

Using fMRI to assess the social competence of an autonomous conversational robot

PhD student: Camilla Di Pasquasio

Supervisors: Dr Thierry Chaminade & Dr Brigitte Bigi

Collaborators: Prof Marc Cavazza (University of Stirling, UK), Dr Bruno Nazarian (CERIMED)



From Human-Like Features to Human-Like Minds: What Triggers Mentalization?

Early HRI research → **morphological human-likeness** seen as key to for an efficient cooperation between humans and robots



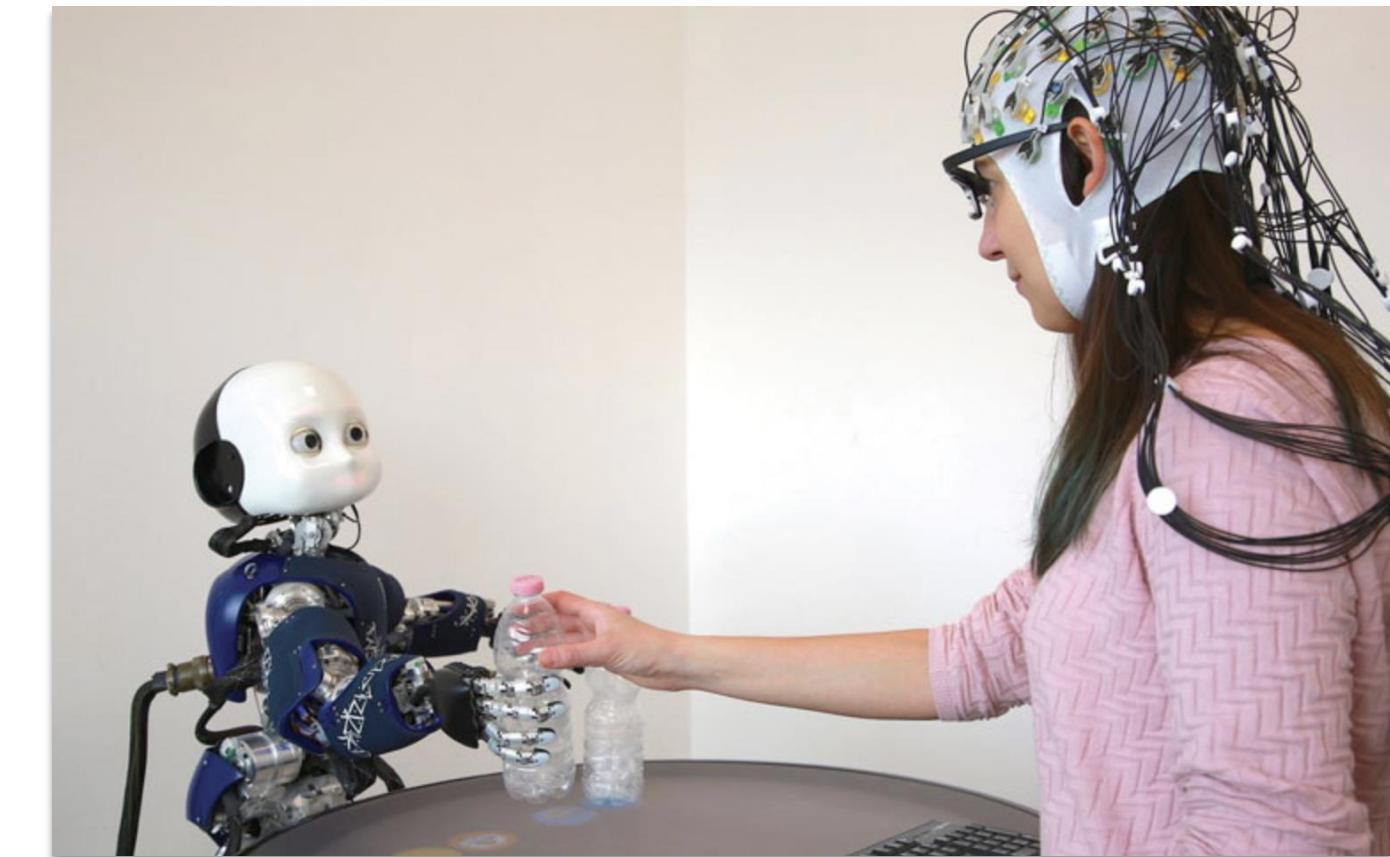
Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human–robot cooperation. *Proceedings of the 12th IEEE International Workshop on Robot and Human Interactive*

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley (translated from 1970). *IEEE Robotics and Automation Magazine*

From Human-Like Features to Human-Like Minds: What Triggers Mentalization?

Thellman, de Graaf, Ziemke, 2022

- Robots rated **more mental than computers but less than humans**
- Attribution of mental states could also depend on the agent's **social behavior** influencing trust, perception, and cognitive load



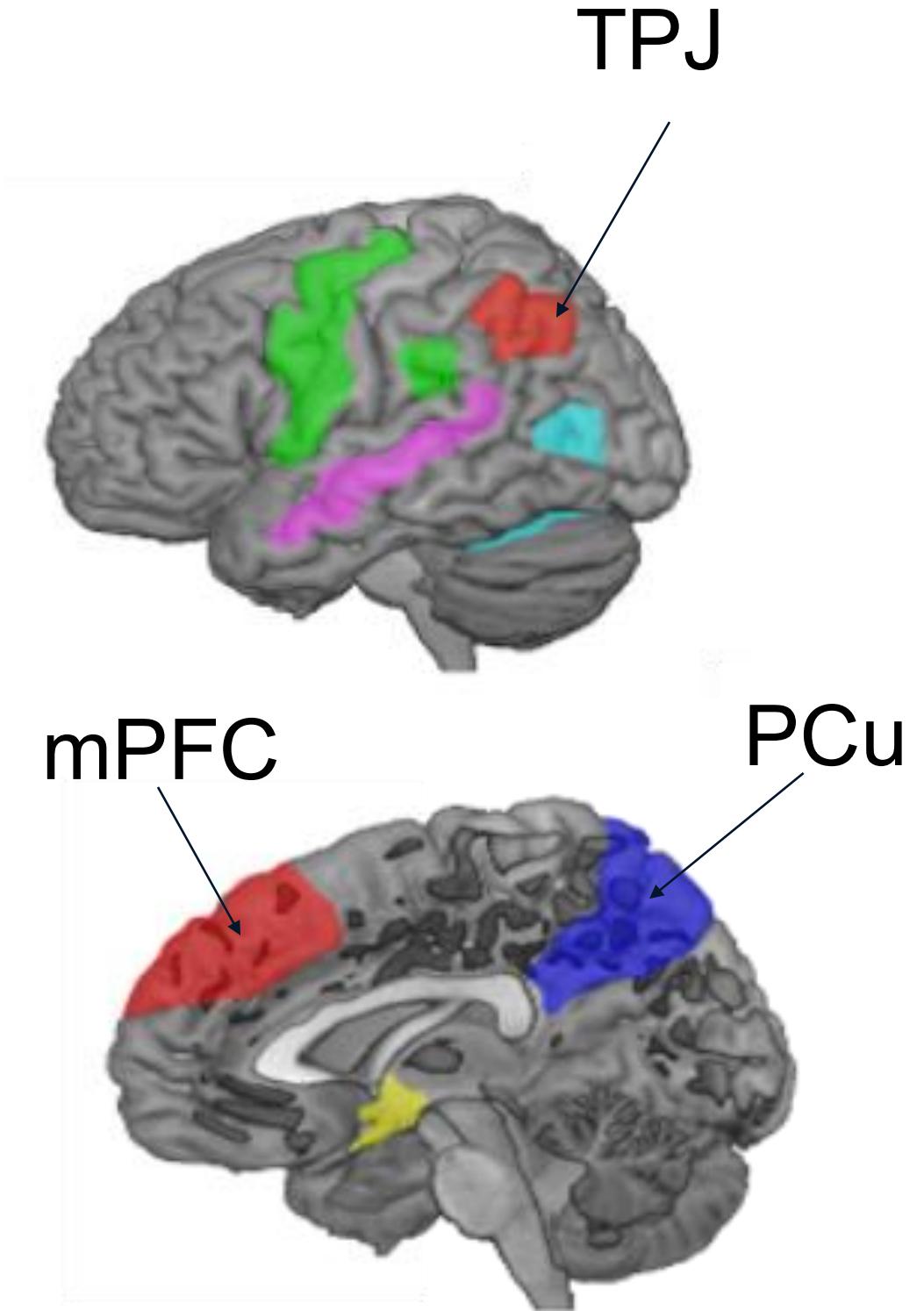
Wykowska et al. Social Cognition in Human-Robot Interaction: Putting the 'H' back in 'HRI'

