Rapid Time-Locked Lexical Processing of Attended but Not of Unattended Continuous Speech

Christian Brodbeck
University of Maryland, College Park

brodbeck@umd.edu

Outline

Speech processing of continuous speech

- Different levels observed with magnetoencephalography (MEG) / electroencephalography (EEG)
  - Acoustic processing
  - Phonetic features (e.g. Di Liberto et al., 2015, but also see Daube et al., 2019)
  - **Lexical processing of phonetic information?**
  - Semantic processing (e.g. Broderick et al., 2018)

Outline

Speech processing of continuous speech

- Different levels observed with magnetoencephalography (MEG) / electroencephalography (EEG)
  - Acoustic processing
  - Phonetic features (e.g. Di Liberto et al., 2015, but also see Daube et al., 2019)
  - **Lexical processing of phonetic information?**
  - Semantic processing (e.g. Broderick et al., 2018)

Lexical processing

- Information from phoneme level information is integrated in a time-locked fashion for word perception (cohort theory)

Outline

Speech processing of continuous speech

- Different levels observed with magnetoencephalography (MEG) / electroencephalography (EEG)
  - Acoustic processing
  - Phonetic features (e.g. Di Liberto et al., 2015, but also see Daube et al., 2019)
  - **Lexical processing of phonetic information?**
  - Semantic processing (e.g. Broderick et al., 2018)

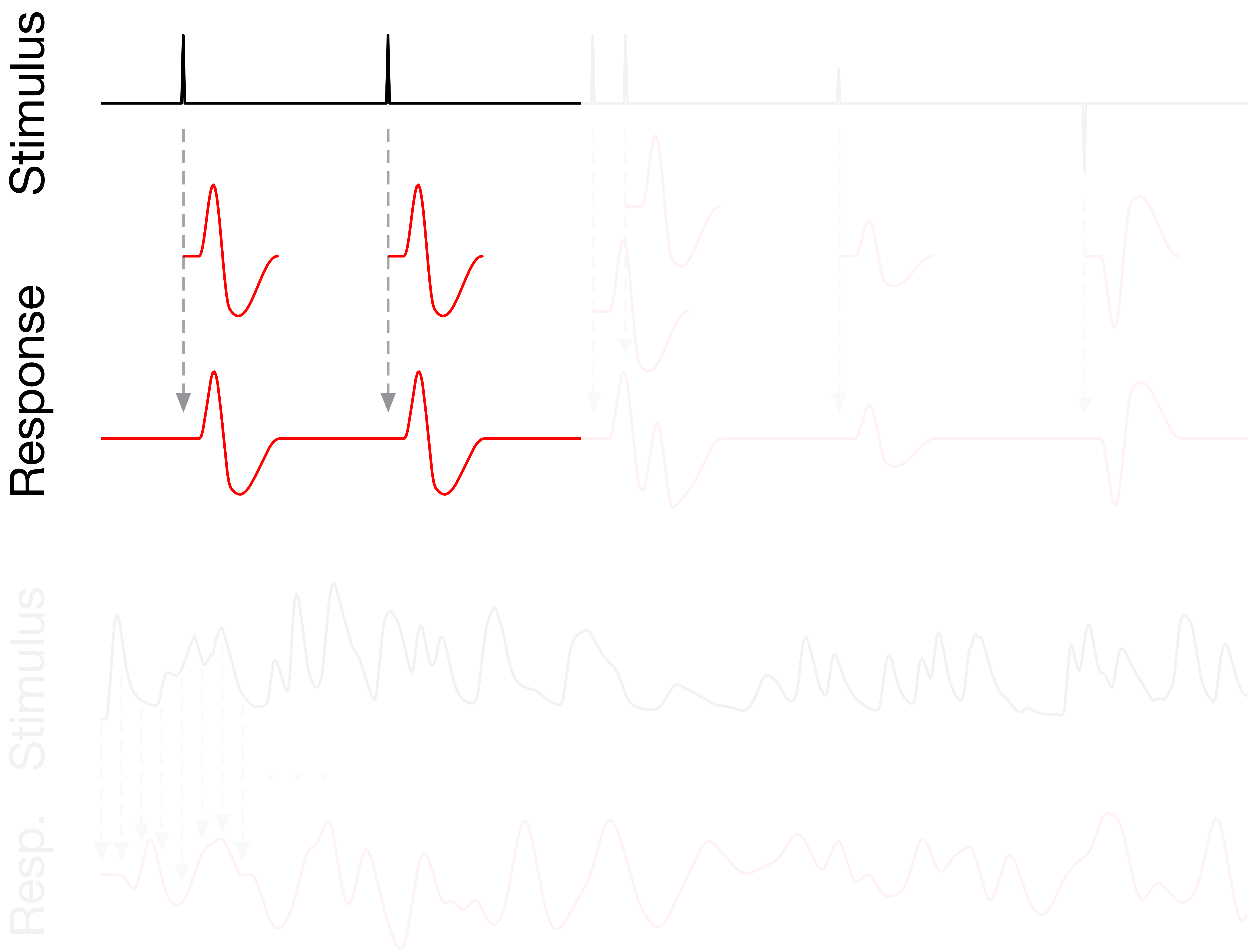
Lexical processing

- Information from phoneme level information is integrated in a time-locked fashion for word perception (cohort theory)

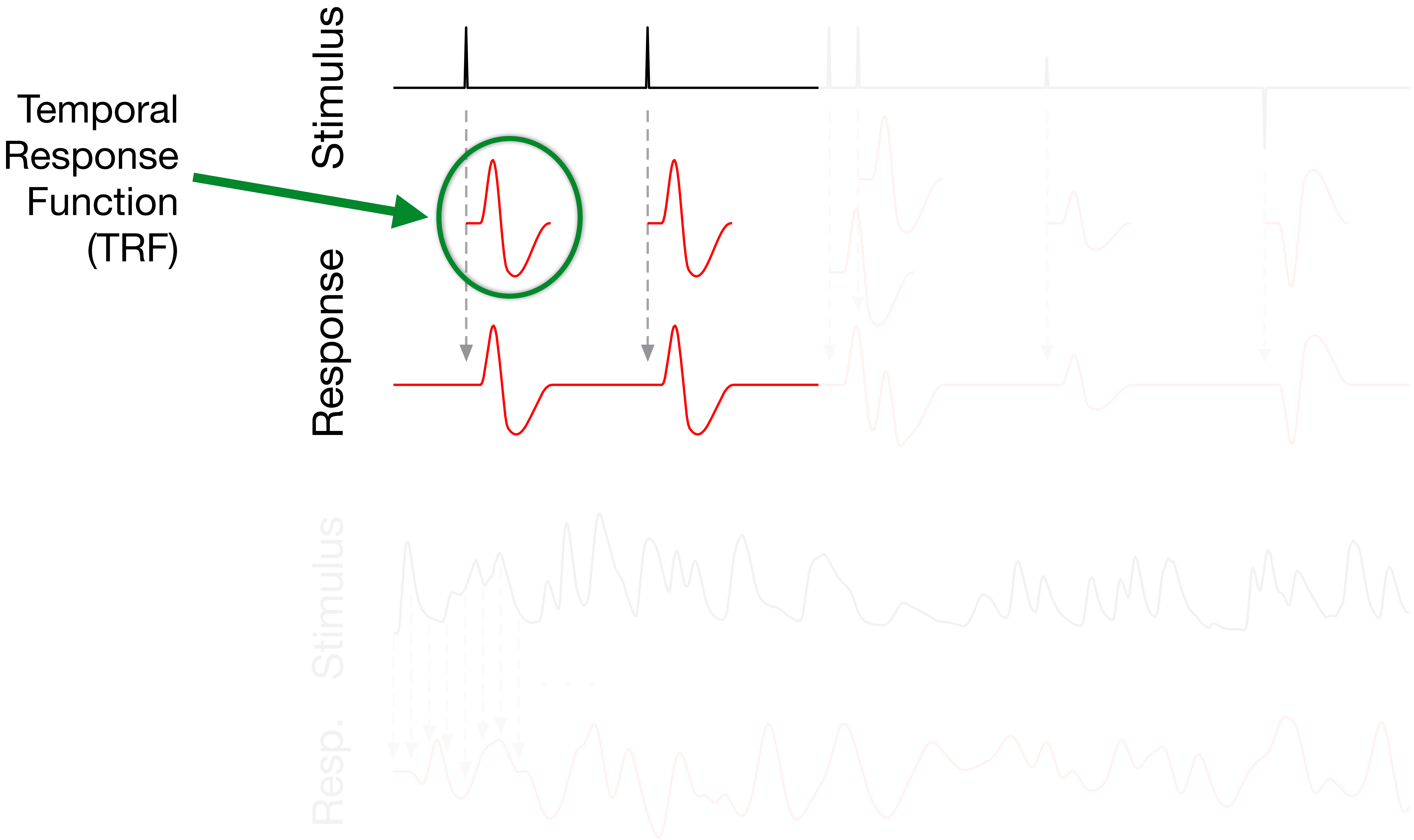
This presentation

- Measure lexical processing of phonetic information with MEG
- Lexical processing in cocktail-party stimuli

