Transformations from Auditory to Lexical Representations, across Auditory Cortex, are Rapid and Attention Dependent

Jonathan Z. Simon

Department of Electrical & Computer Engineering

Department of Biology

Institute for Systems Research

University of Maryland

Outline

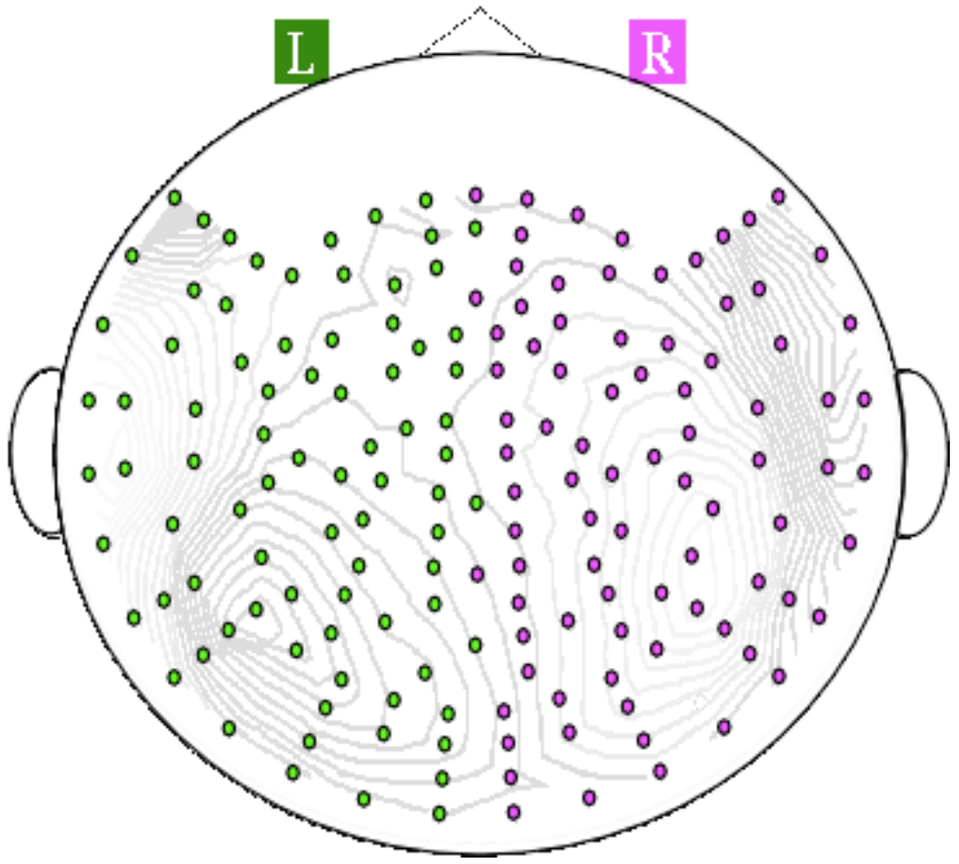
- Background & motivation
  - Neural responses in time
  - Neural responses in time & space
  - Representations: from Acoustic to Linguistic
- Spatiotemporal representation transformation from Acoustic to Lexical

Outline

- Background & motivation
  - Neural responses in time
  - Neural responses in time & space
    - Representations: from Acoustic to Linguistic
- Spatiotemporal representation transformation from Acoustic to Lexical

Magnetoencephalography (MEG)

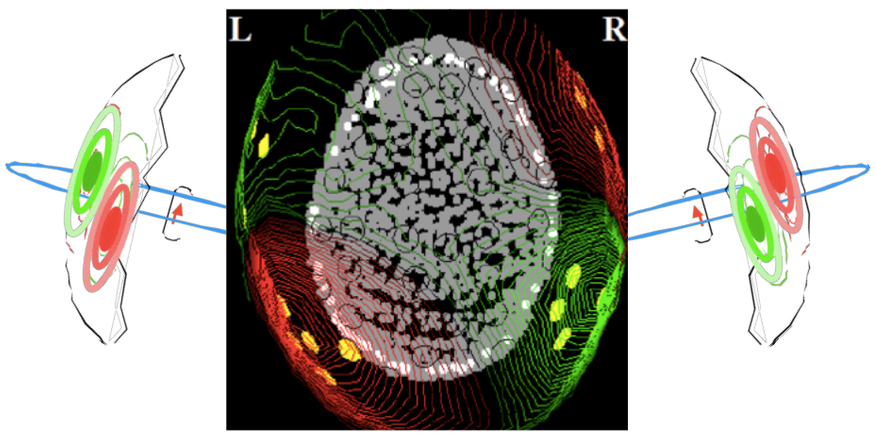
- Non-invasive, passive, silent neural recordings from cortex
- Simultaneous whole-head recording (~200 sensors)
- Sensitivity
  - high: ~100 fT (10^{-13} Tesla)
  - low: ~10^4 – ~10^6 neurons
- Temporal resolution: ~1 ms
- Spatial resolution
  - coarse: ~1 cm
  - ambiguous

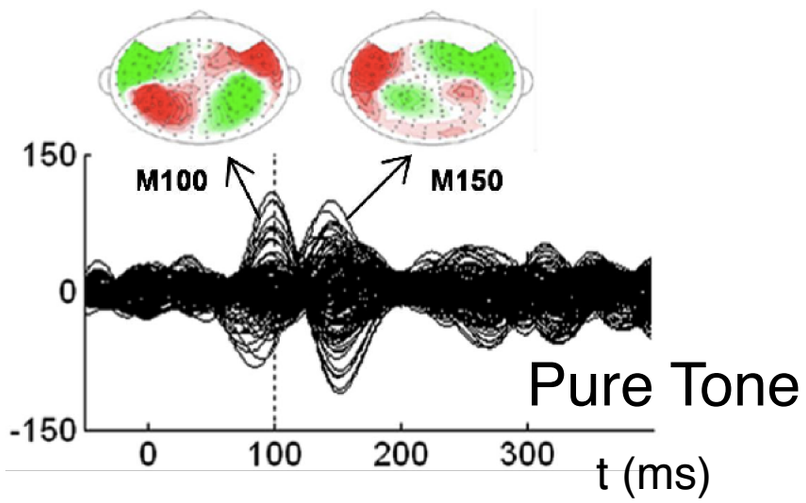


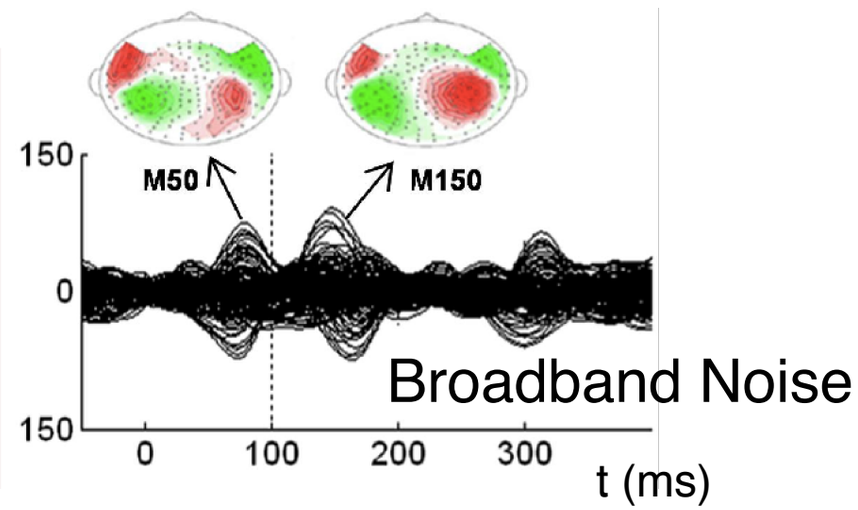
Time Course of MEG Responses

Time-locked auditory responses

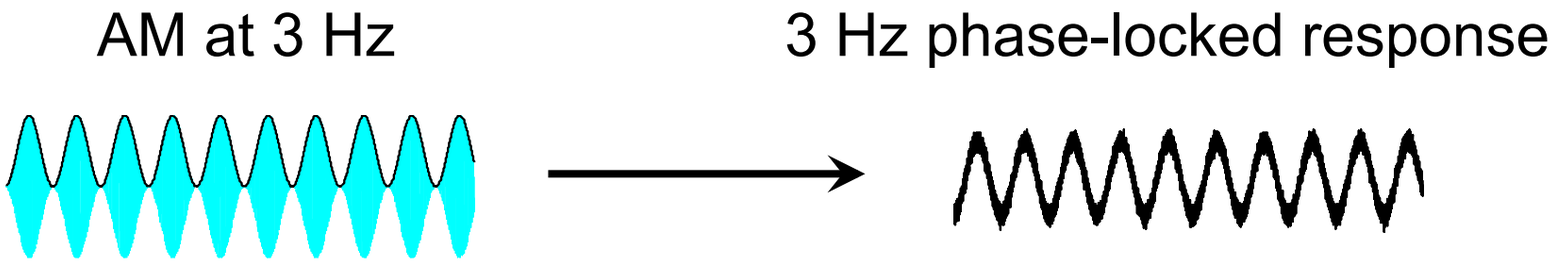
- MEG response patterns time-locked to stimulus events
- Robust
- Strongly lateralized
- Cortical origin



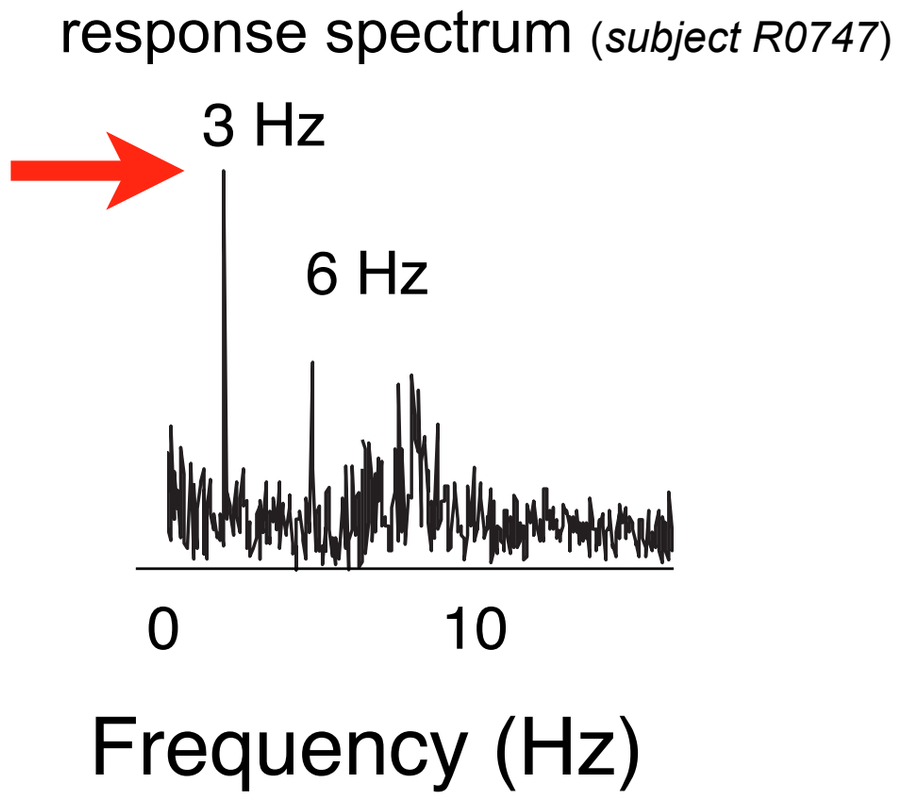




Time Course of MEG Responses



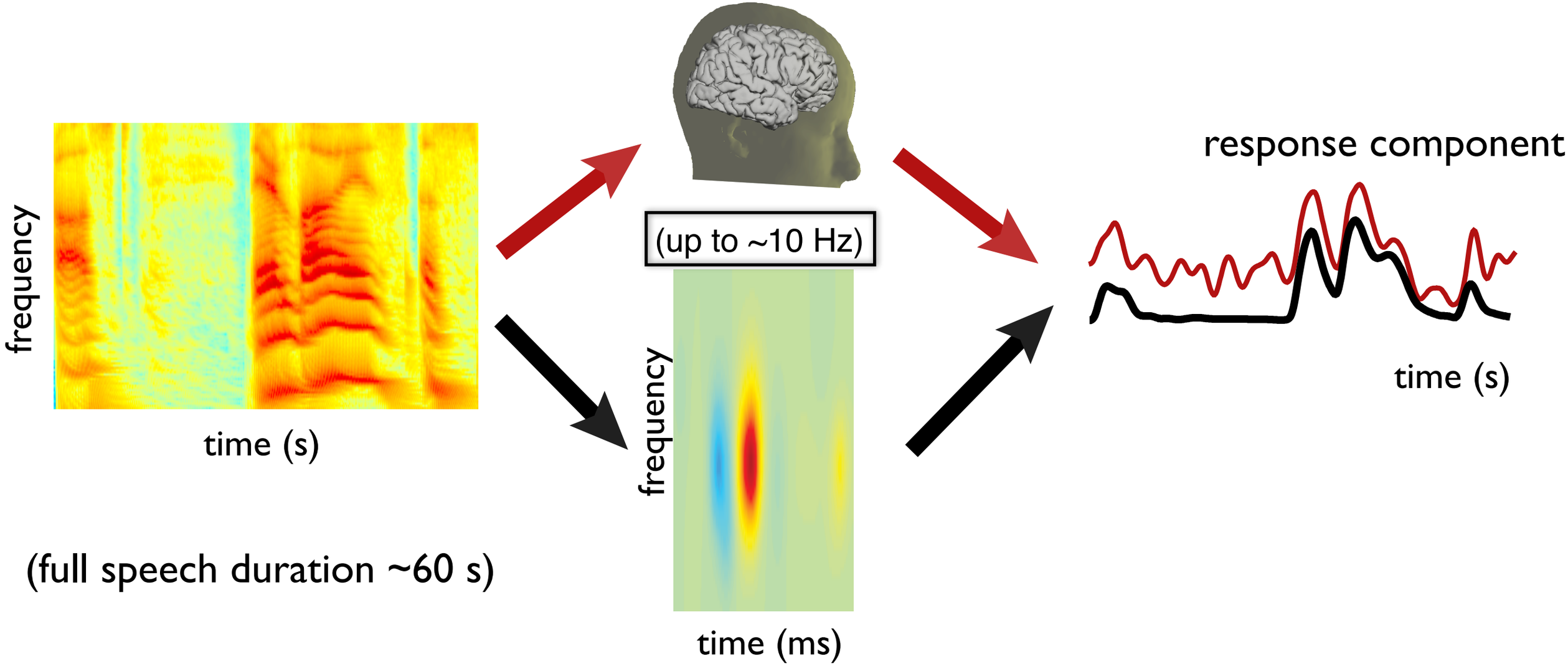
MEG activity is also time-locked to temporal modulations of sound