Areas of Social Cognition

Cognitive neuroscience studies using fMRI have identified specific brain regions associated with mentalisation

- **TPJ** (Temporo-Parietal Junction)
- **mPFC** (Medial Prefrontal Cortex)
- **PCu and PCC** (Precuneus and Posterior Cingulate Cortex)

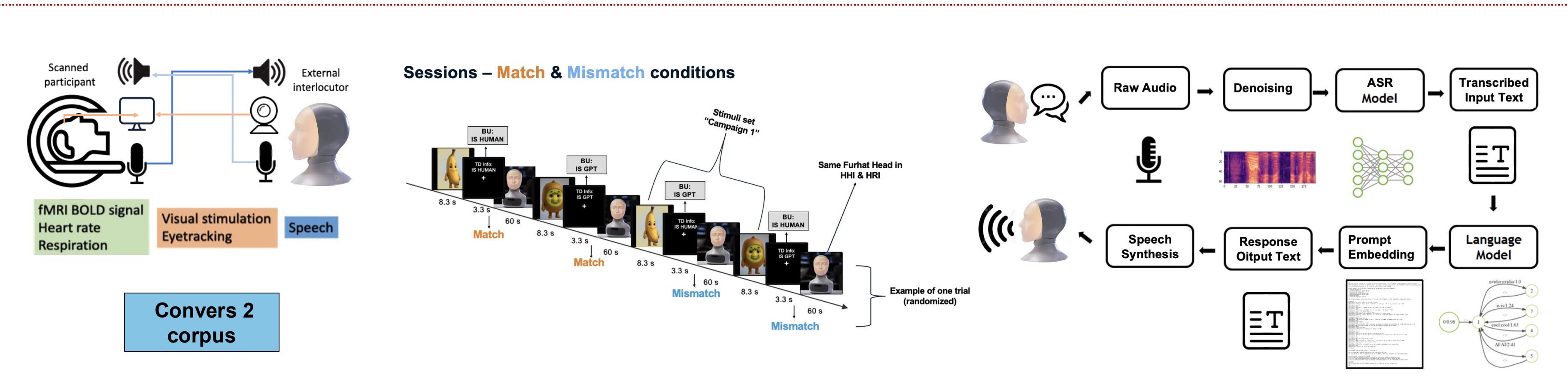
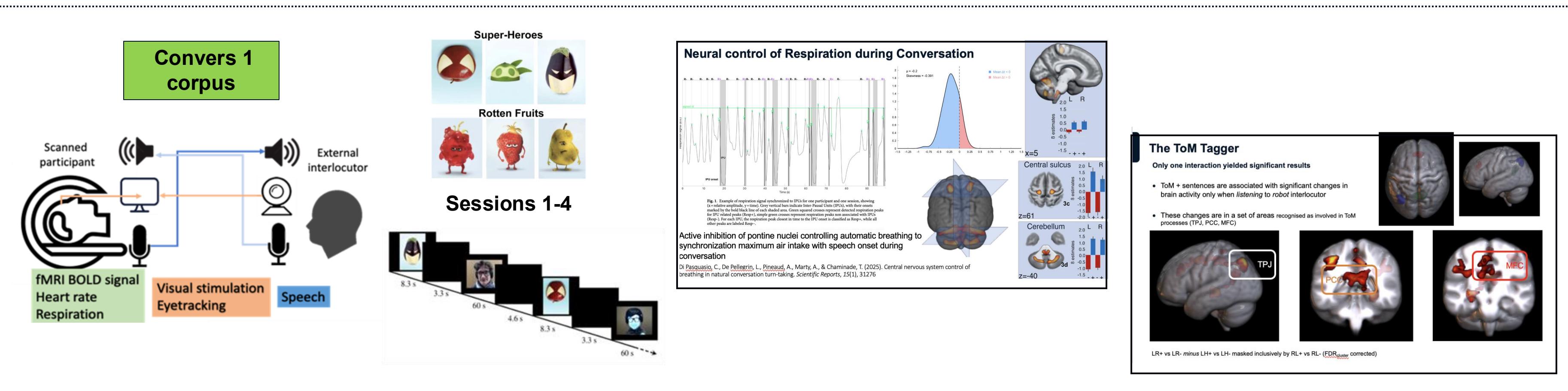
These areas are active during tasks that require the attribution of mental states to others and ToM



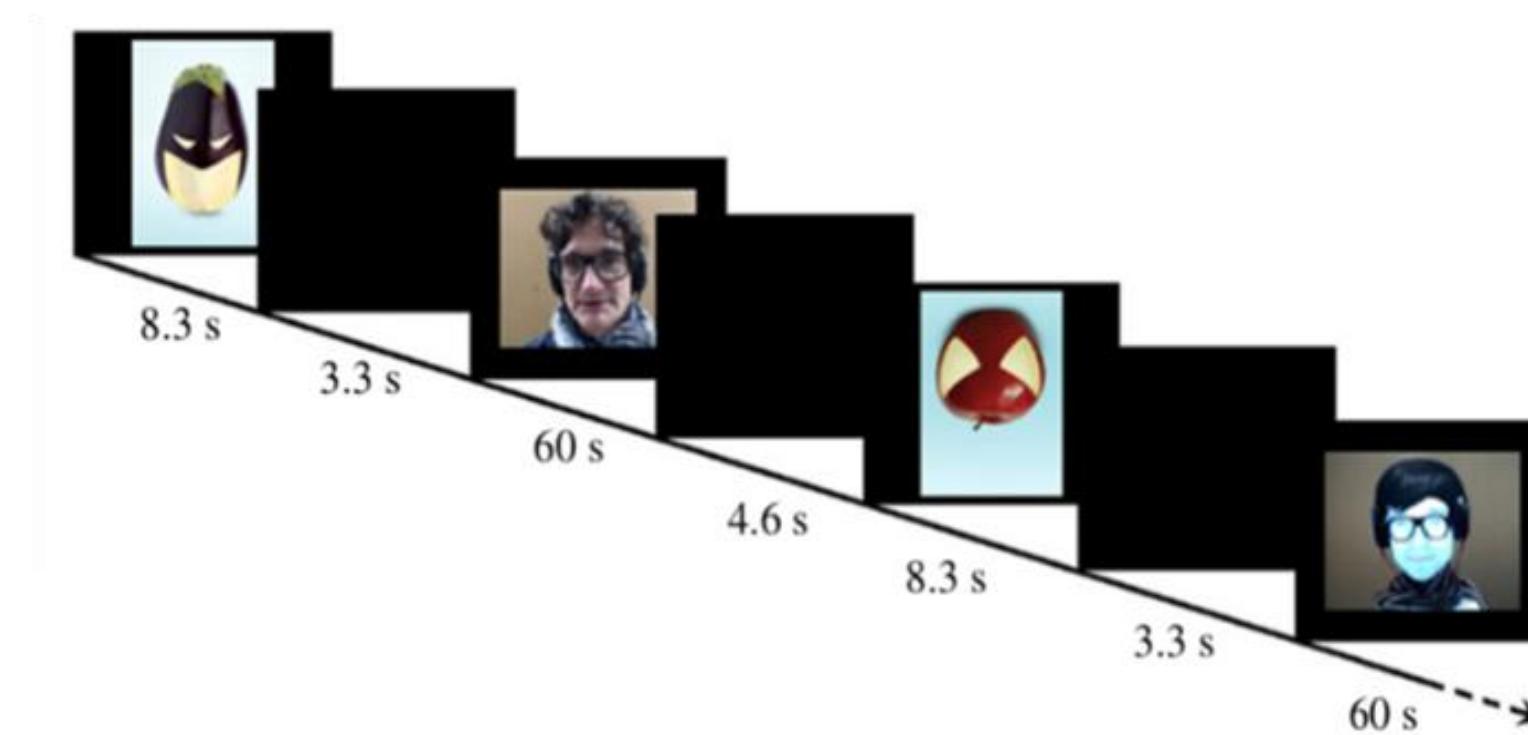
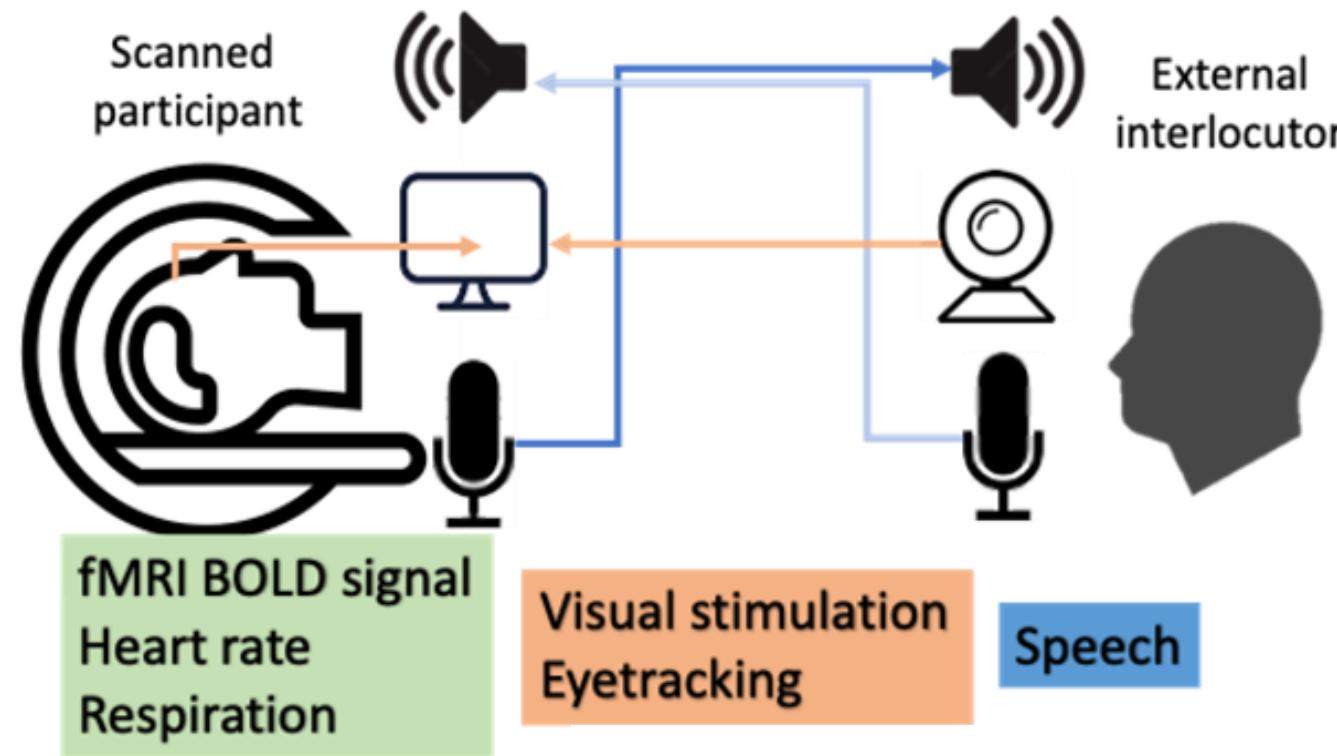
Schurz M, Radua J, Aichhorn M, Richlan F, Perner J. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neurosci Biobehav Rev*. 2014