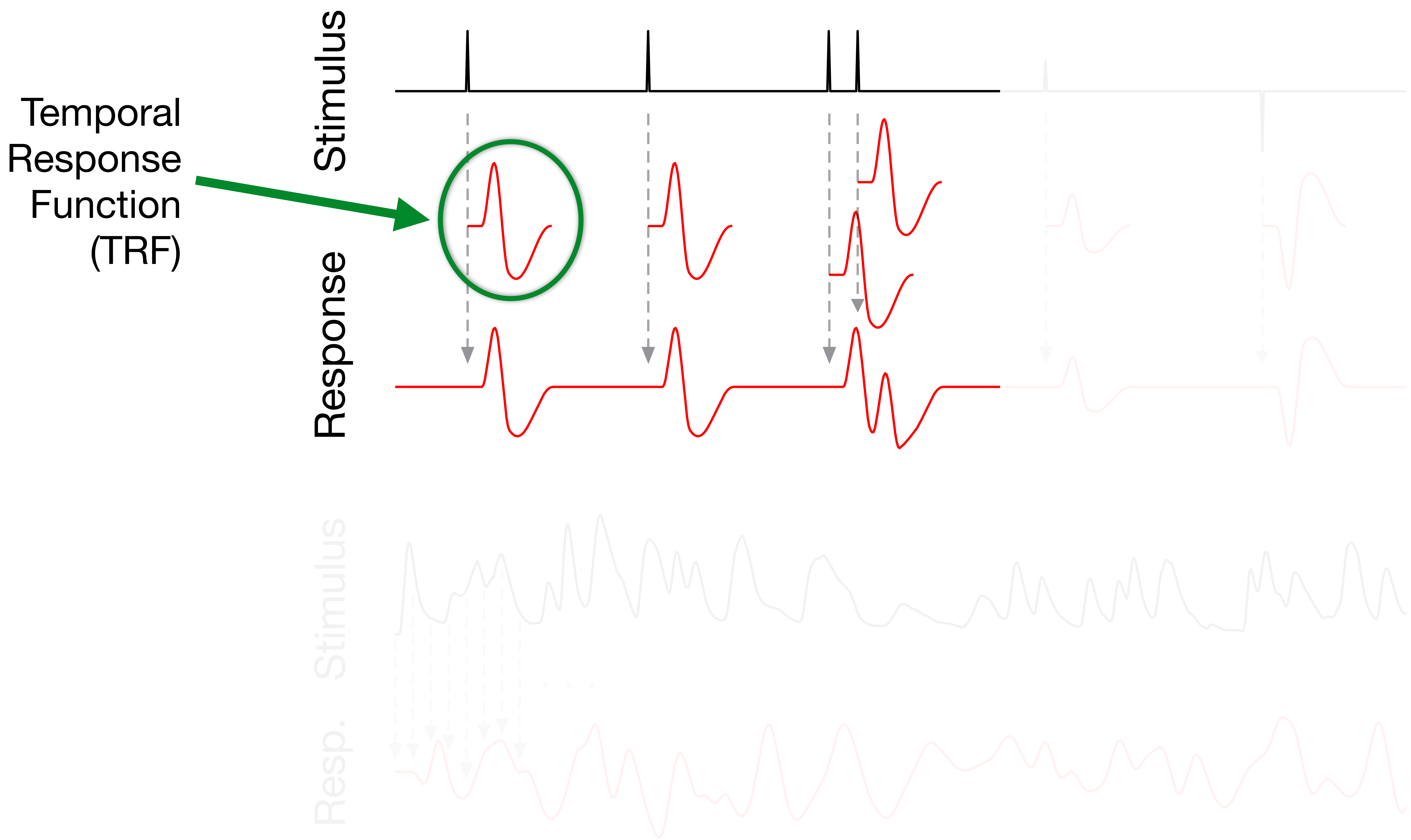
Linear convolution model



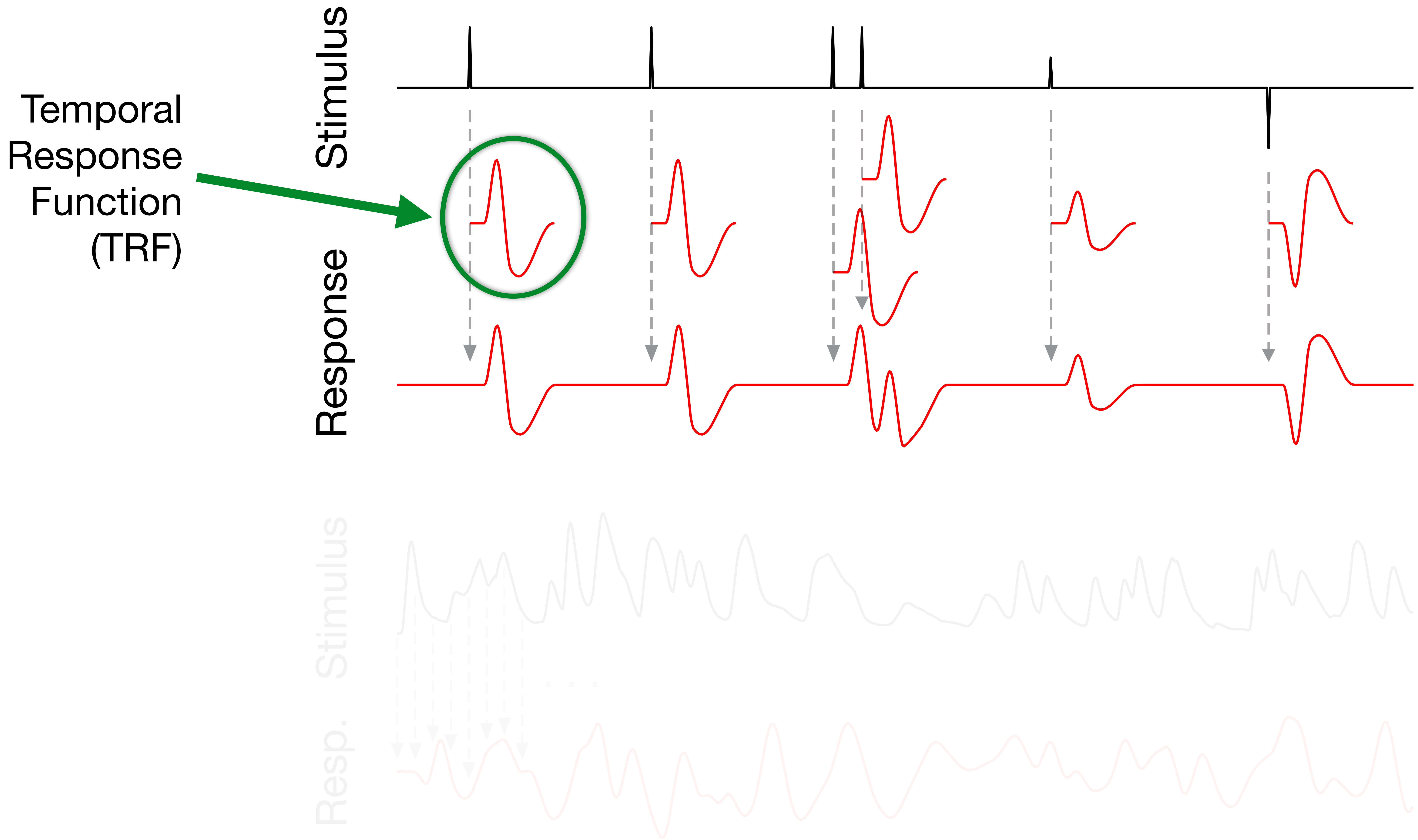
Linear convolution model



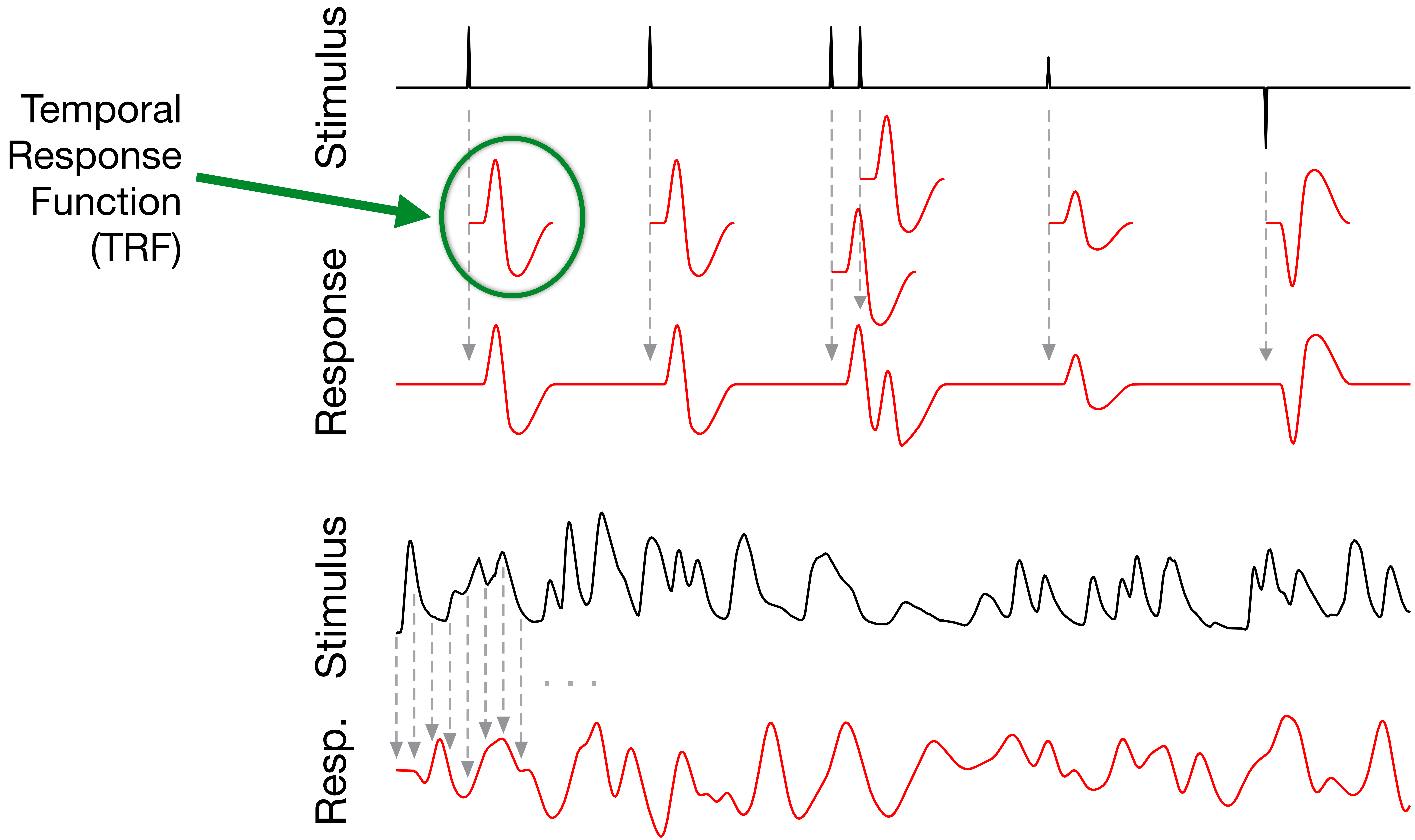
Linear convolution model



Linear convolution model



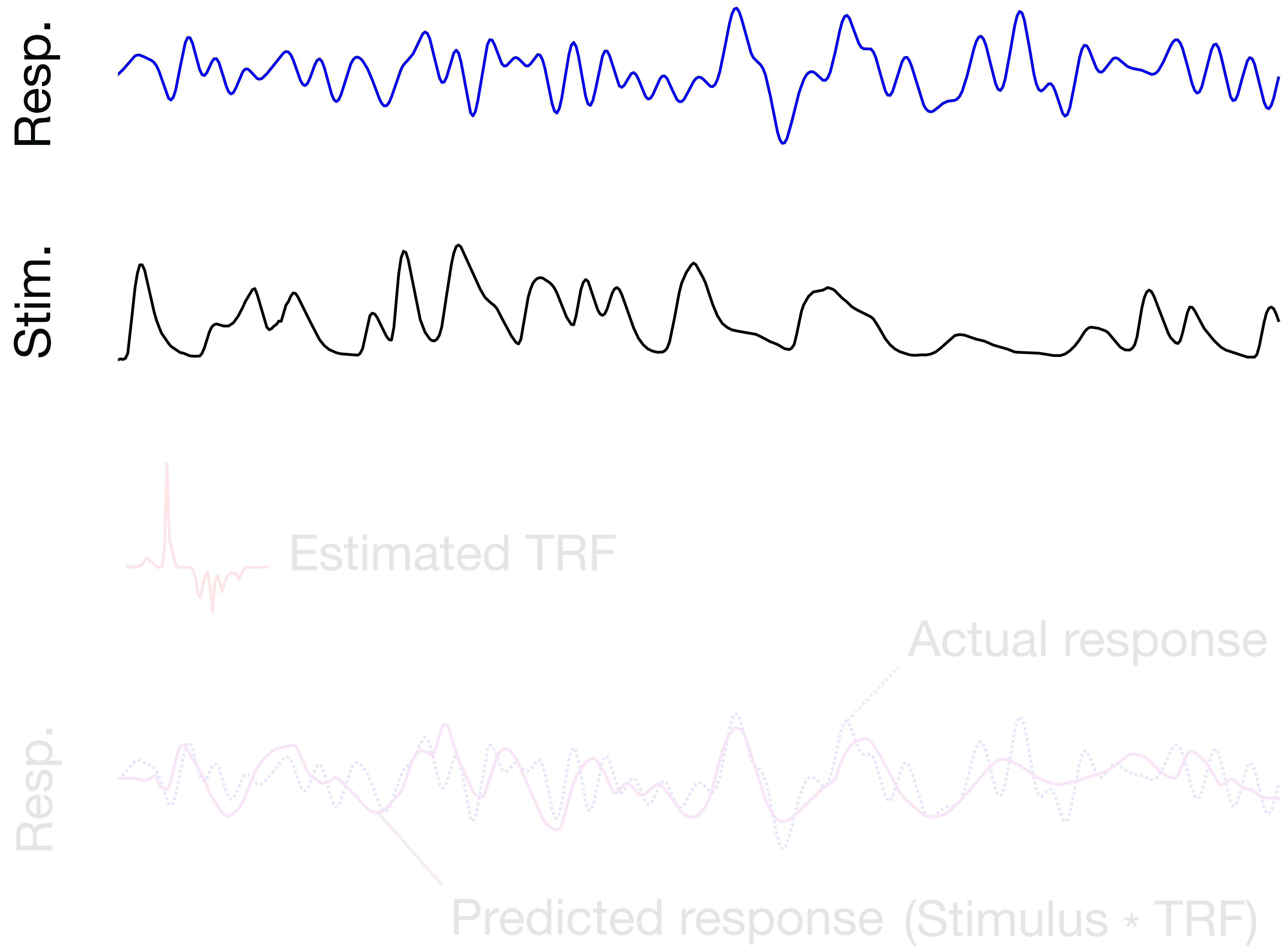
Linear convolution model



Linear convolution model

Temporal Response Function (TRF) estimation:

