Neural Tracking of Linguistic Speech Representations

Marlies Gillis[1], Jonas Vanthornhout[1], Jonathan Z. Simon[2], Tom Francart*[1] & Christian Brodbeck*[3]

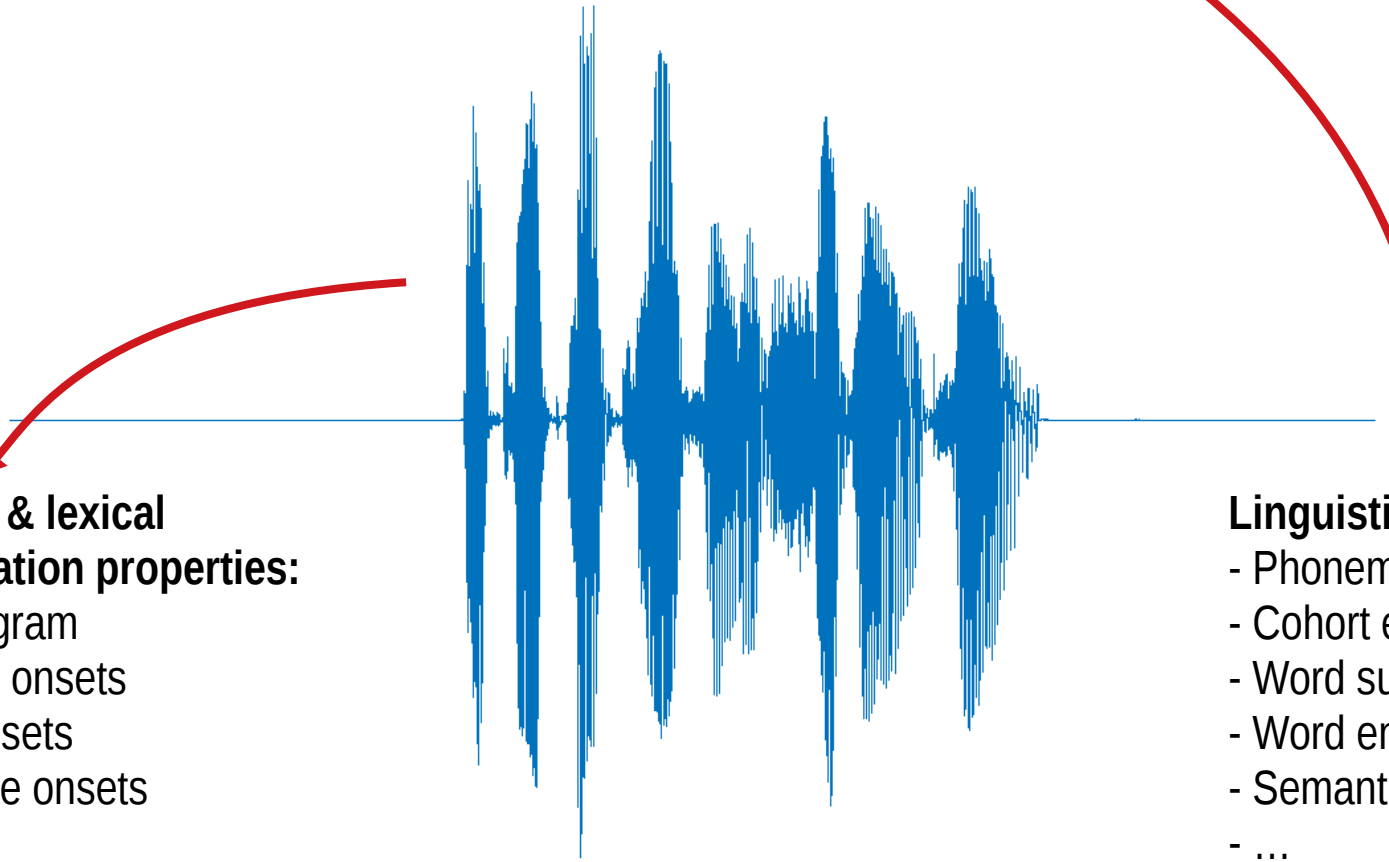
*shared last authorship

[1] KU Leuven, Department of Neurosciences, ExpORL, Belgium
[2] Electrical & Computer Engineering; Institute for Systems Research, University of Maryland, College Park, USA
[3] Department of Psychological Sciences, University of Connecticut, USA

Figure from Gordon Johnson

Neural Tracking

“I take my coffee with sugar and cream.”