Chaminade, T. & al. (2012) How do we think machines think ? An fMRI study of alleged competition with an artificial intelligence. *Frontiers in Human Neuroscience*

Full Project Overview



Experimental Paradigm – Convers 1



- Participants**
25 BOLD acquired in the 3T MRI Scanner
Cover story is provided
- Trials within each fMRI Session**
6 3 x 1 min Trials with a Robotic Interlocutor (HRI)
3 x 1 min Trials with a Human Interlocutor (HHI)
- fMRI Sessions**
4 6 Trials (6 mins) x 4 Sessions = 24 mins x Subject
- Hours of Recordings**
10 of Real Time Conversations

 Check for updates

OPEN **Central nervous system control of breathing in natural conversation turn-taking**

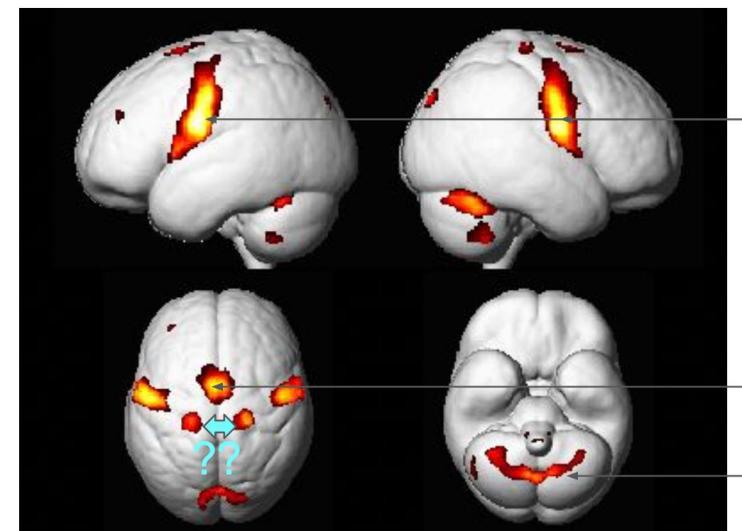
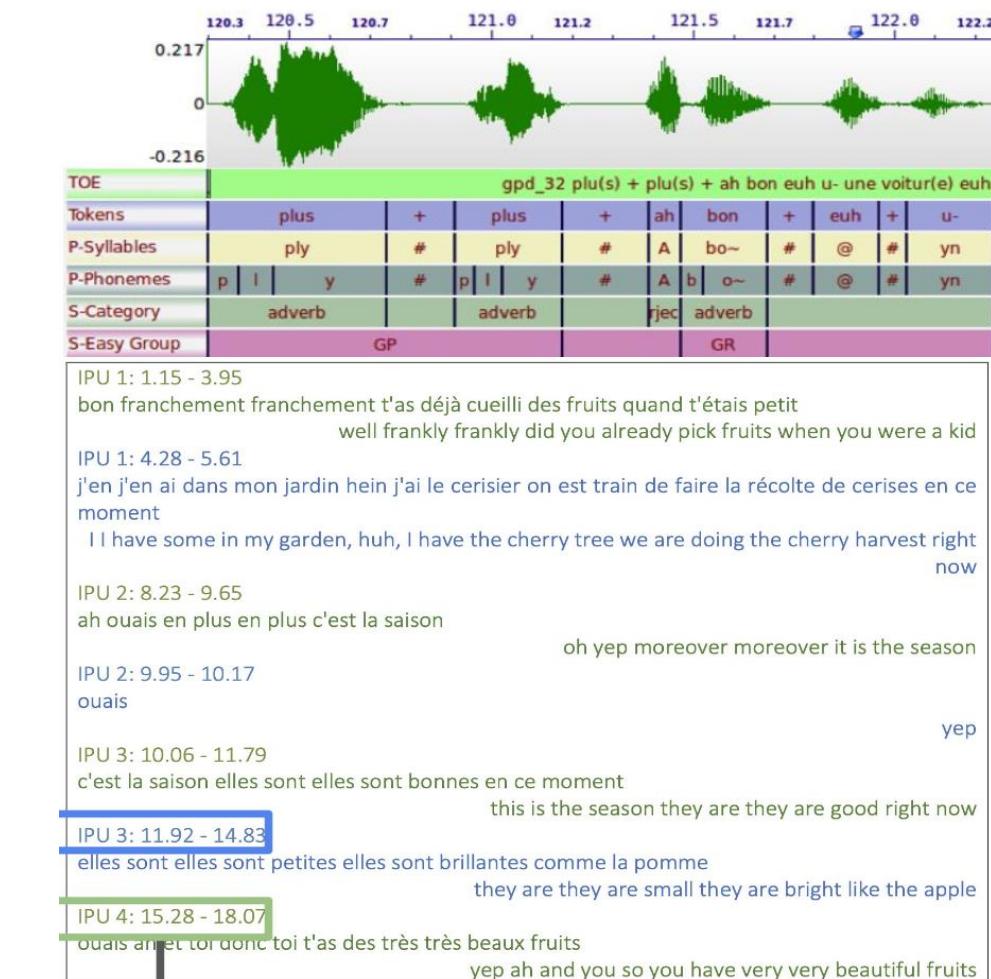
Camilla Di Pasquasio, Lila De Pellegrin, Arthur Pineaud, Antonin Marty & Thierry Chaminade✉

Aim: Understand how the brain controls breathing differently when respiration supports **speech production** versus **physiological survival** during **natural conversation**.

