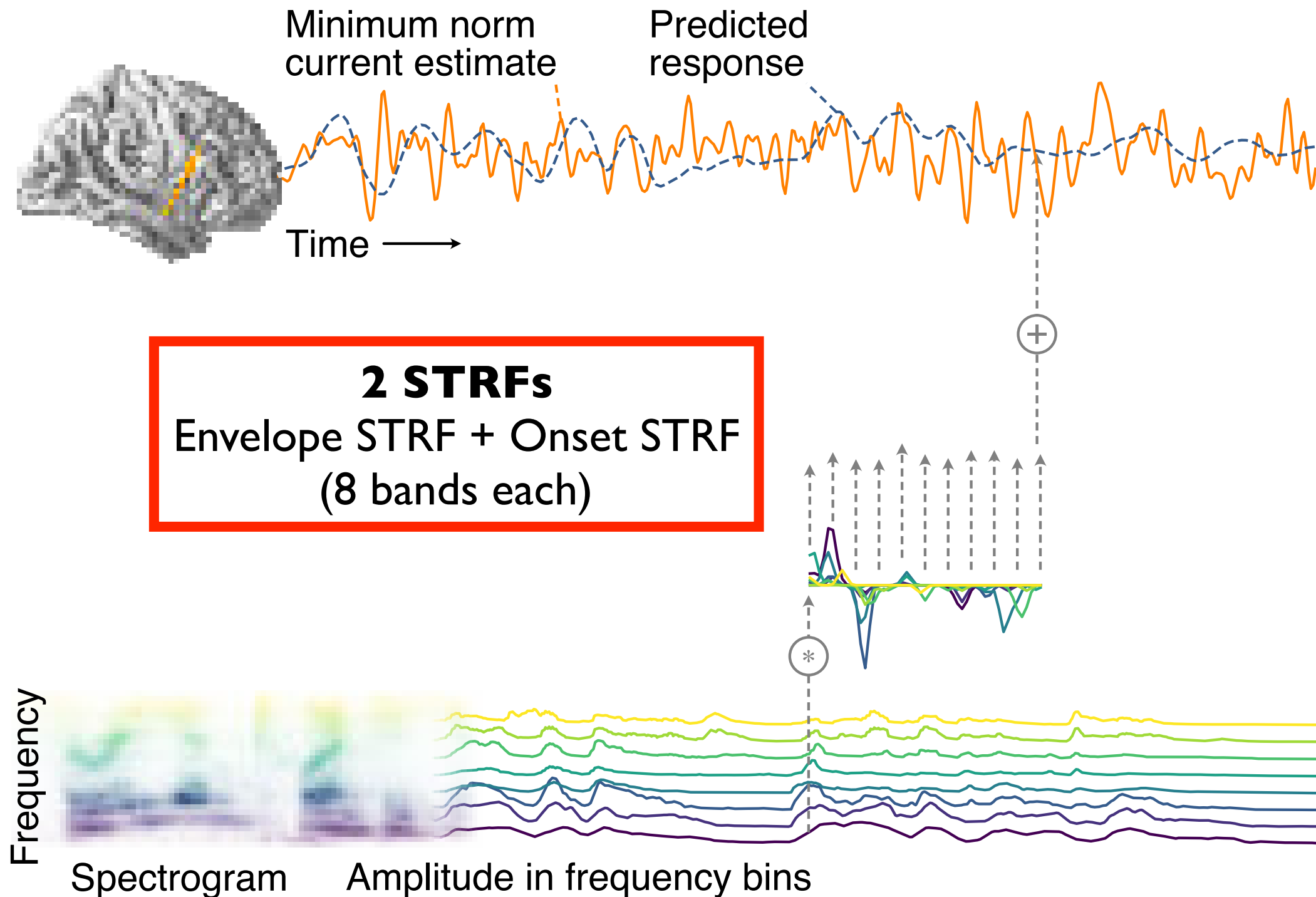
Bregman et al., 1994



Spectro-Temporal Response Functions



Spectro-Temporal Response Functions



Outline

- Cocktail party listening
 - ▶ Speech segregation & cortical processing of ignored speech
 - ▶ **MEG representations of speech**
- Methods
- Results
- Summary

Outline

- Cocktail party listening
 - ▶ Speech segregation & cortical processing of ignored speech
 - ▶ MEG representations of speech
- **Methods**
- Results
- Summary

Methods

Data set already used and described in Brodbeck, 2019

26 adults, mean age 45 (range 22 - 61)

8 one-minute-long segments (4 male + 4 female speakers) from *A Child's History of England* by Dickens

16 one-minute-long segments constructed from the same passages **with two competing speakers**, male + female, equal loudness

- Subjects' instructions: Attend to one, ignore the other (counter-balanced)
- After each segment, answer question about content of the attended stimulus

Distributed MNE source estimates, restricted to Region of Interest (below)

- Sources in *fsaverage* brain (individual anatomical MRI not used)

Multivariable TRF at each source element via boosting (10 ms resolution; 50 ms Hamming window basis)

Significance of each representation with respect to shuffled stimulus x 3

Threshold-free cluster enhancement, 10,000 permutation null distribution

Outline

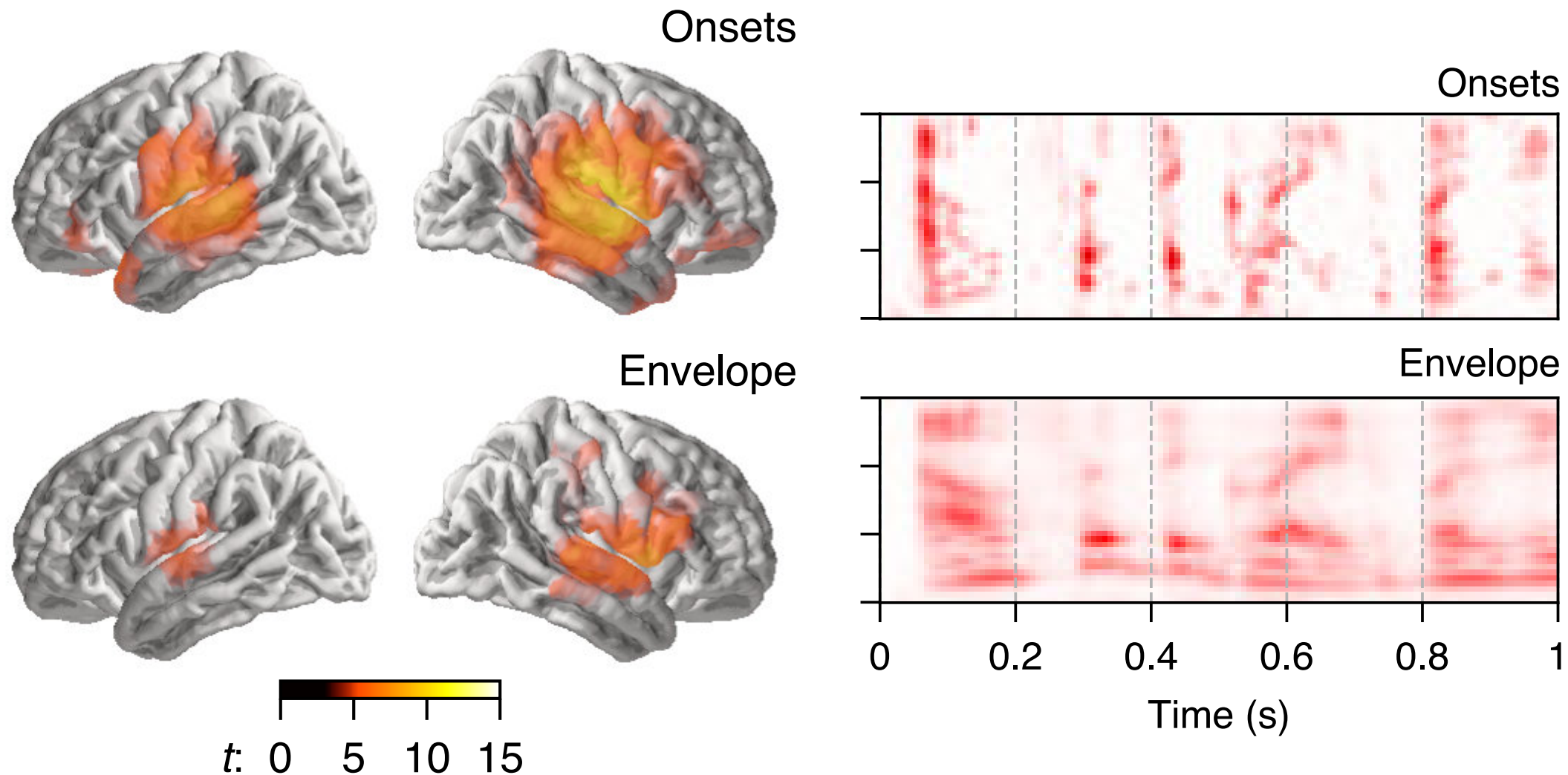
- Cocktail party listening
 - ▶ Speech segregation & cortical processing of ignored speech
 - ▶ MEG representations of speech
- **Methods**
- Results
- Summary

Outline

- Cocktail party listening
 - ▶ Speech segregation & cortical processing of ignored speech
 - ▶ MEG representations of speech
- Methods
- **Results**
- Summary

Successful Response Prediction

Significant prediction

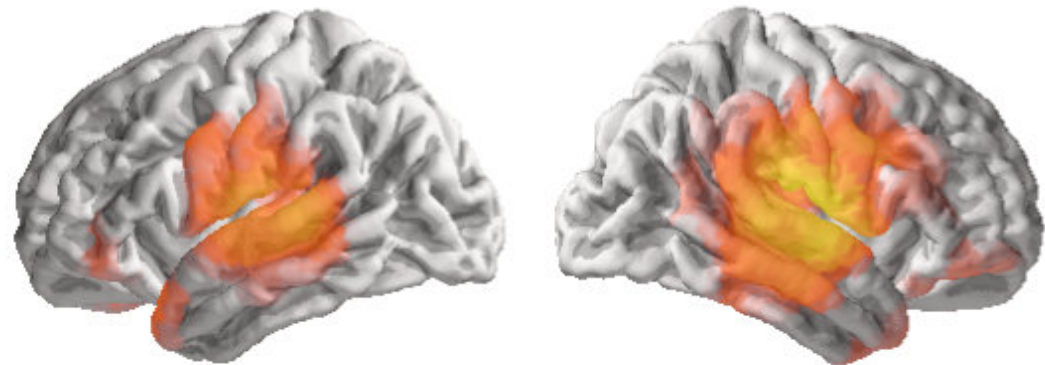


Source Localization consistent with auditory cortex
(Heschl's gyrus, superior temporal Gyrus)