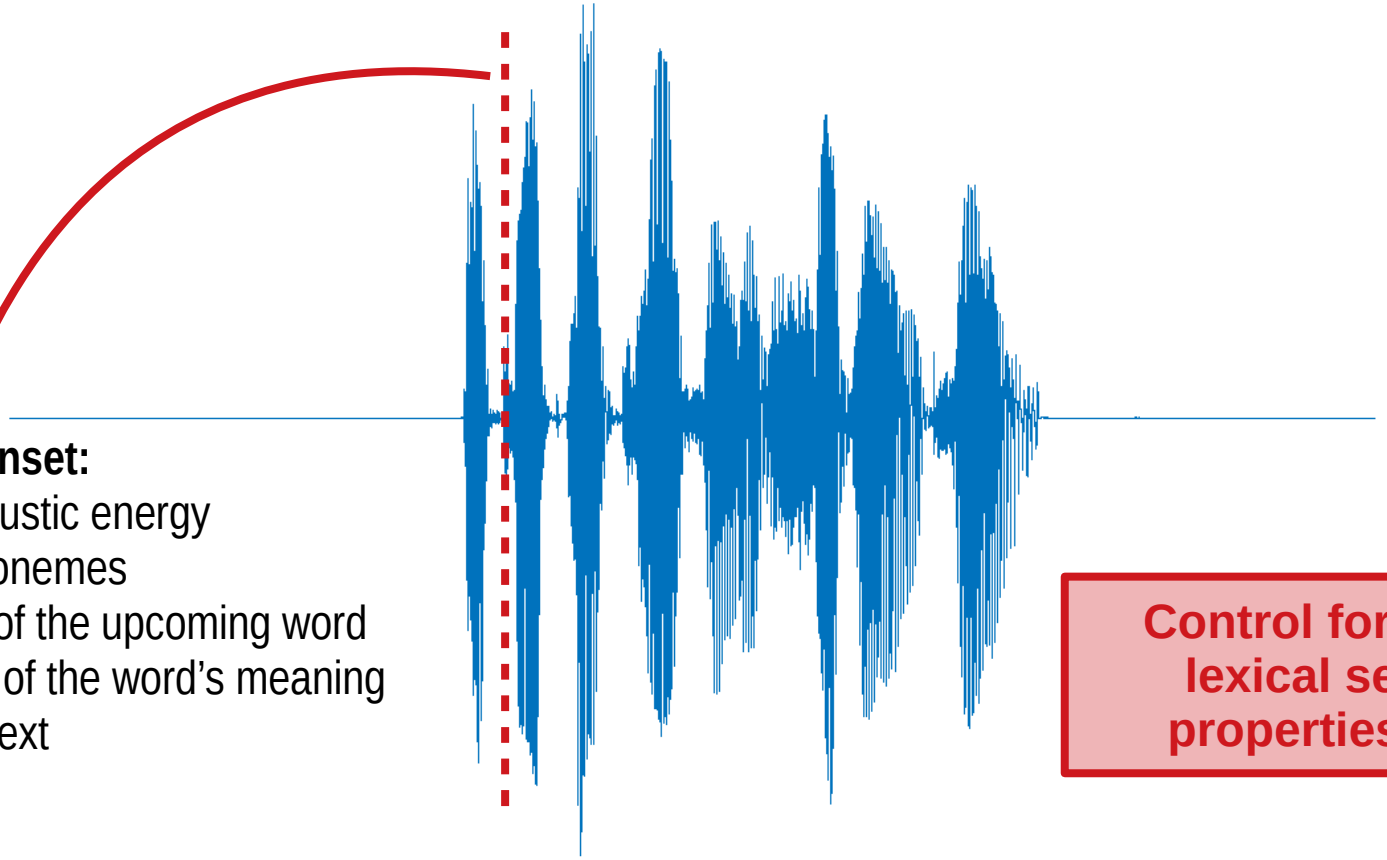


Acoustic & lexical segmentation properties:
- Spectrogram
- Acoustic onsets
- Word onsets
- Phoneme onsets

Linguistic properties:
- Phoneme surprisal
- Cohort entropy
- Word surprisal
- Word entropy
- Semantic dissimilarity
- …

Did he/she understand the content of the presented speech?

“I take my coffee with sugar and cream.”