Ding & Simon, J Neurophysiol (2009)
Wang et al., J Neurophysiol (2012)

Time Course of MEG Responses to Speech

& STRF model predictions



(full speech duration ~60 s)

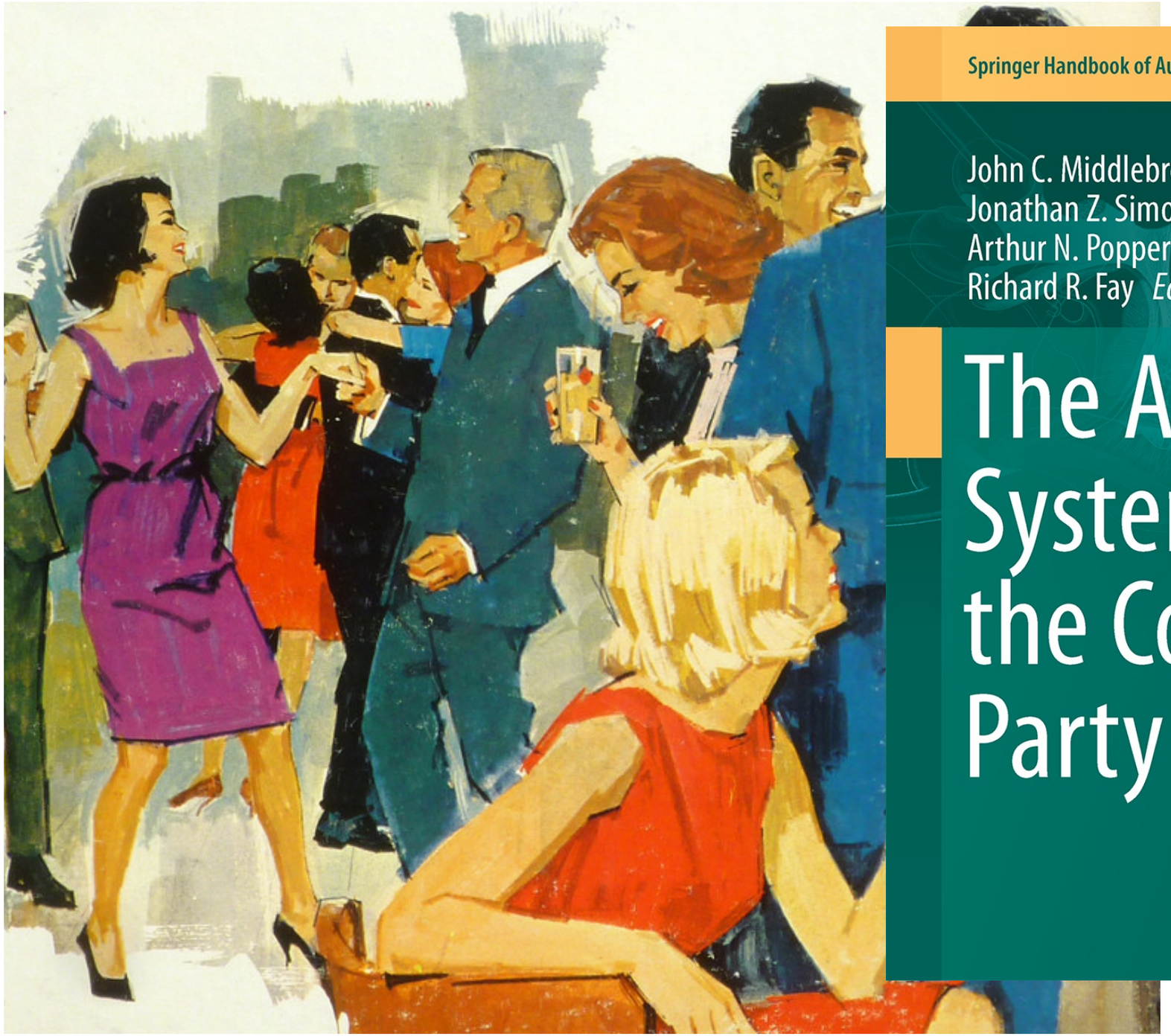
Ding & Simon, J Neurophysiol (2012)

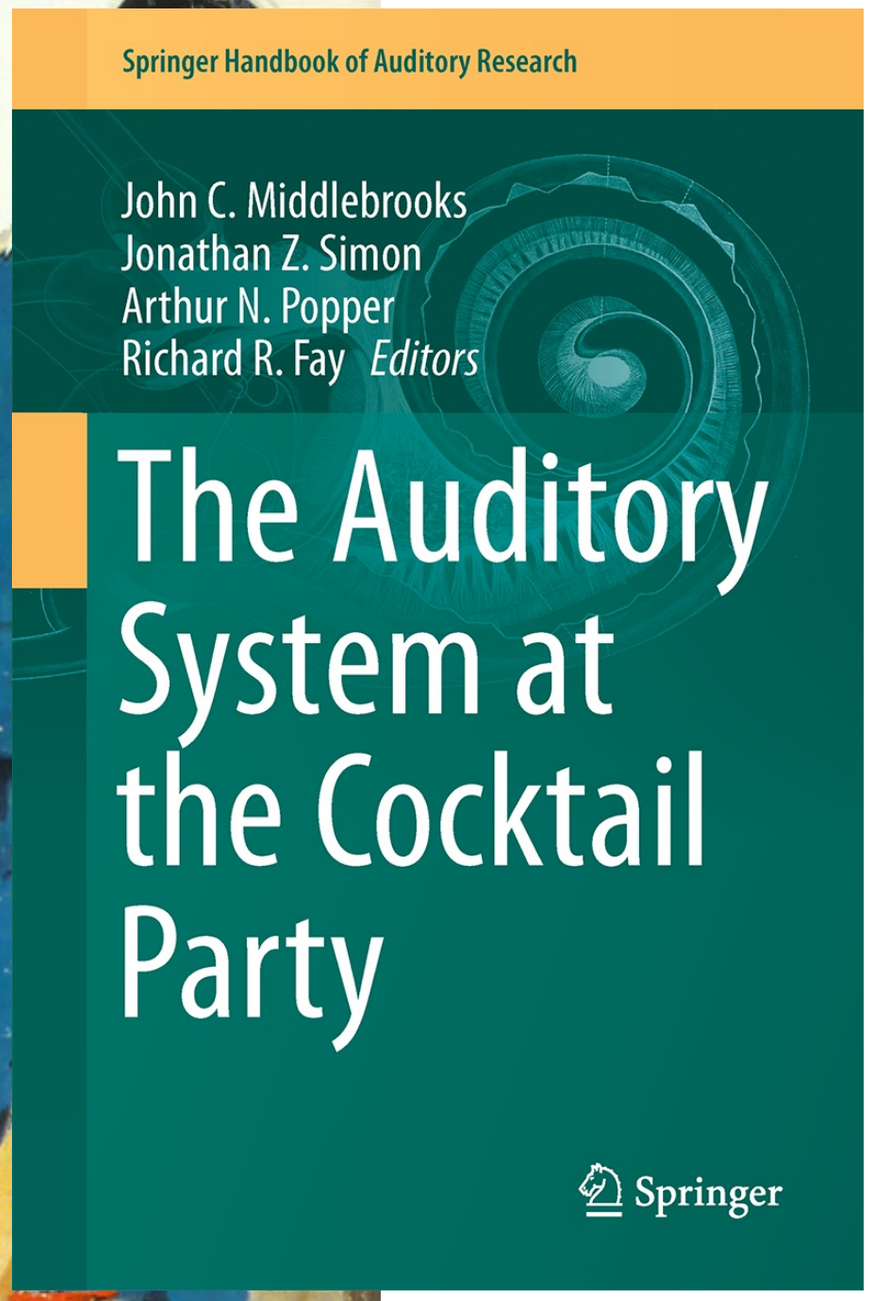
"Spectro-Temporal Response Function"

Cortical Speech Representations

- Neural representation: encoding
- Linear model
- Speech spectrotemporal **envelope** only
- Envelope rates: ~ 1 - 10 Hz
- Sensor-space based

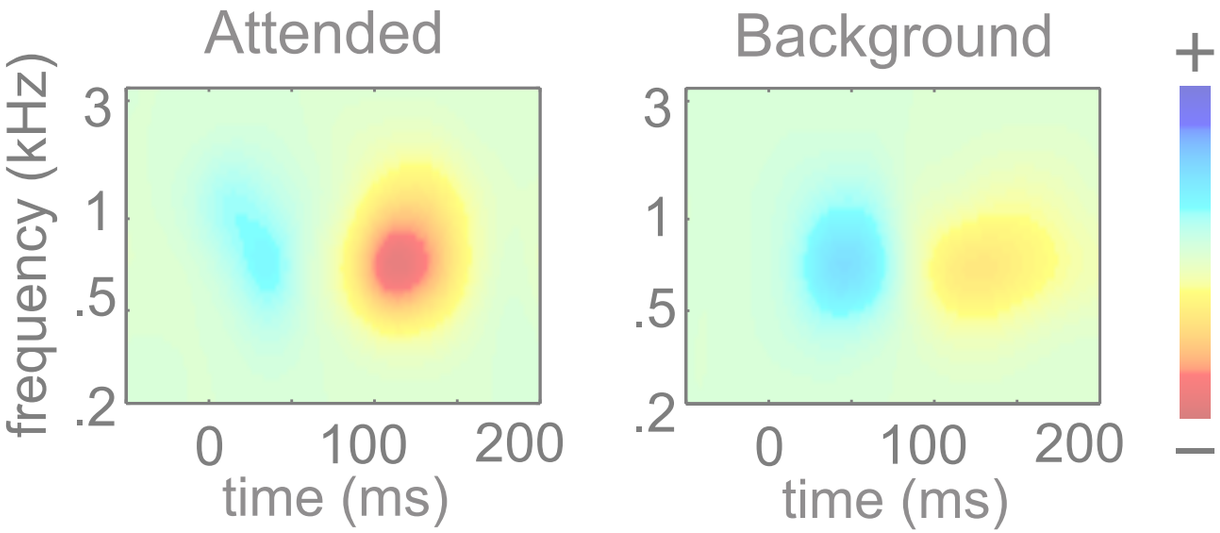
Listening to Speech at the Cocktail Party

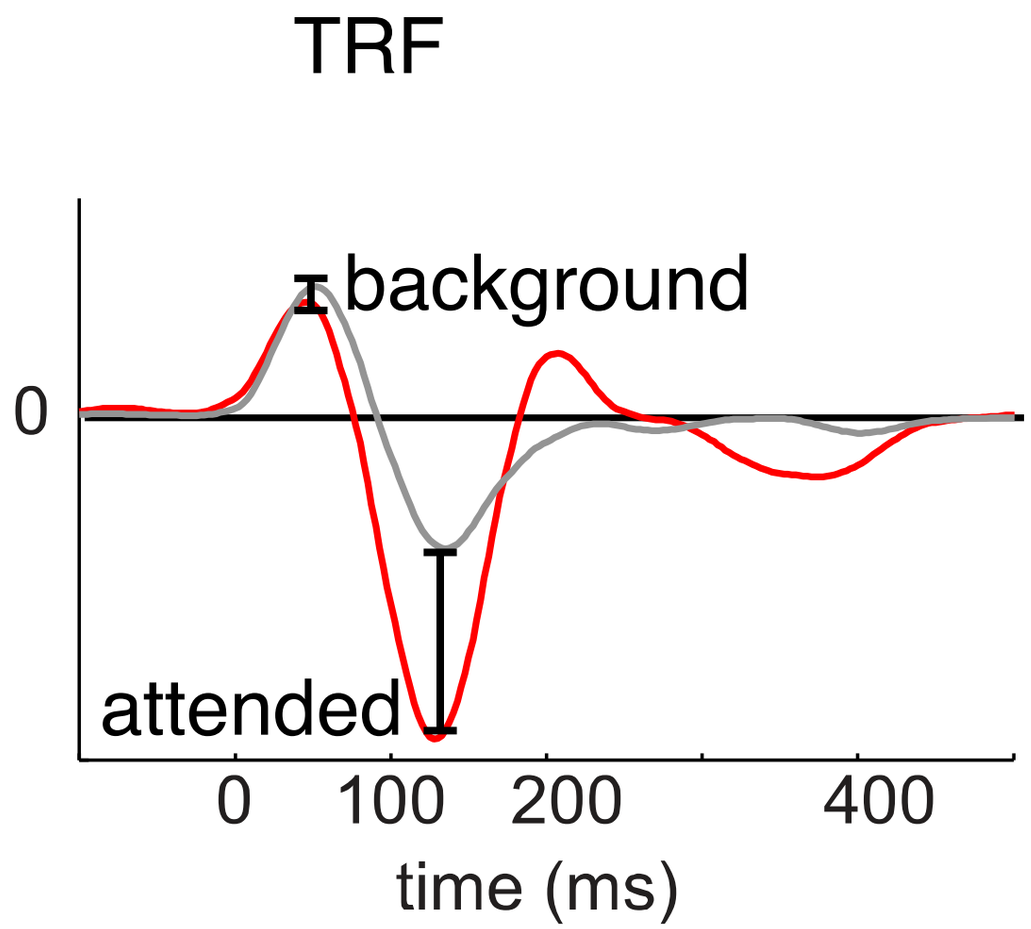




Listening to Speech at the Cocktail Party

Multispeaker STRFs





- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- $M50_{STRF}$ positive peak
- $M100_{STRF}$ negative peak
- **$M100_{STRF}$ strongly modulated by attention, *but not $M50_{STRF}$***

