Neural Speech Tracking During Attention Switching Between Speakers in a Cocktail Party

Renana Vaknin Harel[1], Paz Har-shai Yahav[1], Eshed Rabinovitch[1], Adi Korisky[1], Roi Gueta[2], Sharon Gannot[2], Elana Zion Golumbic[1]

[1] The Leslie and Susan Gonda Multidisciplinary Brain Research Center, Bar-Ilan University, Ramat-Gan, Israel
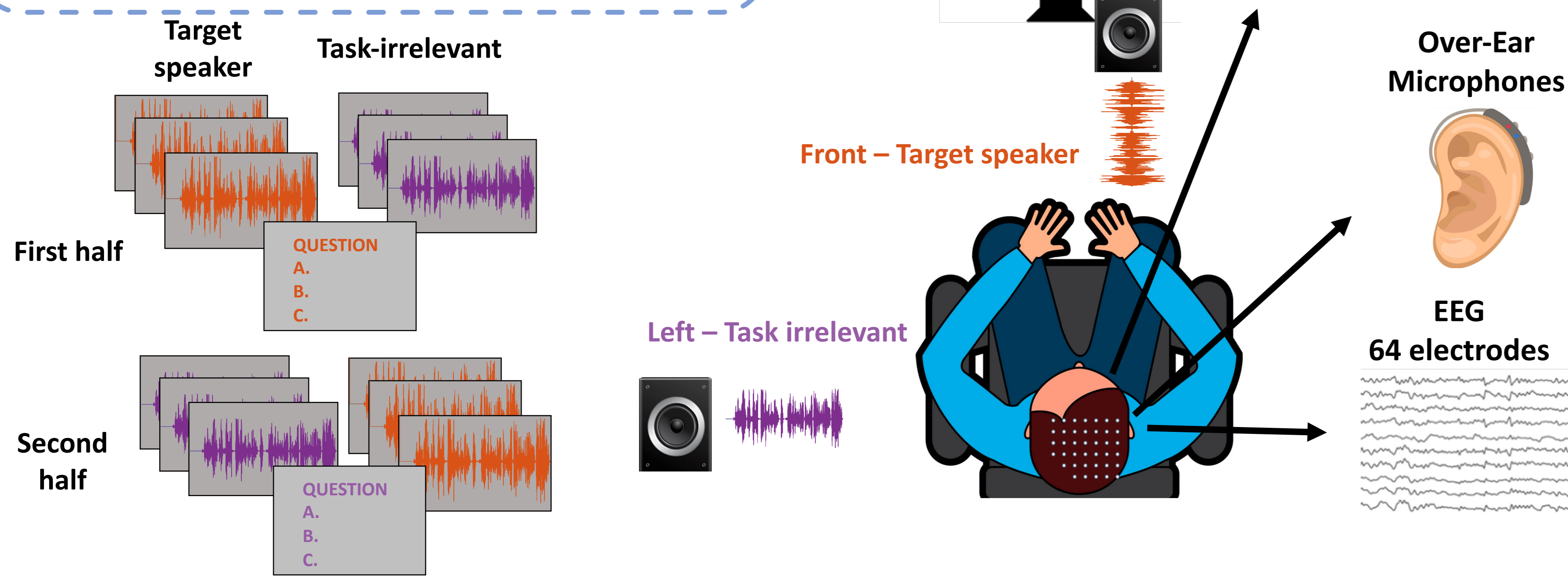[2] The Alexander Kofkin Faculty of Engineering, Bar-Ilan University, Ramat-Gan, Israel