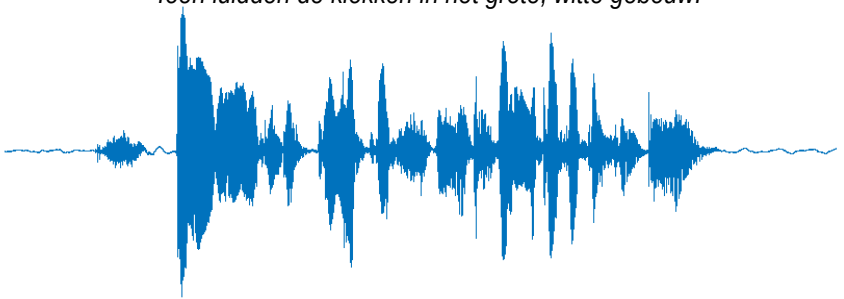
At a word onset:
- A lot of acoustic energy
- Identify phonemes
- Prediction of the upcoming word
- Integration of the word’s meaning into the context
- ...

Control for acoustic and lexical segmentation properties is required.

A lot of processing to understand language which are modelled with **correlated speech representations**

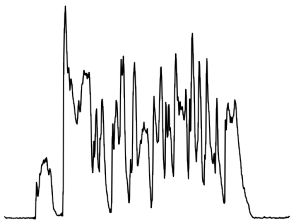
“Then the bells rang in the great, white building.”

“Toen luidden de klokken in het grote, witte gebouw.”

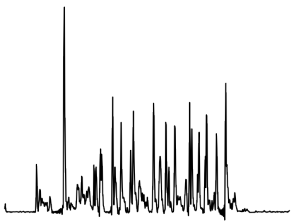


Acoustical Properties

Spectrogram



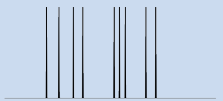
Acoustic onsets



Lexical Segmentation

Phoneme onsets

Word onsets



Content word onsets

Function word onsets

Linguistic Properties

Phoneme surprisal

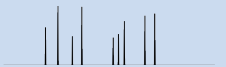
Cohort entropy

Word surprisal

Word entropy

Word precision

Word frequency



Semantic dissimilarity

At the phoneme level

Sound
/b/

Activation of word cohort
Activated lexical items

- Baby
- Back
- Barbecue
- Basilisk
- Badminton
- Back-flip
- Bakery
- Bandit
- Bathtub
- Belly
- Bed
- Besides
- Boomer
- Boost
- Bid
- ...

Phoneme surprisal
= - log(P(next phoneme | previous phonemes))

Cohort entropy
= Shannon entropy of phonemes compatible with the given input
→ "degree of competition"

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

Based on Brodbeck et al. (2018)
Probabilities derived from SUBTLEX-NL

At the word level

“But you know, happiness can be found even in the darkest of times, if one only remembers to turn on the light.”

— J.K. Rowling, Harry Potter and the Prisoner of Azkaban

Does not take sentence boundaries into account!

Word surprisal
= - log(P(next word | previous 4 words))