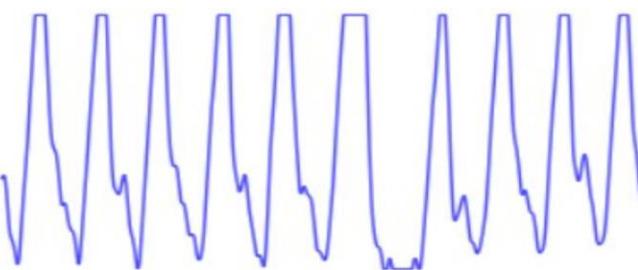
Identified respiration maxima either:

- **Speech-related** (occurring just before speech onset)
- **Non-speech-related** (likely driven by metabolic demand)

Exploratory approach to characterize neural circuits and behavioral aspects underlying **conversational breathing**

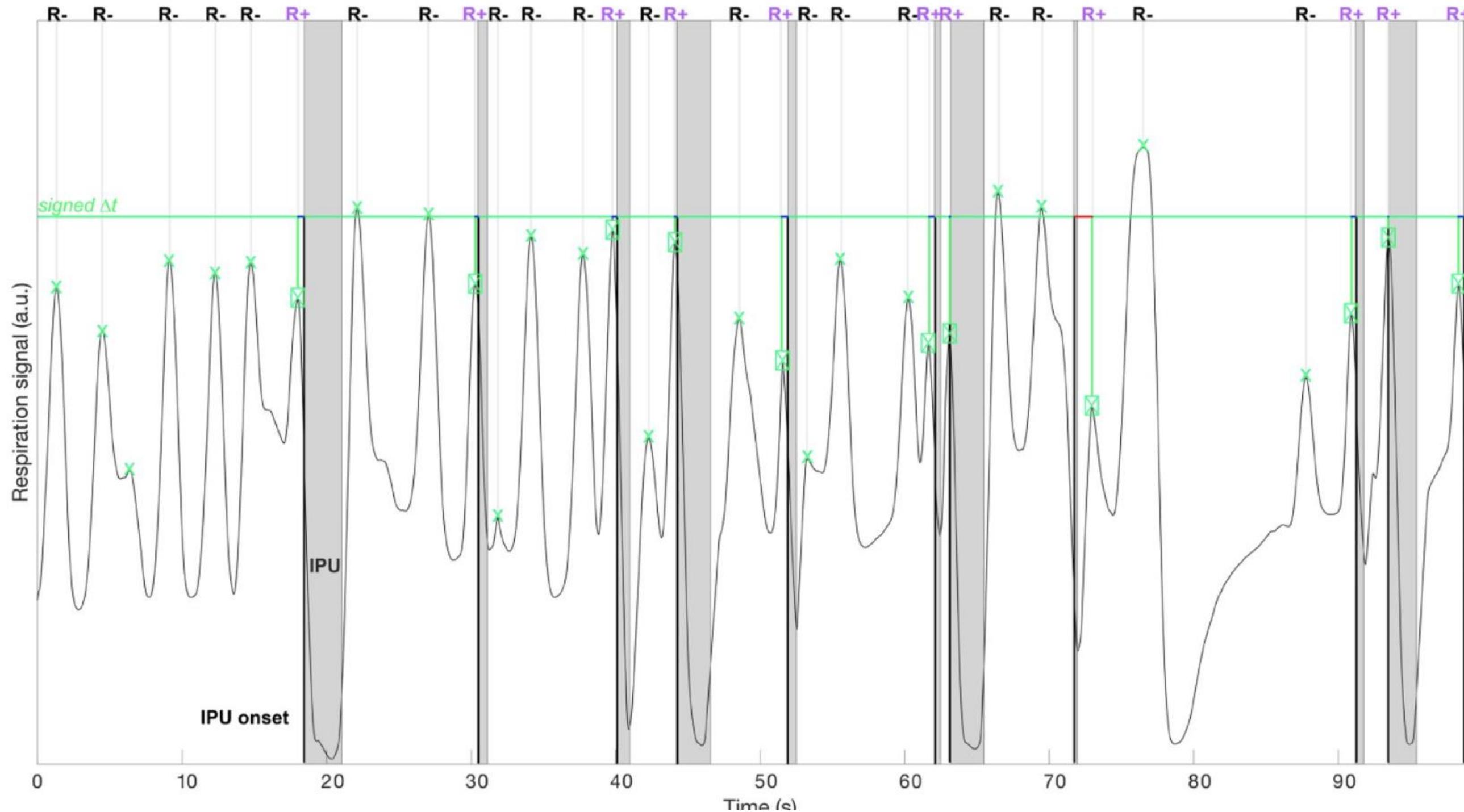


Neurophysiological data
(BOLD signal)



Physiological signal
(breathing, Physio toolbox)

Respiration data



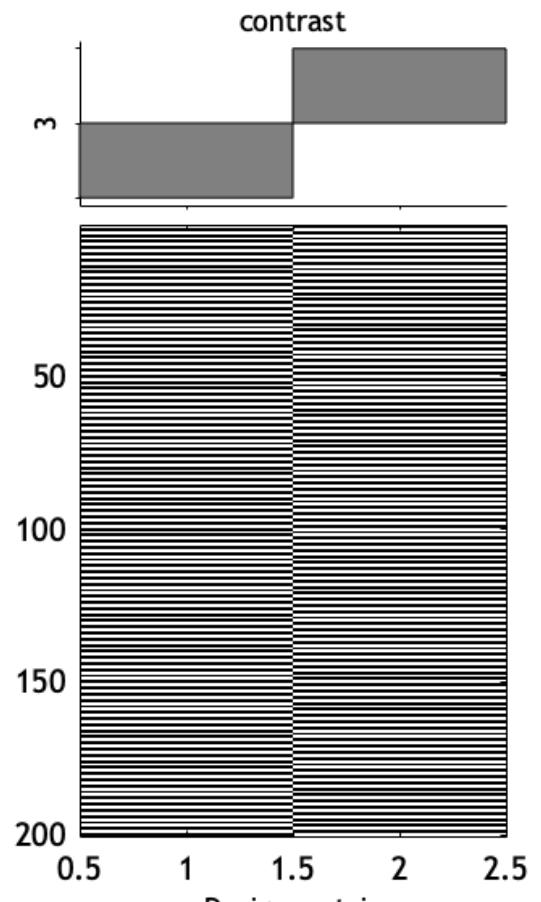
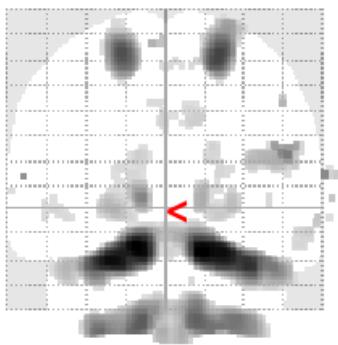
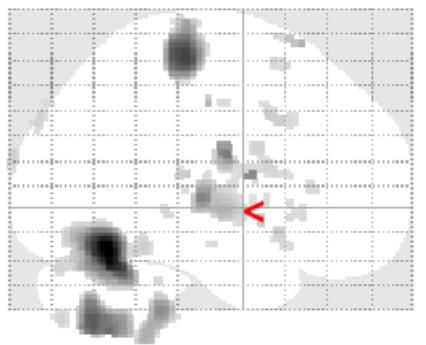
- Identification of **local respiration maxima** associated with neuronal events
- Distinction whether or not single maximum are **temporally associated with speech (operationalized as IPU onset)**.
- Identification of **Respiration Max associated with IPUs**

Analyses

Contrast definition

- SPM mat Resp+ Resp -
- Exclusion mask IPU+ IPU-

IPU (masked [excl.] by IPU_Exclusive_Mask.nii)