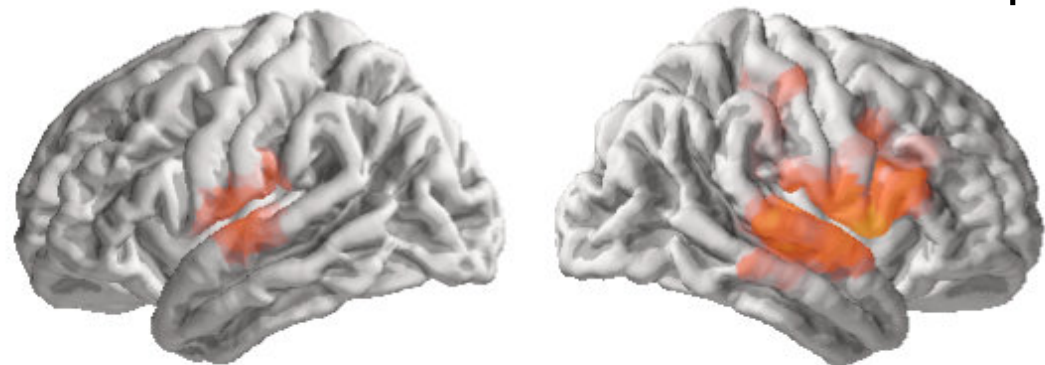
Successful Response Prediction

Significant prediction

Onsets

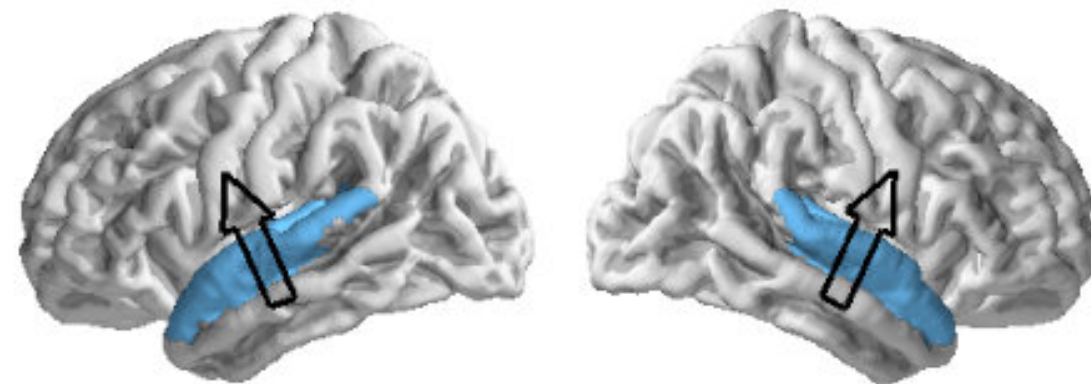
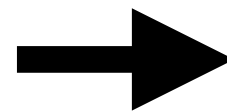


Envelope



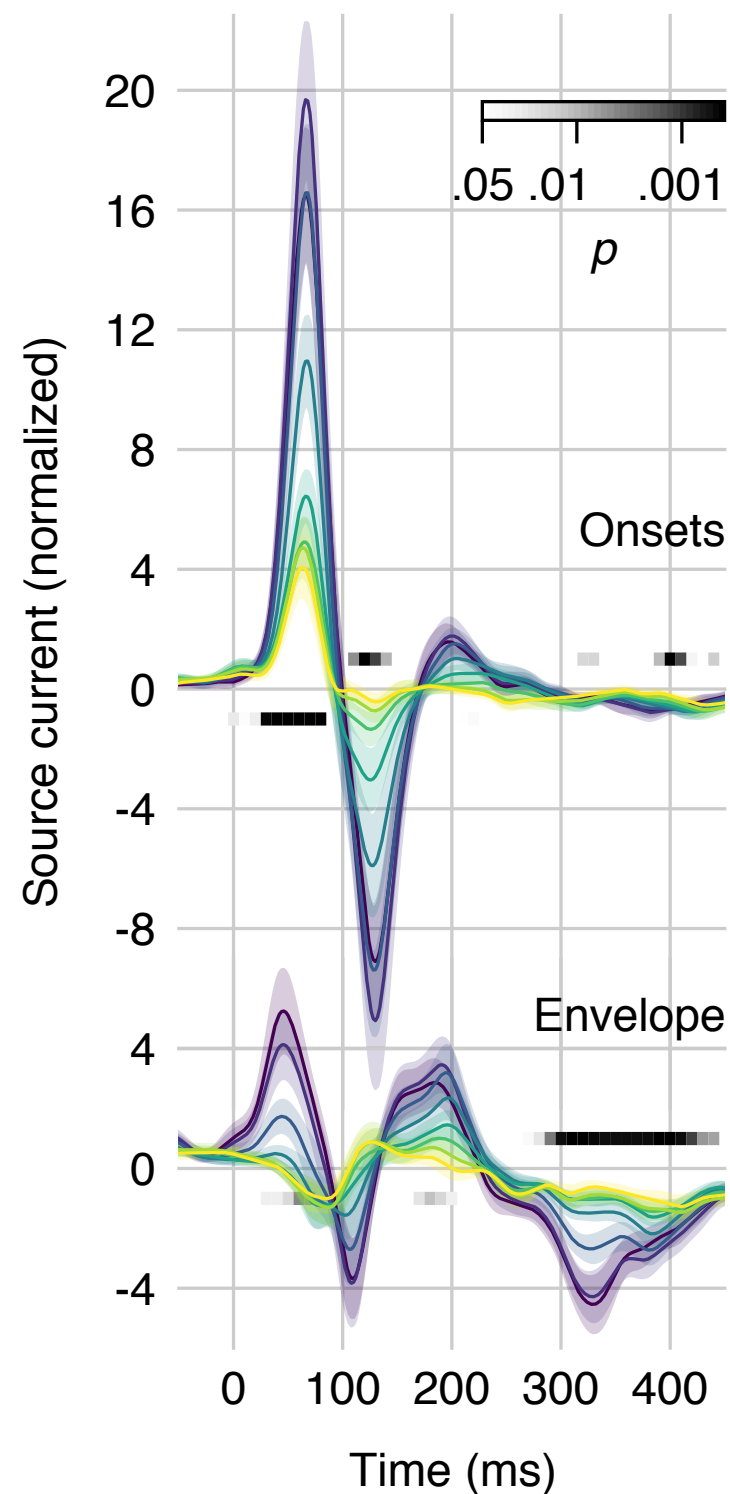
t: 0 5 10 15

ROI for TRF analysis



Source Localization consistent with auditory cortex
(Heschl's gyrus, superior temporal Gyrus)

Onset & Envelope STRFs

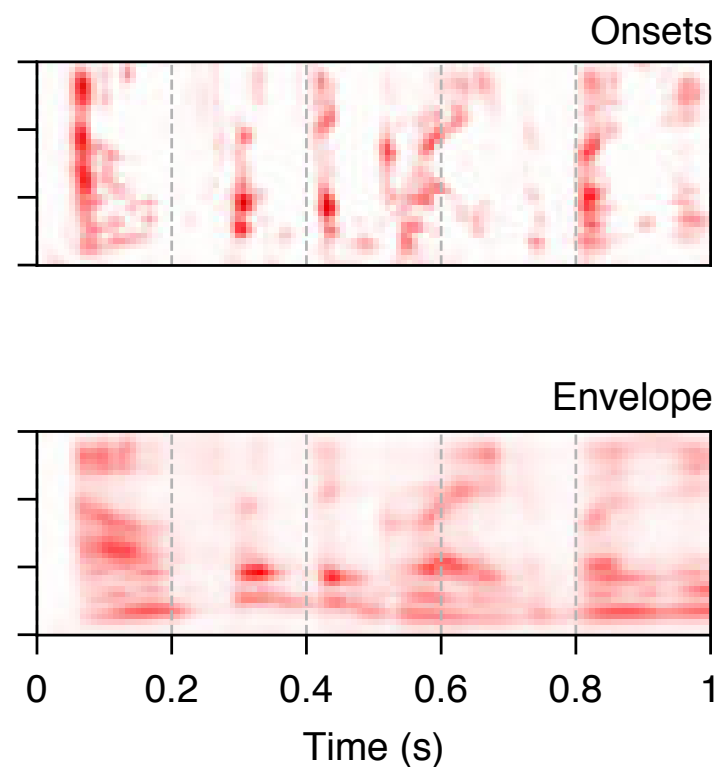


Onset STRF

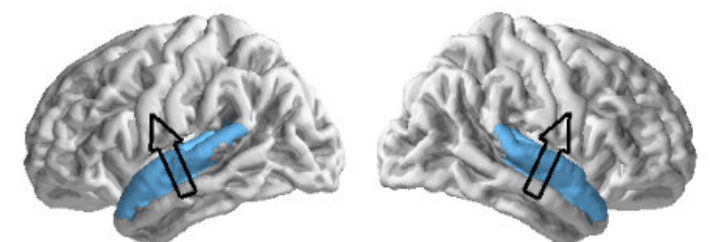
- Typical response pattern + early peak – later peak