Word entropy
= Shannon entropy

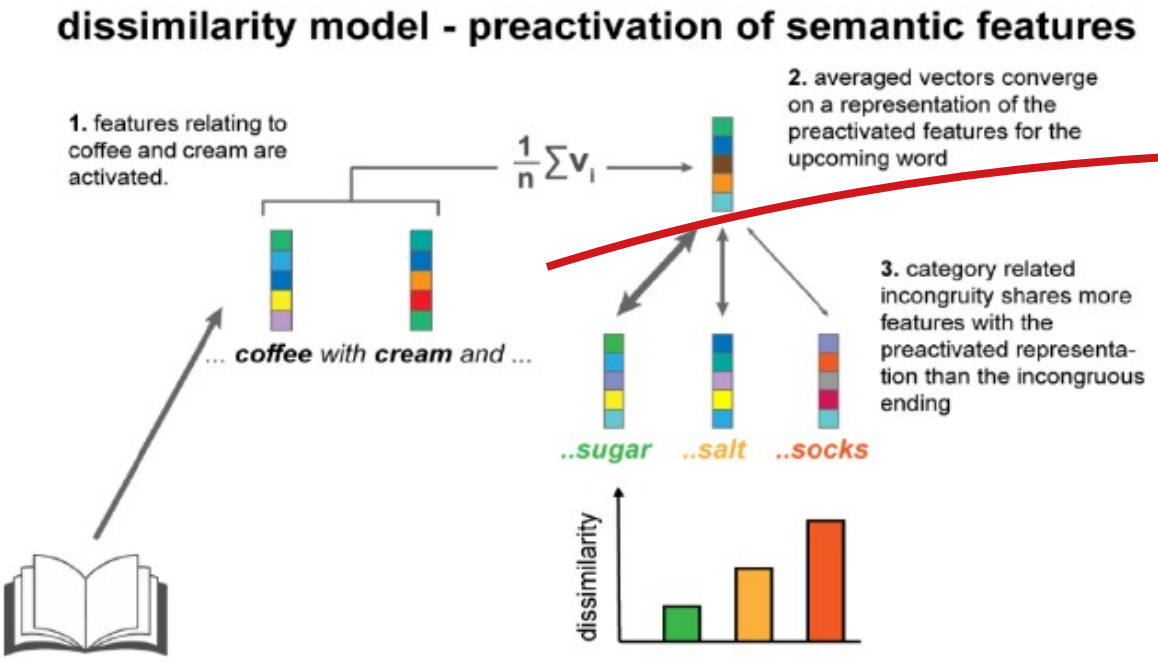
Word precision
= 1/entropy

Word frequency
= P(word)

Based on Weissbart et al. (2019)
Derived from a 5-gram model

At the contextual level

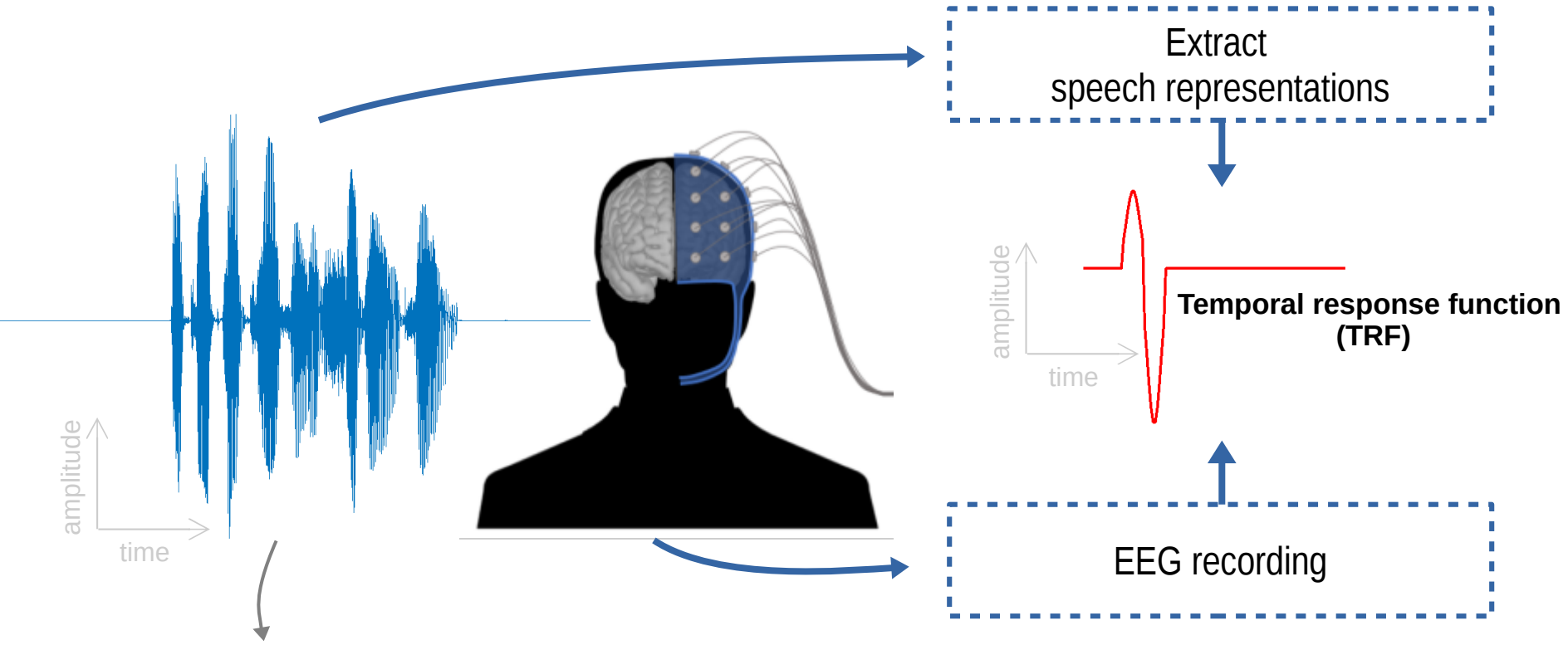
*Example sentence: I take my coffee with cream and **sugar** / **salt** / **socks***