SPMresults: ./RespIPU/Second/Final

Height threshold T = 5.044407 {p<0.05 (FWE)}

Extent threshold k = 0 voxels

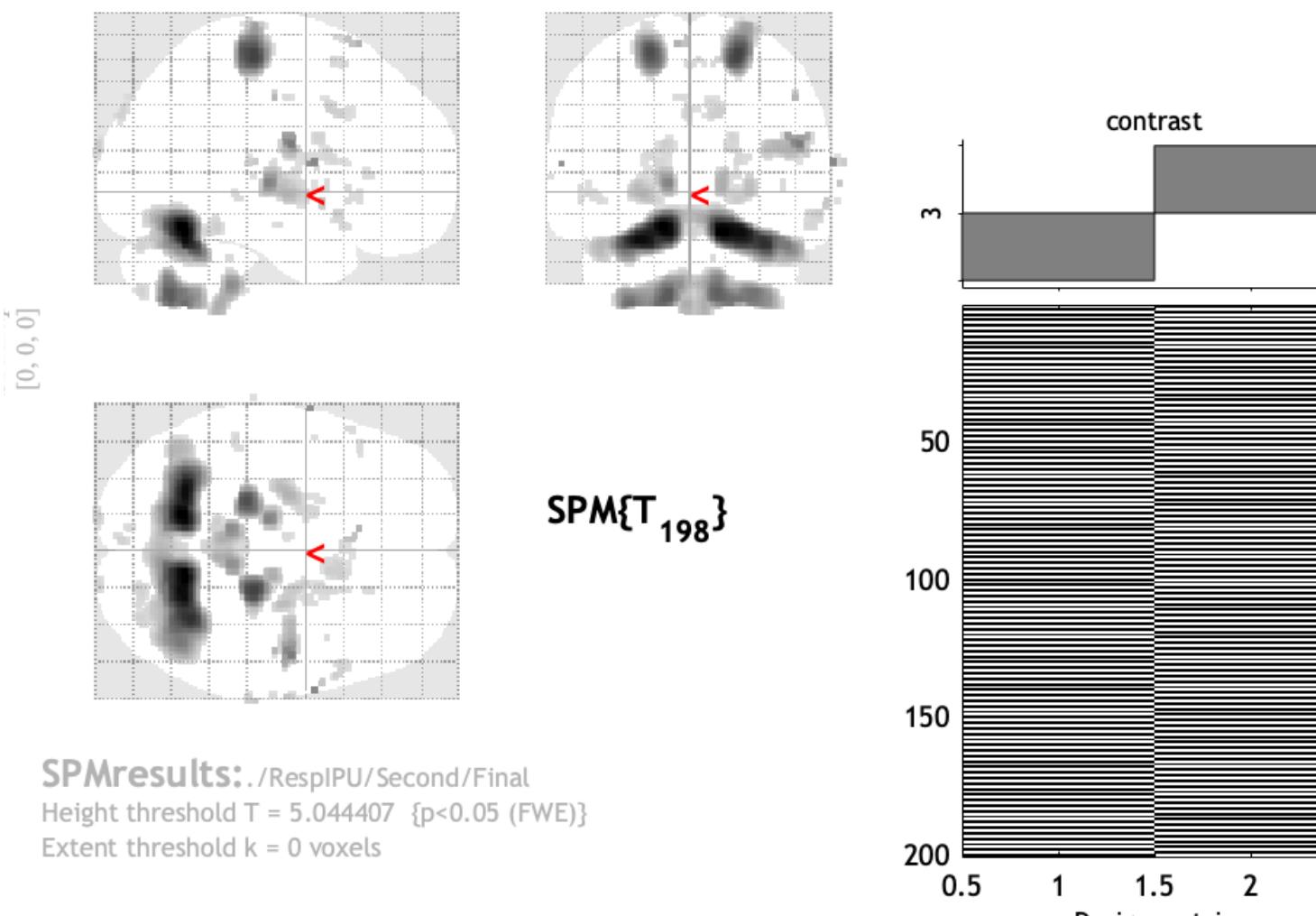
Analyses

- Attribution of **single Resp+ events to every IPU event** allowed us to calculate the relative time difference between these two events

Contrast definition

- SPM mat Resp+ Resp -
- Exclusion mask IPU+ IPU-

IPU (masked [excl.] by IPU_Exclusive_Mask.nii)



SPMresults: ./ResPIPU/Second/Final
Height threshold T = 5.044407 {p<0.05 (FWE)}
Extent threshold k = 0 voxels

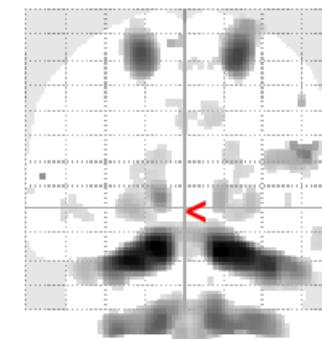
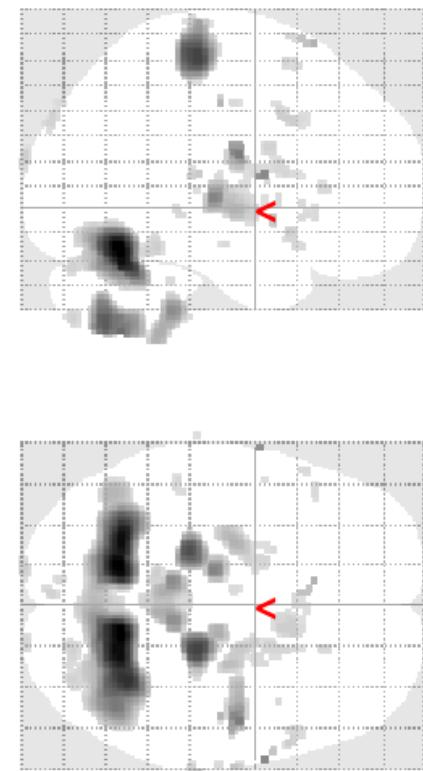
Analyses

- Attribution of **single Resp+** events to **every IPU** event allowed us to calculate the relative time difference between these two events

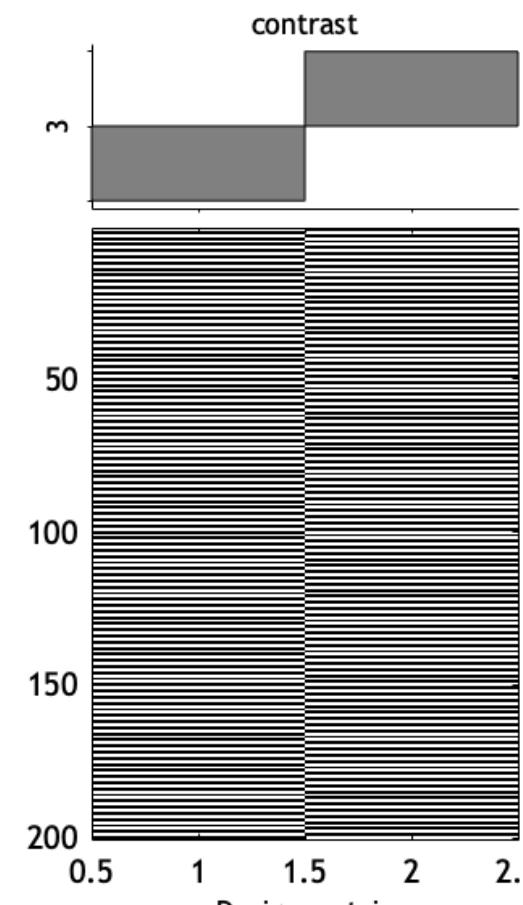
Contrast definition

- SPM mat Resp+ Resp –
- Exclusion mask IPU+ IPU-

IPU (masked [excl.] by IPU_Exclusive_Mask.nii)