Envelope STRF

- Less well-defined

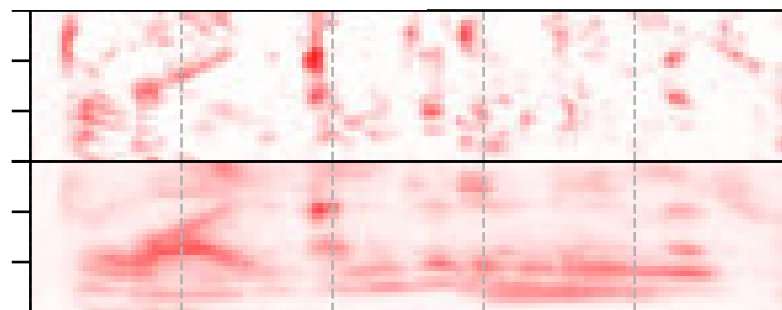


ROI for TRF analysis

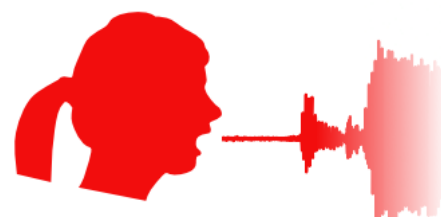
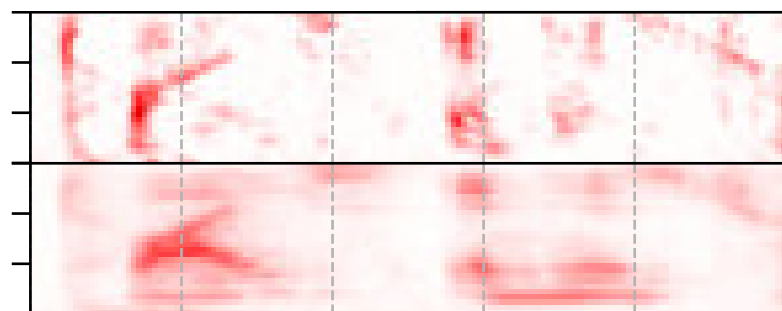


Cocktail Party Listening

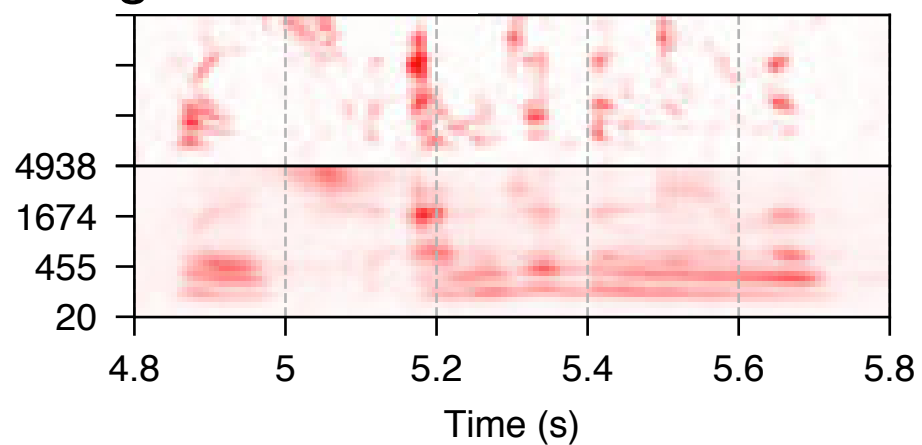
Acoustic mixture



Attended source



Ignored source

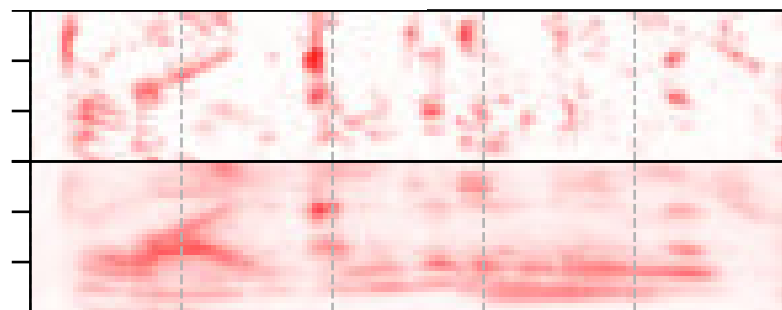


Potential representations

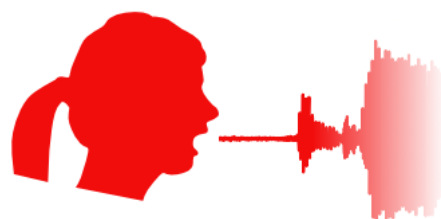
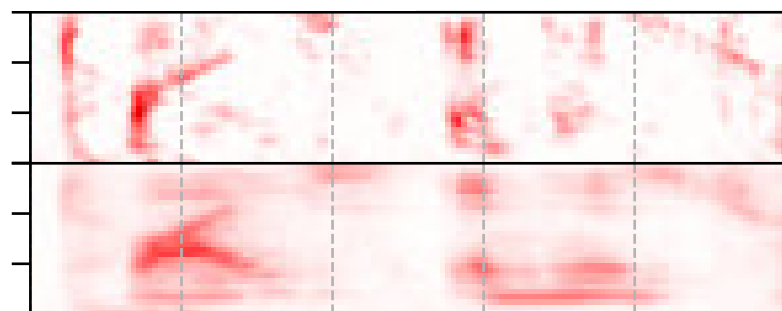
- Acoustic mixture (input)
- Recovered source signals
 - Attended source
 - Ignored source(?)

Cocktail Party Listening

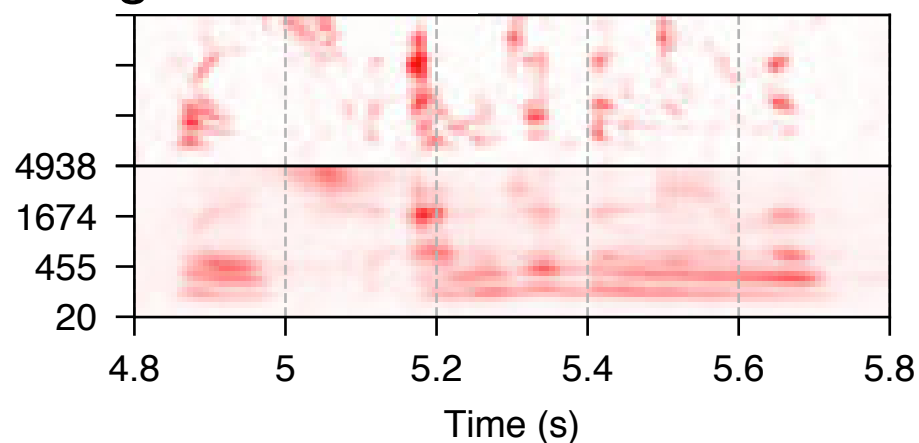
Acoustic mixture



Attended source



Ignored source



Potential representations

- Acoustic mixture (input)
- Recovered source signals
 - Attended source
 - Ignored source(?)

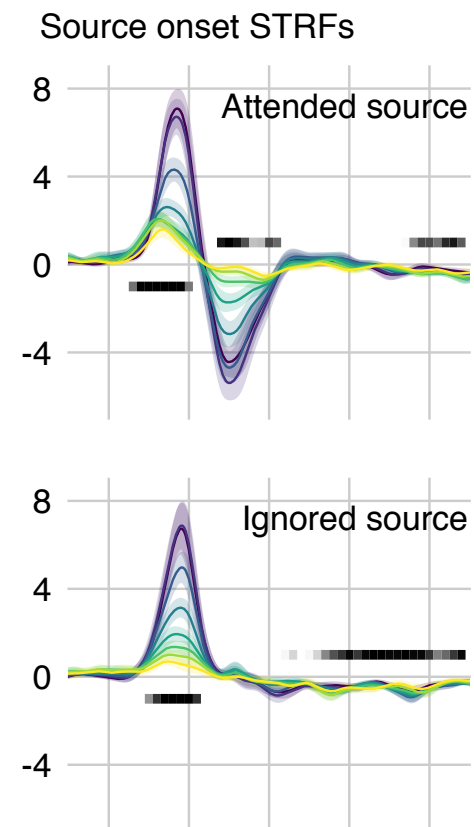
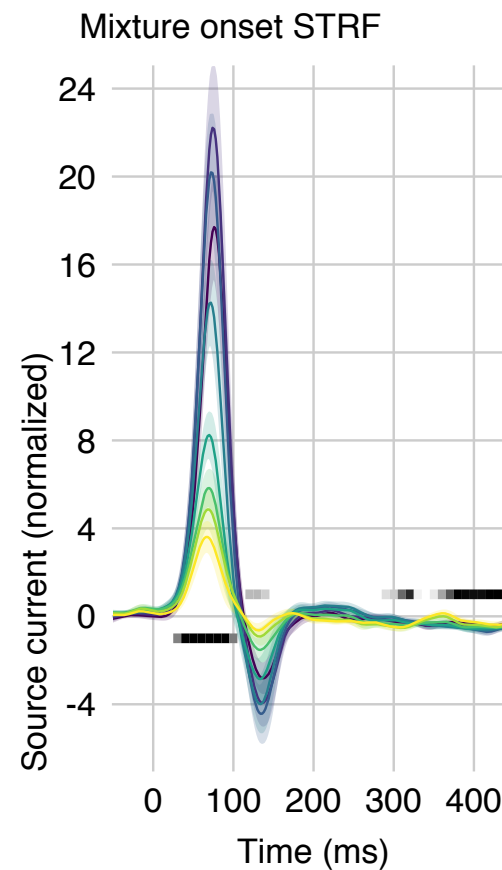
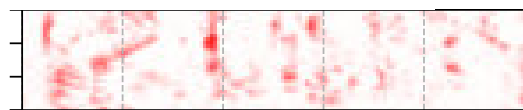
6 STRFs(!)