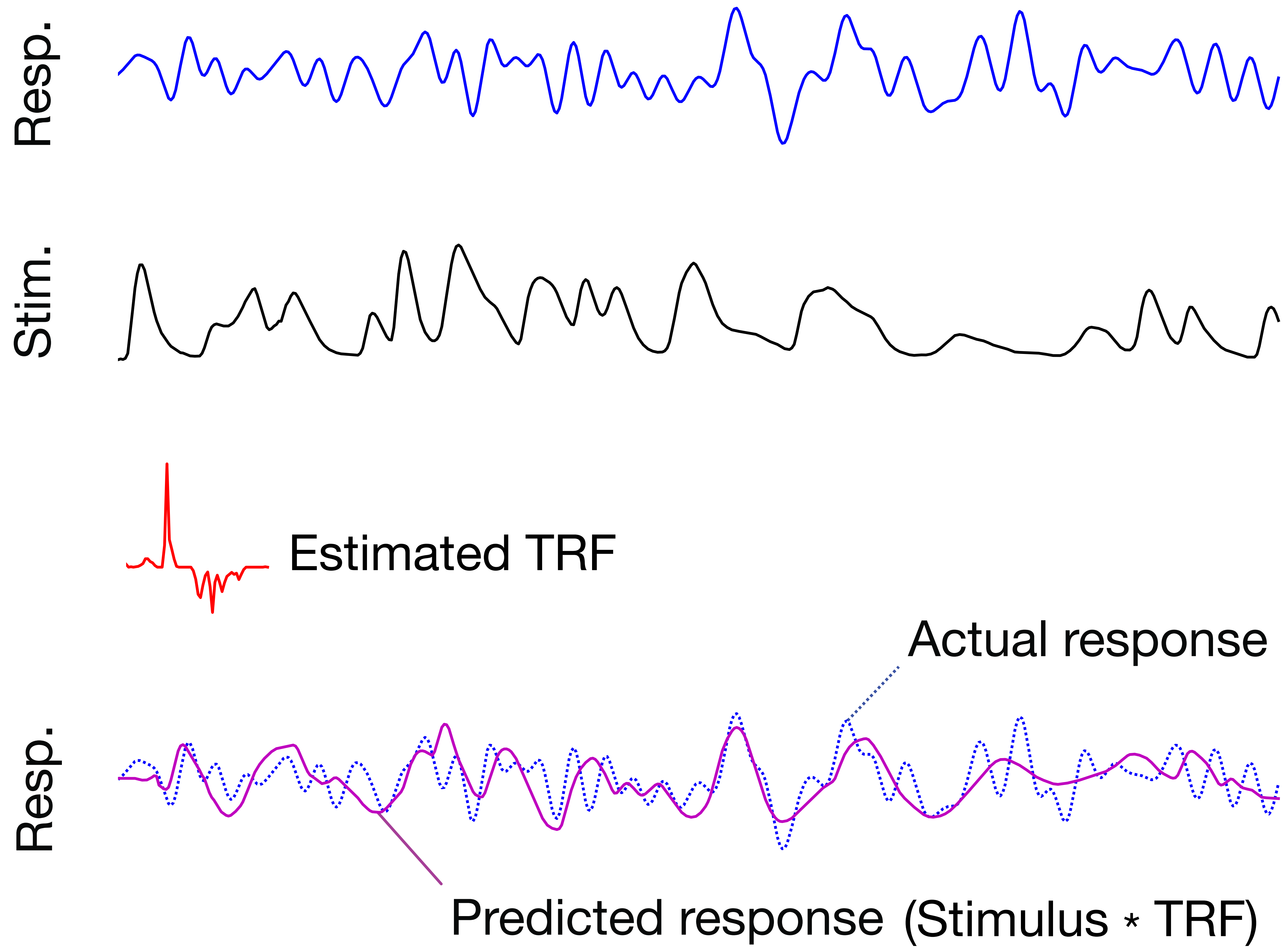
Stimulus and response are known; find the best TRF to produce the response from the stimulus:



Linear convolution model

Temporal Response Function (TRF) estimation:

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



Statistics

Evaluate model:

- Pearson correlation:
  r(predicted response, measured response)

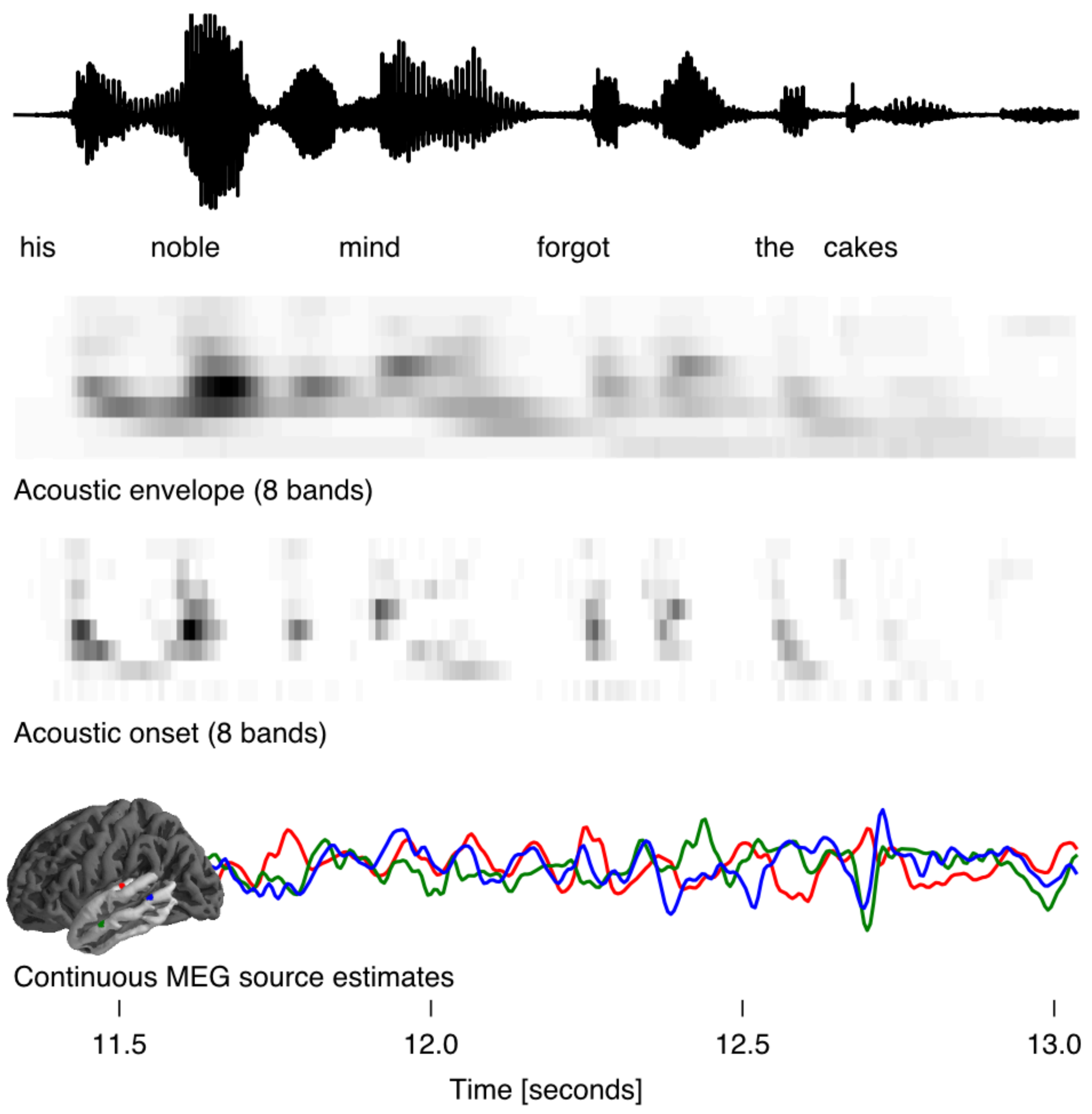
Evaluate a predictor, bias-corrected (e.g., word frequency):

- *R* of the full model
  Envelope + Word frequency + Semantic composition
- *R* of a model with word frequency permuted
  Envelope + Permute(word frequency) + Semantic composition
- Test for significant improvement across subjects

Significance test:

- Mass-univariate *t*-test
- Threshold-free cluster enhancement
- Max statistic distribution with 10,000 permutations

Model: Acoustics



Stimulus

- Audiobook excerpt (8 minutes)

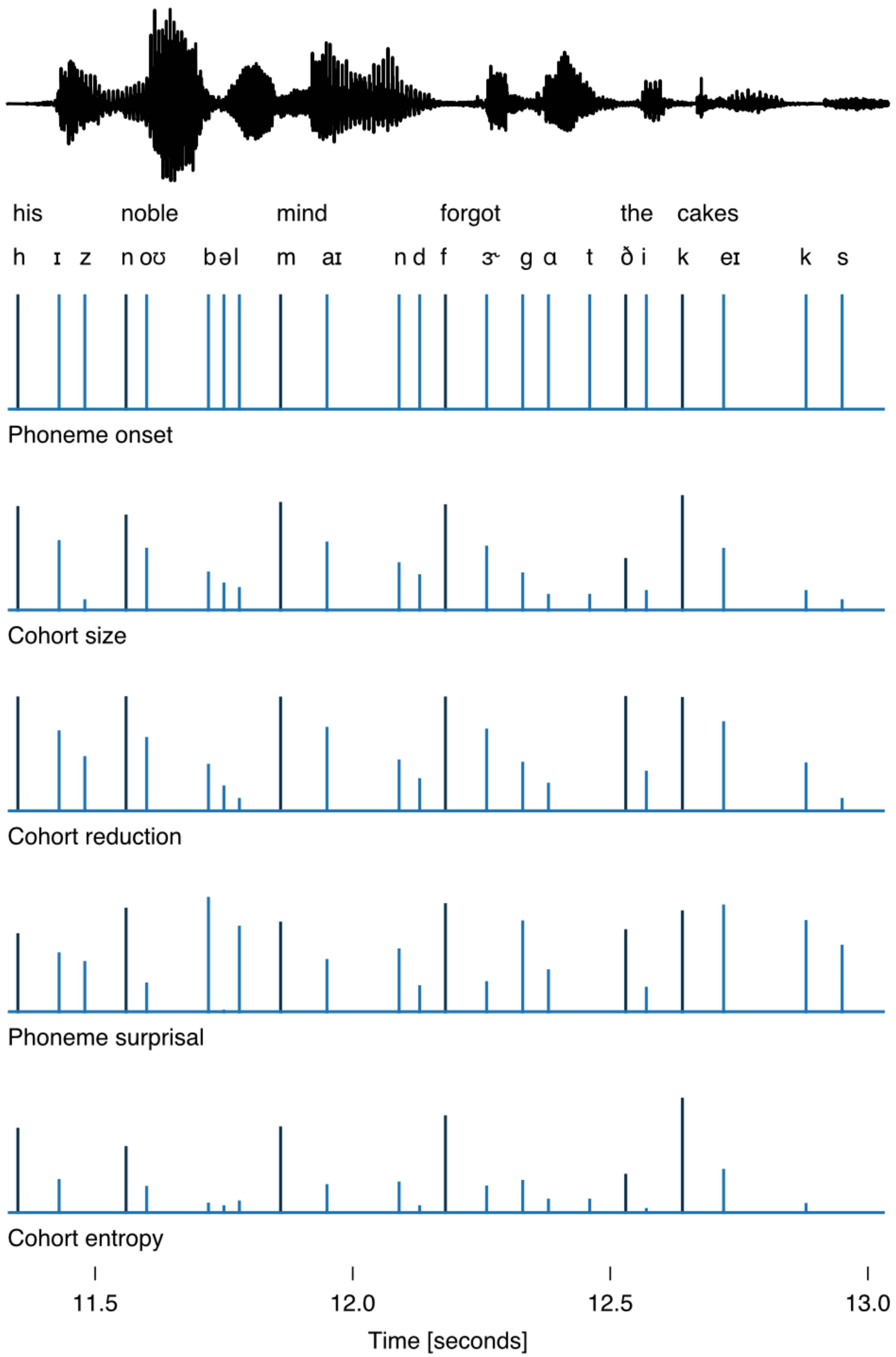
Predictors:

- Acoustic spectrogram (8 bands)
- Acoustic onsets (8 bands)
- Phonemes (next slide)

Predicted:

- Continuous, *source-localized* MEG responses

Model: Phonemes



Phonemes

- Modeled as impulses at phoneme onset

Phoneme processing

- Impulses *scaled by relevant variable*

Word onset

- Separately from word-internal phonemes

Cohort model

/k.../

Graphs	Pronunciation	SUBTLEX Count
ca	K AH	109
	K AA	
cab	K AE B	1826
caba	K AA B AH	2
cabal	K AH B AA L	13
caballero	K AE B AH Y EH R OW	21
cabana	K AH B AE N AH	46
cabanas	K AH B AE N AH Z	2
cabaret	K AE B ER EY	115
cabarets	K AE B ER EY Z	13
cabbage	K AE B AH JH	148
	K AE B IH JH	
cabbages	K AE B IH JH IH Z	37
cabbie	K AE B IY	71
...		
4447		1811951

Cohort model

/keɪ…/

Graphs	Pronunciation	SUBTLEX Count
cable	K EY B AH L	1108
cabled	K EY B AH L D	19
cablegram	K EY B AH L G R AE M	10
cables	K EY B AH L Z	110
cade	K EY D	11
cadence	K EY D AH N S	15
cadences	K EY D AH N S IH Z	1
cady	K EY D IY	64
caesarean	K EY S ER IY N	10
caesareans	K EY S ER IY N Z	1
cage	K EY JH	1034
	K EY JH IH	
caged	K EY JH D	83
…		
90		52908

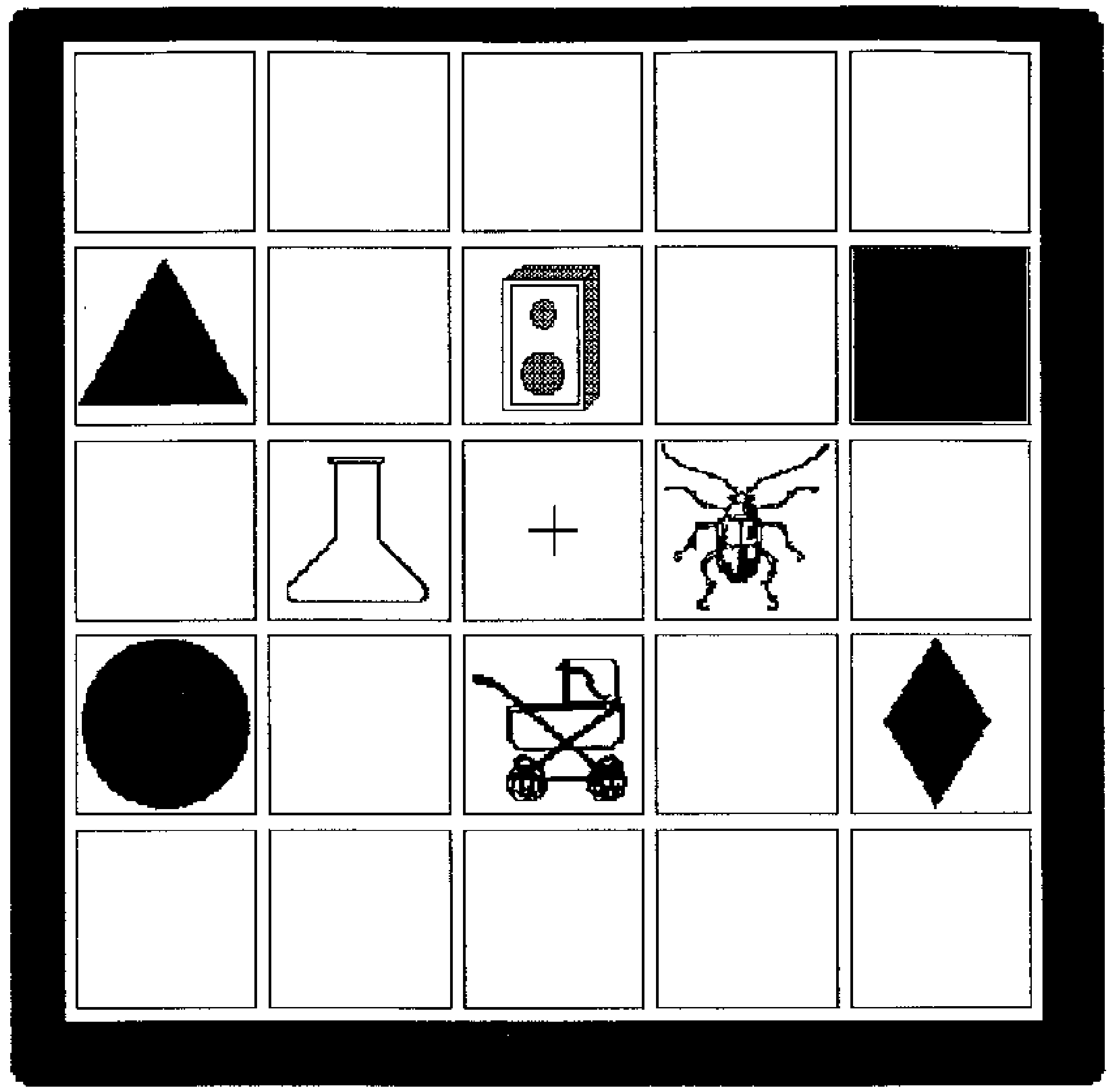
Cohort model

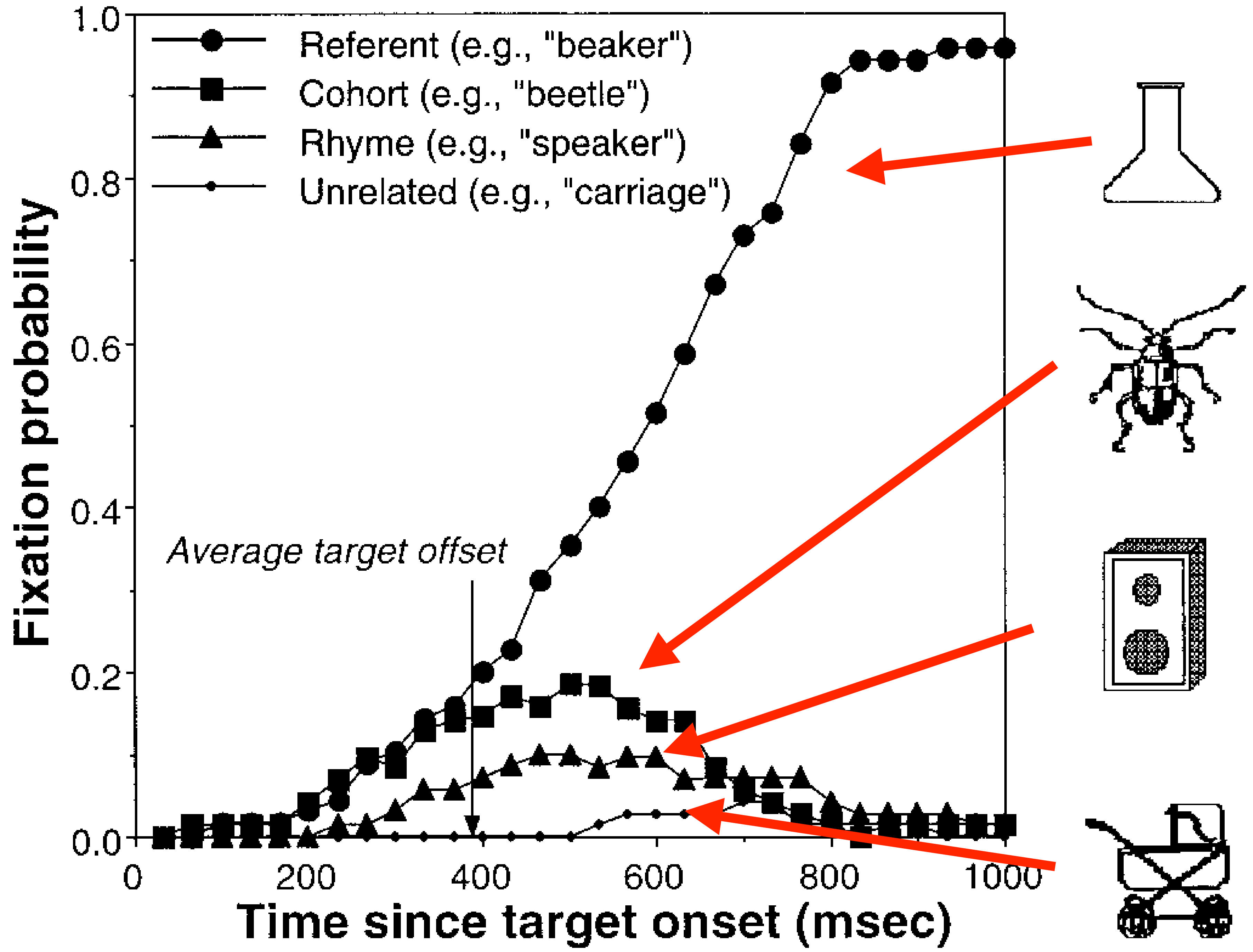
/keɪk…/

Graphs	Pronunciation	SUBTLEX Count
cake	K EY K	2298
caked	K EY K T	9
cakes	K EY K S	291
3		2598

Cohort model

- Activation of multiple candidates
- Competition for recognition





“Pick up the <u>beaker</u>. Now put it above the diamond.”