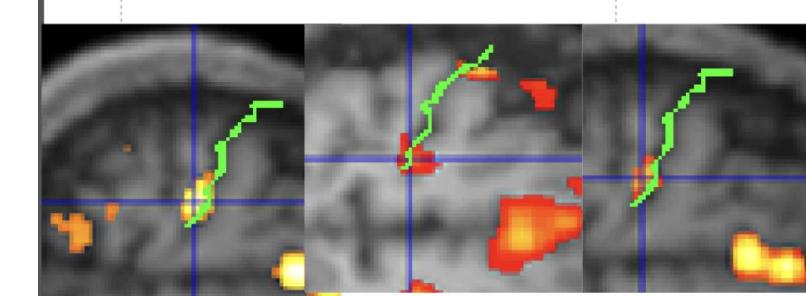
SPM{T₁₉₈}



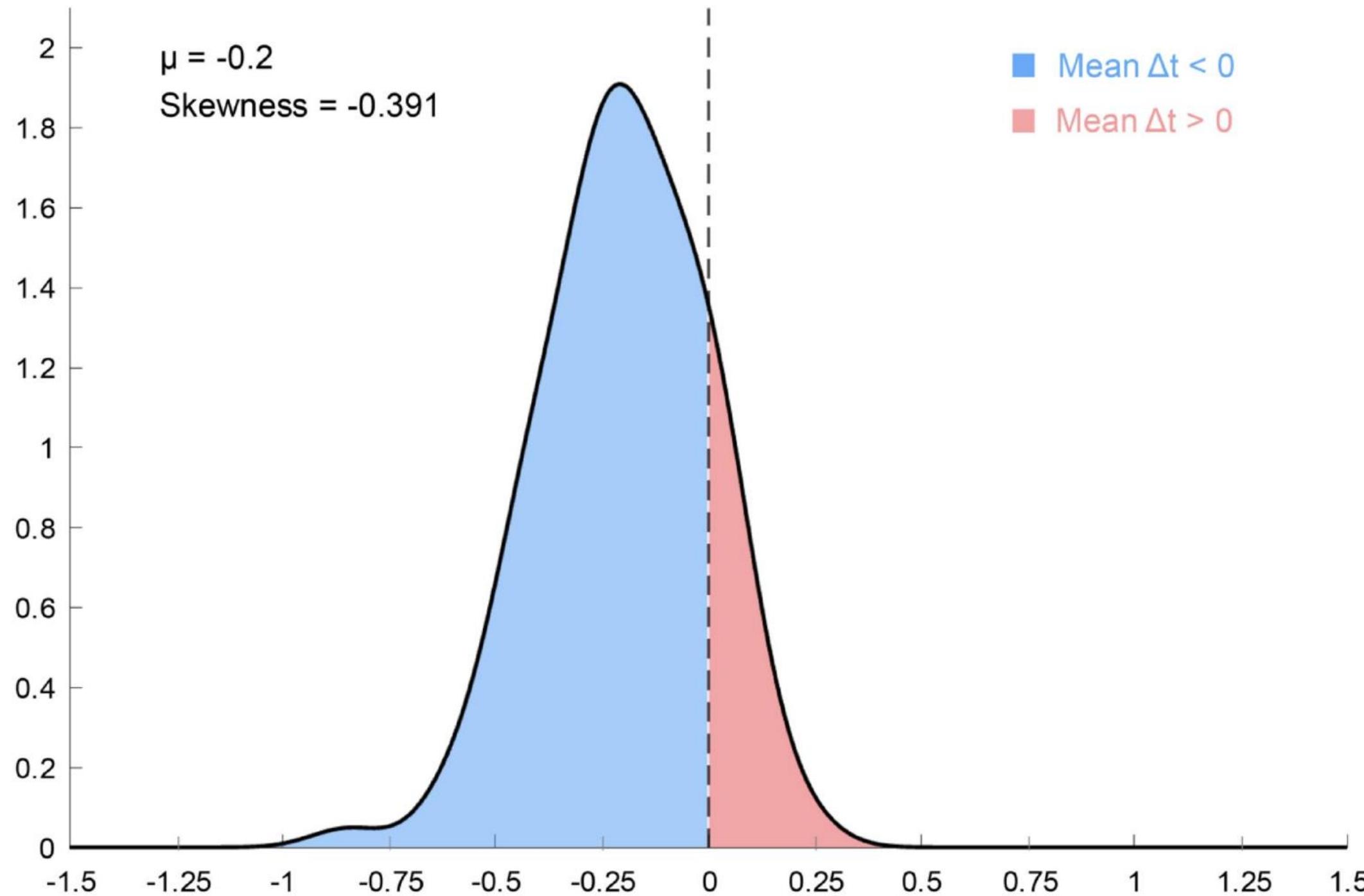
SPMresults: ./RespIPU/Second/Final
Height threshold T = 5.044407 {p<0.05 (FWE)}
Extent threshold k = 0 voxels



Central sulcus localisation



Results: Behavioral

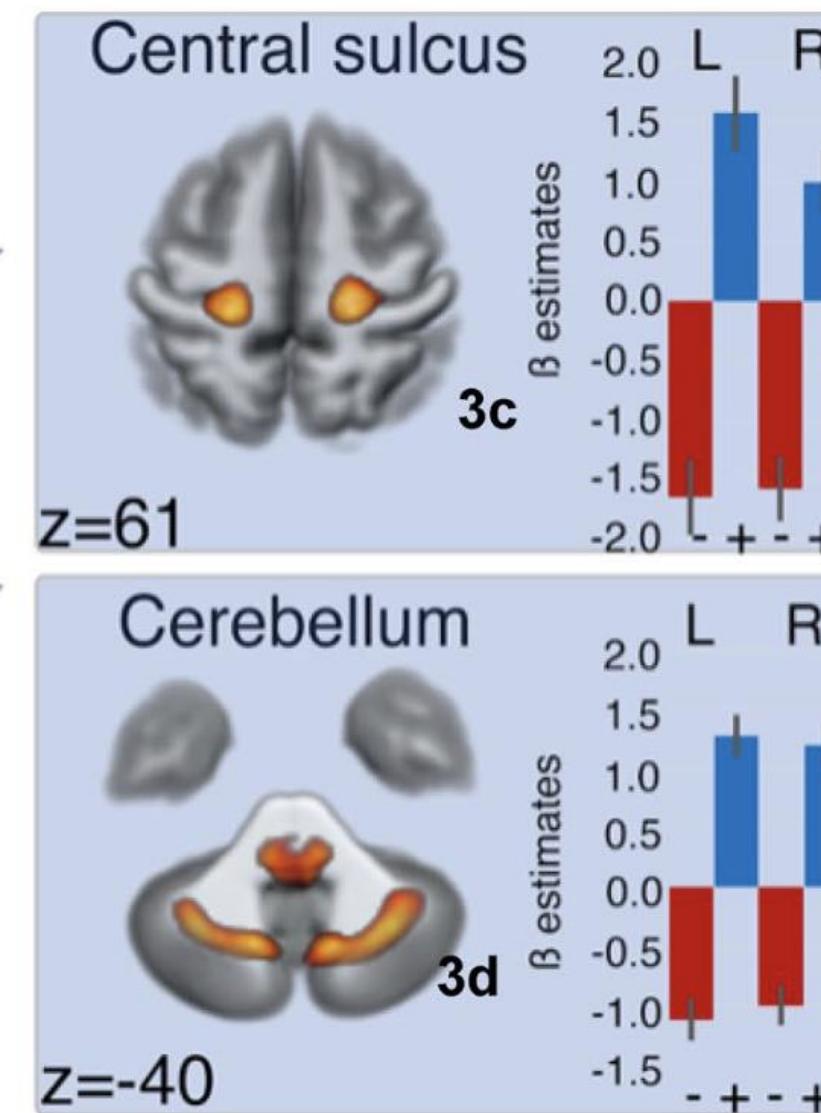
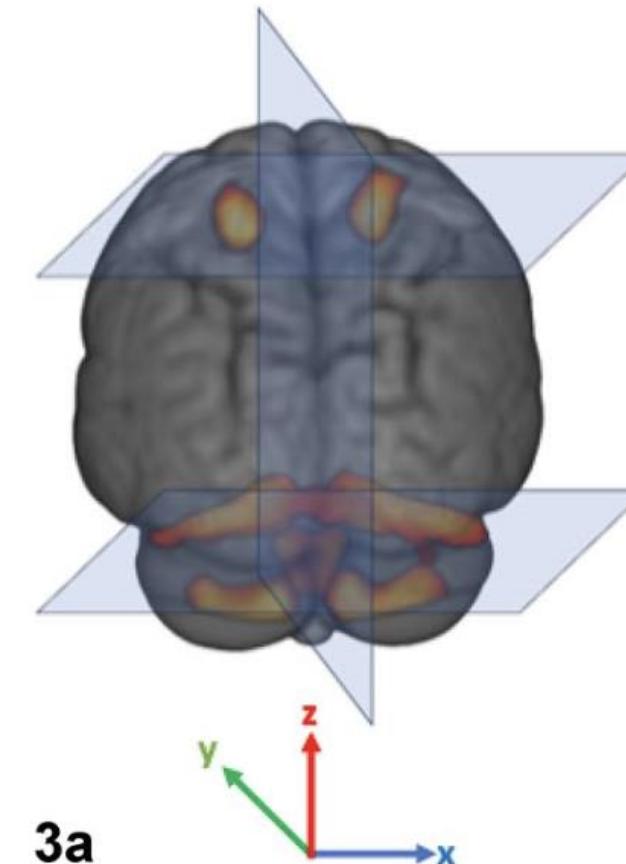
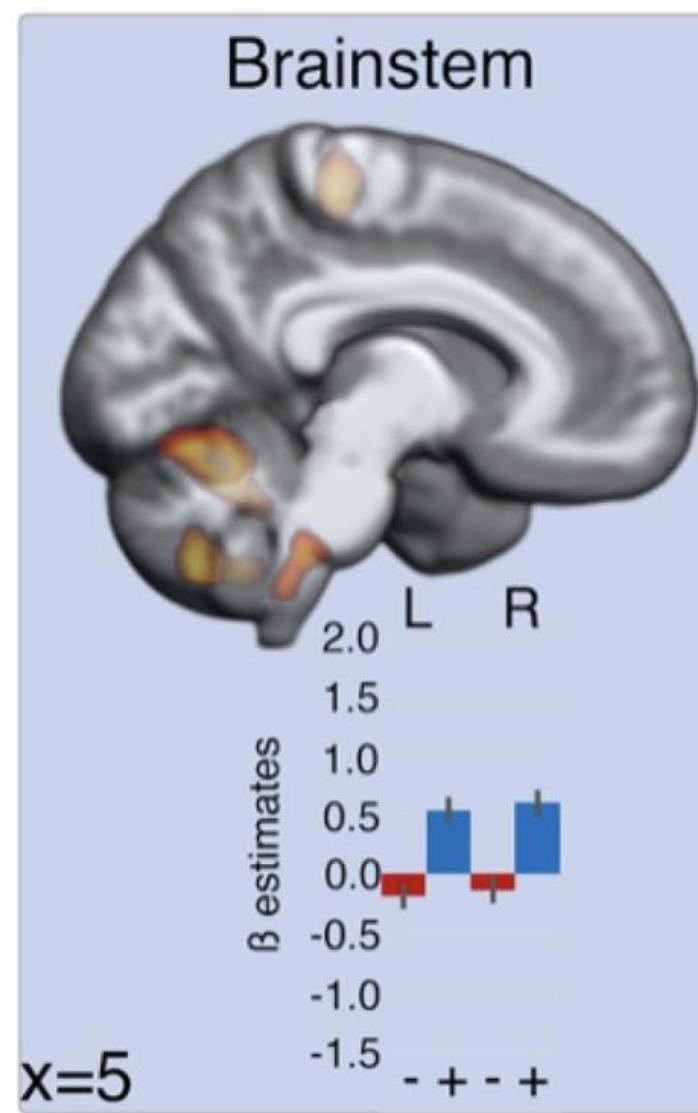