Envelope STRF + Onset STRF
for each of
Mixture, Attended, Ignored
(8 bands each)

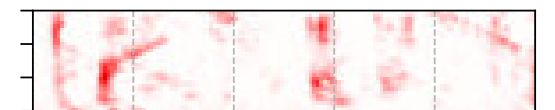
Cocktail Party STRFs

Onset STRFs

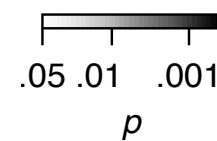
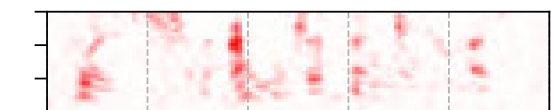
Acoustic mixture



Attended source



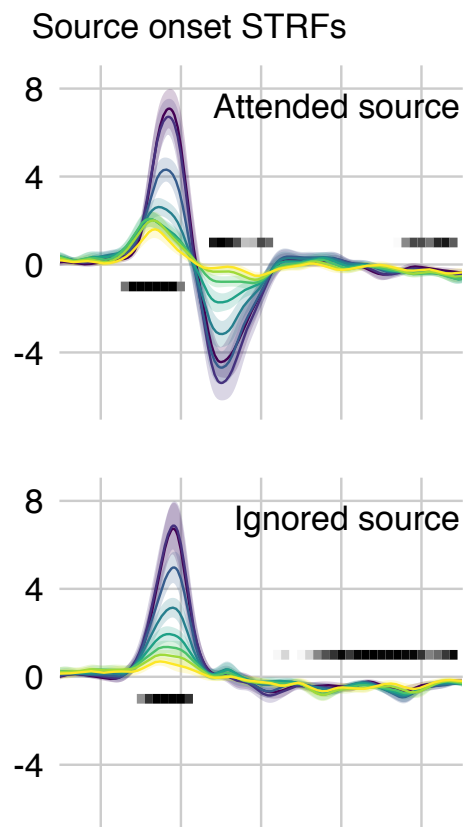
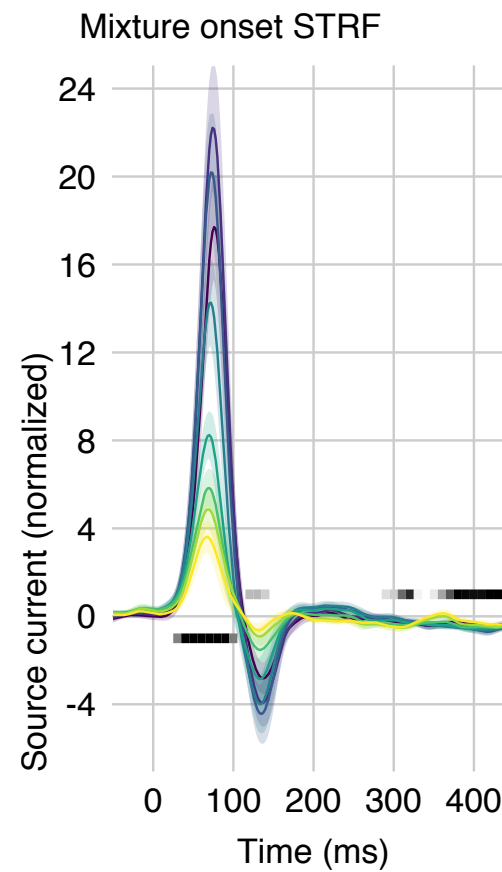
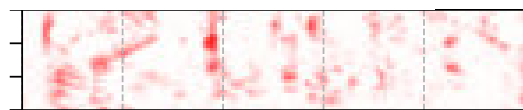
Ignored source



Cocktail Party STRFs

Onset STRFs

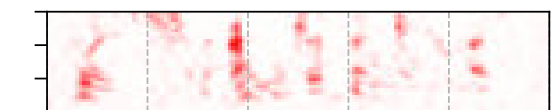
Acoustic mixture



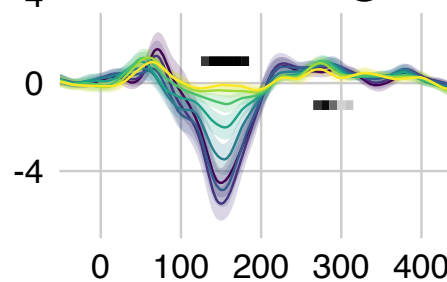
Attended source



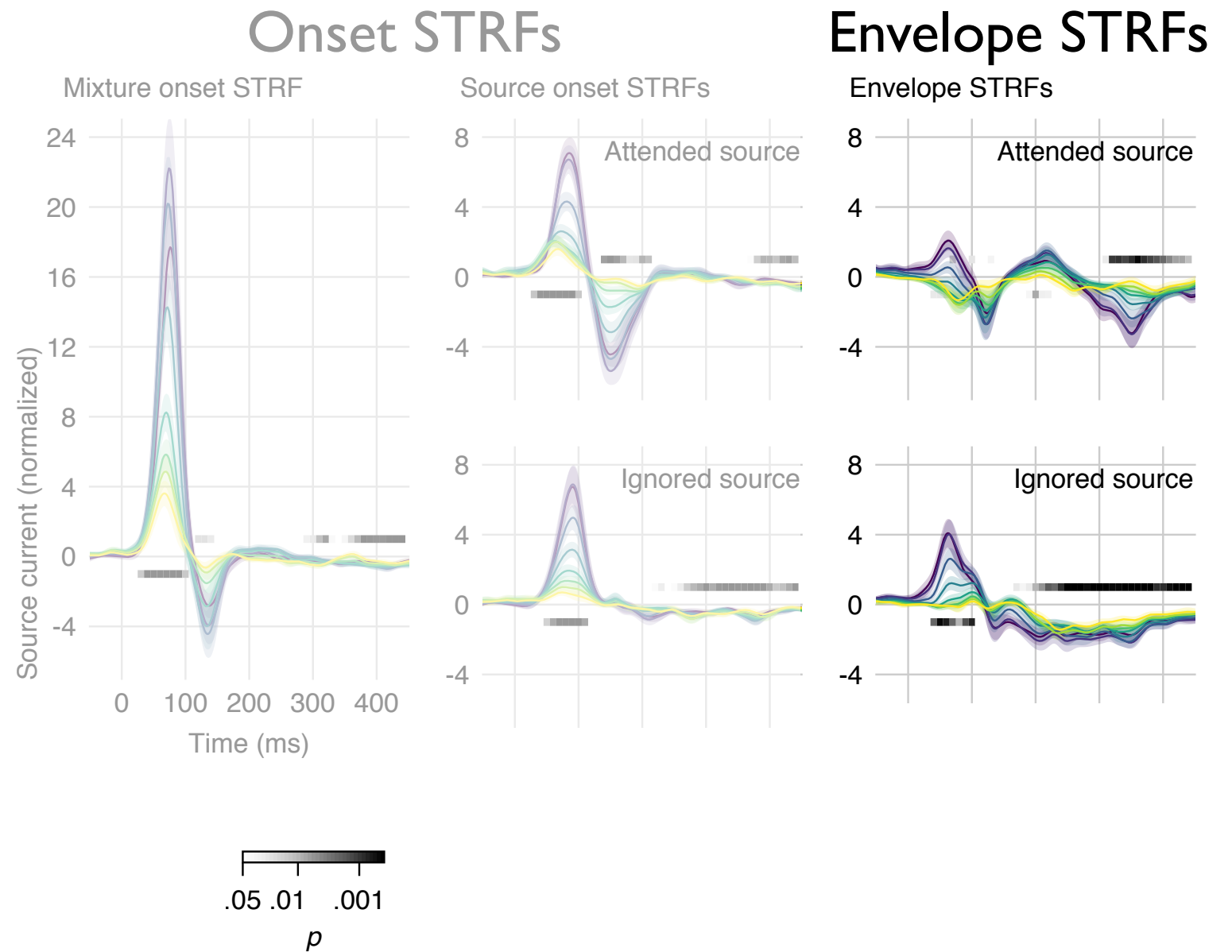
Ignored source



Attended - Ignored



Cocktail Party STRFs

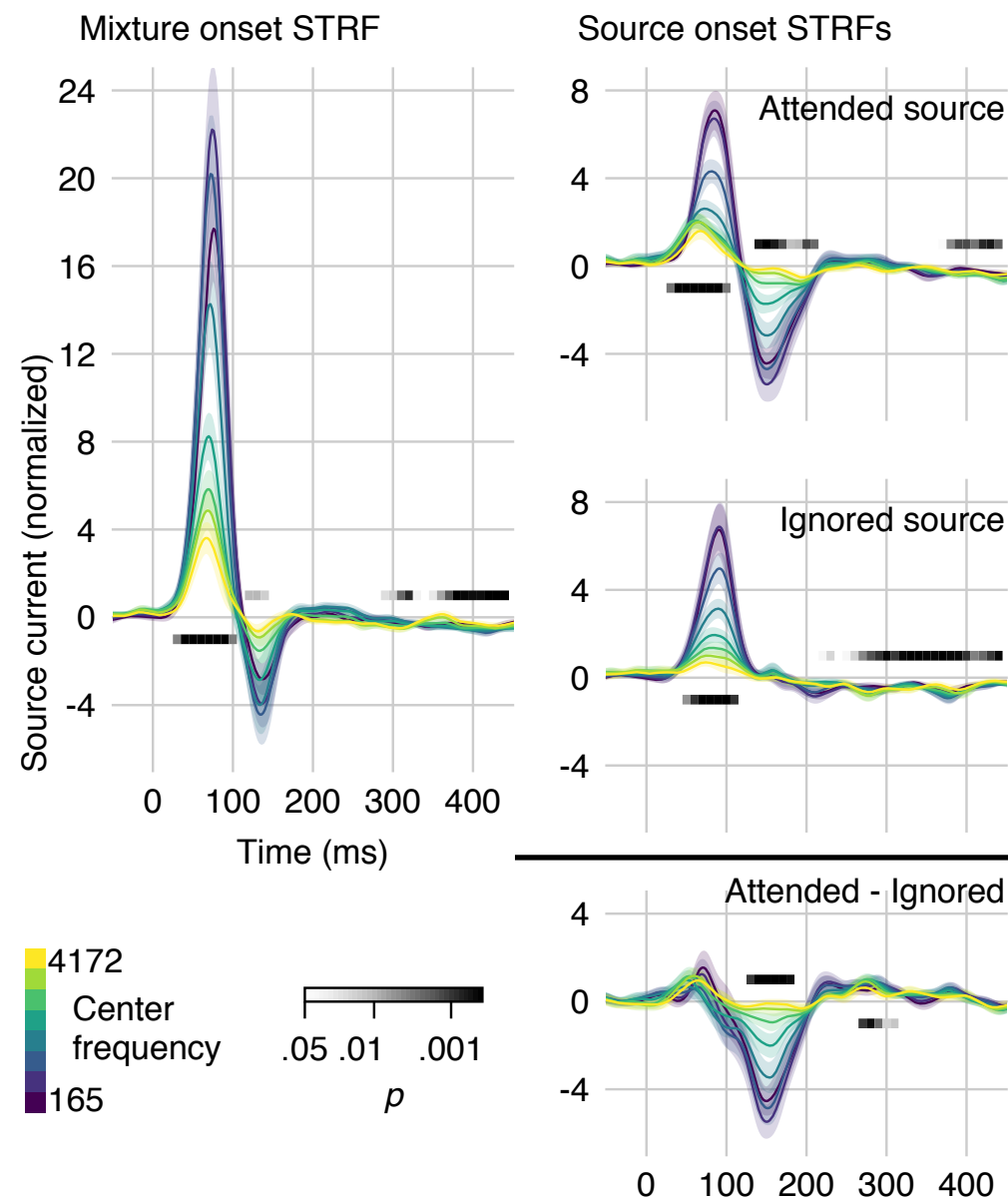


Cocktail Party Onset STRFs

STRF for mixture

- + large peak (72 ms)
- – smaller peak (126 ms)