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

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", …

⋮

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

K EY M …
23875 (45%)
(4 words)

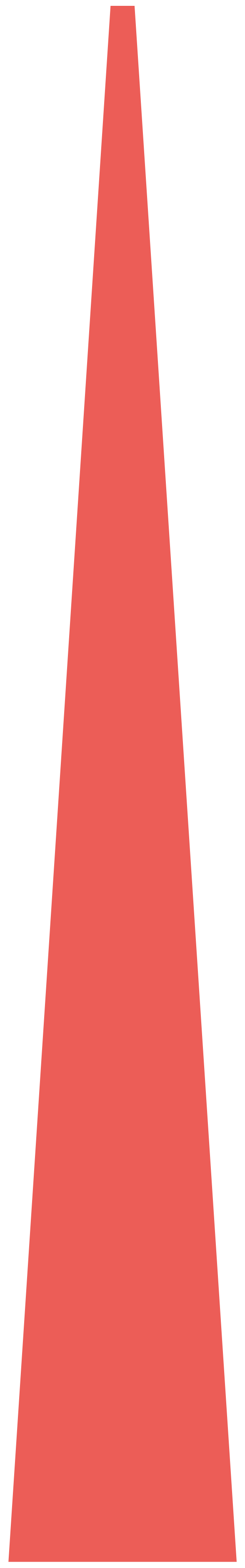
K EY S …
16048 (30%)
(13 words)

K EY K …
2598 (5%)
(3 words)

K EY N …
1337 (3%)
(13 words)

⋮

Surprisal



"came", "Cambridge", …

"case", "cases", "caseworker", "casein", …

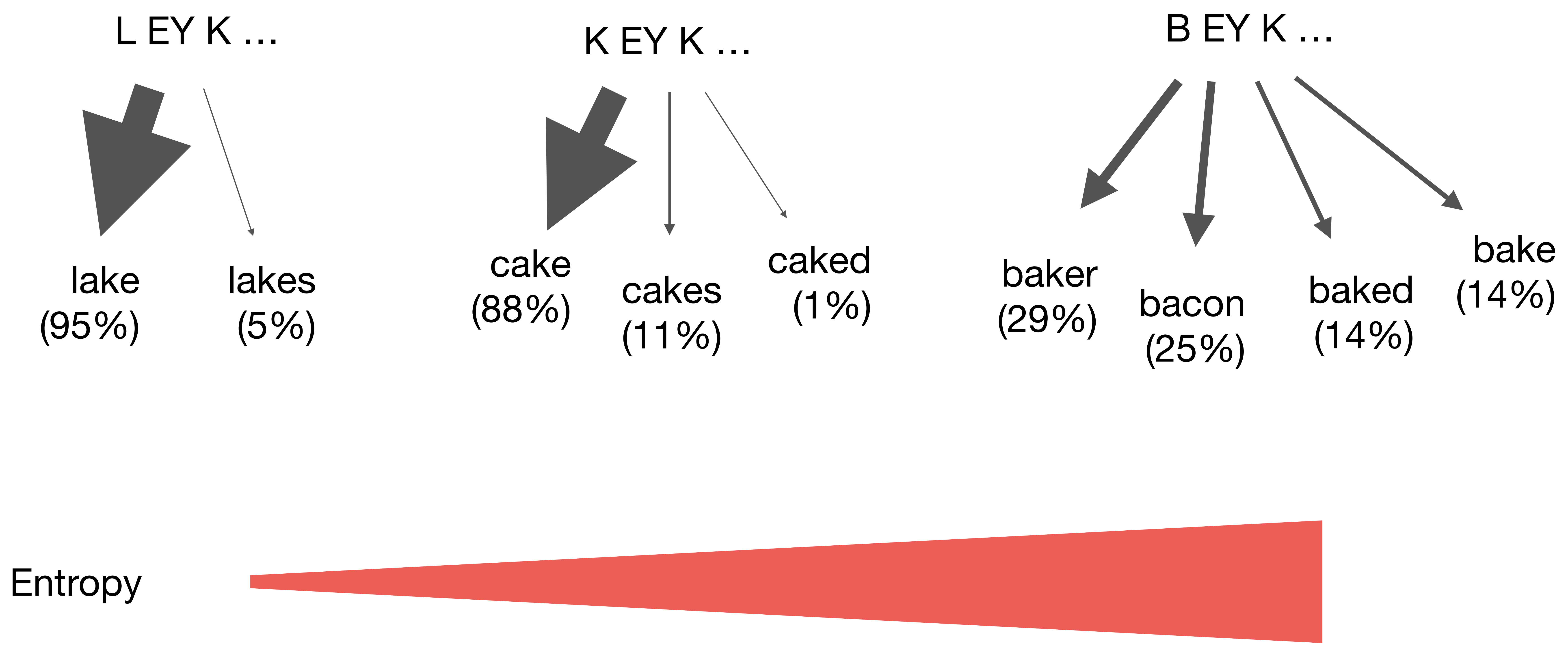
"cake", "caked", "cakes"

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

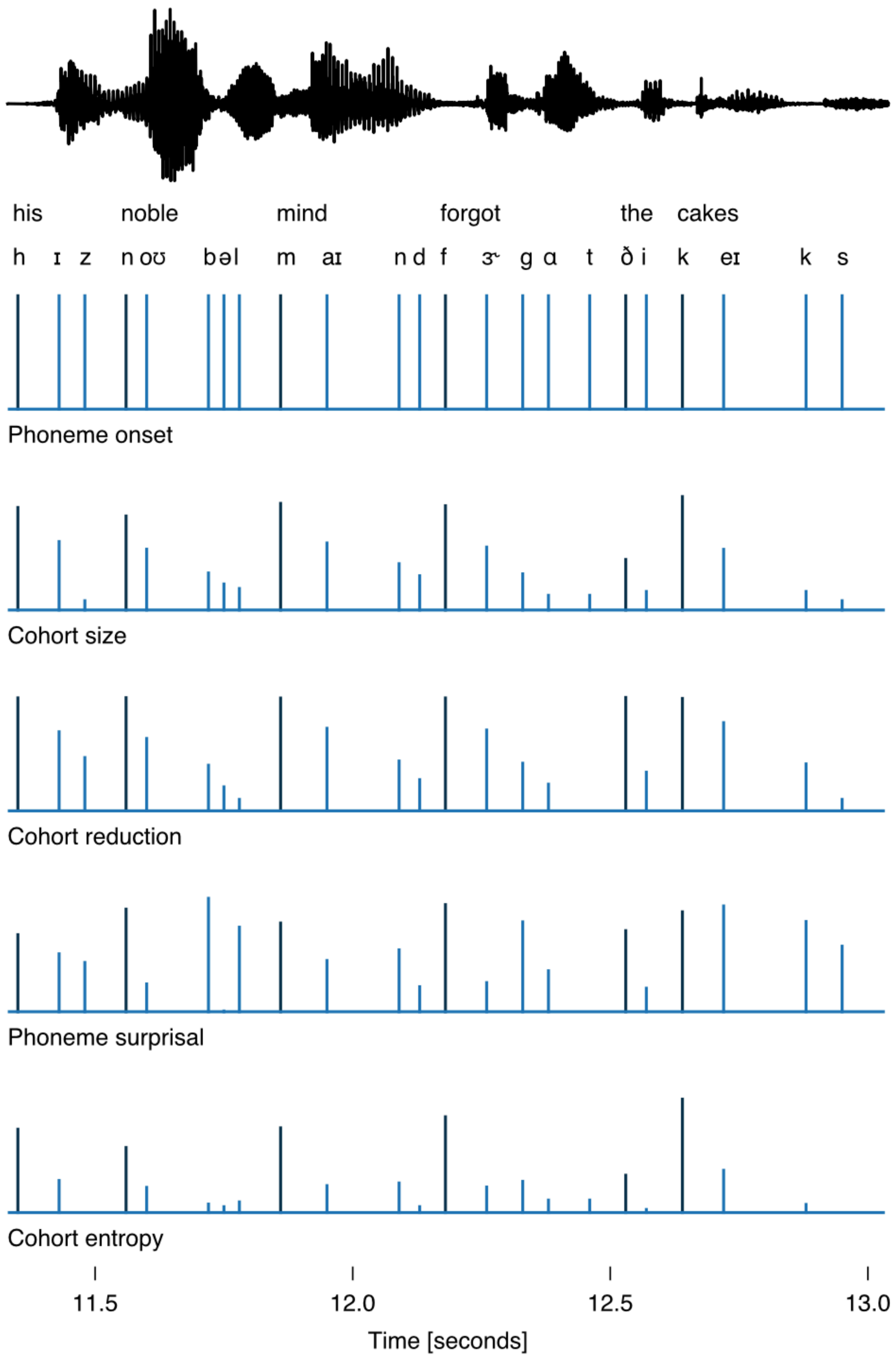
Entropy

Cohort entropy

- How unpredictable is the current word?



Model: Phonemes



Phoneme onset

- Impulse at every phoneme onset

Cohort size

- Number of words in cohort (log)