→ density plot of means of time differences
for all participants between IPU onset and max
breathings (Δt)

Results: fMRI

→ Resp+ vs Resp- exclusively masked to exclude speech related areas (IPU vs non-IPU)

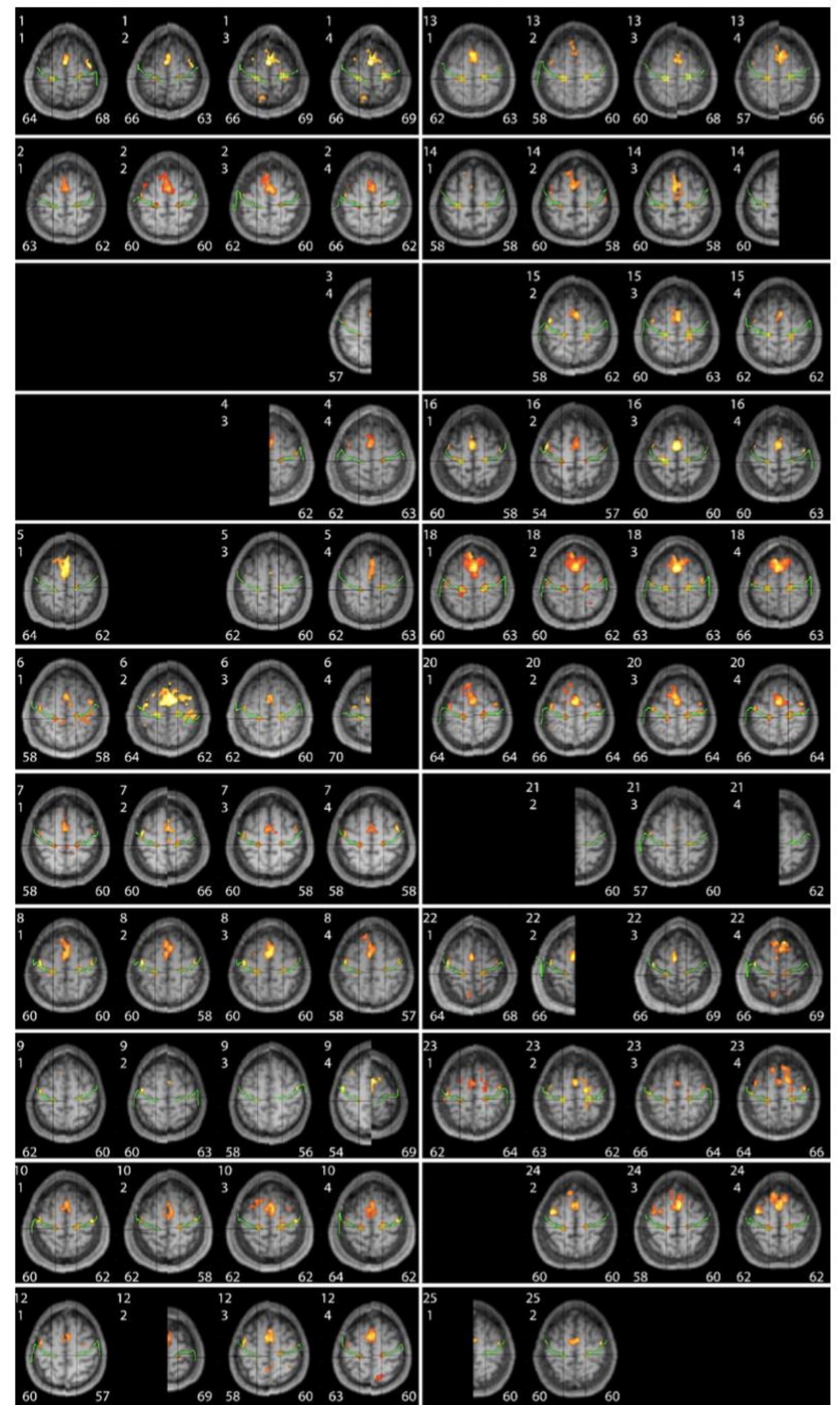
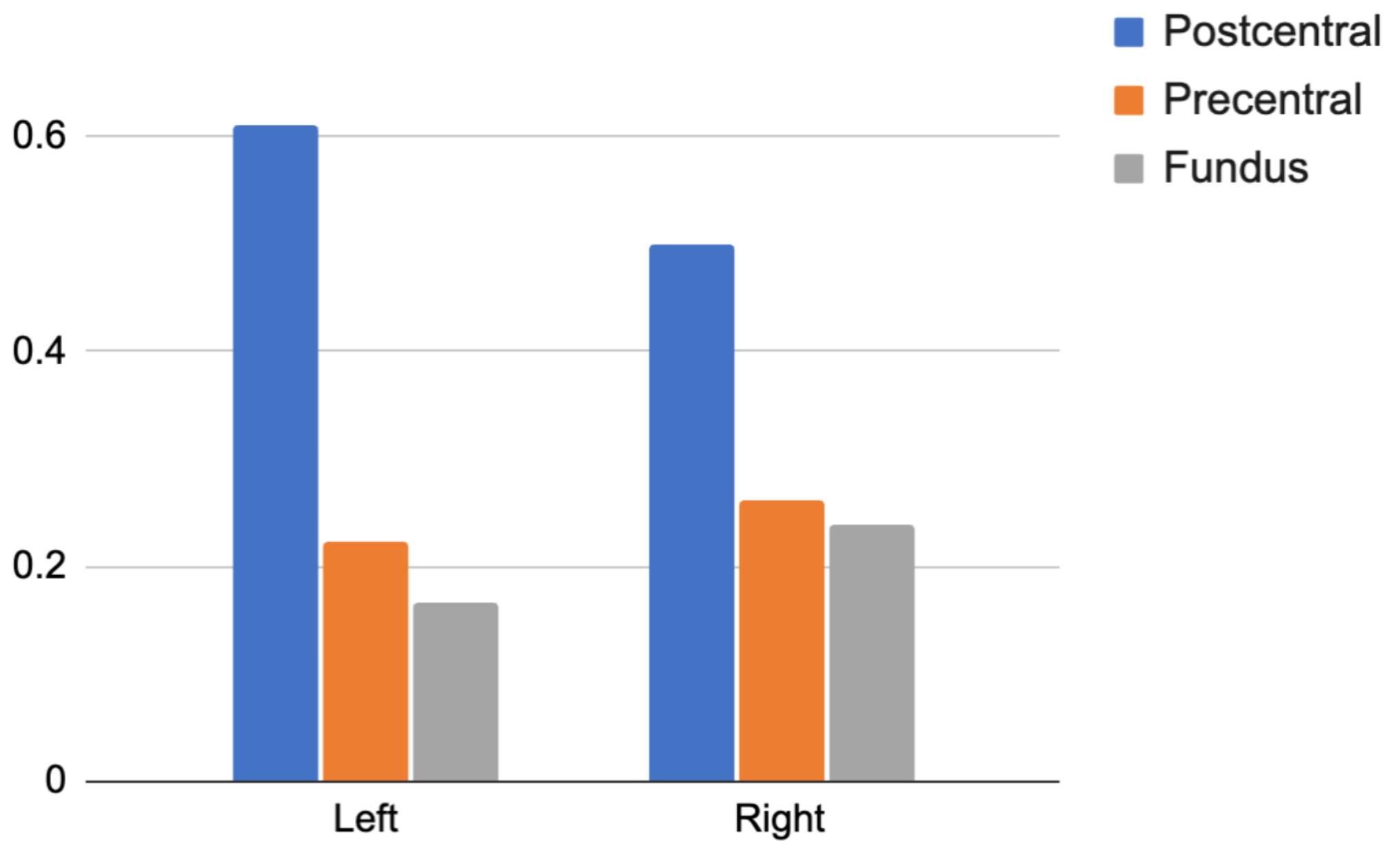
3b