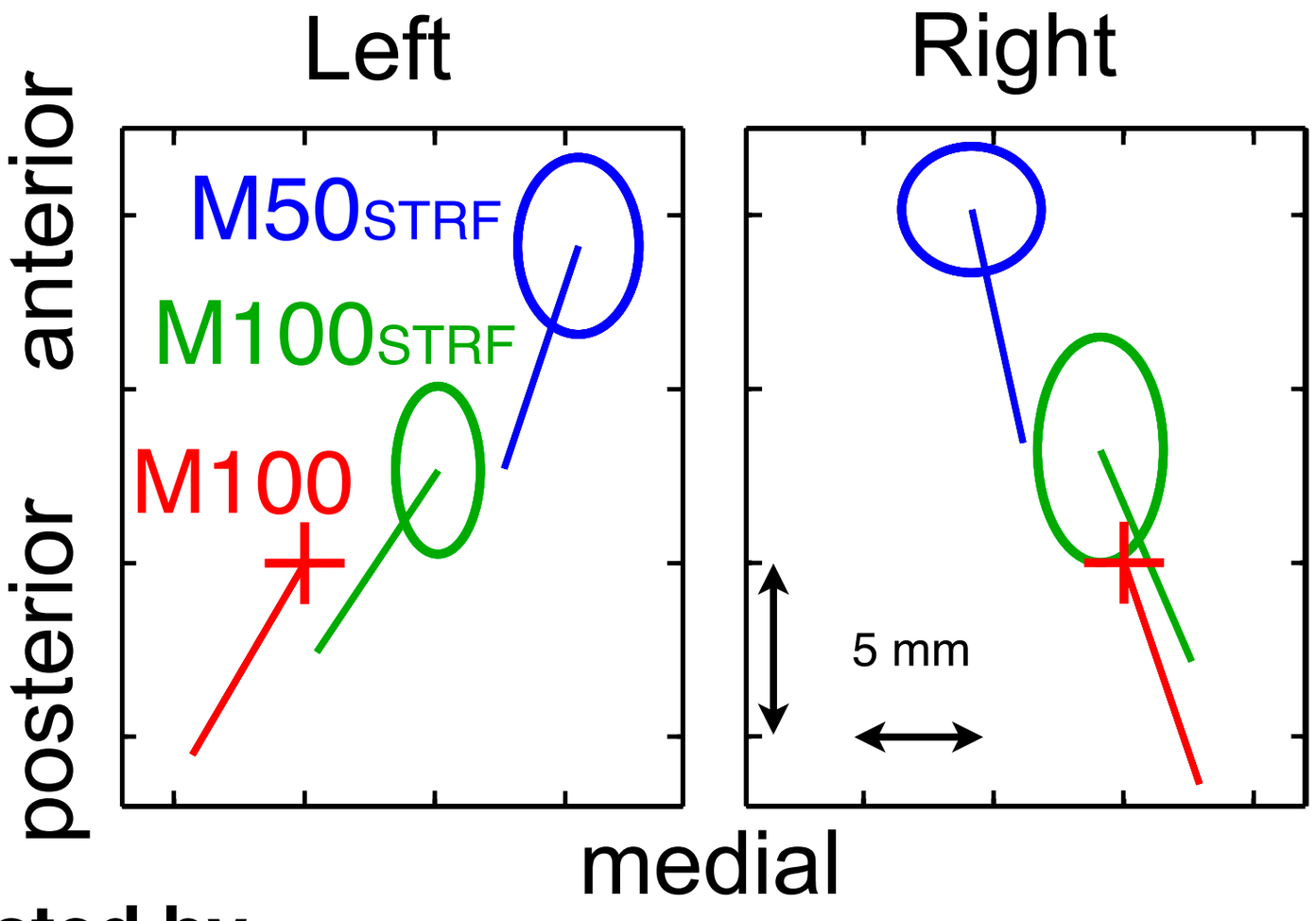
Ding & Simon, PNAS (2012)

Neural Sources

of STRF peaks

- **$M100_{STRF}$** source near M100 source: **Planum Temporale**

- **$M50_{STRF}$** source is anterior and medial to M100: **Heschl's Gyrus**

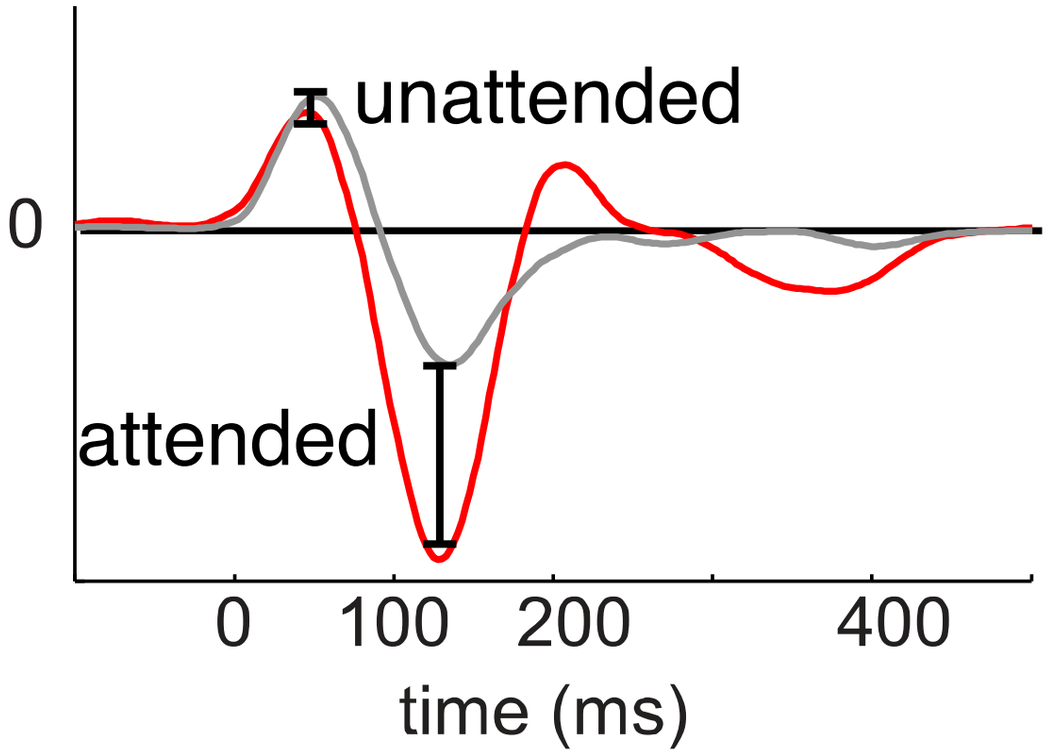


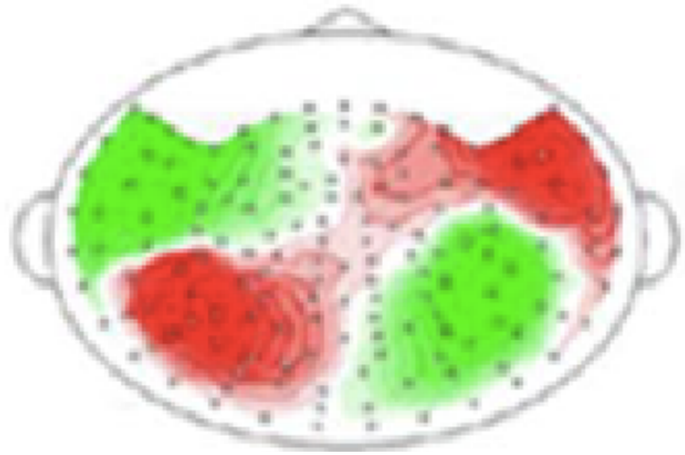
- **PT source strongly affected by attention, *but not HG source***

Ding & Simon, PNAS (2012)

Time Course of MEG Response

Temporal Response Function
of dominant auditory component





Time course analysis of single response component is

- useful
- simplifying
- a good start

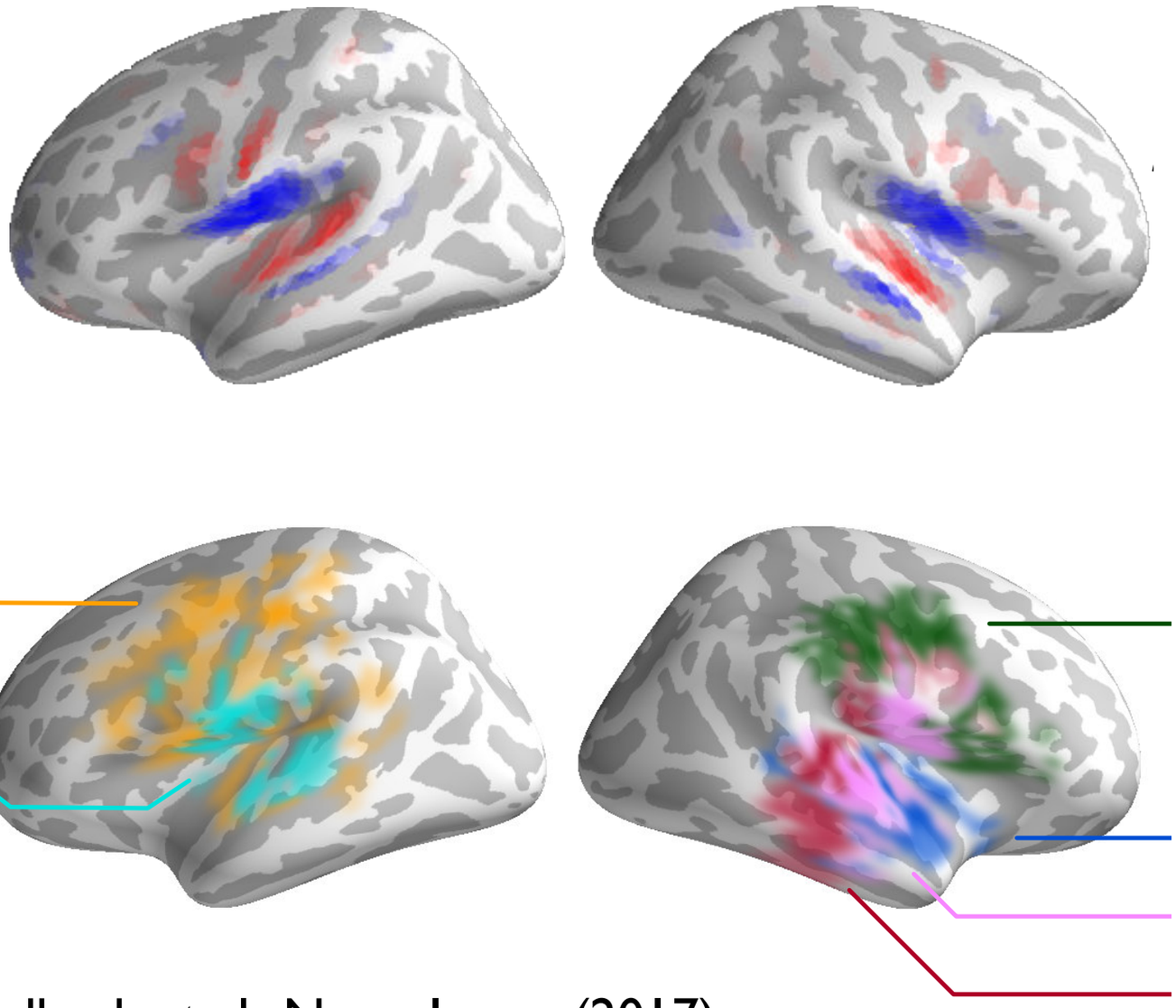
- $M100_{TRF}$ strongly modulated by attention, *but not* $M50_{TRF}$

Ding & Simon, PNAS (2012)

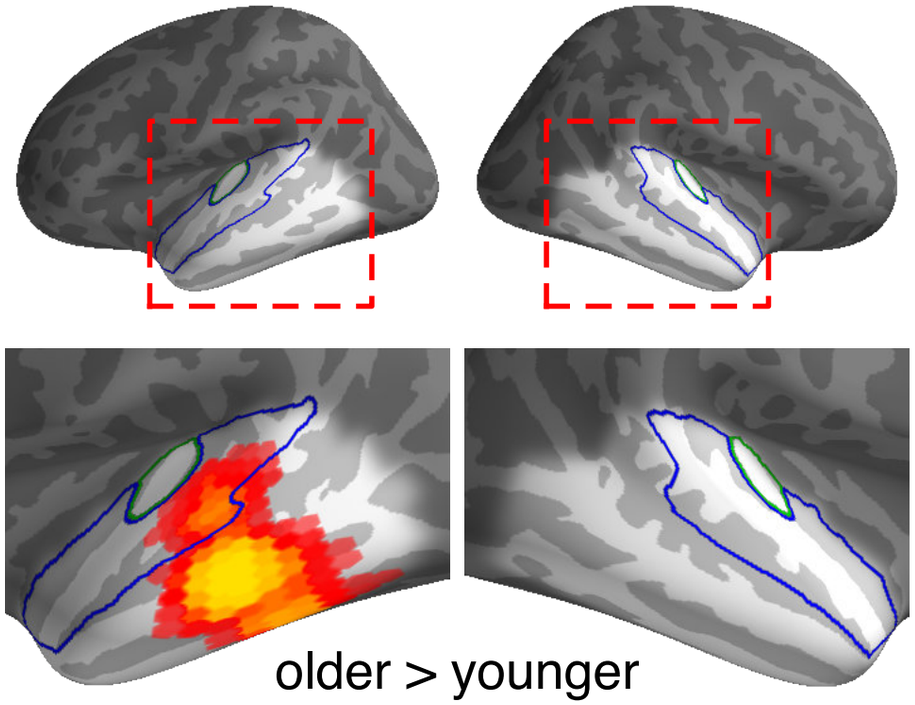
Outline

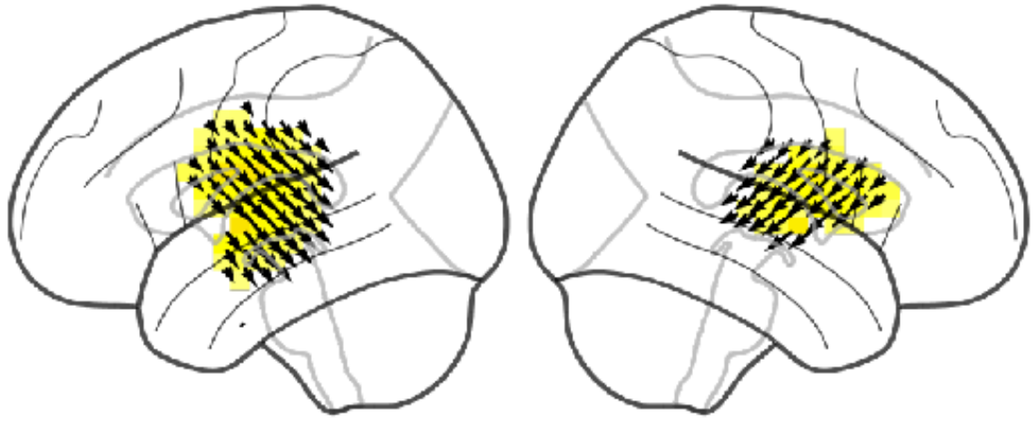
- Background & motivation
  - Neural responses in time
  - Neural responses in time & space
    - Representations: from Acoustic to Linguistic
- Spatiotemporal representation transformation from Acoustic to Lexical