Cohort reduction

- Number of words that are removed from the cohort

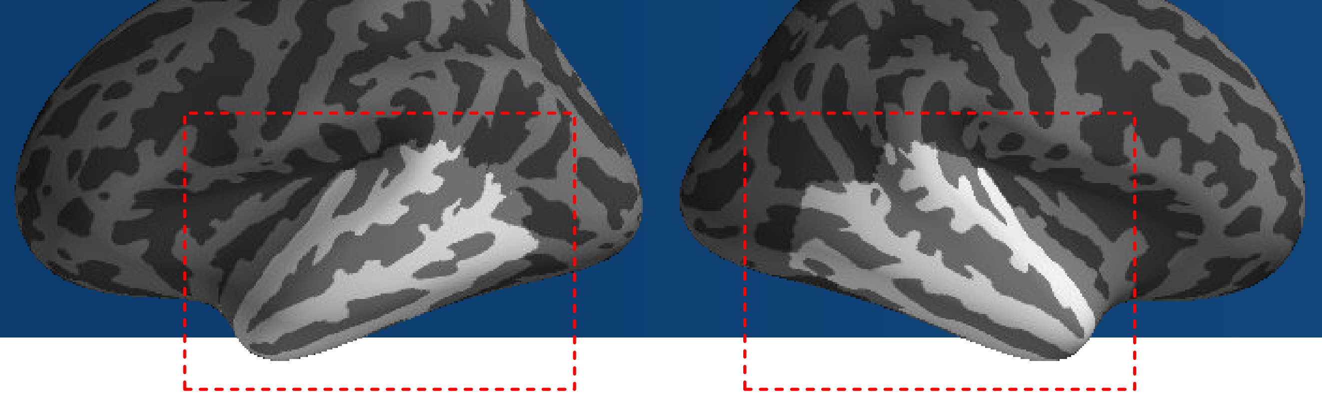
Phoneme surprisal

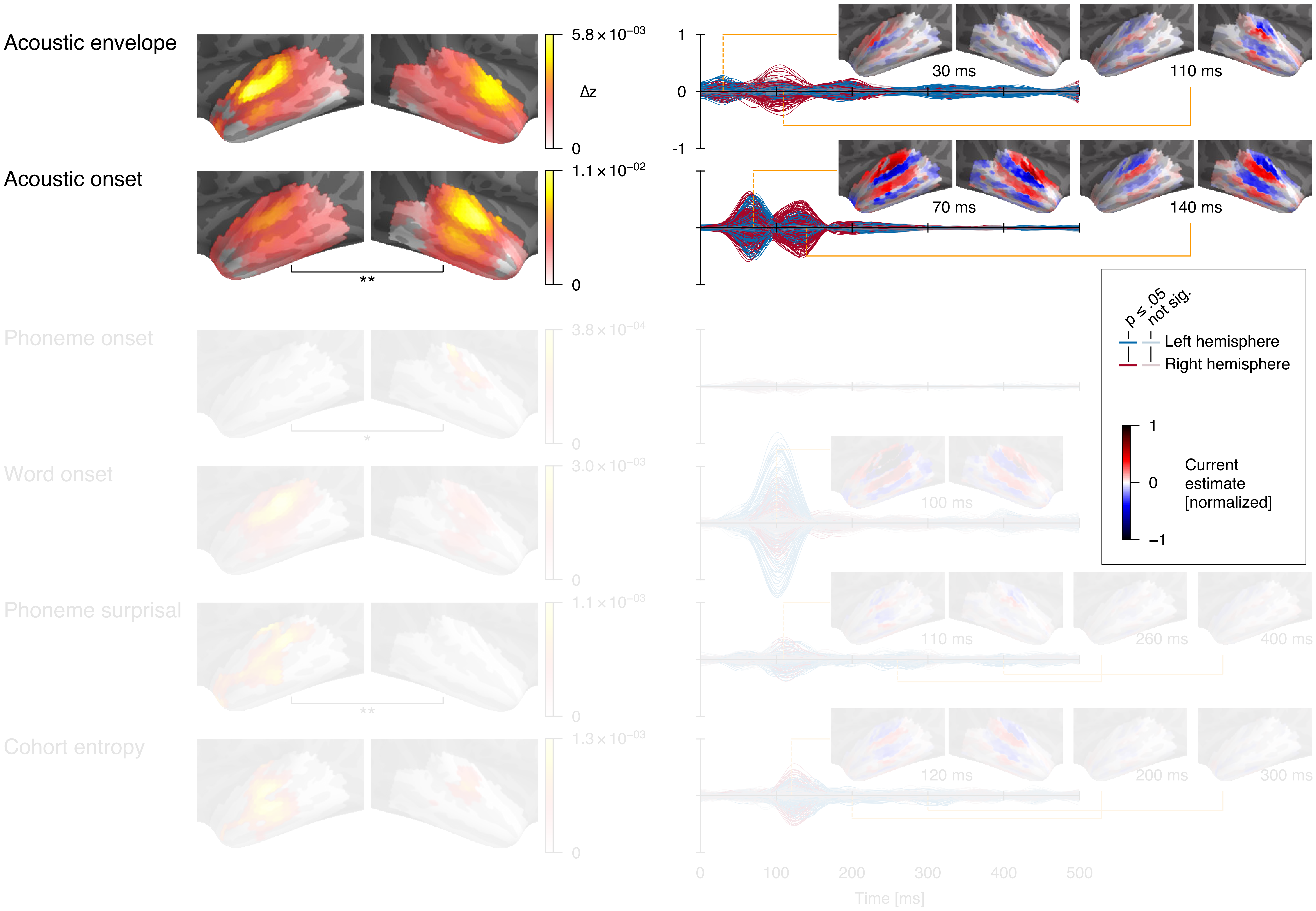
- Related to prediction error

Cohort entropy

- Related to lexical competition

Results: Acoustics

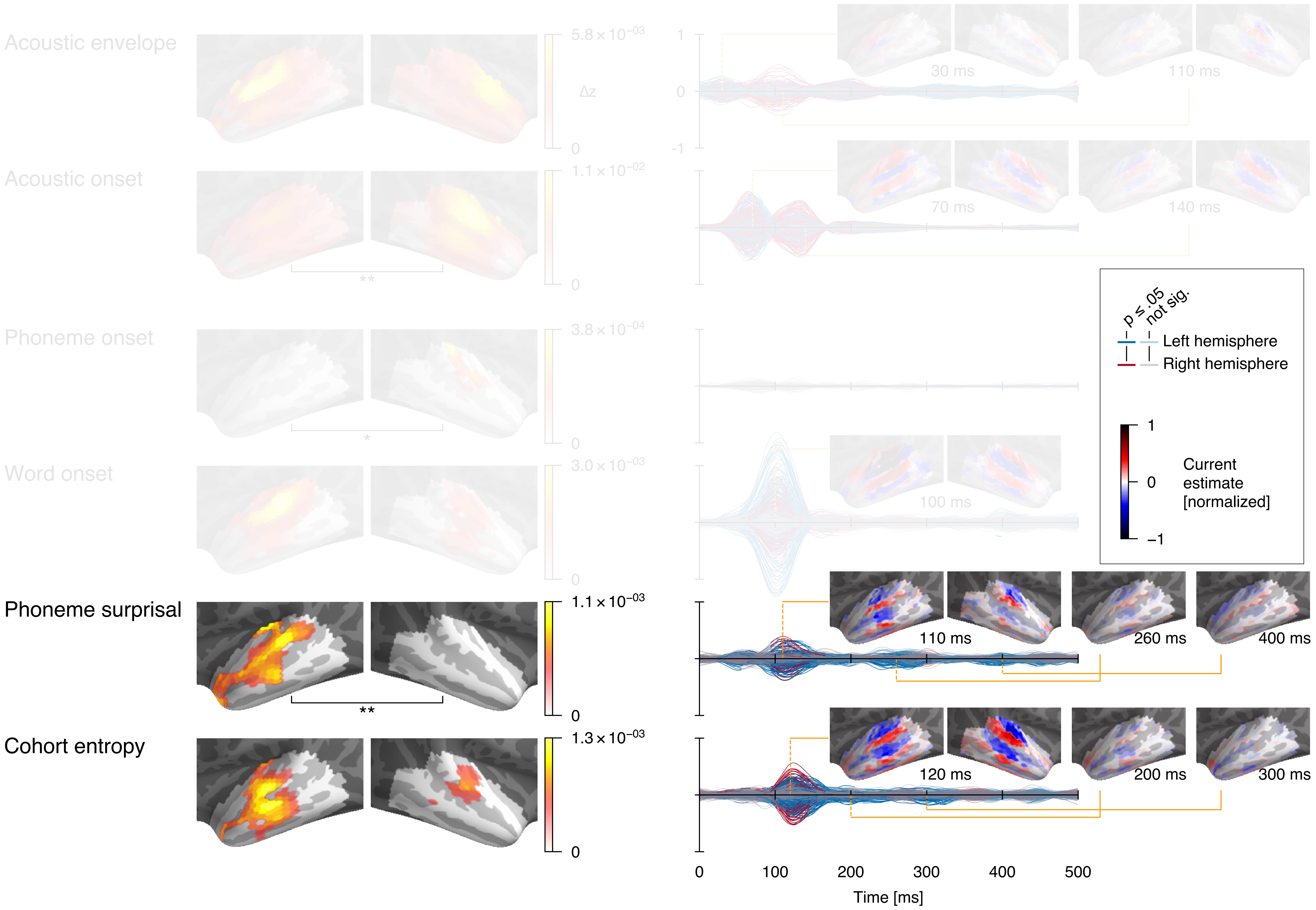




Acoustic features

- Strong bilateral responses
- Two main response peaks (as expected)
- Strong responses to acoustic onsets

Results: Phonemes



No significant effect:

- Cohort size
- Cohort reduction
- Any modulation of word-onset

Phoneme surprisal

- Left-lateralized
- Related to prediction error

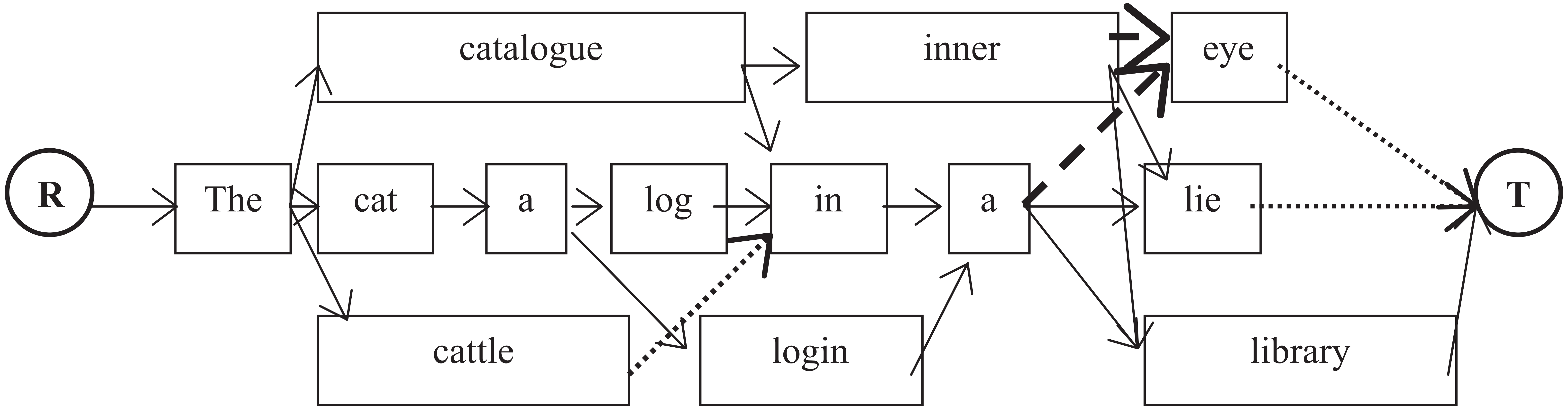
Cohort entropy

- Left-lateralized after excluding right-handers
- Slightly longer latency than surprisal (2 stages?)
- Related to lexical competition

Word onsets

Do we…

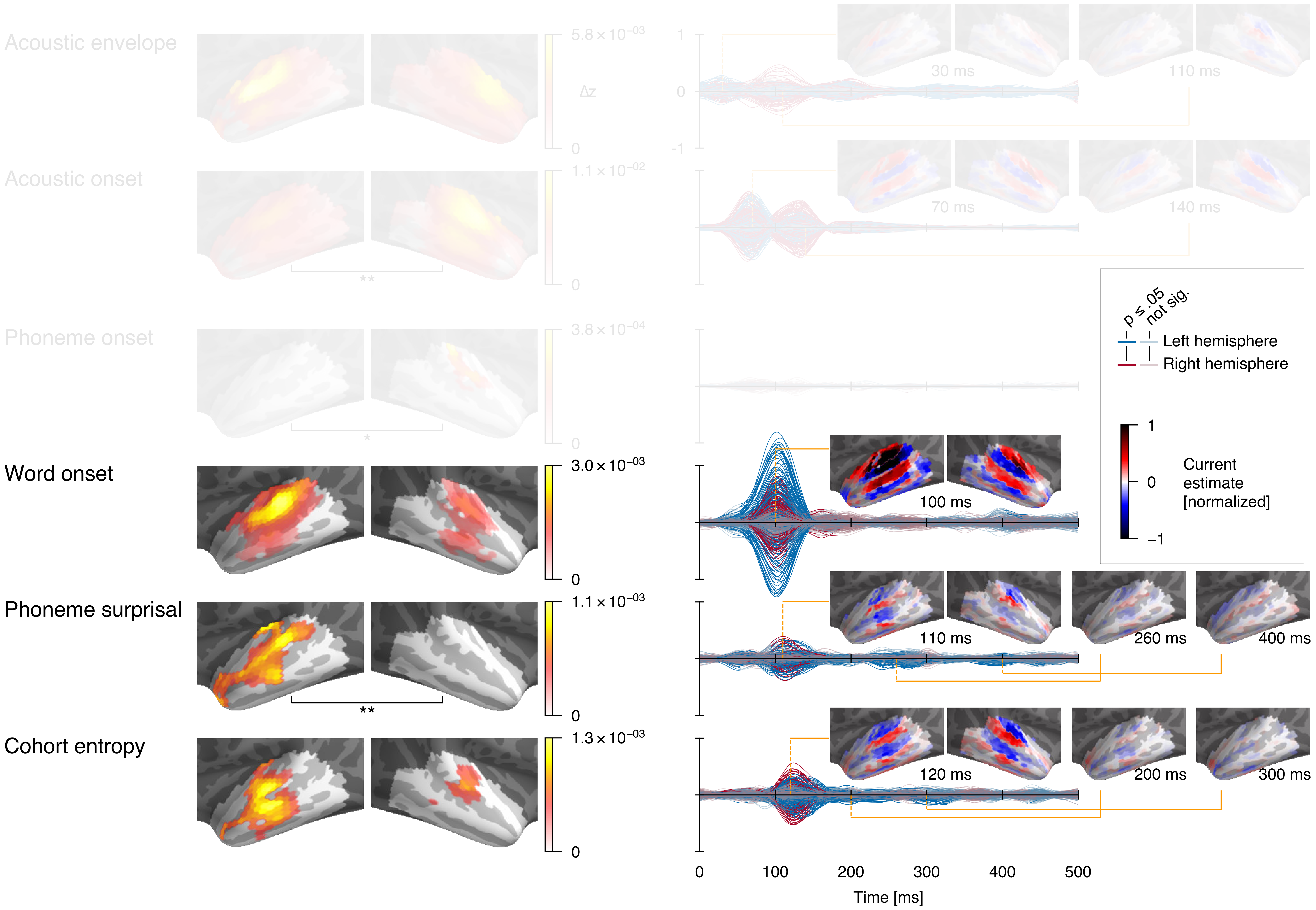
- Anticipate word boundaries based on preceding context?
- Infer them later based on consistency with subsequent context?