Onset STRFs



STRF for attended

- + peak (81 ms)
- – peak (150 ms)

STRF for ignored

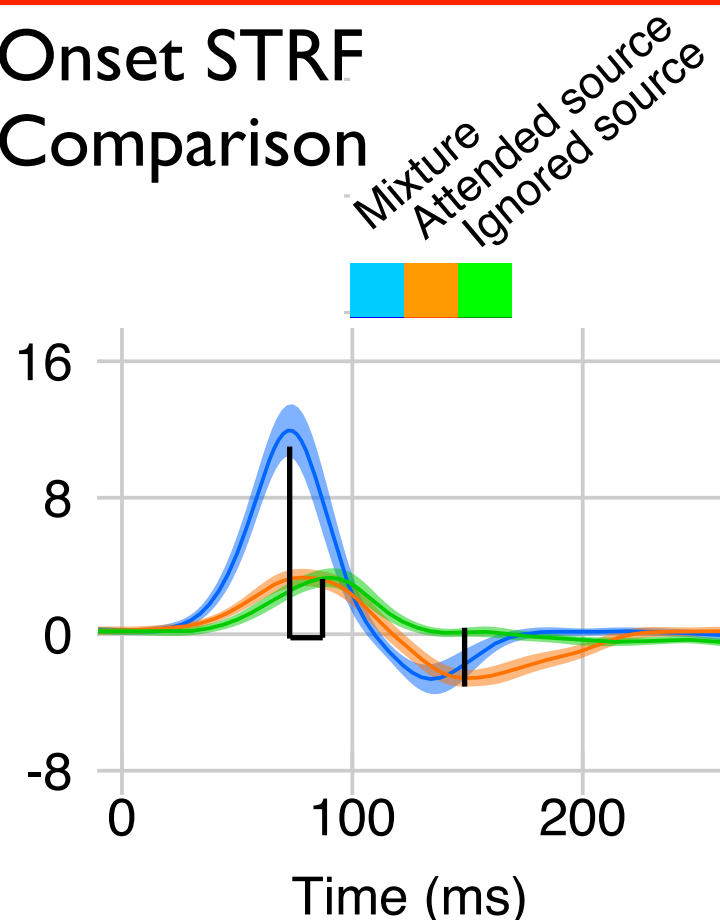
- + peak (88 ms)
- – peak absent

Cocktail Party Onset STRFs

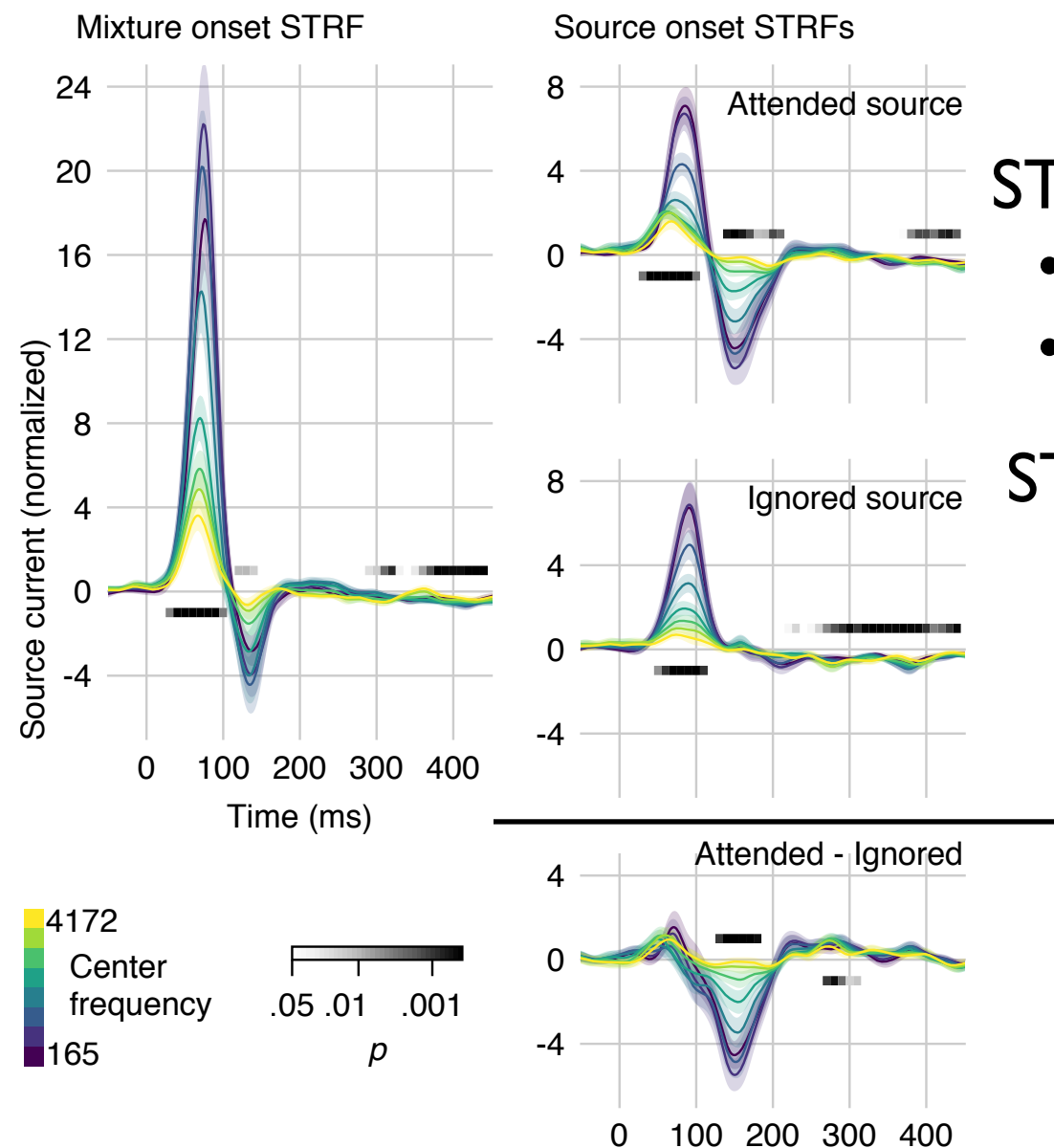
STRF for mixture

- + large peak (72 ms)
- – smaller peak (126 ms)

Onset STRF
Comparison



Onset STRFs



STRF for attended

- + peak (81 ms)
- – peak (150 ms)

STRF for ignored

- + peak (88 ms)
- – peak absent

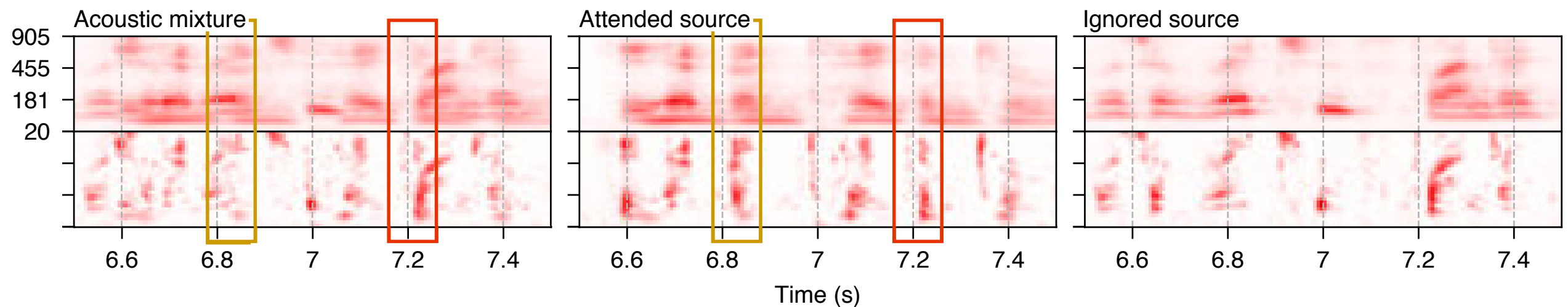
Early response to mixture greater & earlier than sources
Early response to sources not distinguishable
Later response only to attended source, not ignored

Not all onsets are the same

- Source onsets can be masked by other source
 - “Masked onset”
 - Typically occurs when other (masking) source sustained
 - → No onset apparent in mixture despite source onset
- Source onsets may not be masked by other source
 - “Overt onset”
 - Onset apparent in both mixture and source
 - Other source does not interfere
- Overt onsets allow segregation via filtering
- Covert onsets more difficult to unmix