Spatial Distributions of MEG Neural Currents



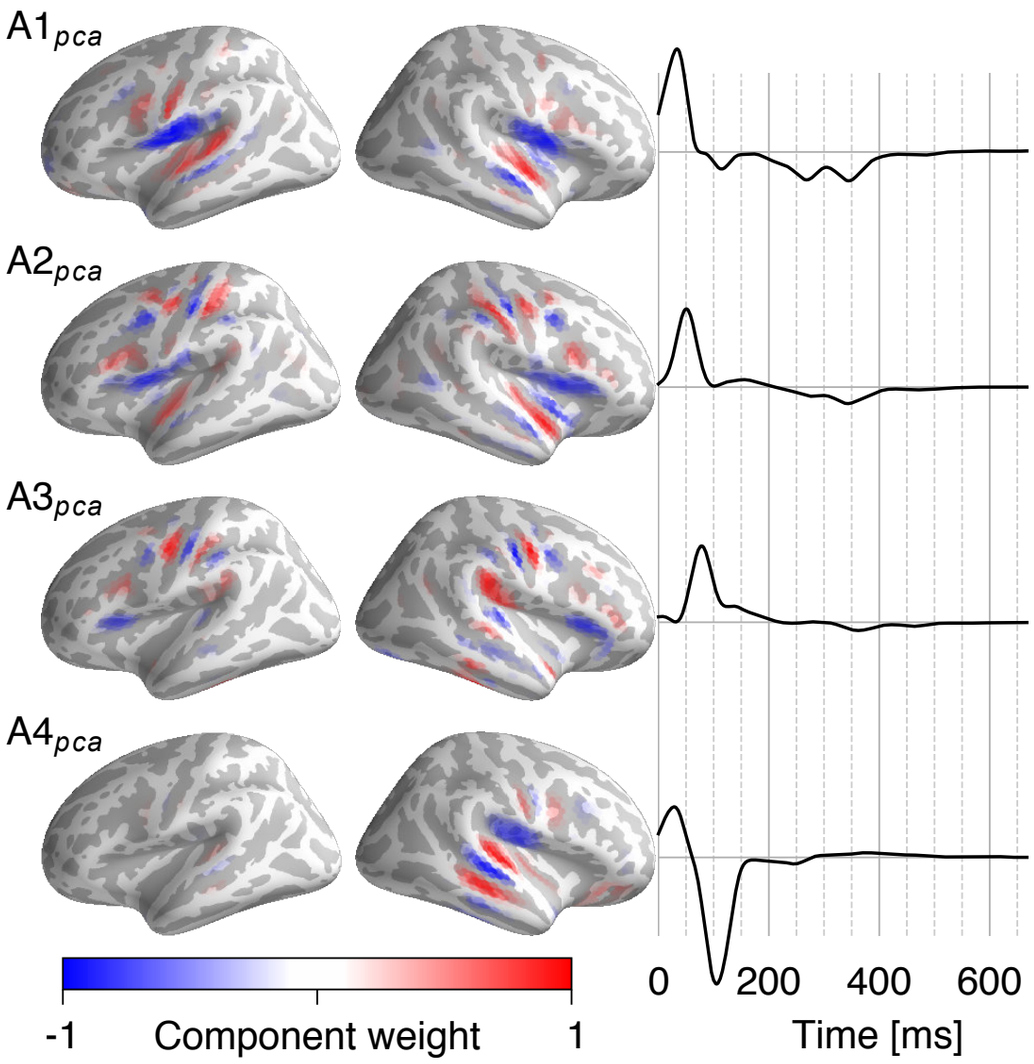
Brodbeck et al., NeuroImage (2017)
Brodbeck et al., Acta Acust united Ac (2018)

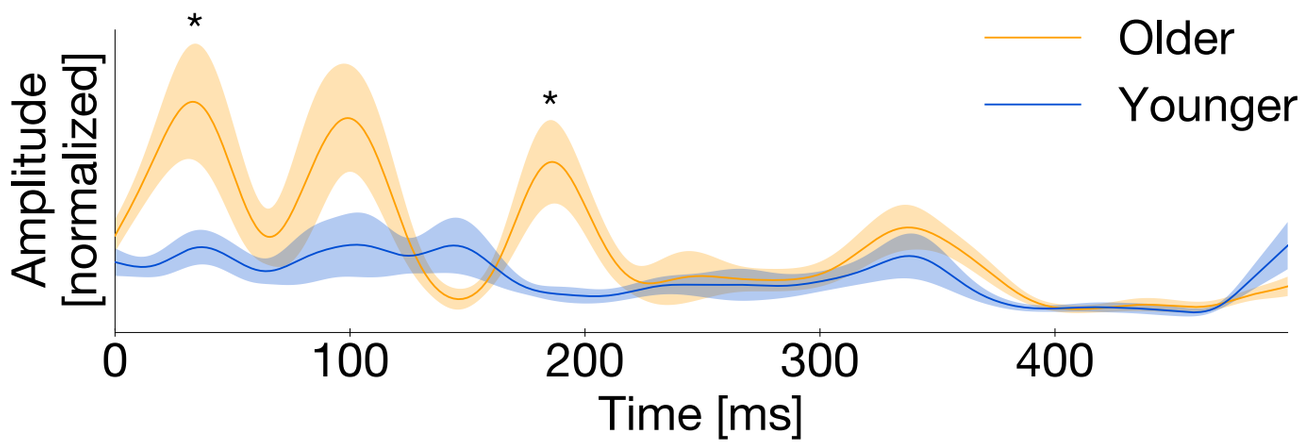


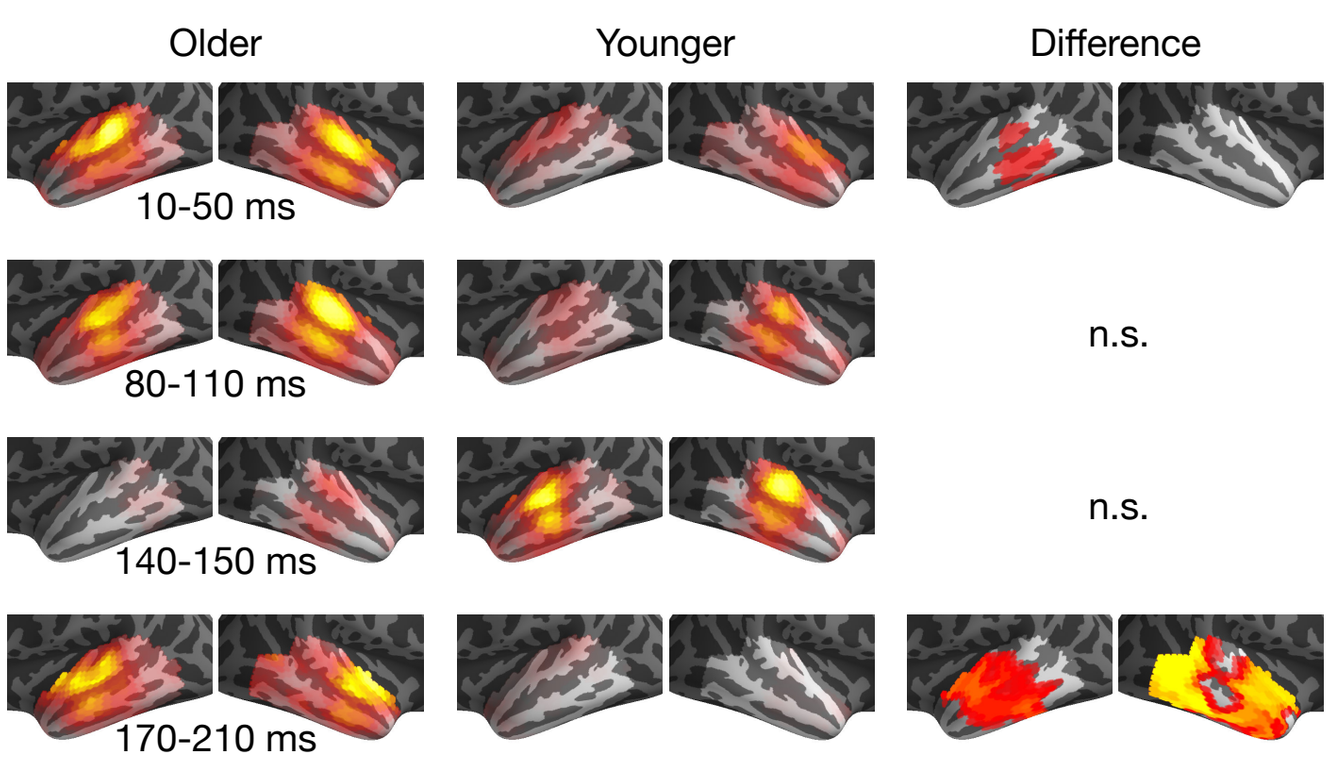


Das et al., Asilomar (2018)

Spatiotemporal Distribution of Neural Currents







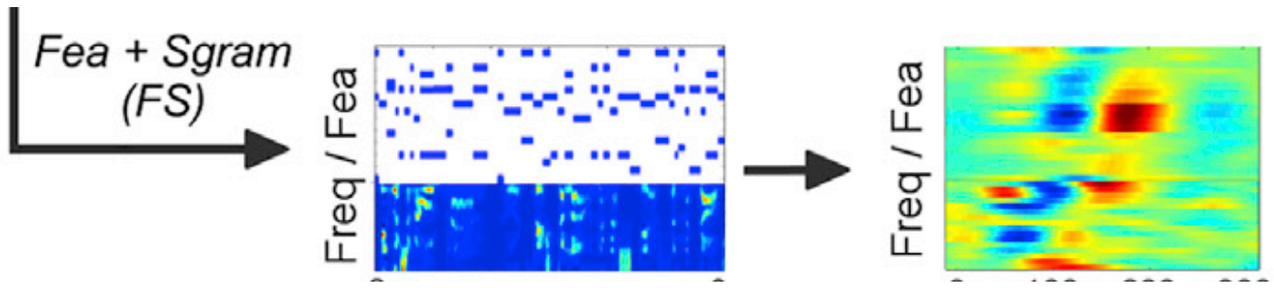
Brodbeck et al., NeuroImage (2017)
Brodbeck et al., Acta Acust united Ac (2018)

Outline

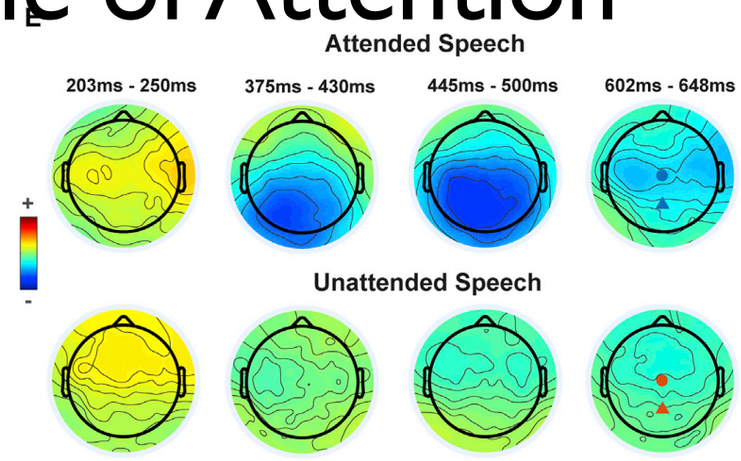
- Background & motivation
  - Neural responses in time
  - Neural responses in time & space
  - Representations: from Acoustic to Linguistic
- Spatiotemporal representation transformation from Acoustic to Lexical

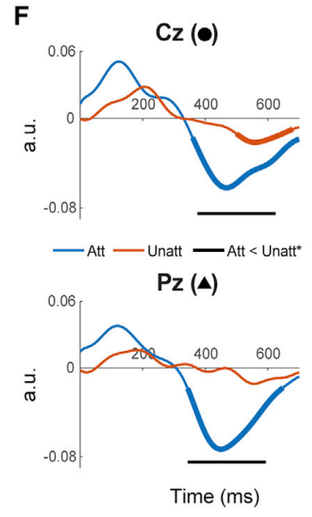
Acoustic Speech to Linguistic Speech

- Phonemes
  - Mesgarani et al., Science (2014)



  - Di Liberto et al., Curr Biol (2015)
- Semantic Information & Role of Attention
  - Broderick et al., Curr Biol (2018)





- But see also Daube et al., Curr Biol (2019)

Outline

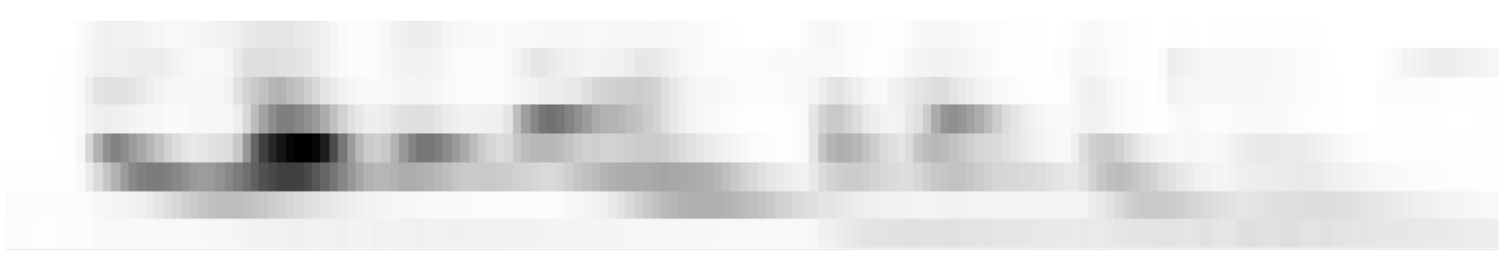
- Background & motivation
  - Neural responses in time
  - Neural responses in time & space
  - Representations: from Acoustic to Linguistic
- Spatiotemporal representation transformation from Acoustic to Lexical

Brodbeck et al., Curr Biol (2018)

Acoustic to Lexical Speech Processing

his noble mind forgot the cakes

Acoustic Envelope (8 bands)



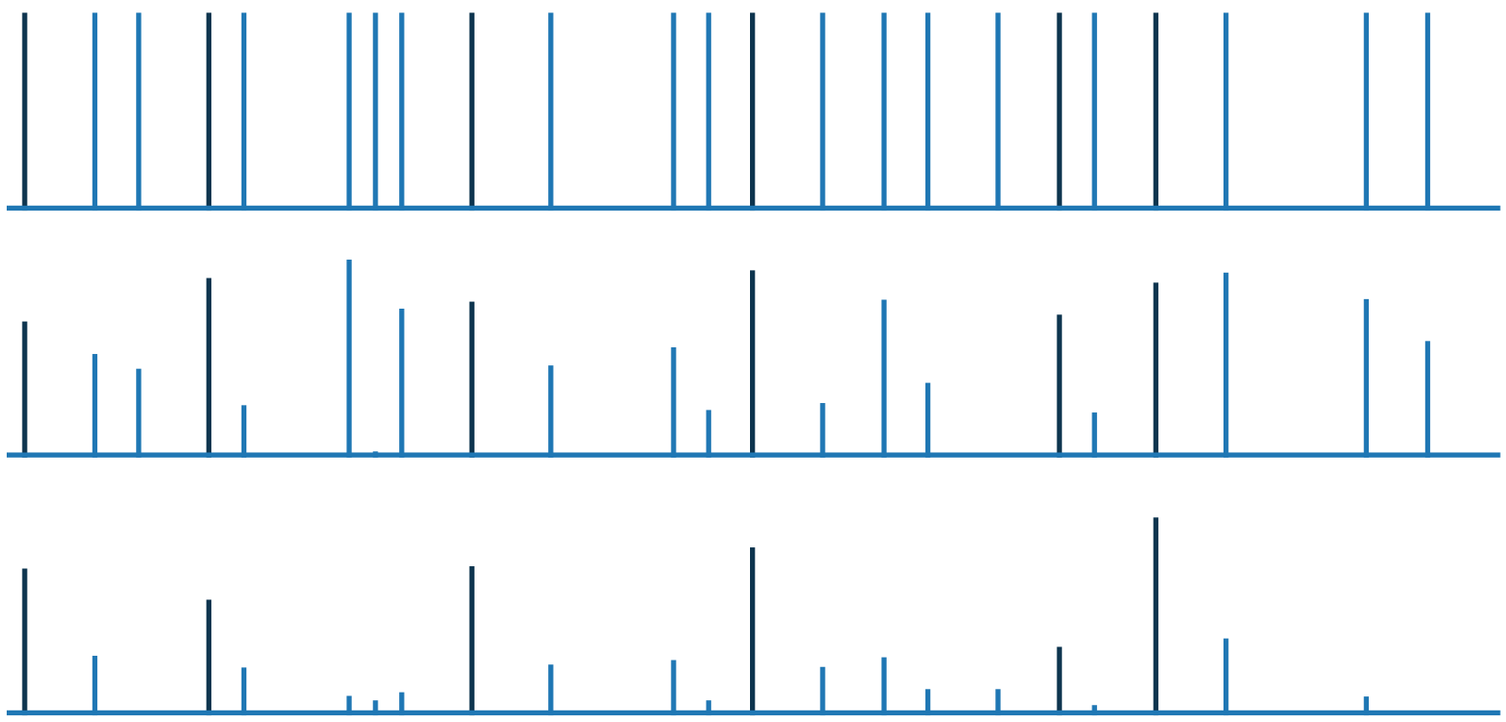
Acoustic envelope (8 bands)