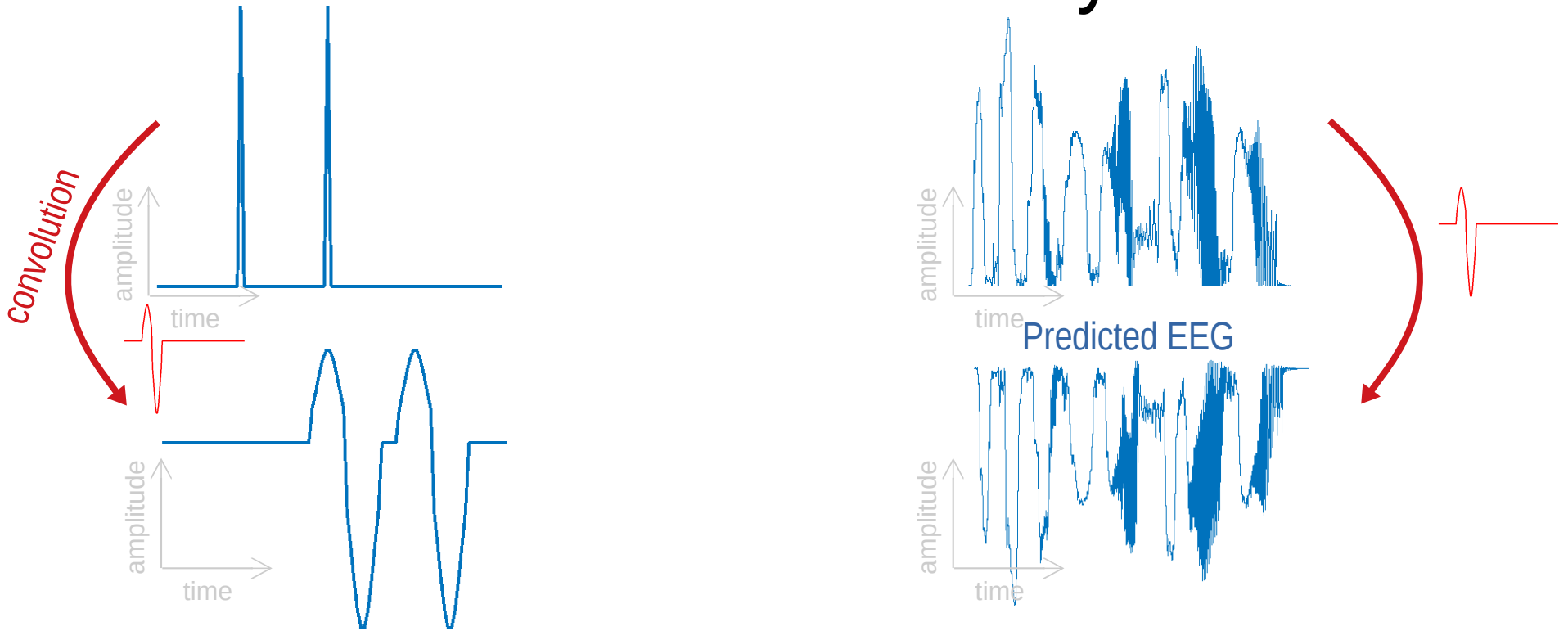
300 dimensional vector derived from Word2Vec

Based on Broderick et al. (2018)
Figure from Broderick et al. (2020)
Using Dutch version of Word2Vec

Temporal response function (TRF)



Prediction accuracy



Prediction accuracy = correlation between predicted and actual EEG

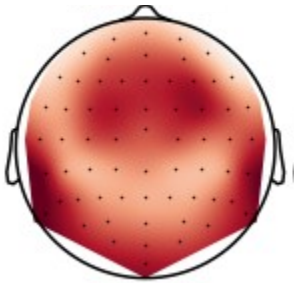
Methods

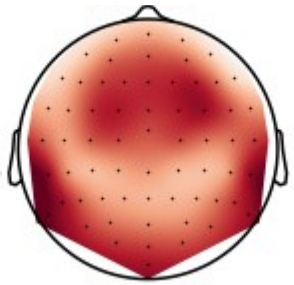
- Participant Details
  - 29 normal-hearing participants
  - Listening to 5 stories
    duration between 13 to 45 minutes
- Modelling
  - Forward modelling approach
  - Does the representation have an added value?
    → increase in prediction accuracy
  - Do they reflect separate stages of language processing?
    → looking at the TRF