Background

- The ability to focus our attention on one speaker in common noisy situations is a challenging feat in our busy real-life environments.
- In this project, we attempt to take the next step towards more ecological set-ups using **loudspeakers**, which allows us to conduct laboratory experiments that more closely resemble everyday challenges of selective attention.
- In such 'Cocktail Party' scenarios, top-down attention operates to selectively amplify one voice (**target**) and suppress other competing voices (**task irrelevant**).
- Using the neural signatures of this selective attention (**TRF**), we investigate this competition for processing resources.
- **Research question: How is the ability to attend-to or to ignore speech affected when the target and task-irrelevant speakers switch roles.**
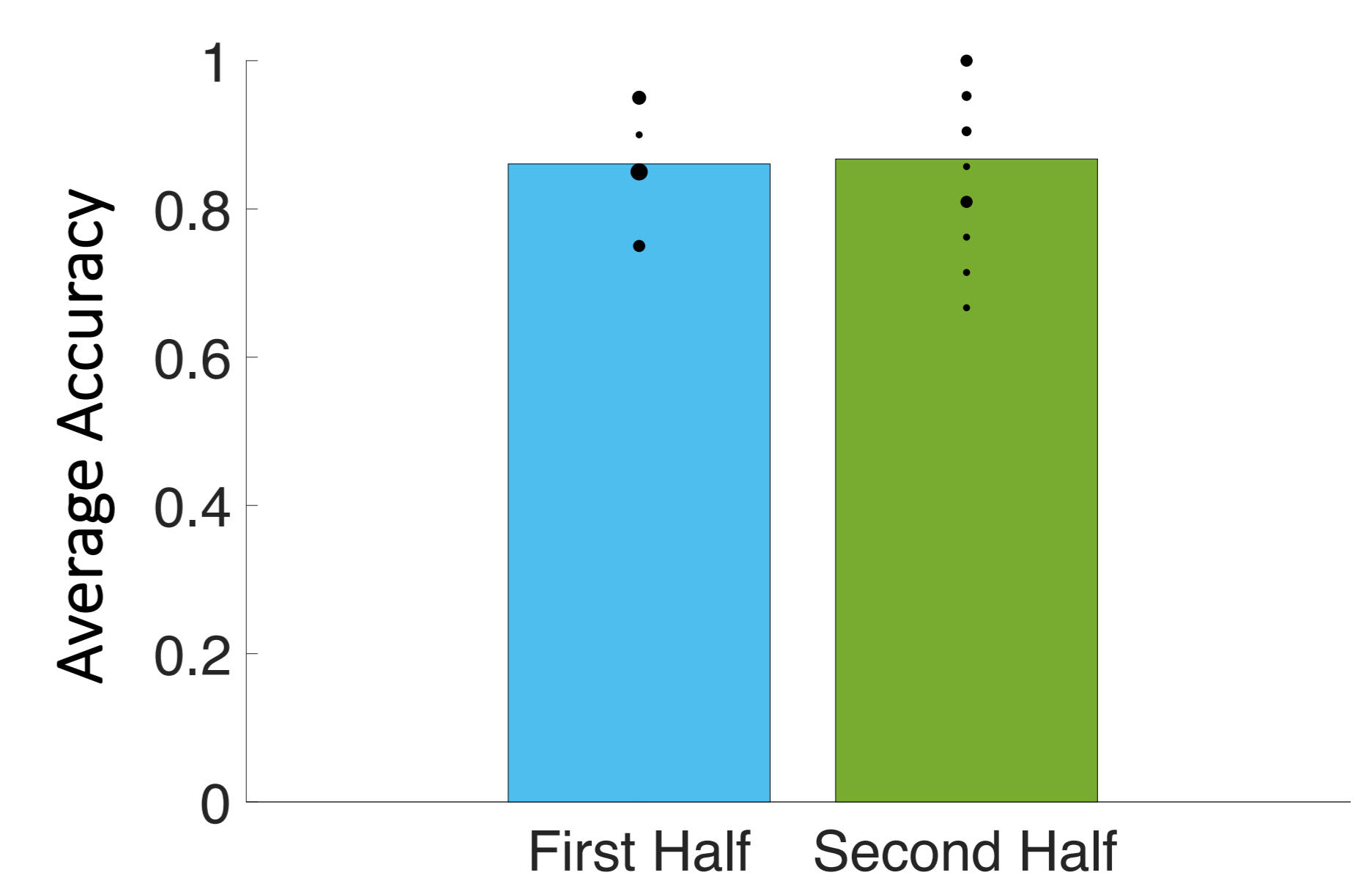
Experiment Design