Acoustic Onset (8 bands)

Acoustic to Lexical Speech Processing

his noble mind forgot the cakes

h ɪ z noʊ bəl m aɪ nd f ɝ g ɑ t ðə k eɪ k s



Acoustic to Lexical Speech Processing

his noble mind forgot the cakes

h ɪ z noʊ bəl m aɪ nd f ɝ g ɑ t ðə k eɪ k s

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

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

Methods

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

Acoustic time-frequency representation: 8-band auditory spectrogram

Word frequencies: **SUBTLEX: 51 million words** movie subtitle database (stress info stripped)

Distributed MNE source estimates, restricted to temporal lobe (314 L, 313 R)

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

Model reduction: iteratively remove largest *p*-value (non-significant) variable

Word perception

Levels of representation

- Phonemes: based on acoustic properties, related acoustic patterns
- Words: discrete linguistic entities (lexical item)

Word perception

Levels of representation

- Phonemes: based on acoustic properties, related acoustic patterns
- Words: discrete linguistic entities (lexical item)

Cohort model of lexical processing (Marslen-Wilson, 1987)

- The cohort is a set of activated words
- The first phoneme activates all words starting with that phoneme
- Each subsequent phoneme is used to narrow down the cohort
- **Separable from acoustics**

Word perception

Levels of representation

- Phonemes: based on acoustic properties, related acoustic patterns
- Words: discrete linguistic entities (lexical item)

Cohort model of lexical processing (Marslen-Wilson, 1987)

- The cohort is a set of activated words
- The first phoneme activates all words starting with that phoneme
- Each subsequent phoneme is used to narrow down the cohort
- **Separable from acoustics**

Influence of distribution frequencies?

- Some words are heard more frequently than others

"the", "cat", "chrysalis"

- How do we measure this?
  - SUBTLEX: Corpus with subtitles from movies and TV shows
- Does the brain take this into account?
  - Lexical decision experiments

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

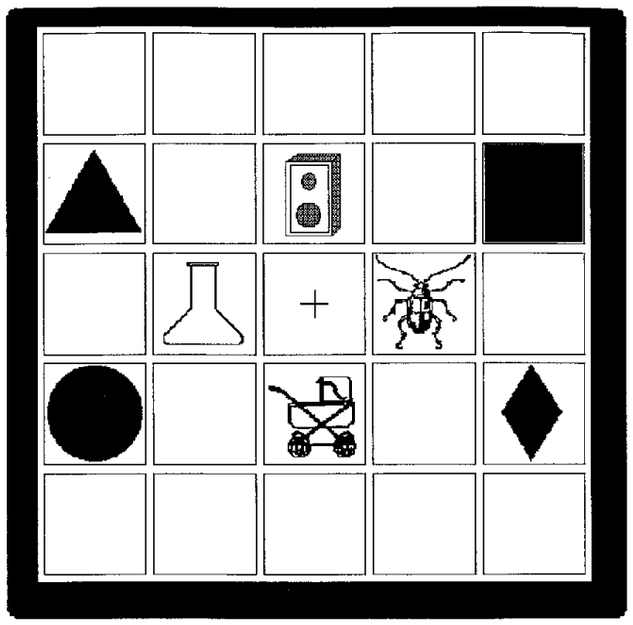
Cohort model

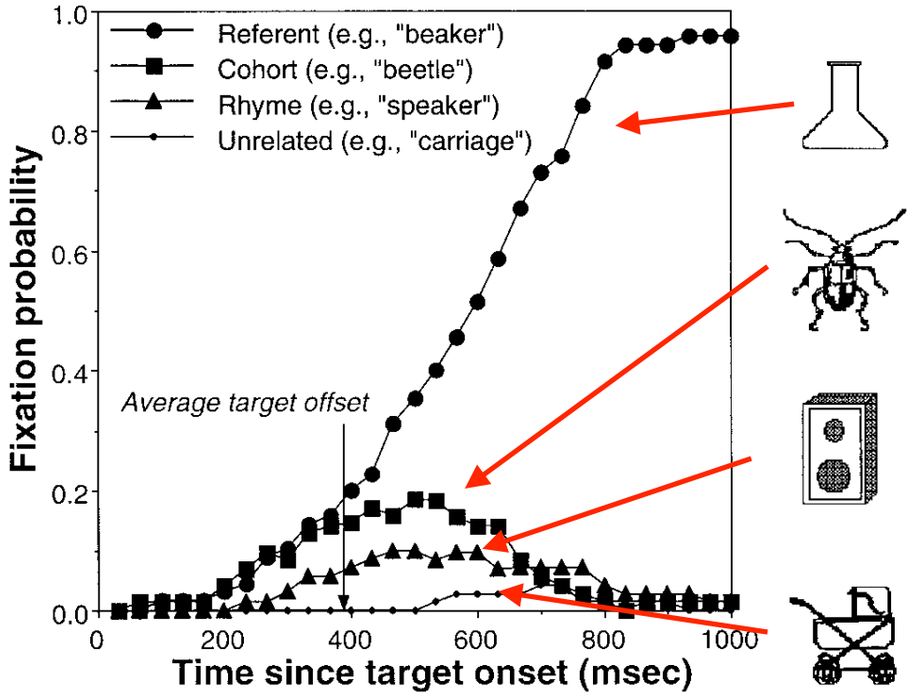
/keɪk…/

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 beaker. Now put it above the diamond."

(Allopenna, Magnuson, & Tanenhaus, 1998)

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)

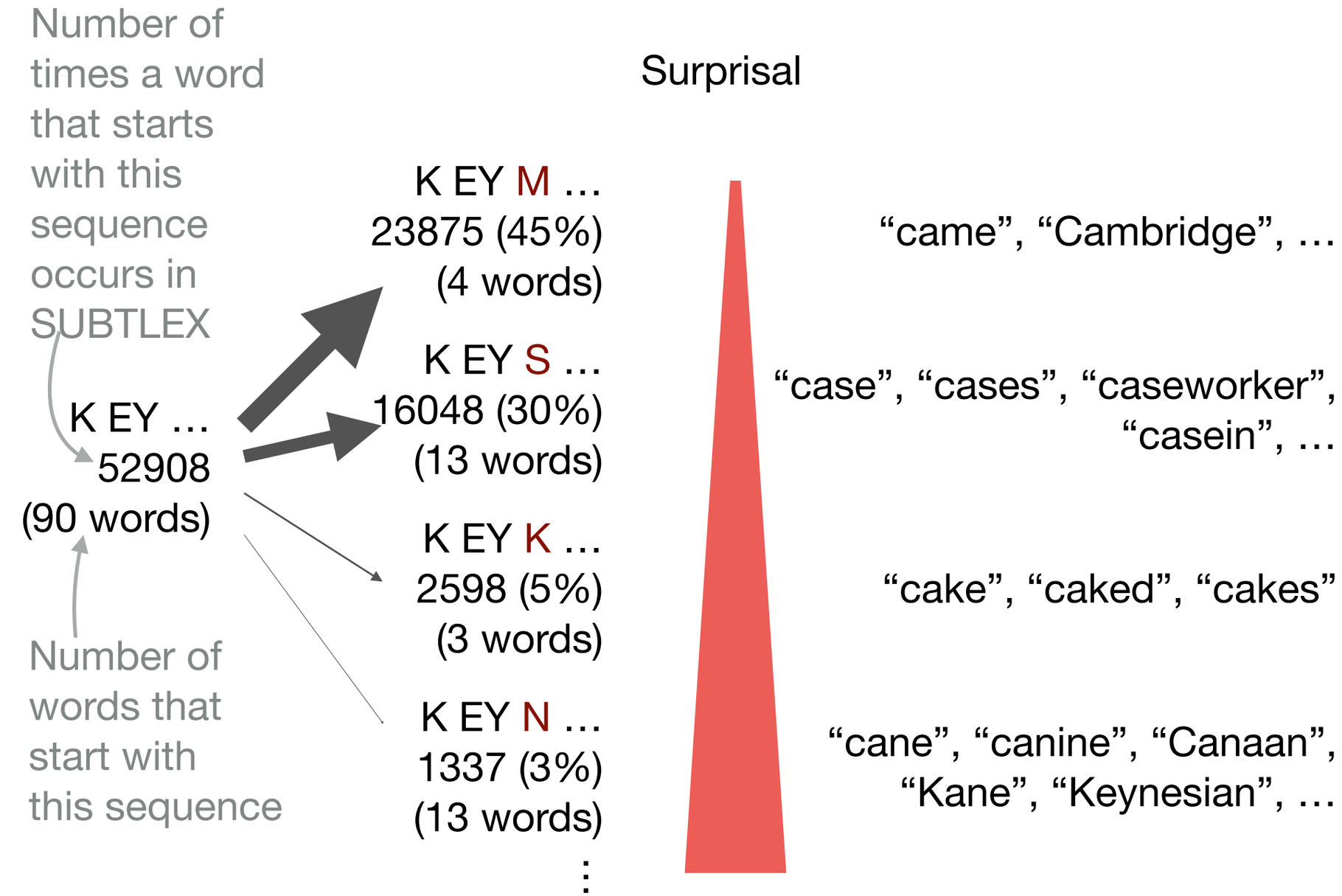
“cake”, “caked”, “cakes”

K EY N … 1337 (3%) (13 words)

“cane”, “canine”, “Canaan”, “Kane”, “Keynesian”, …

⋮

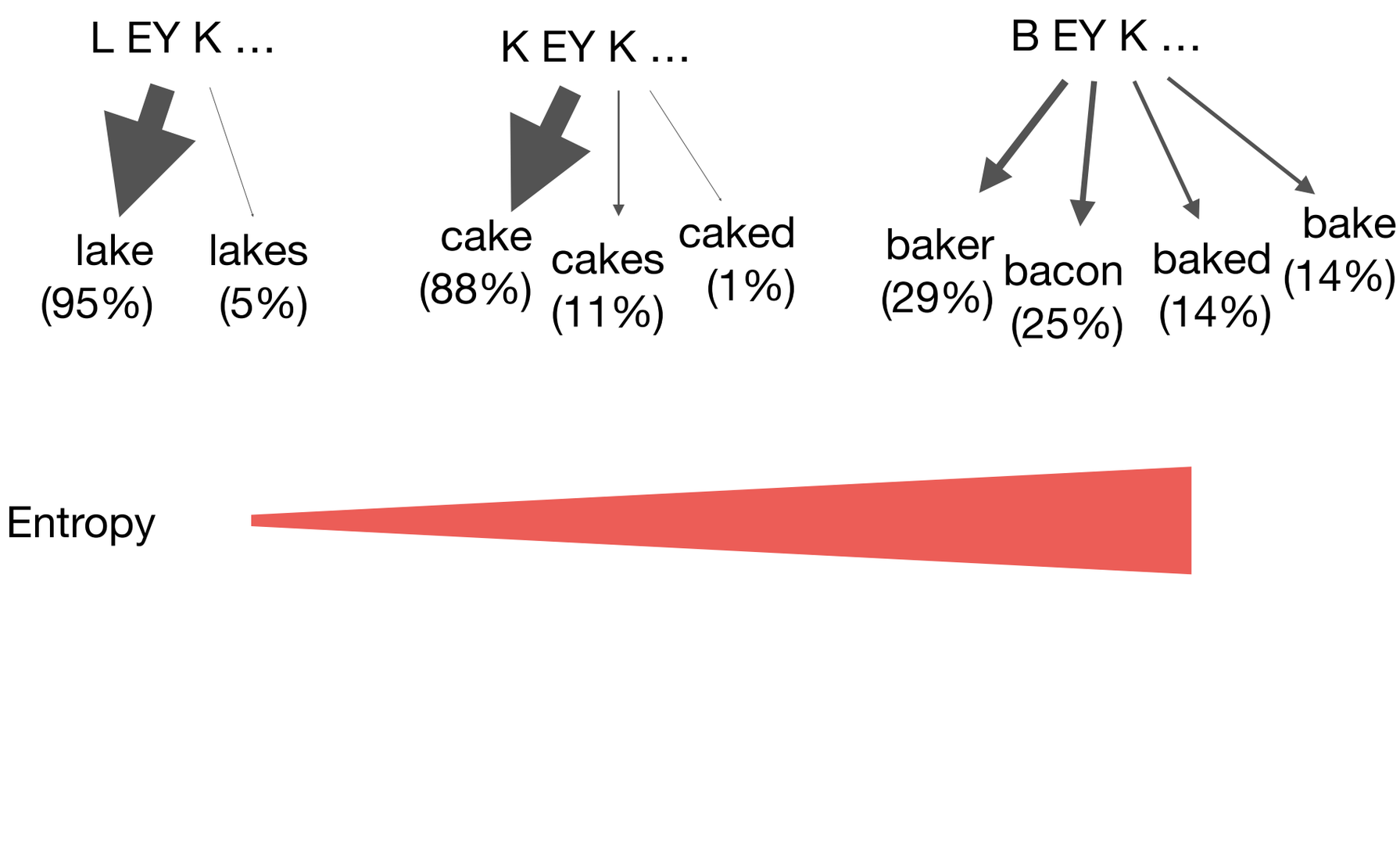
Surprisal


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)
⋮
“came”, “Cambridge”, …
“case”, “cases”, “caseworker”, “casein”, …
“cake”, “caked”, “cakes”
“cane”, “canine”, “Canaan”, “Kane”, “Keynesian”, …