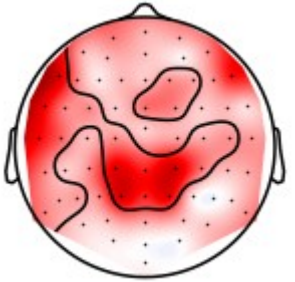
- Within one story:
  - Which linguistic representations are tracked?
  - Testing & training on one story
  - One story of 45 minutes

- Across Story:
  - Do these linguistic representations have an added value when applied across stories?
  - Training on above story of 45 minutes
  - Applied on stories or story parts of 15 minutes

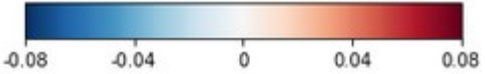
Does the representation have an added value?







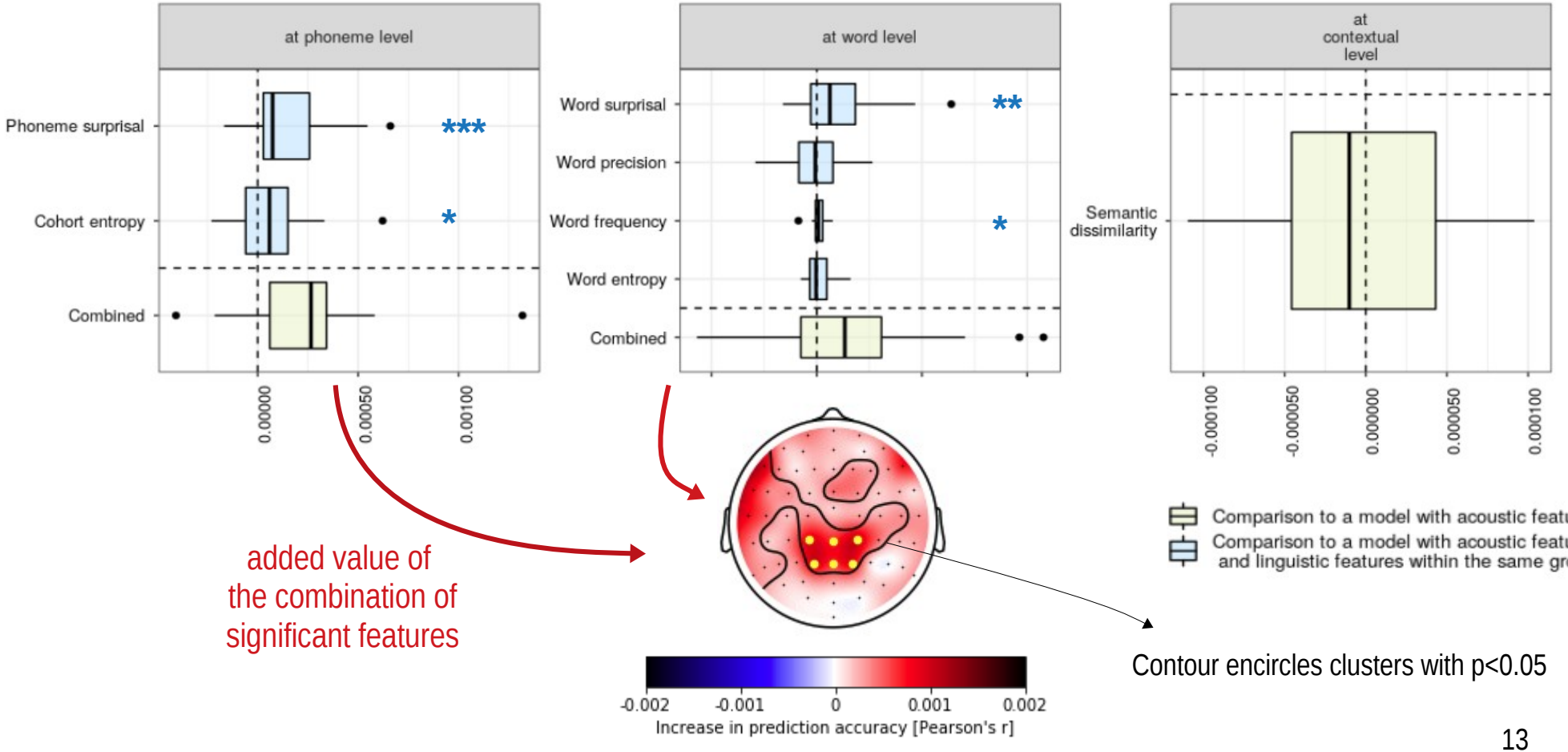
“complete model”		“baseline model”		difference in correlation	
Acoustic properties	v	Acoustic properties	v	Acoustic properties	x
Lexical segmentations	v	Lexical segmentations	v	Lexical segmentations	x
Linguistic properties	v	Linguistic properties	x	Linguistic properties	v