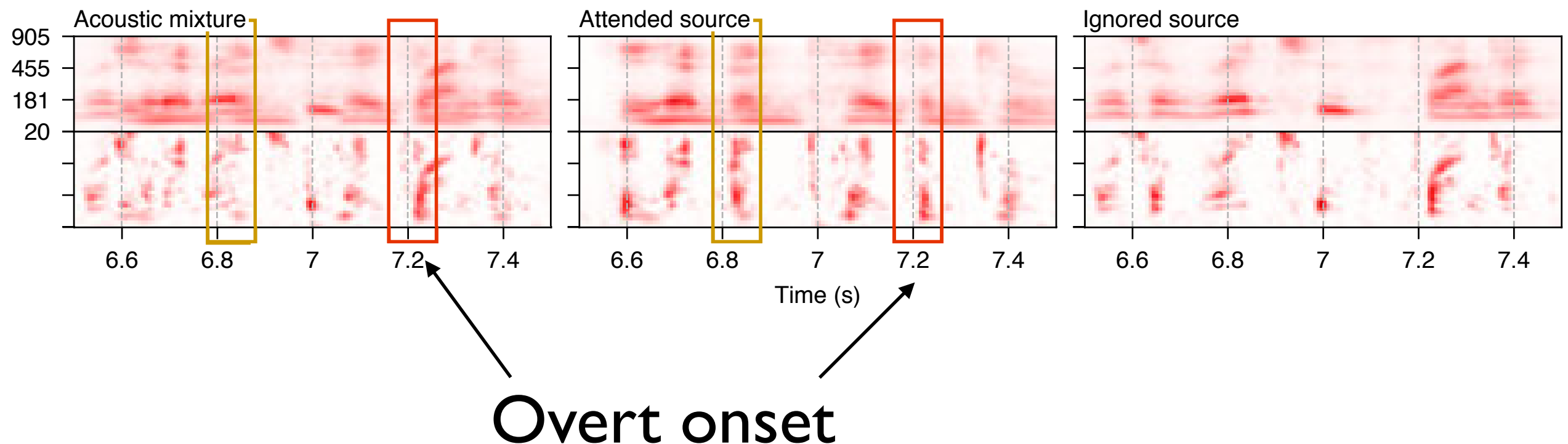
Not all onsets are the same

- Onsets in one source can be masked by other
 - Typically occurs when masking source sustained
 - → No onset apparent in mixture despite source onset



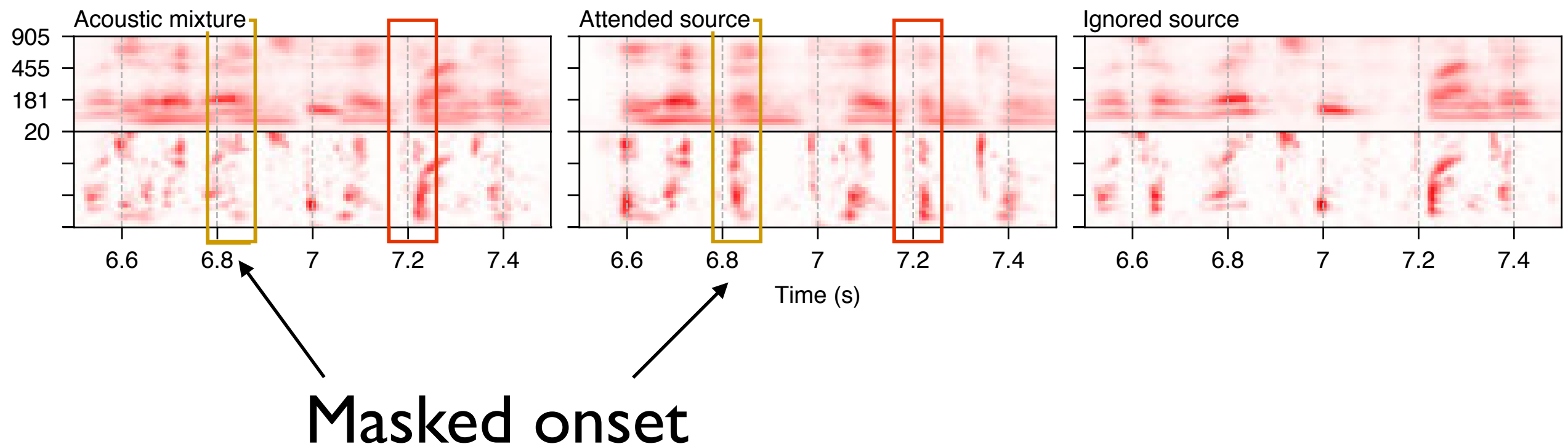
Not all onsets are the same

- Onsets in one source can be masked by other
 - Typically occurs when masking source sustained
 - → No onset apparent in mixture despite source onset



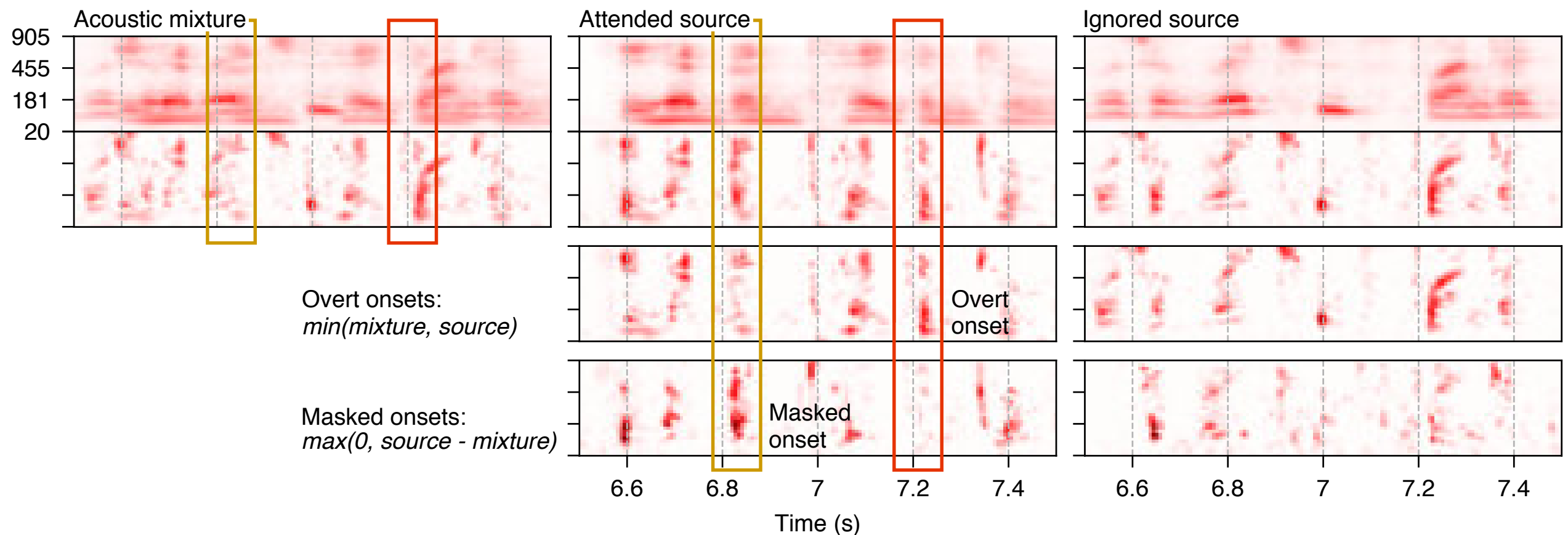
Not all onsets are the same

- Onsets in one source can be masked by other
 - Typically occurs when masking source sustained
 - → No onset apparent in mixture despite source onset



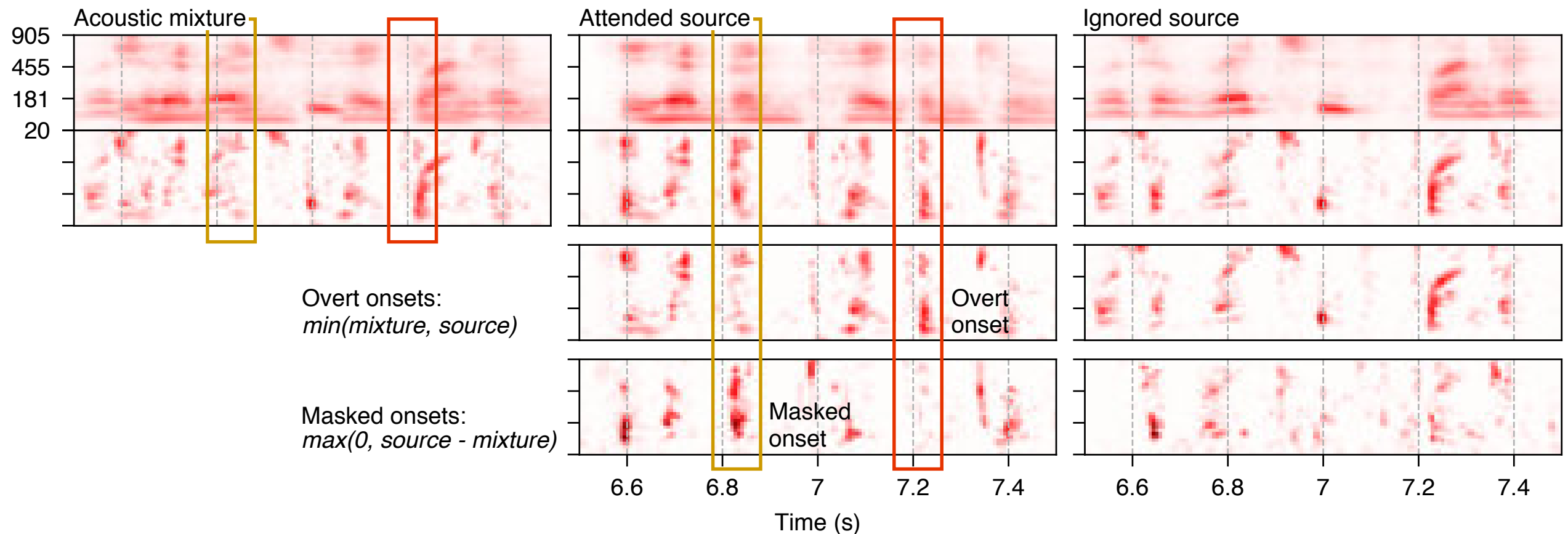
Not all onsets are the same

- Onsets in one source can be masked by other
 - Typically occurs when masking source sustained
 - → No onset apparent in mixture despite source onset