Entropy

Cohort entropy

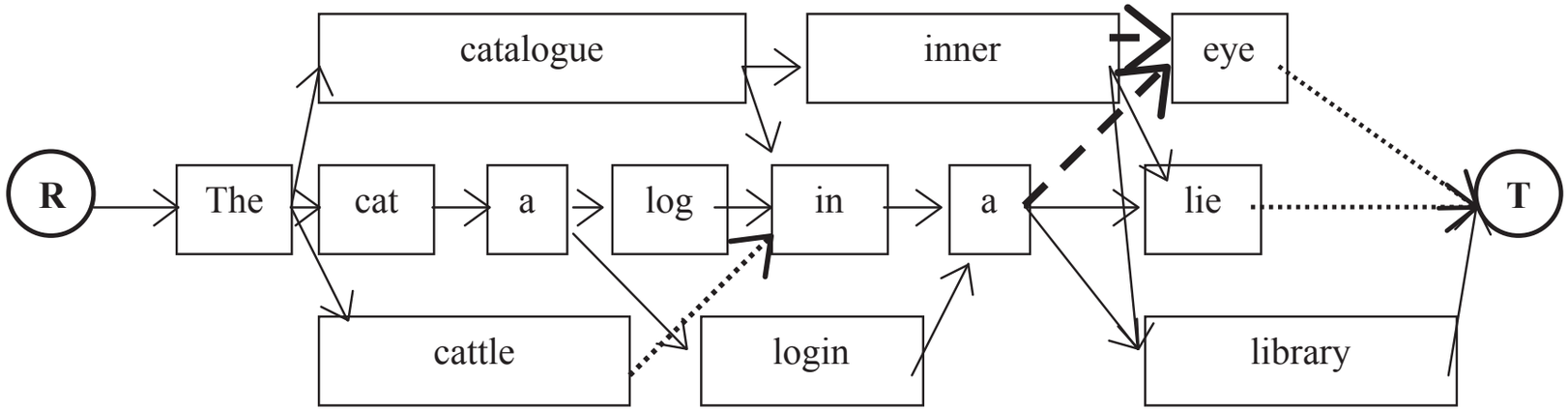
- How unpredictable is the current word?



Word onsets

Do we…

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

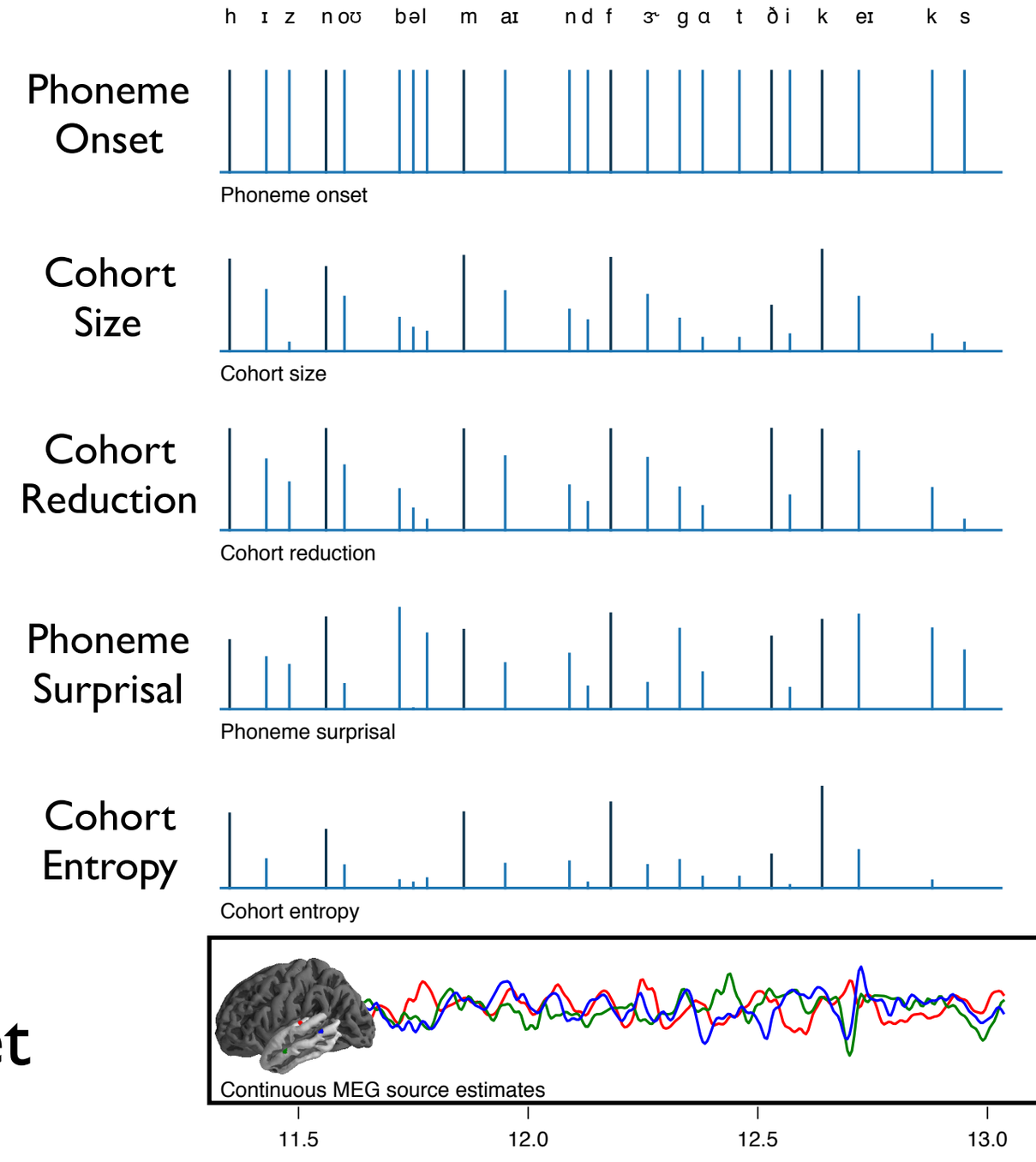


“The catalogue in a library”

(Norris & McQueen, 2008)

Acoustic to Lexical Speech Processing

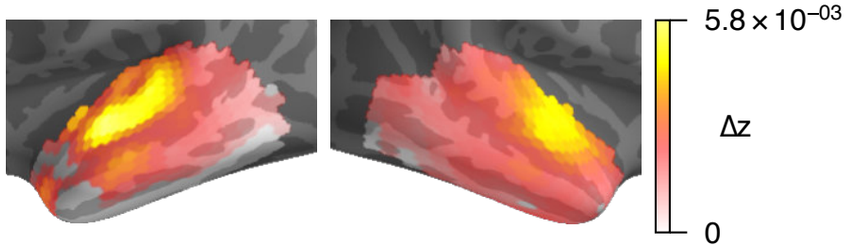
- 16 acoustic
- 8 lexical
  - 4 medial (+ 4 initial)
- 1 word onset
- 1 (non-initial) phoneme onset



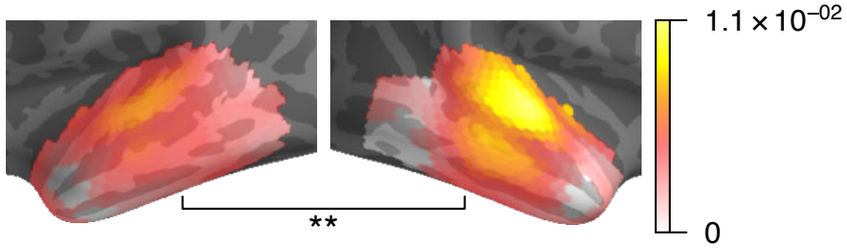
Brodbeck et al., Curr Biol (2018)

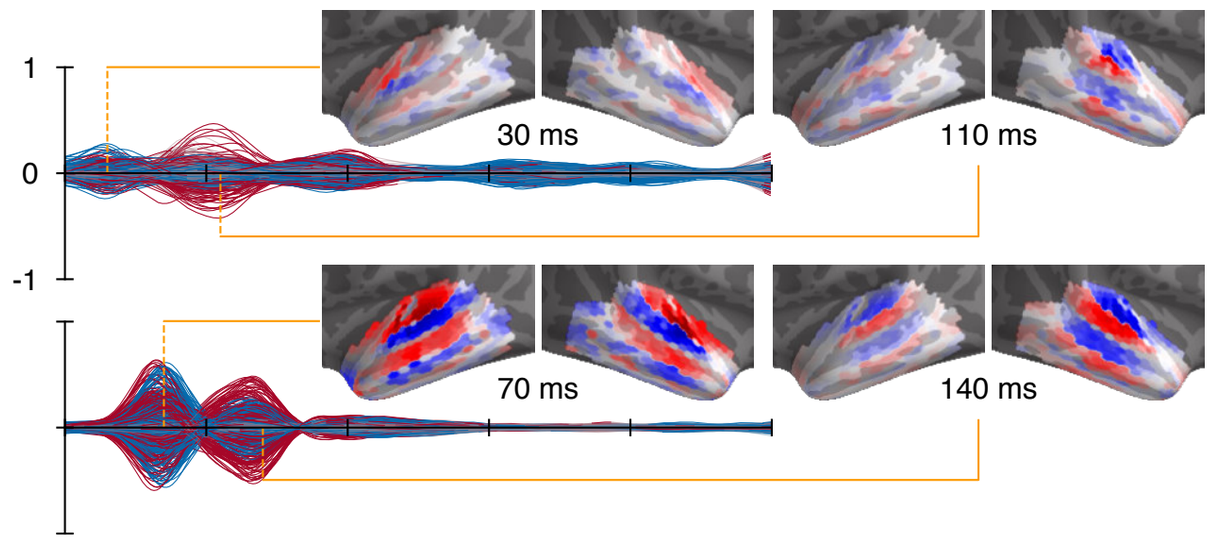
Acoustic Results

Acoustic Envelope



Acoustic Onset





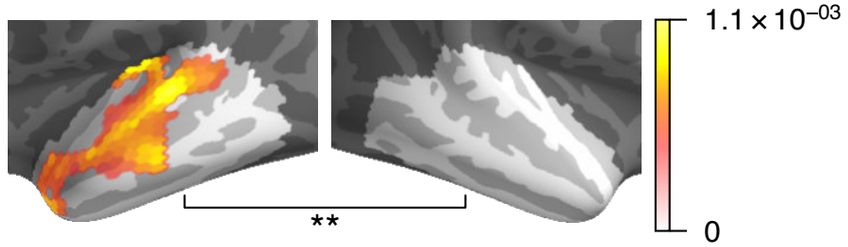
cf. Hamilton et al., 2018,
Daube et al., Curr Biol (2019)

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

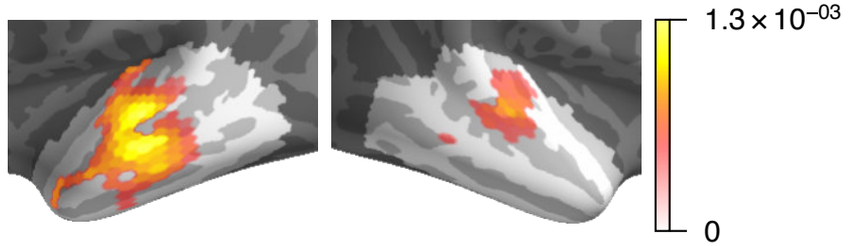
Brodbeck et al., Curr Biol (2018)

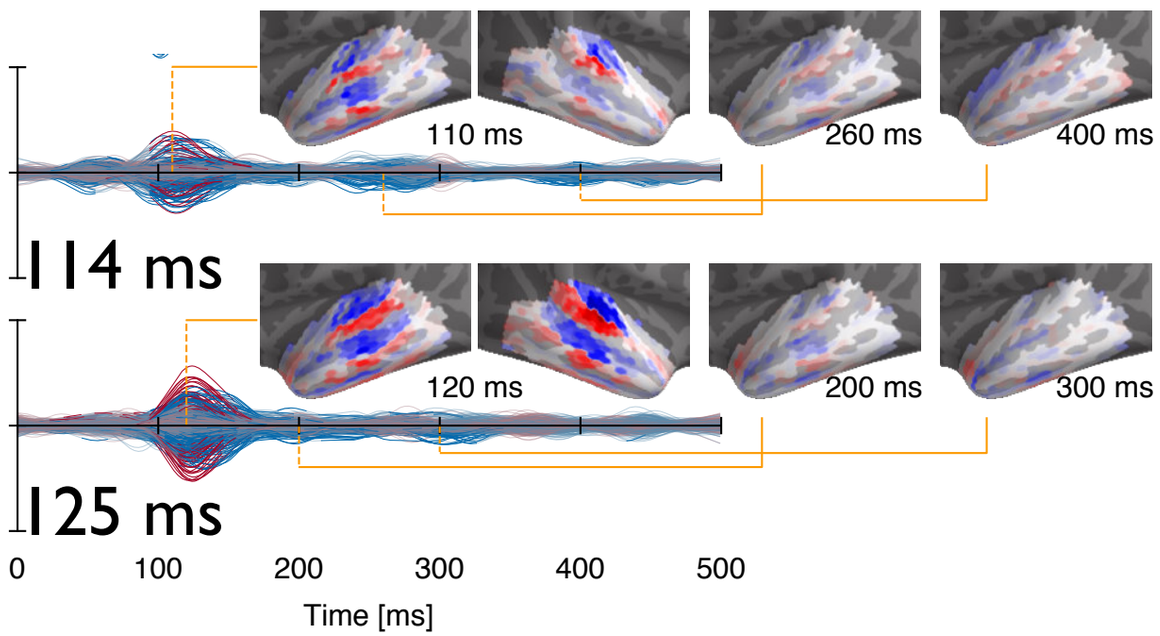
Lexical Results

Phoneme Surprisal



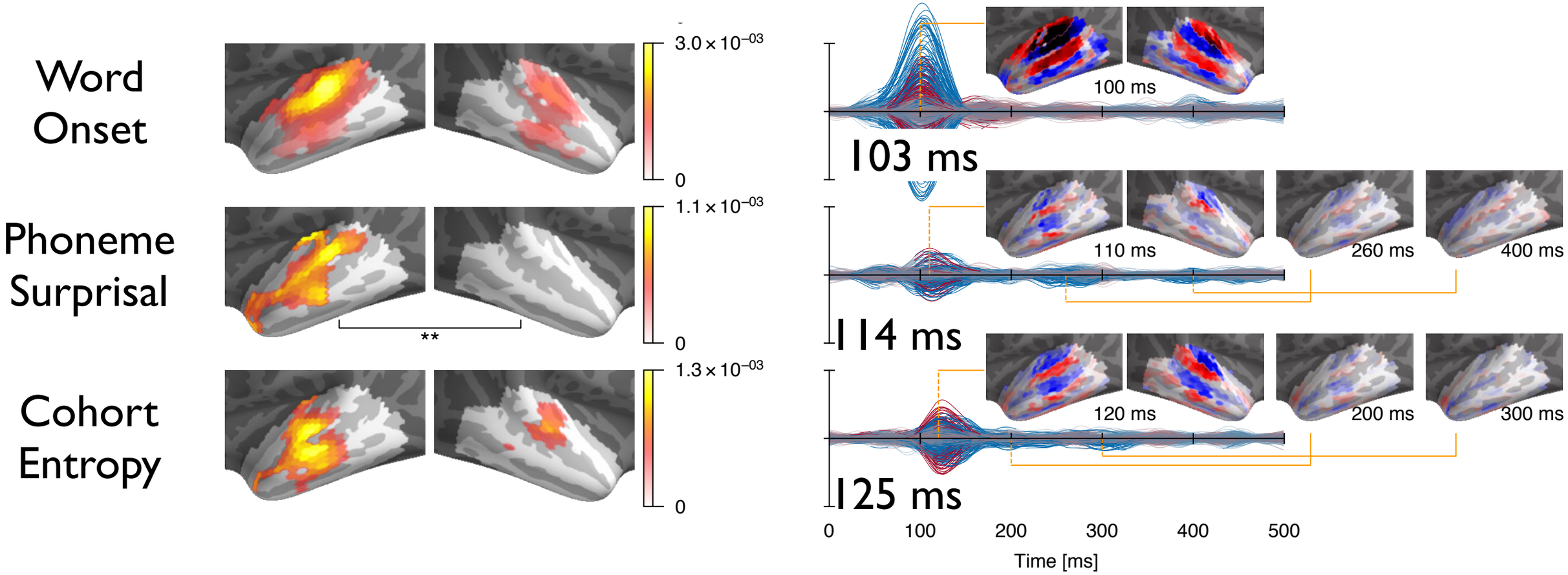
Cohort Entropy





Brodbeck et al., Curr Biol (2018)