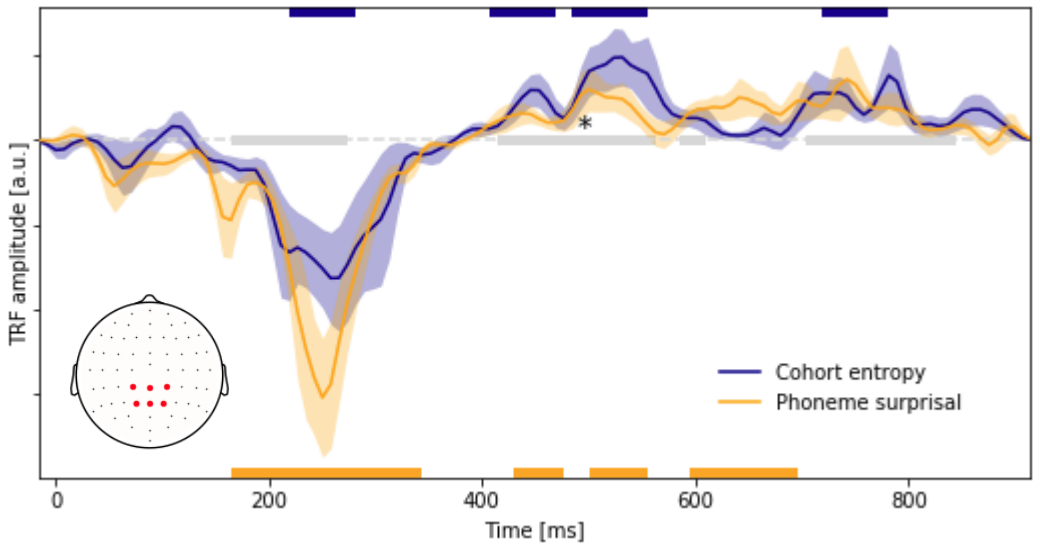
Prediction accuracy [Pearson’s r]

Difference in prediction accuracy [Pearson’s r]

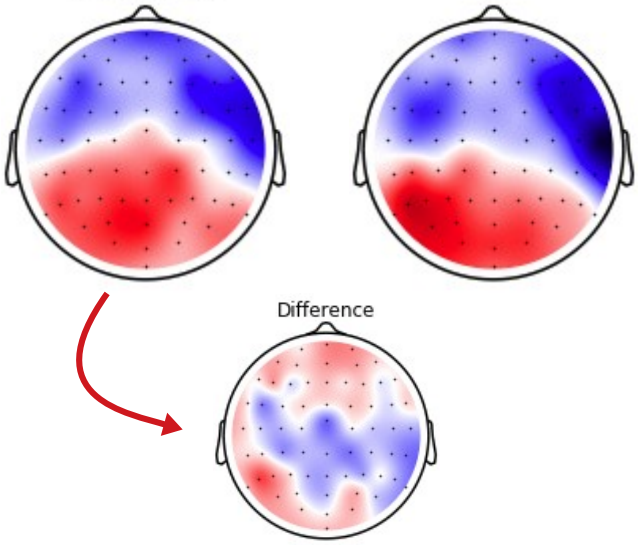
Which linguistic representations are tracked?



Do they reflect separate stages of language processing?