“The catalogue in a library”

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. Psychological Review, 115(2), 357–395.

Results: Word onsets



Response at word onsets

- Suggests that on average, word onsets are processed immediately
- Localization similar to acoustic responses
- Opposite current direction of surprisal

Two speakers (“Cocktail Party”)

Lexical processing of unattended speech?

- Hearing your name attracts attention (Cherry, 1953)
- Attending to a conversation is easier when you don’t know the language in the background
- Do we process words in unattended speech?

Two speakers ("Cocktail Party")

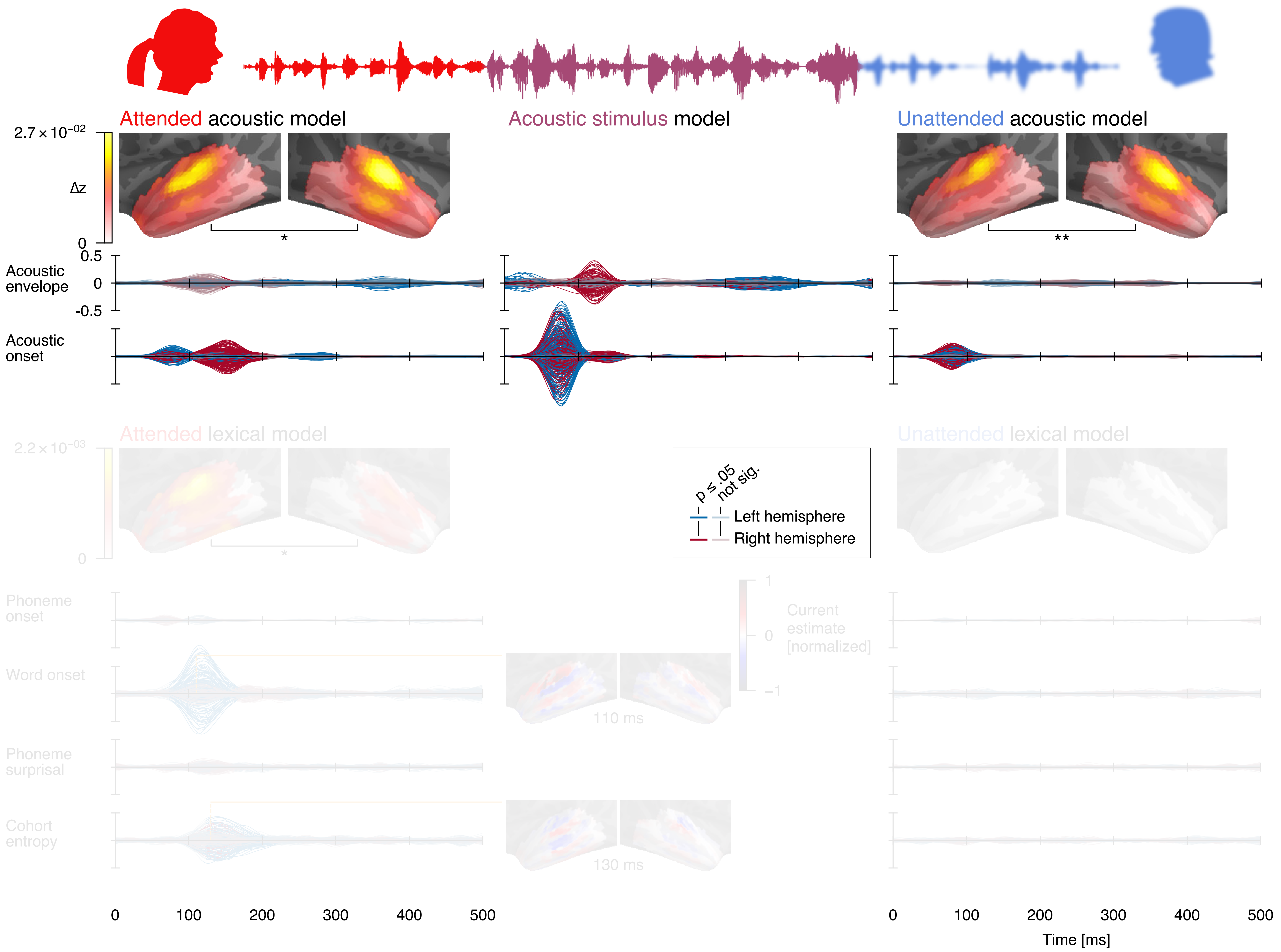
Lexical processing of unattended speech?

- Hearing your name attracts attention (Cherry, 1953)
- Attending to a conversation is easier when you don't know the language in the background
- Do we process words in unattended speech?

Stimuli

- Two speakers, equal loudness
- Instructions: Attend to one, ignore the other
- After each segment, answer a question about the content of the attended stimulus

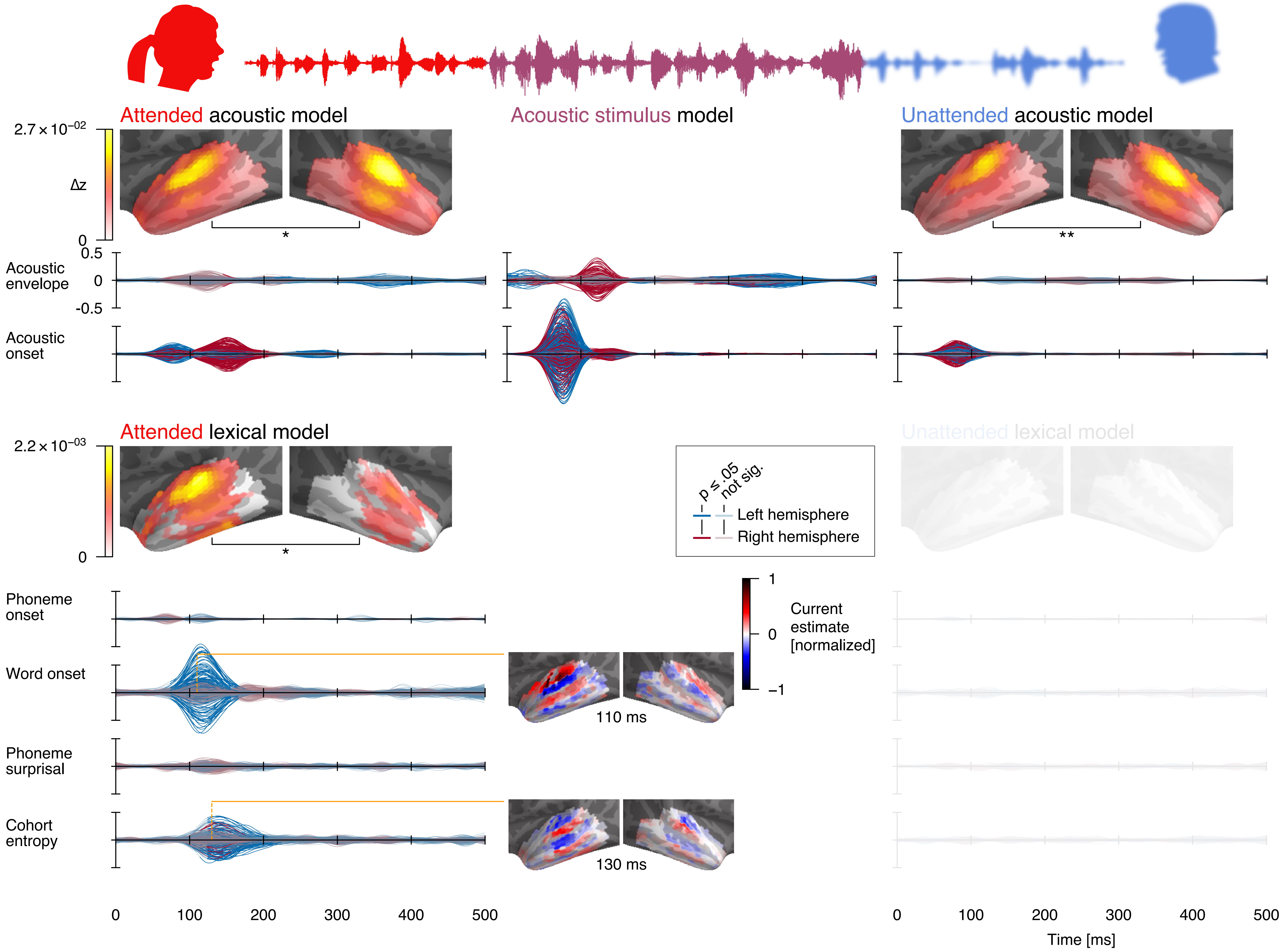
Results: Two speakers



Acoustic responses

- Neural representation of attended and unattended speech
- Amplification of attended features (M100)