- **Participants**
  - N = 14
  - Mean age: 23.3
  - Normal hearing, Native Hebrew speakers

- **40 trials** (~30 sec each)
- 3 multiple choice questions after each 3 trials

Target speaker / Task-irrelevant
First half
Second half
QUESTION A. B. C.

Eye Tracking
Front – Target speaker
Left – Task irrelevant
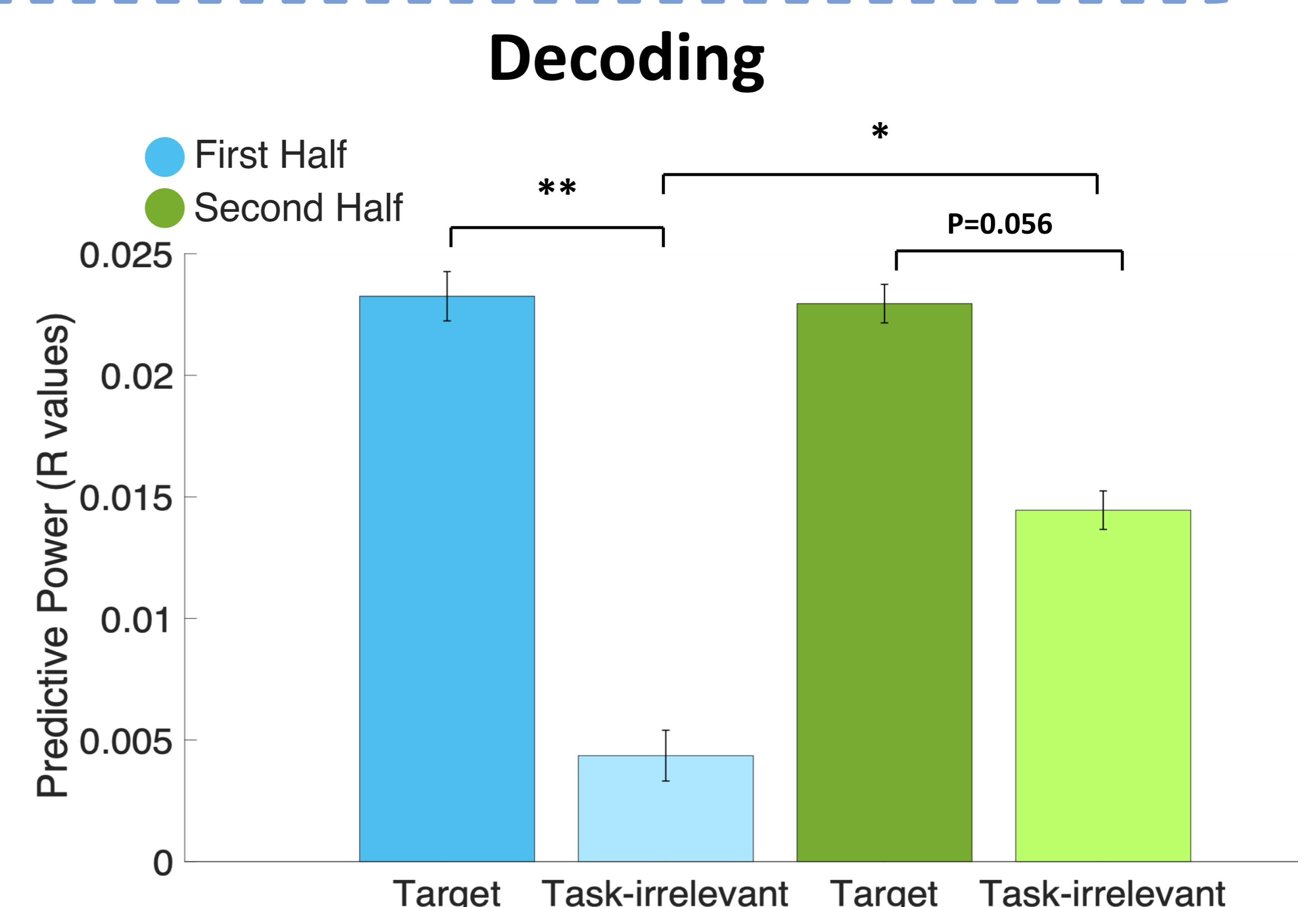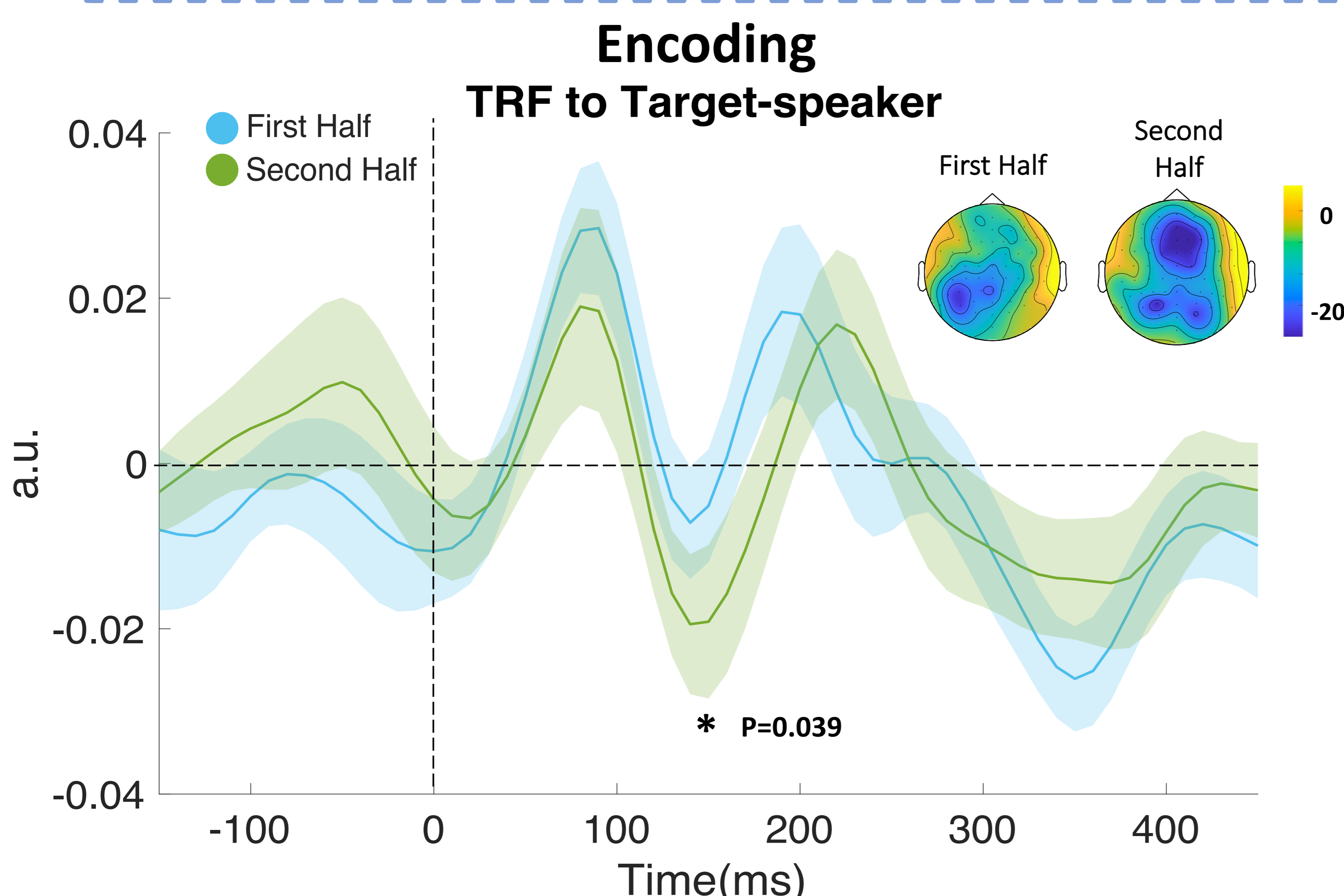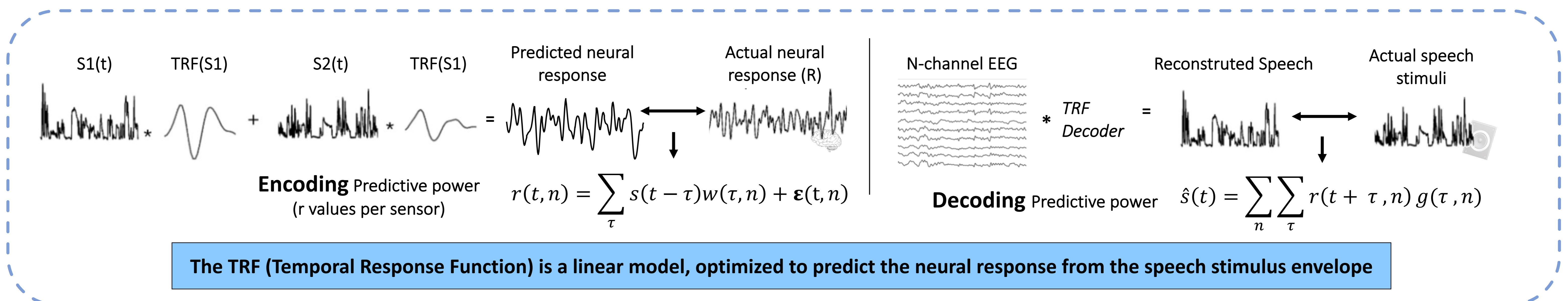Over-Ear Microphones
EEG 64 electrodes

