Tracking of Continuous Speech in Noisy Auditory Scenes with 7T fMRI

Lars Hausfeld [a,b], Federico De Martino [a,b], Elia Formisano [a,b,c]

[a] Dept. Cognitive Neuroscience, Maastricht University, The Netherlands
[b] Maastricht Brain Imaging Centre (M-BIC), The Netherlands
[c] Maastricht Centre for Systems Biology, The Netherlands

MOTIVATION

- Speech/sound envelope tracking is a highly replicated phenomenon in MEG/ECoG/EEG studies reported during the last decade (refs. 1-3)
- Localized to primary and secondary auditory cortical (AC) regions (refs. 1, 2, 4)
- fMRI-EEG study presented evidence that EEG tracking correlates with activation in right TPJ on a group level but no significant trial-by-trial relationship (highest correlation in left anterior STS) (ref. 5)

EXPERIMENTAL DESIGN

- Speech (audiobooks) of two voices (v1 and v2, diotically presented; from ref. 6)
- Silent periods for each voice < 300ms
- Conditions: Single speaker (*clean speech*) and auditory scene *(noisy speech*)
- Participants (*N* = 10) performed (selective) listening tasks

FUNCTIONAL MRI (fMRI)

- 7 Tesla
- TR = 1000ms, 1.5x1.5x1.5 mm^3 (GRAPPA 2, Multi-band 3)
- Whole cortex coverage
- Two 5min-trials per run + response and rest periods (~ 720 volumes/run)
- Continuous acquisition (i.e., no silent gap between image acquisitions)
- 6 runs/participant

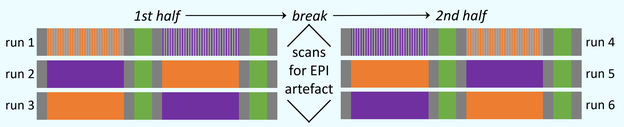
Functional run: single speaker (2x)

Functional run: auditory scene (4x)

RP: response period

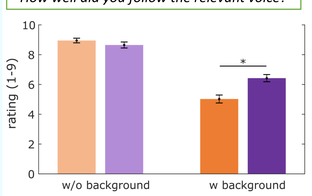
Presentation sequence

presentation order within runs reversed for every other participant

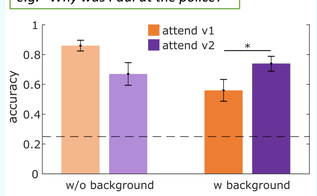


BEHAVIOR

Rating
"How well did you follow the relevant voice?"



Content questions (4AFC)
e.g. "Why was Paul at the police?"



OBJECTIVES

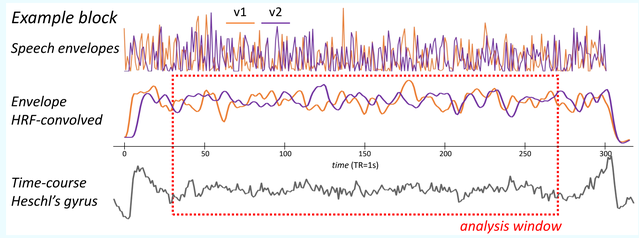
1. Does the BOLD signal track the speech envelope amplitude?
2. Which areas track the speech envelope?
   i. Single speaker: clean speech
   ii. Auditory scene: relevant and non-relevant speech

TRACKING ANALYSIS

- Voxel-wise *General Linear Model* (GLM) framework (≈ forward models, below)
- Envelope time courses convolved with hemodynamic response function (HRF)
- Data was cut to 4min/trial removing first and final 30s
- Design matrix $\mathbf{X}$ includes predictors for the speech envelope(s), participant's motion, level difference and offset

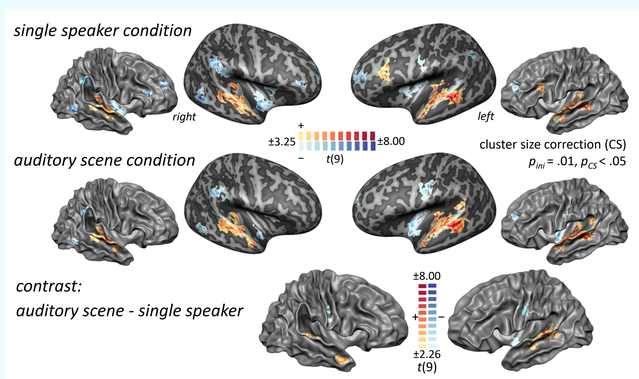
$$y = \mathbf{X}\beta + \varepsilon$$

where y: voxel time course, $\mathbf{X}$: design matrix, β: coeffients of $\mathbf{X}$, ε: error term