Results: Two speakers



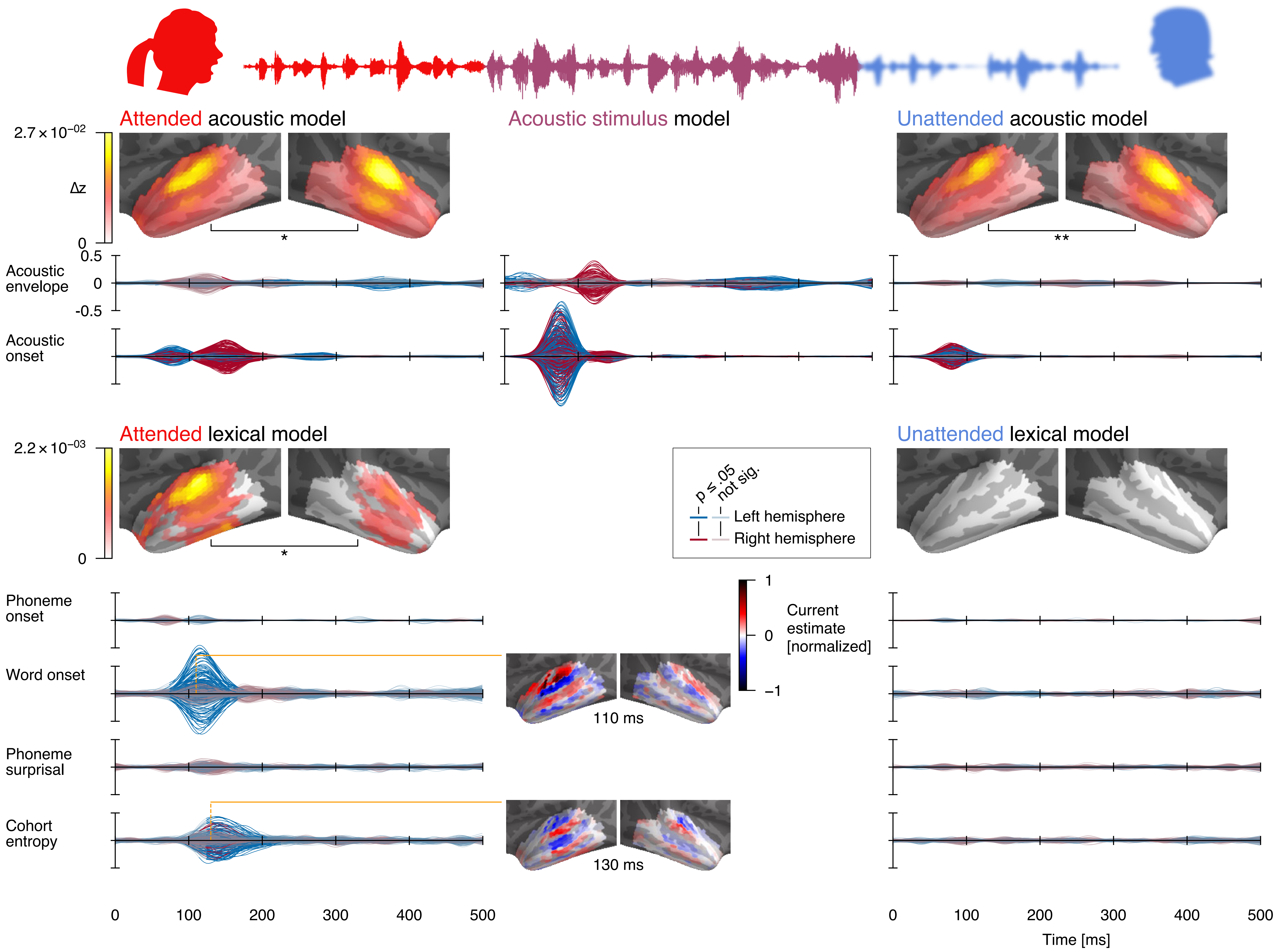
Acoustic responses

- Neural representation of attended and unattended speech
- Amplification of attended features (M100)

Attended: Lexical processing

- Response patterns consistent with single speaker responses
- Delayed responses (~15 ms)

Results: Two speakers



Acoustic responses

- Neural representation of attended and unattended speech
- Amplification of attended features (M100)

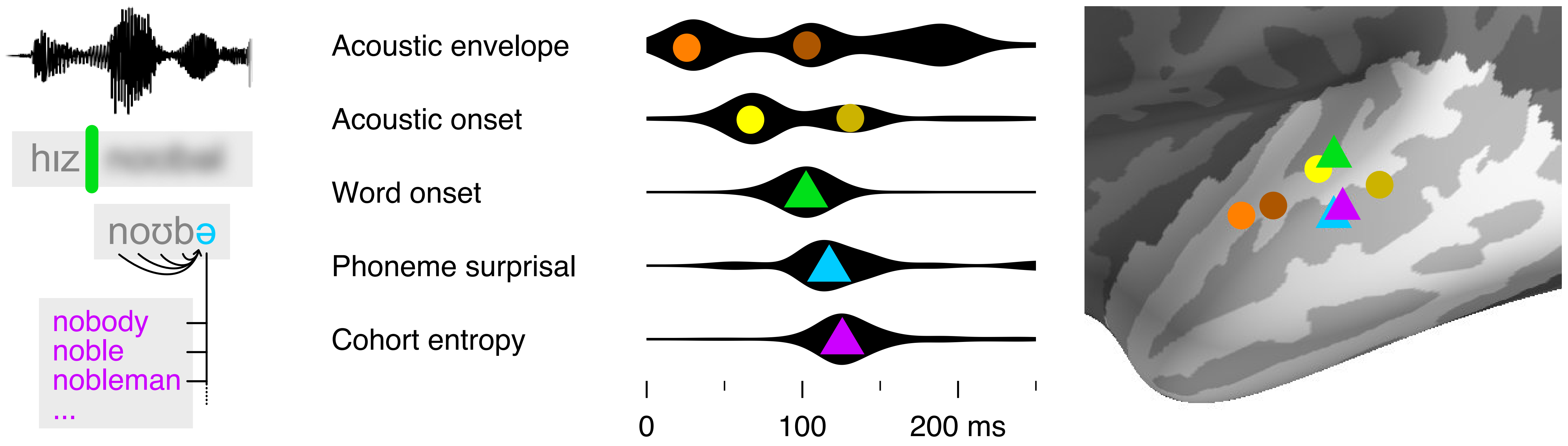
Attended: Lexical processing

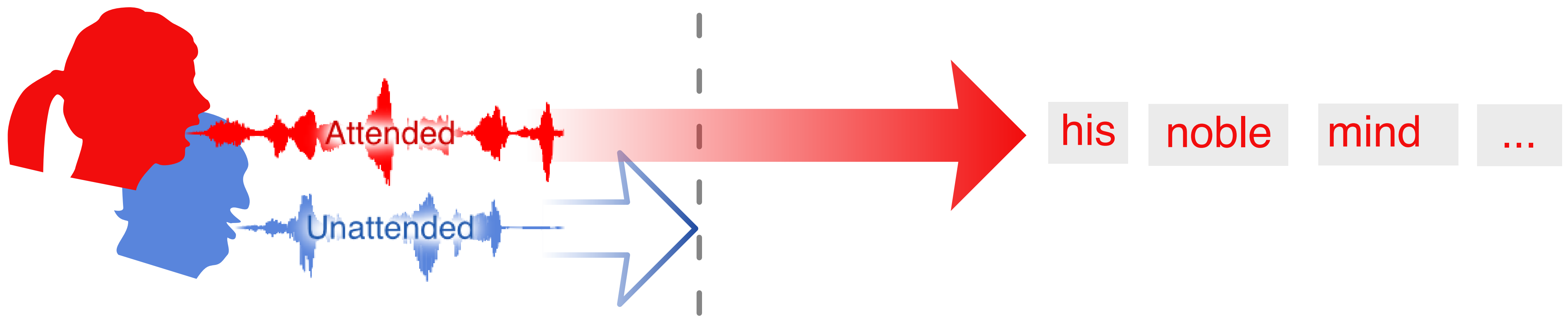
- Response patterns consistent with single speaker responses
- Delayed responses (~15 ms)

Unattended: No time-locked lexical processing

- Lexical processing could still happen but in non time-locked fashion

Summary





Levels of speech

- Acoustic features (acoustic envelope, onsets)
- Phonemes: categories or features
- **Lexical processing** of phonemes: transformation from phonemes to words
  - Cohort-model: activate multiple words compatible with phonemes that are perceived

Time-locked lexical processing of phonetic information

- Modeled as information content of individual phonemes
  - Word onsets (lexical segmentation)
  - Phoneme surprisal (phoneme information content, predictive coding) ~110 ms
  - Lexical entropy (lexical competition) ~130 ms

“Cocktail party”

- Two concurrent speakers, attend to one and ignore the other
- Time-locked lexical processing of only the attended speech

Thank you!

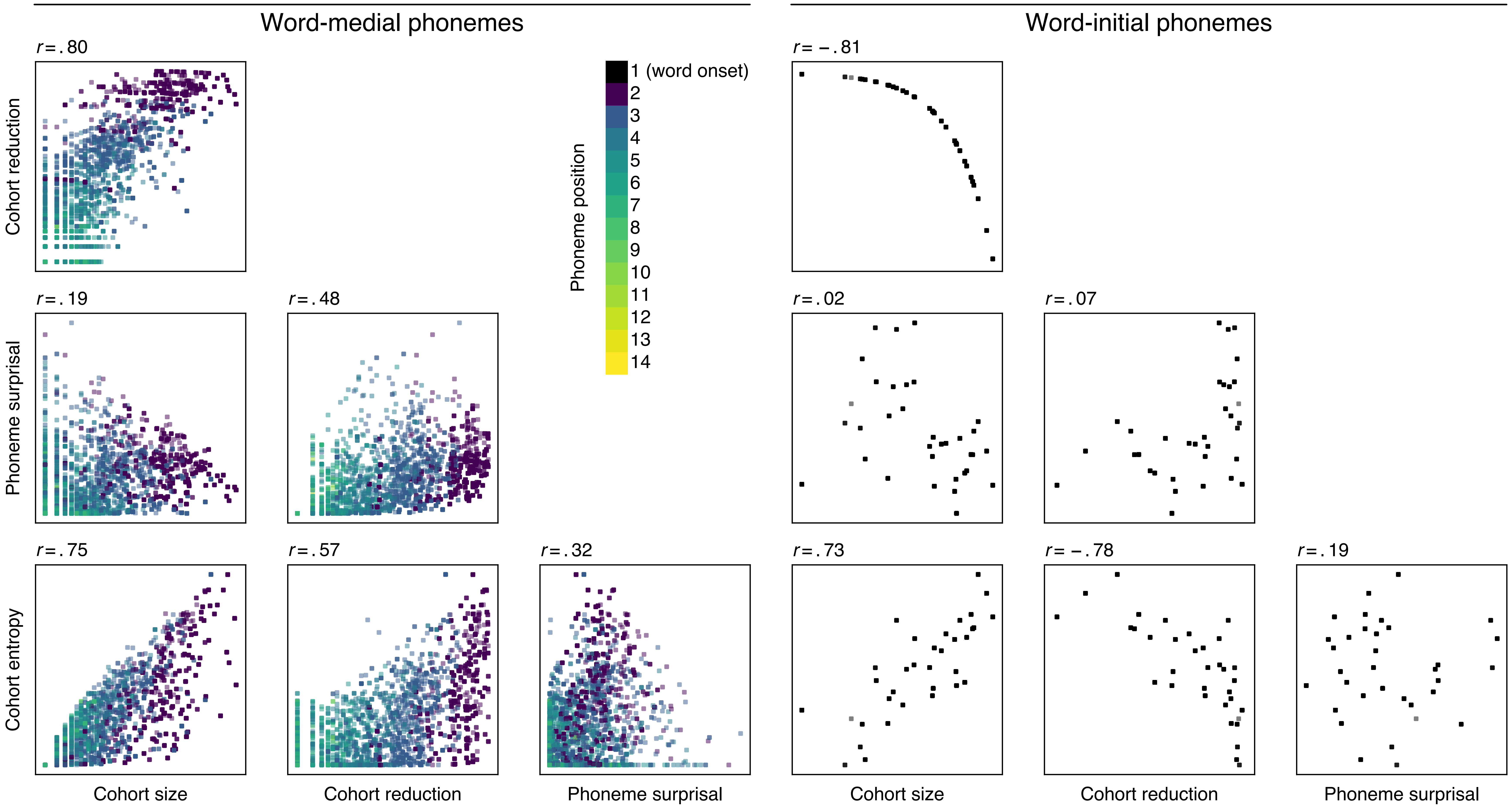
Acknowledgements

- Jonathan Simon, Elliot Hong
  Co-authors
- Krishna Puvvada
  Stimulus design and data collection
- Natalia Lapinskaya
  Data collection
- NIH, University of Maryland
  Funding

Published as

- Brodbeck, C., Hong, L. E., & Simon, J. Z. (2018). Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech. Current Biology, 28(24), 3976-3983.e5. 10.1016/j.cub.2018.10.042

Pairwise Correlations



Spatial Separation