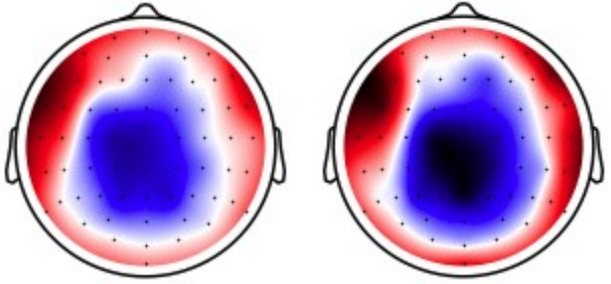


Cohort entropy Phoneme surprisal



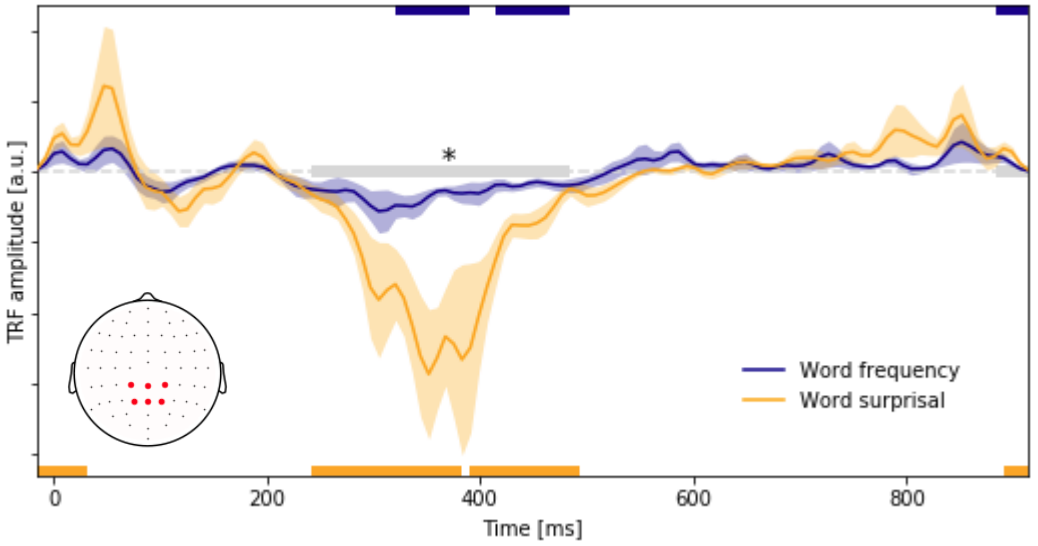
→ Statistical difference (p=0.035)
→ Representations might represent distinct speech processing stages

Cohort entropy Phoneme surprisal

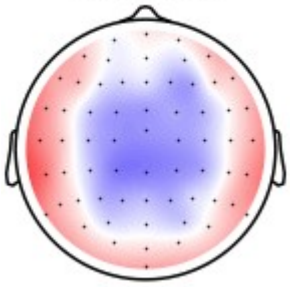


→ No statistical difference in topography (p=0.309)
→ Suggest similar underlying neural generators

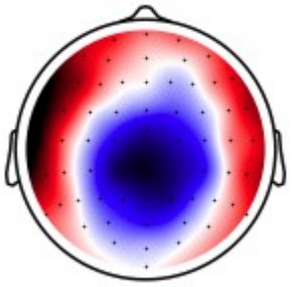
Do they reflect separate stages of language processing?



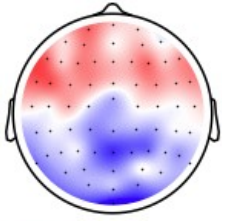
Word frequency



Word surprisal



Difference



→ Statistical difference ($p<0.001$)
→ Representations might represent distinct speech processing stages
→ N400 responses might reflect multiple processes

Take-home message

- ✔ Determining the added value of a speech representations?
  → Control for acoustic and lexical segmentation properties
- ✔ Cohort entropy and phoneme surprisal reflect different speech processing stages around 400 to 500 ms after the phoneme onset.
- ✔ Word surprisal and word frequency might explain different language processes, captured by the N400-response.

Questions? Remarks? Suggestions? Don't hesitate to contact me: marlies.gillis@kuleuven.be

This research has received funding from ERC (No. 637424; Tom Francart), from NIH (R01-DC014085; Jonathan Simon) and NSF (1754284; to University of Connecticut - J. Magnuson, PI; Christian Brodbeck). Financial support is provided to Jonas Vanthornhout and Marlies Gillis funded by FWO (SB 1SA0620N, 1290821N)