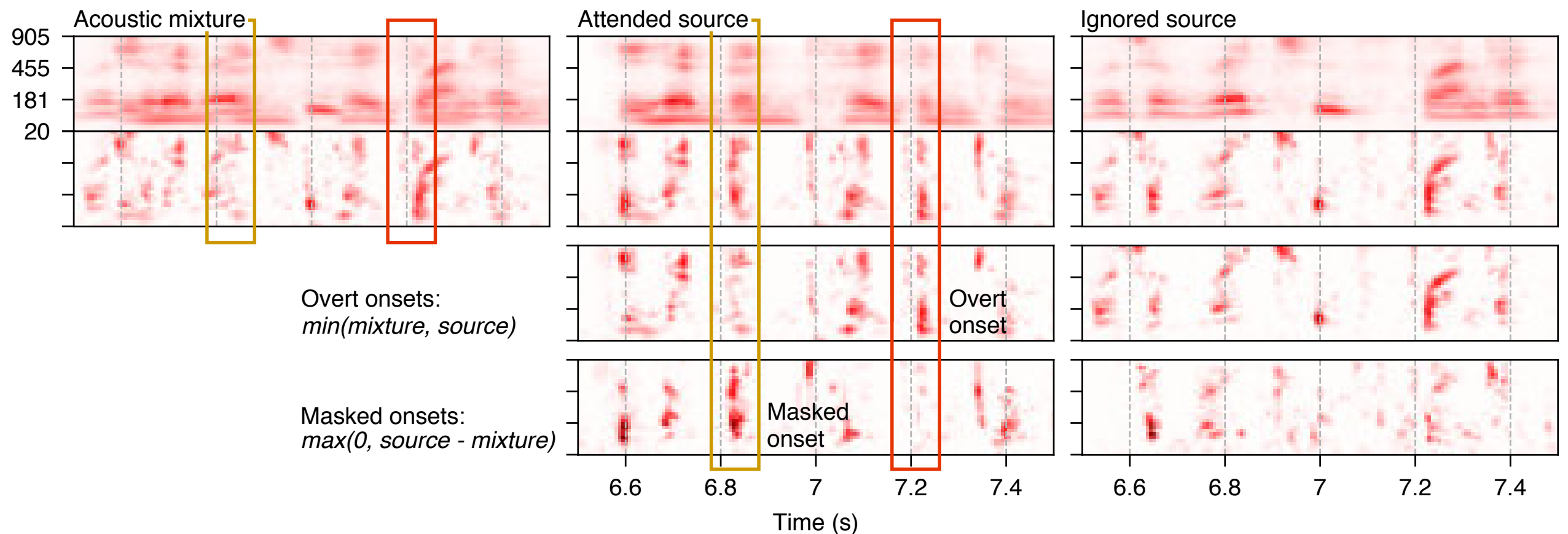
Masked onsets and active vs. passive segregation

- Can we distinguish active segregation mechanisms from passive?
 - Masked onsets cannot be processed passively



Masked onsets and active vs. passive segregation

- Can we distinguish active segregation mechanisms from passive?
 - Masked onsets cannot be processed passively

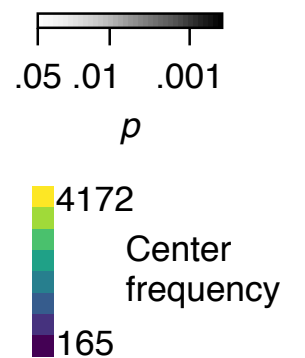
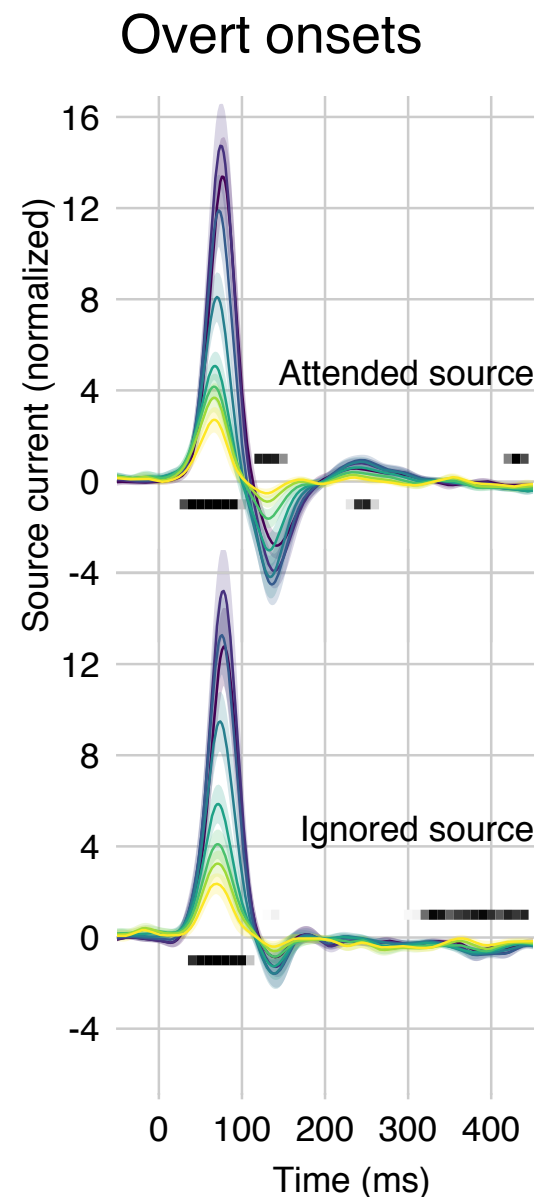


Masked & Overt onset streams generate separate STRFs

Auditory Cortex and Overt vs. Masked Onsets

Overt Onset Responses

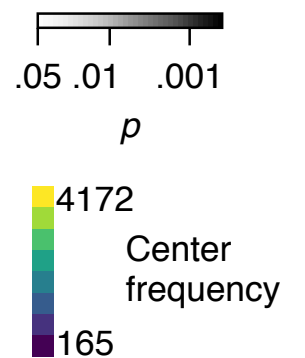
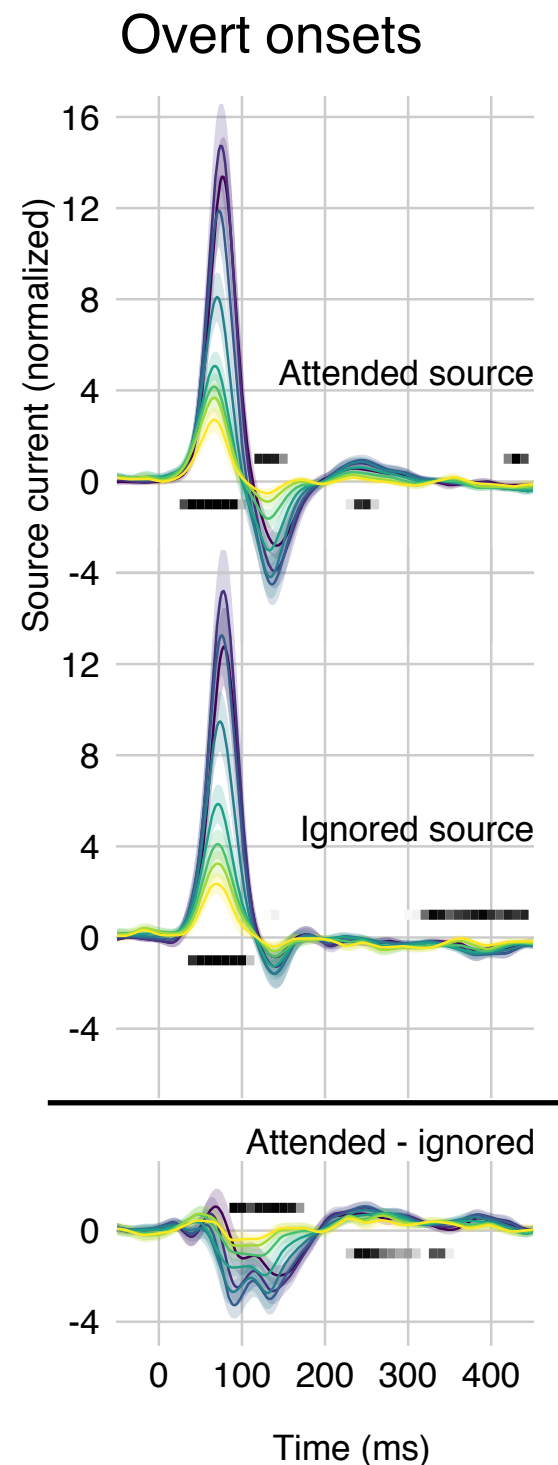
- Similar to onset responses above
- Early response to sources not distinguishable
- Later response only to attended source, not ignored



Auditory Cortex and Overt vs. Masked Onsets

Overt Onset Responses

- Similar to onset responses above
- Early response to sources not distinguishable
- Later response only to attended source, not ignored