Hypothesis

- Task-irrelevant in the second half (the familiar voice) will be more interfering and serve as an 'attention grabber', leading to an **enhanced speech tracking response of the target speech and possibly reduced behavioral performance.**

Behavioral results

EEG Results – Speech Tracking Analysis

$S1(t)$ TRF(S1) $S2(t)$ TRF(S1) Predicted neural response Actual neural response (R)

Encoding Predictive power (r values per sensor)

$$r(t,n) = \sum_{\tau} s(t-\tau)w(\tau,n) + \varepsilon(t,n)$$

N-channel EEG *TRF Decoder* Reconstructed Speech Actual speech stimuli

Decoding Predictive power

$$\hat{s}(t) = \sum_{n}\sum_{\tau} r(t+\tau,n)\,g(\tau,n)$$

The TRF (Temporal Response Function) is a linear model, optimized to predict the neural response from the speech stimulus envelope

Encoding
TRF to Target-speaker

First Half / Second Half

* P=0.039

Decoding

First Half / Second Half

** / * / P=0.056

Target / Task-irrelevant / Target / Task-irrelevant

There is a difference between the first and the second half regarding the process of the Task-irrelevant speaker

Conclusions

- The neural representation of **task-irrelevant speech and of target speech** is affected by previous exposure.
- Neural decoding of **task-irrelevant speech** was enhanced in the second vs. first half, suggesting it may be harder to ignore.
- Neural encoding of the **target speech** was also enhanced in the second half, which may reflect higher listening effort (to avoid distraction) or effects of familiarity of the previously task-irrelevant voice.
- These results indicate that familiarity with voice and previous experience shape neural selectivity to speech in multi-talker environments.