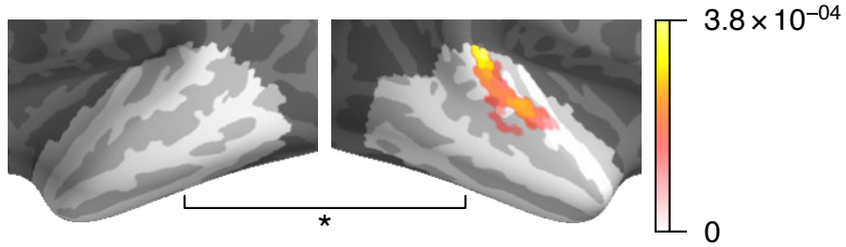
Lexical Results



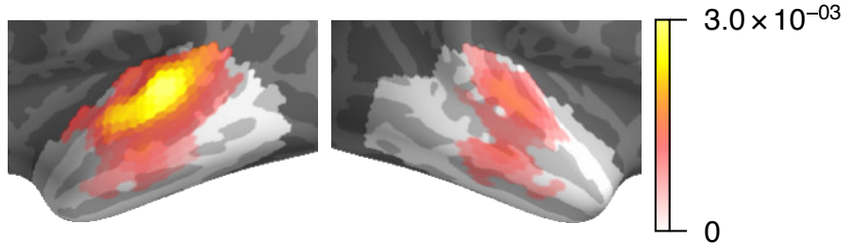
Brodbeck et al., Curr Biol (2018)

Lexical Results

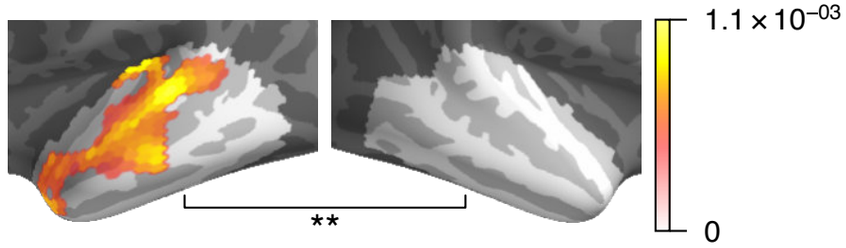
Phoneme Onset



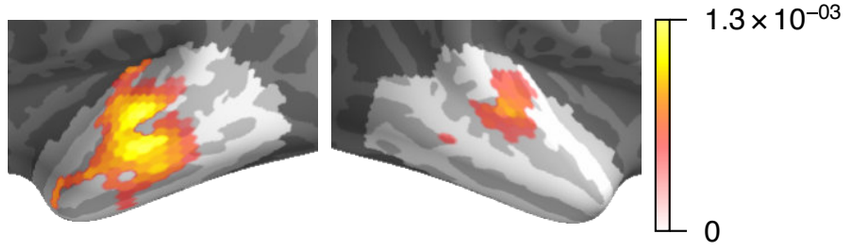
Word Onset

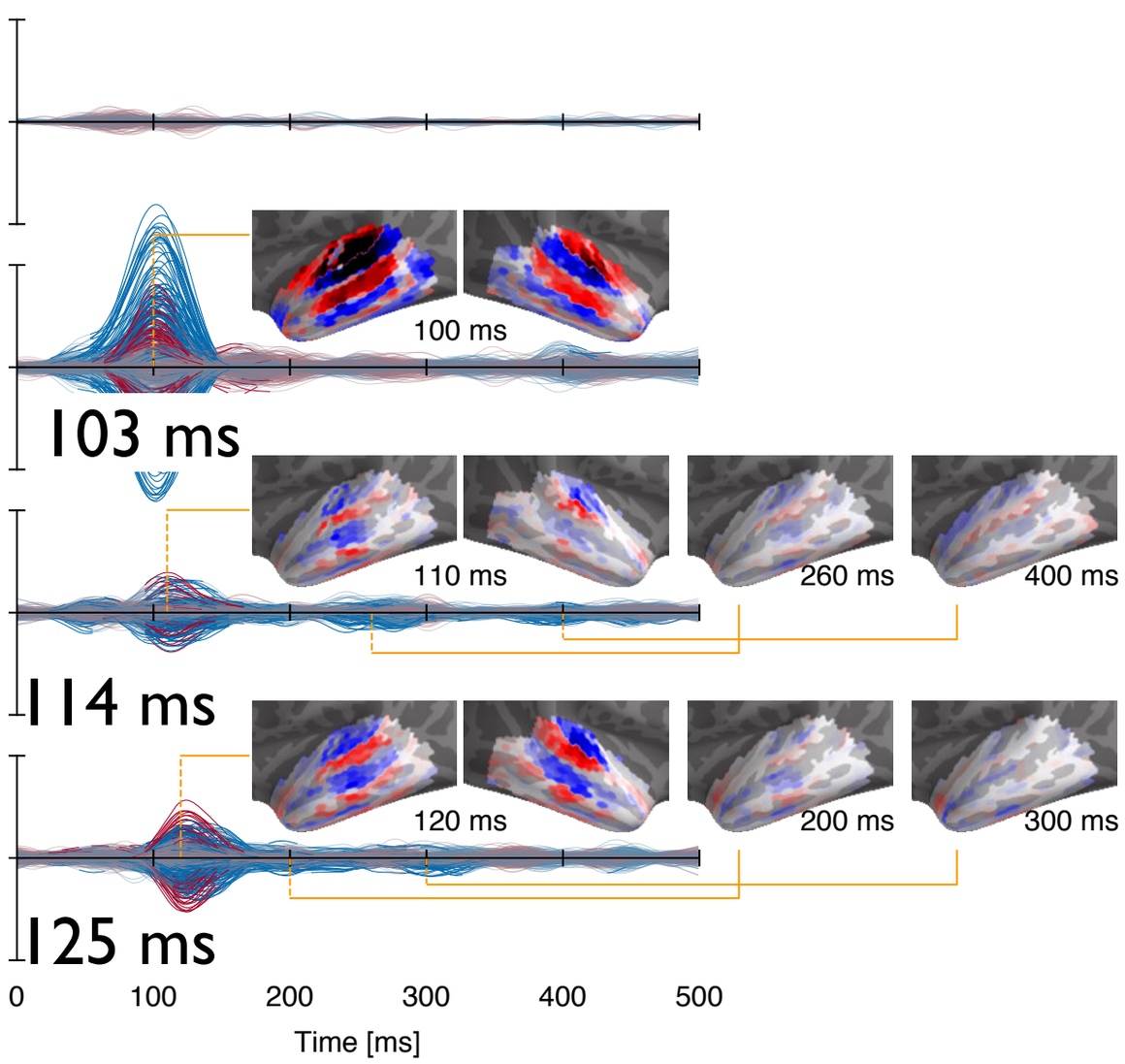


Phoneme Surprisal



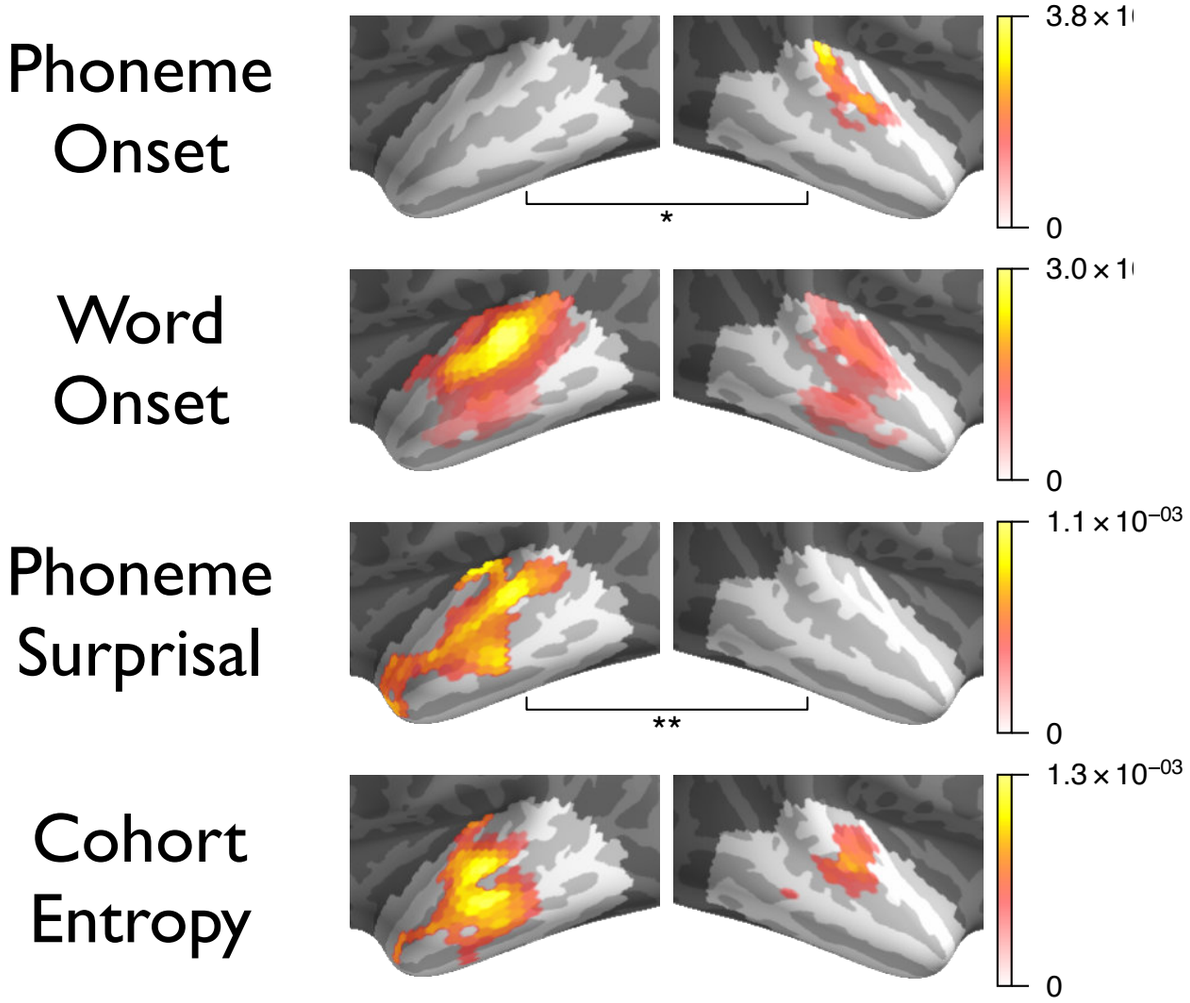
Cohort Entropy



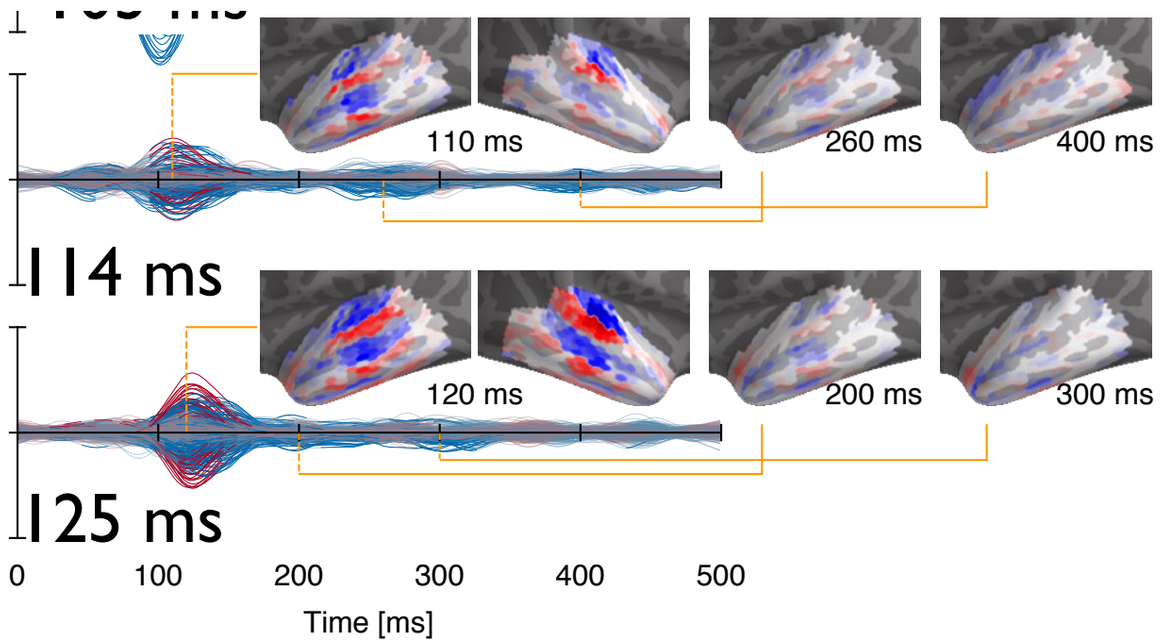


Brodbeck et al., Curr Biol (2018)