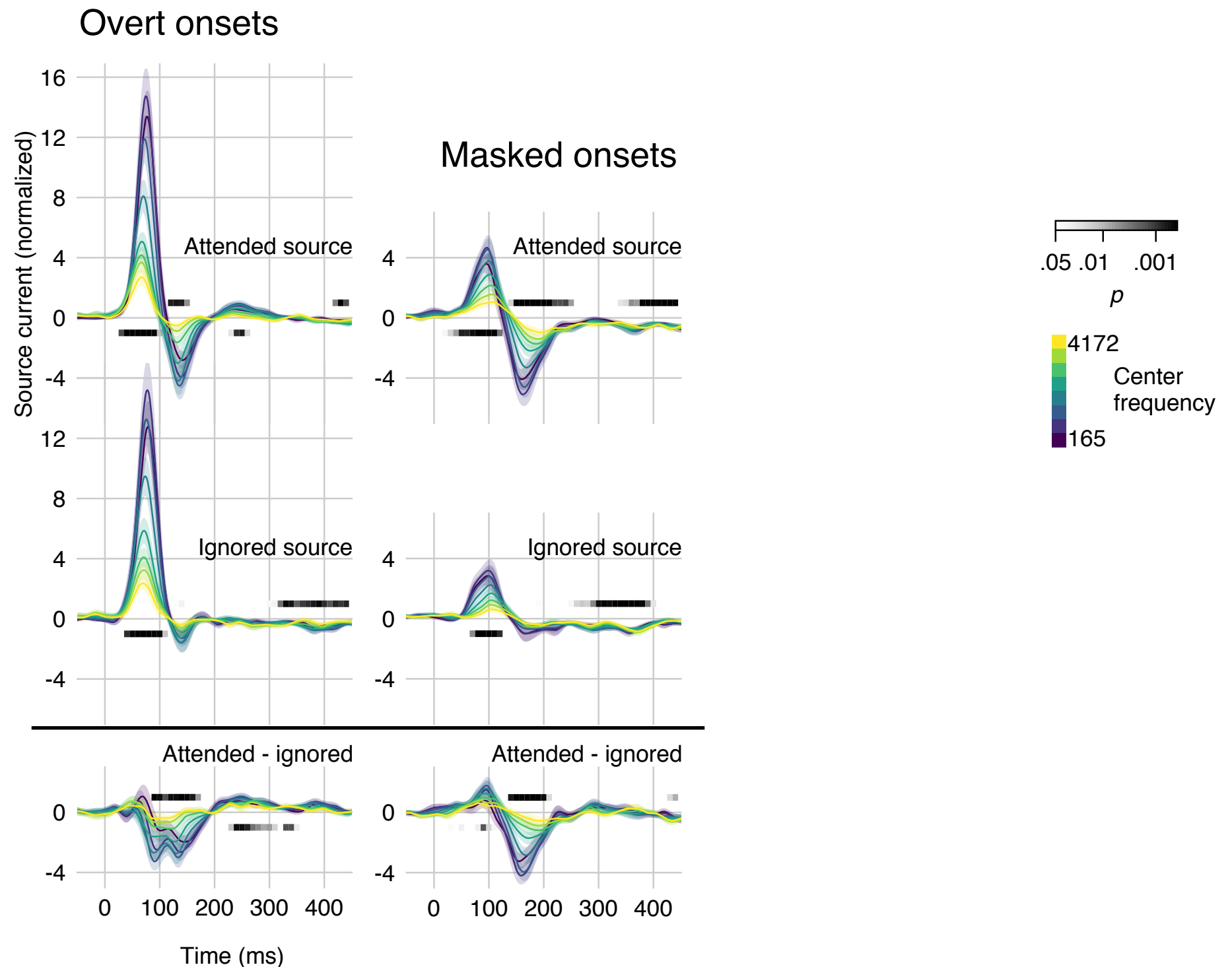
Auditory Cortex and Overt vs. Masked Onsets

Overt Onset Responses

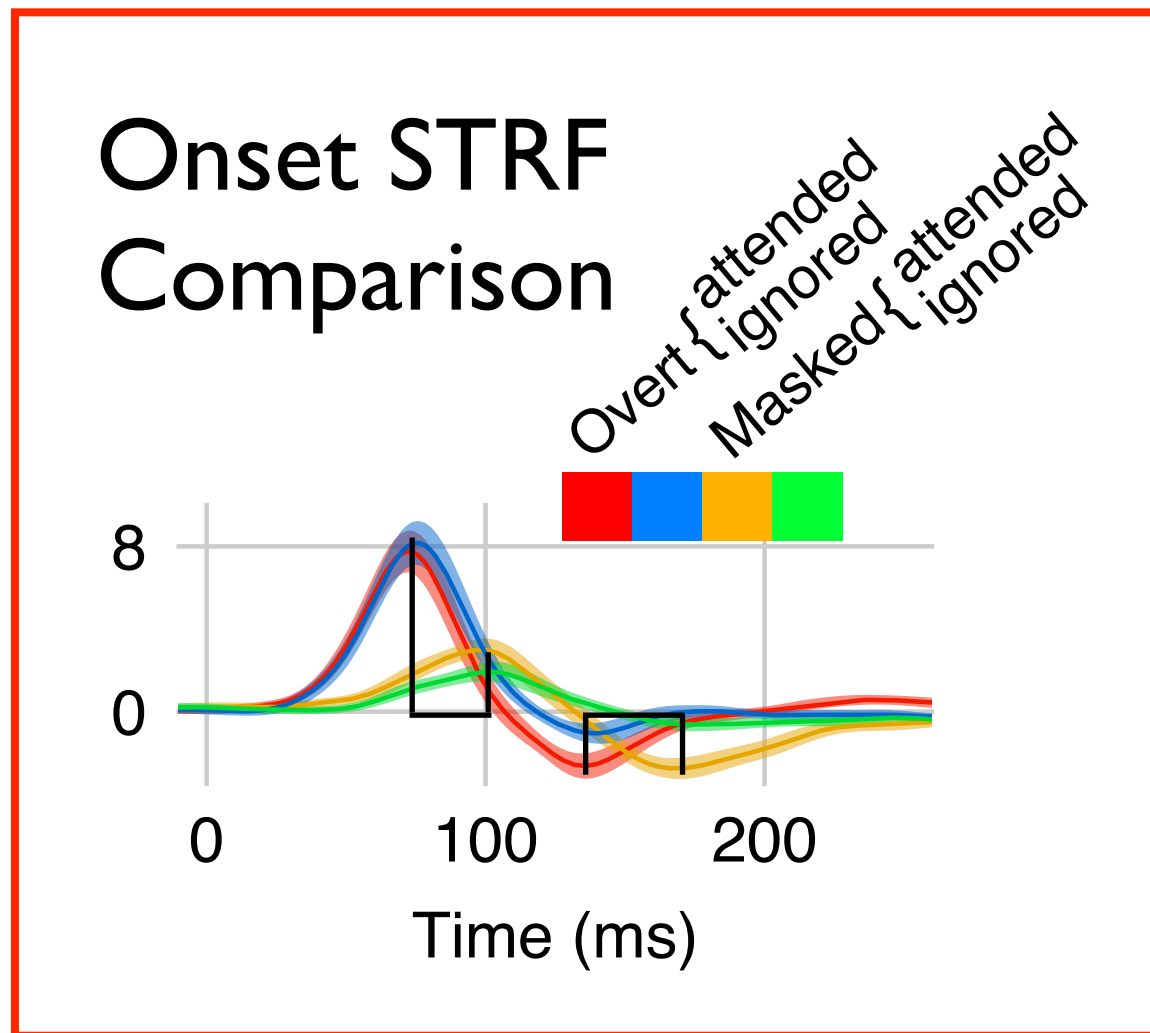
- Similar to onset responses above
- Early response to sources not distinguishable
- Later response only to attended source, not ignored

Masked Onset Responses

- Smaller peaks ($\neq 0$)
- Early peak shows effect of attention
- and ...



Masked onsets engage extended cortical processing



- Masked onset peaks delayed relative to overt onset peaks
 - early masked peaks delayed ~20 ms
 - later (attended) masked peak delayed ~45 ms
- More time spent processing masked peaks
- Evidence for early active processing in segregation

Outline

- Cocktail party listening
 - ▶ Speech segregation & cortical processing of ignored speech
 - ▶ MEG representations of speech
- Methods
- **Results**
- Summary

Outline

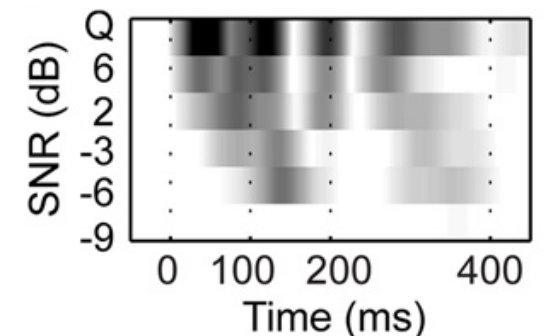
- Cocktail party listening
 - ▶ Speech segregation & cortical processing of ignored speech
 - ▶ MEG representations of speech
- Methods
- Results
- **Summary**

Summary I

- Spectrotemporal acoustic onsets robustly represented in auditory cortex
 - Onsets explain more variance when onsets and envelopes are allowed to compete
- Onsets in both attended and ignored speech represented in auditory cortex, in addition to onsets in acoustic mixture
- Early onset processing does not distinguish between attended and ignored speech
 - **except** masked onsets (attended > ignored): early effects of selective attention

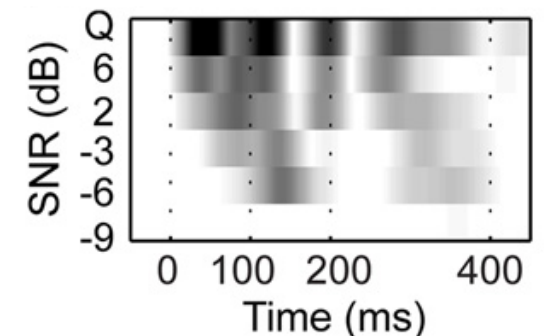
Summary II

- Auditory cortex “un-masks” masked onsets
 - Related to neural filling-in?
- Unmasking requires additional processing time
 - SNR-dependent delays well known, but here shown to be dynamic
- Scene segregation not merely passive spectrotemporal filtering
 - Scene segregation employs active processing



Summary II

- Auditory cortex “un-masks” masked onsets
 - Related to neural filling-in?
- Unmasking requires additional processing time
 - SNR-dependent delays well known, but here shown to be dynamic
- Scene segregation not merely passive spectrotemporal filtering
 - Scene segregation employs active processing



Thank You

Acknowledgements

Current Lab Members & Affiliates

Christian Brodbeck

Alex Presacco

Proloy Das

Lien Decruij

Jason Dunlap

Theo Dutcher

Kevin Hu

Dushyanthi Karunathilake

Joshua Kulasingham

David Nahmias

Peng Zan

Julian Jenkins

Pirazh Khorramshahi

Natalia Lapinskaya

Huan Luo

Sina Miran

Mahshid Najafi

Krishna Puvvada

Jonas Vanthornhout

Ben Walsh

Yadong Wang

Juanjuan Xiang

Jiachen Zhuo

Didier Depireux

Mounya Elhilali

Tom Francart

Jonathan Fritz

Michael Fu

L. Elliot Hong

Stefanie Kuchinsky

Steven Marcus

Cindy Moss

David Poeppel

Shihab Shamma

These slides
available at:
ter.ps/simonpubs



Past Lab Members & Affiliates

Nayef Ahmar

Sahar Akram

Murat Aytakin

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

Past Undergraduate Students

Nicholas Asendorf

Ross Baehr

Anurupa Bhonsale

Elizabeth Camenga

Katya Dombrowski

Kevin Hogan

Alex Jiao

James Williams

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