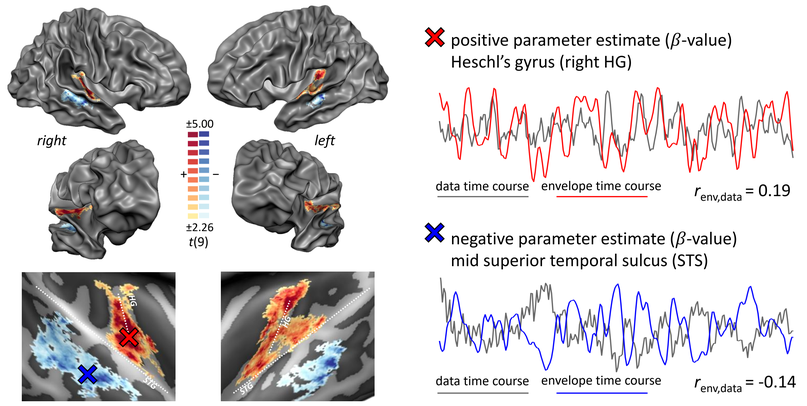
"TRADITIONAL" ANALYSIS

- Design matrix $\mathbf{X}$ (see above) includes box-car predictors for task and response periods (convolved with HRF) as well as the participant's motion and offset

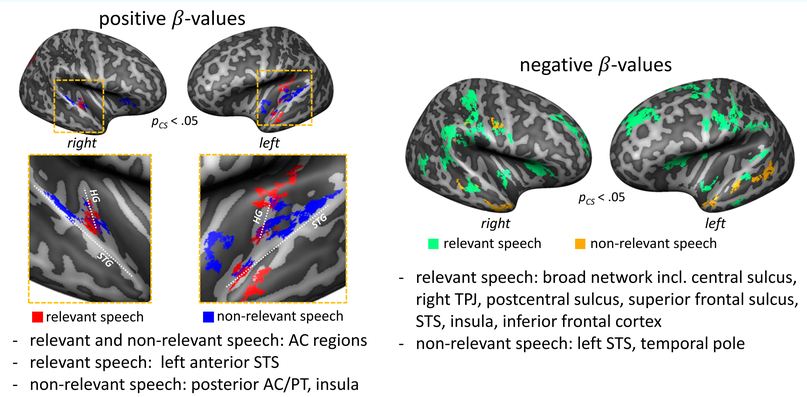


RESULTS SPEECH ENVELOPE TRACKING

single speaker condition (clean speech)



auditory scene condition (noisy speech)



- relevant and non-relevant speech: AC regions
- relevant speech: left anterior STS
- non-relevant speech: posterior AC/PT, insula

- relevant speech: broad network incl. central sulcus, right TPJ, postcentral sulcus, superior frontal sulcus, STS, insula, inferior frontal cortex
- non-relevant speech: left STS, temporal pole

CONCLUSIONS

OBJECTIVE 1: YES

OBJECTIVE 2: DEPENDS (condition and sign of β)

- pos β -> Regions in AC
- neg β -> STS + broad network

OUTLOOK

- Examine neural process of positive and negative β
  - Excitation/inhibition, output/input (ref. 7)
- Connectivity (be aware of stimulus related effects)
- Role of posterior AC/PT (cf. ref. 4)
- Dependency of results on HRF model

[1] Ding & Simon, 2012. *PNAS*
[2] Zion-Golumbic et al., 2013. *Neuron*
[3] O'Sullivan et al., 2015. *CerebCortex*
[4] O'Sullivan et al., 2019. *Neuron*
[5] Puschmann et al.., 2017. *JNeurosci*
[6] Hausfeld et al., 2018. *NeuroImage*
[7] Logothetis, 2008. *NatRevNeurosci*


Lars Hausfeld


Federico De Martino


Elia Formisano

Correspondence
M: lars.hausfeld@maastrichtuniversity.nl
P: +31 - 43 38 84521

Auditory Perception and Cognition
Department of Cognitive Neuroscience
Faculty of Psychology and Neuroscience
Maastricht University