Lexical Results



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



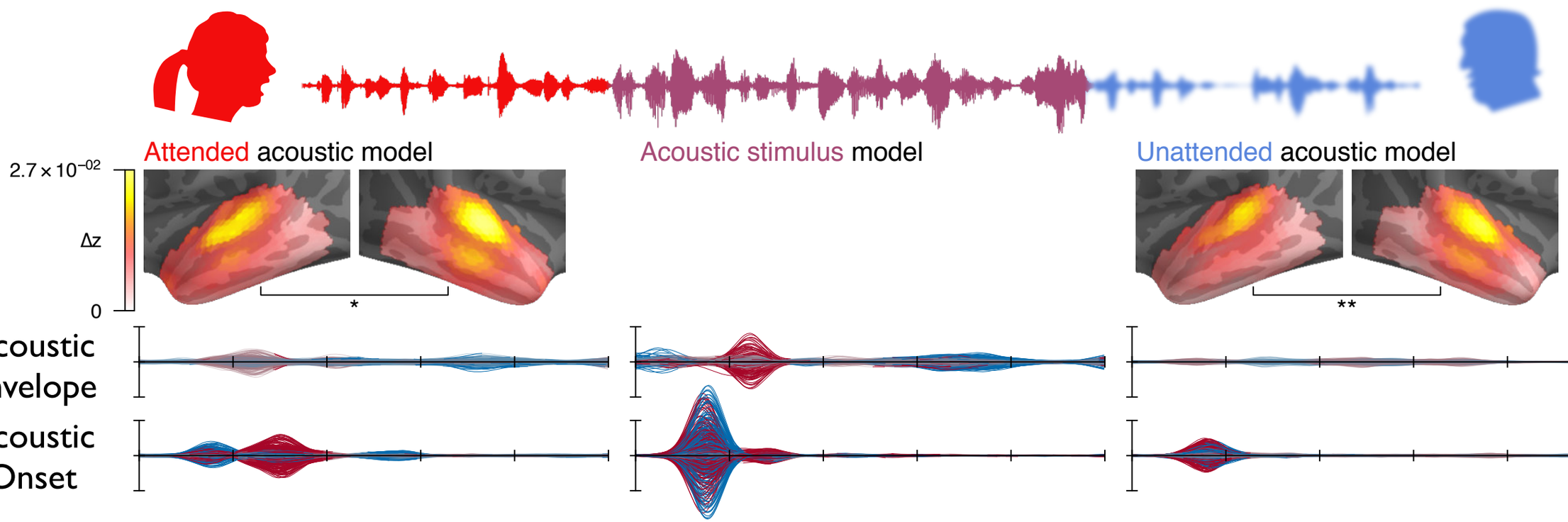
Brodbeck et al., Curr Biol (2018)

Cocktail Party Listening

Methods

- 16 one-minute-long segments constructed from the same passages from *A Child's History of England* by Dickens
- Two competing speakers, male + female, equal loudness
- Instructions: Attend to one, ignore the other, counter-balanced
- After each segment, answer a question about the content of the attended stimulus

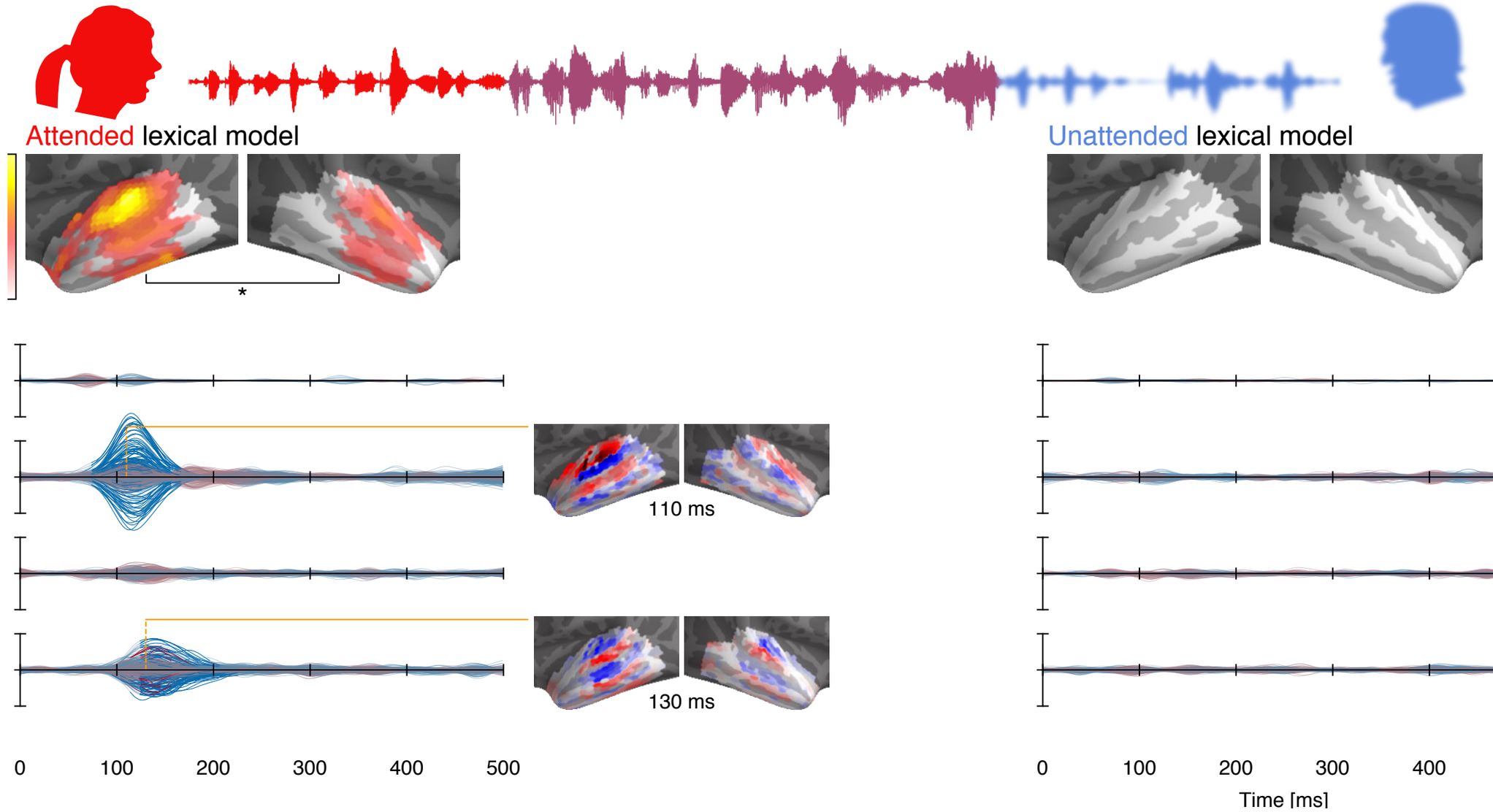
Acoustic Attention



- Onset Representation Dominates
- Attended Dominates Later

Brodbeck et al., Curr Biol (2018)

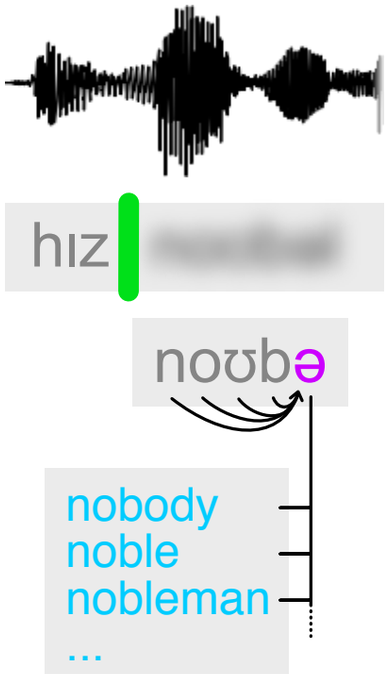
Lexical Attention

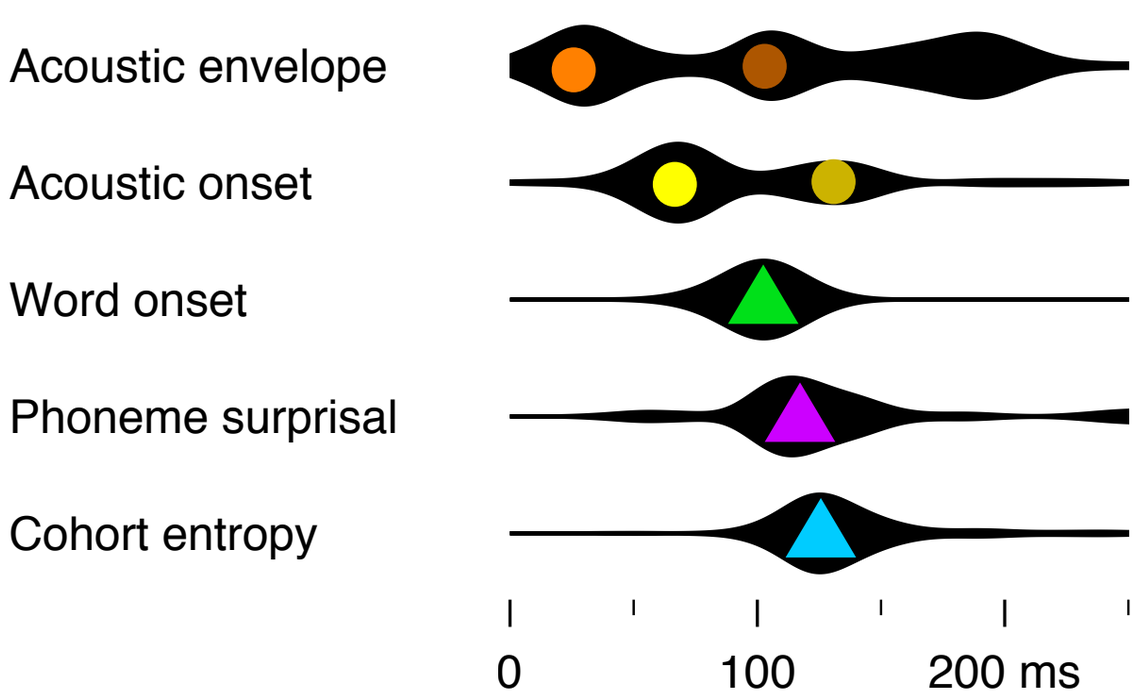


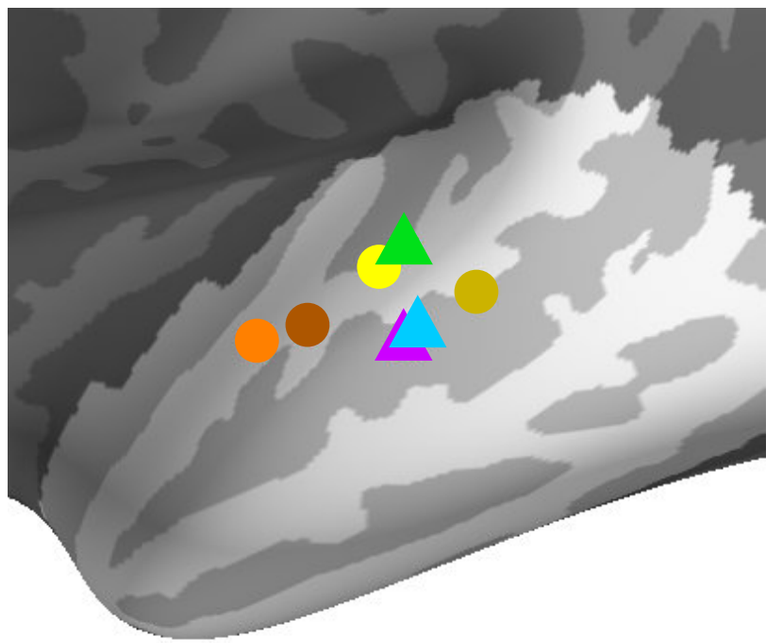
- Only attended speech processed lexically
- Lexical processing slowed by ~15 ms

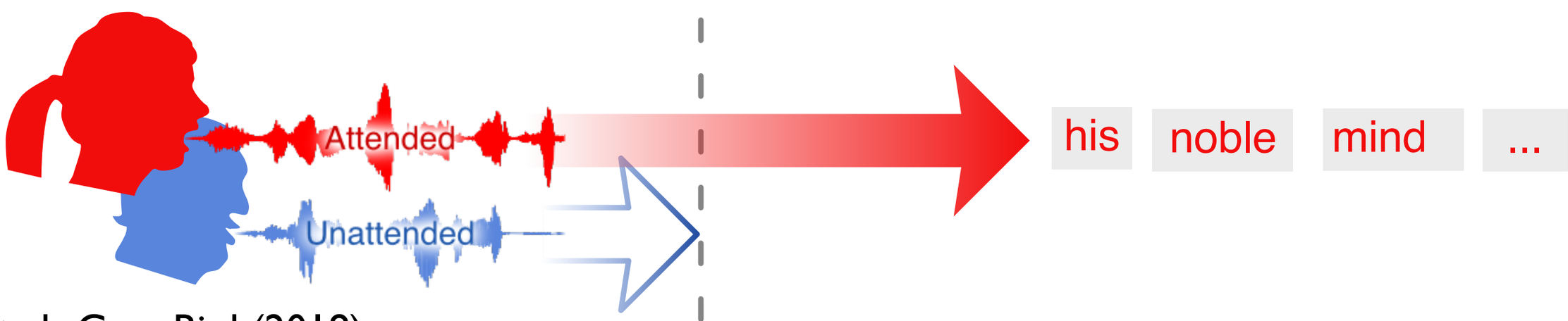
Brodbeck et al., Curr Biol (2018)

Acoustic to Lexical Speech Processing









Brodbeck et al., Curr Biol (2018)

Summary I

- Acoustic processing—Envelope vs. Onset
  - Allowed to compete against each other
  - Onset explains more response variance
    - Strongly bilateral with right-bias for onset
    - Similar latencies, but possibly different neural populations
- Evidence for rapid transformation from acoustic to lexical representations

Summary II

- Fast Lexical Phoneme-based processing
  - Surprisal (114 ms), local measure of phoneme prediction error (predictive coding?)
  - Cohort entropy (125 ms), global measure of lexical competition across cohort
  - Left hemisphere dominant
  - Strongly attention-dependent (bottleneck?)

Summary III

- Low latencies
  - Coarticulation; prediction using context
  - ~15 ms extra delay from interfering speech
- Word Onset
  - Early (103 ms) detection of lexical boundaries
  - Robust, also attention-dependent
- Caveats
  - Time-locked responses only
  - Task/attentional state somewhat intense

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

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
Julian Jenkins
Pirazh Khorramshahi
Huan Luo
Mahshid Najafi
Krishna Puvvada
Jonas Vanthornhout
Ben Walsh
Yadong Wang
Juanjuan Xiang
Jiachen Zhuo

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