Location	Anatomy	Side	Extent (voxels)	MNI coordinates (mm)		
				x	y	z
Central Sulcus (fundus, dorsal part)		R	819	16.13	20	-27 60
		L	681	15.28	-21	-30 63
Cerebellum (centred on Primary Fissure)		R	5676*	21.36	14	-60 -20
		L		20.96	-15	-60 -18
Brainstem (Ventral Respiratory Column)		L	706*	10.82	-8	-38 -45
		R		10.47	8	-38 -45
Cerebellum (hemispheric Lobule VIII)		L	763	14.38	-12	-66 -50
		R	1141	14.16	10	-68 -48

Results: Central Sulcus Localization

- Single-subject and single run analysis to identify the location of the central sulcus cluster in each hemisphere, in particular to distinguish between precentral and postcentral localizations suggested that clusters maxima are mainly found in the postcentral gyrus

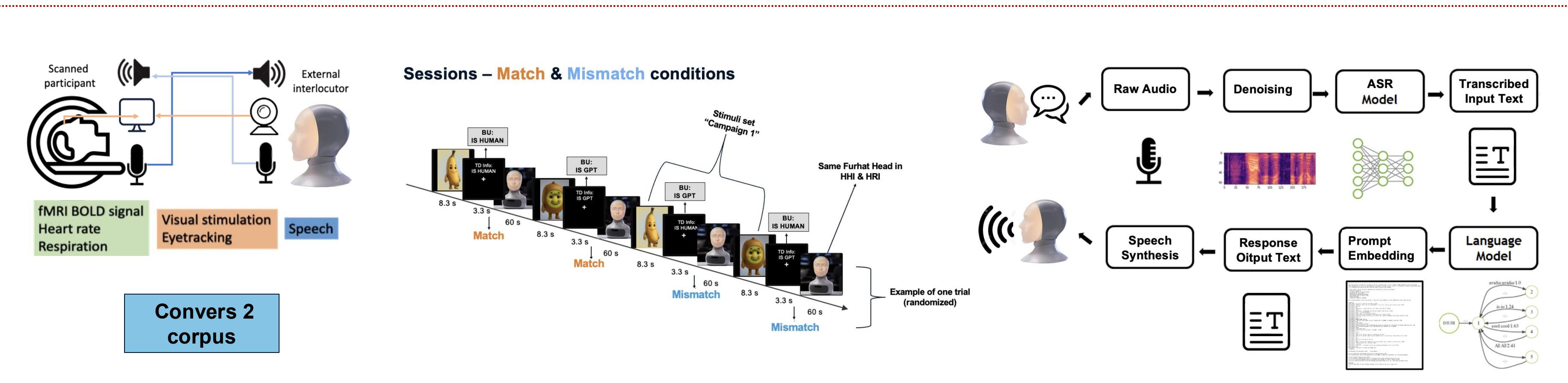
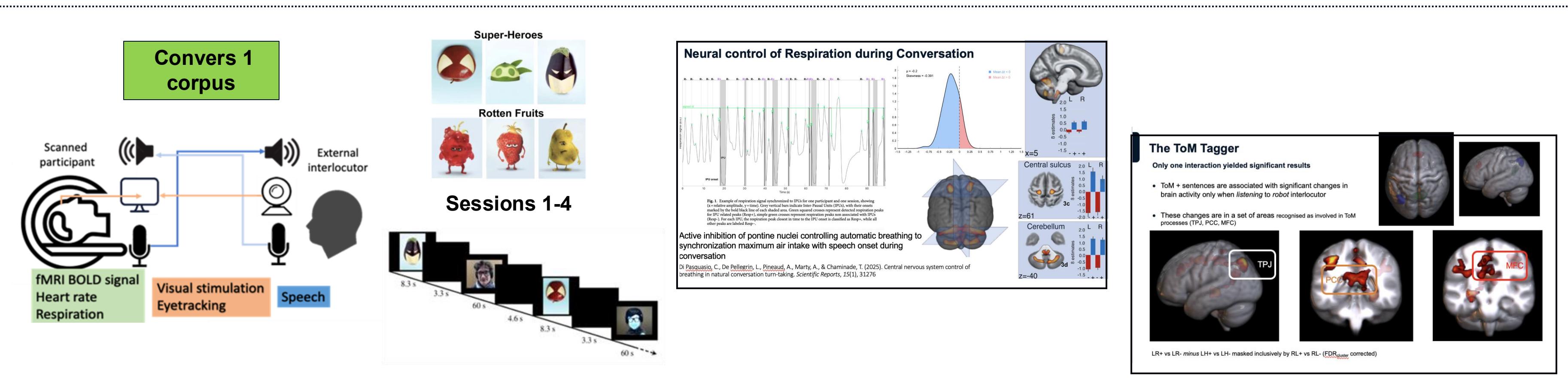


Conclusions

- Temporal dynamics between respiration and conversation (200ms) at a behavioural level
- **Central sulcus (mostly postcentral):** activation reflects *sensory processing* of chest/diaphragmatic input during maximum inspiration before speaking.
- **Brainstem (VRC):** increased response during speech-associated inhalation maxima; involved in generating and modulating the respiratory rhythm to permit speech
- **Cerebellum (lobules V–VI, VIII):** contributes to *timing* and preparatory aspects of vocal motor control; may help inhibit automatic respiratory rhythms

→ Speech preparation relies on an integrated sensory–motor network coordinating bodily feedback, respiratory rhythm regulation, and motor timing

Full Project Overview



Conversational Behavior Data - Hypotheses

01

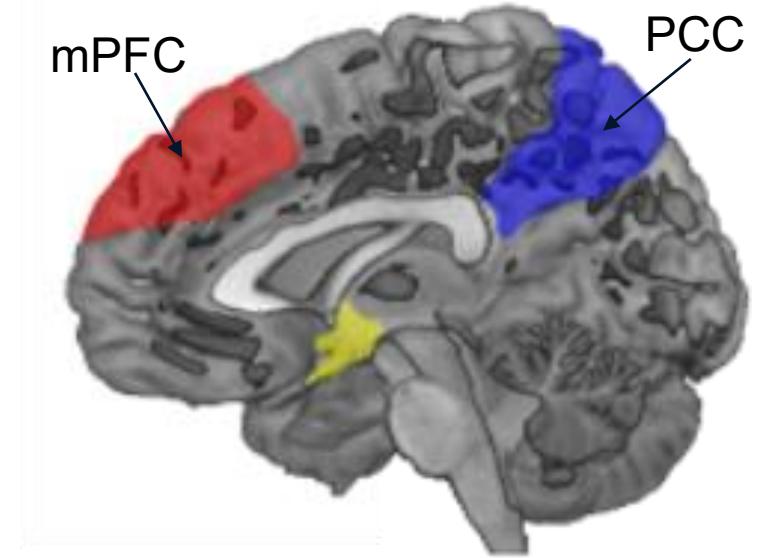
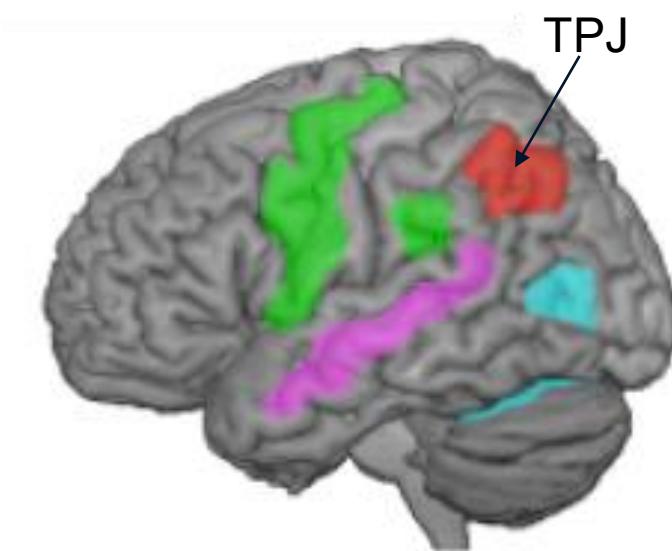
Does the human brain respond to Theory of Mind (ToM) utterances produced by robots during natural conversation?

03

How do neural responses to robot-produced ToM utterances differ from those elicited by human-produced ToM utterances?

02

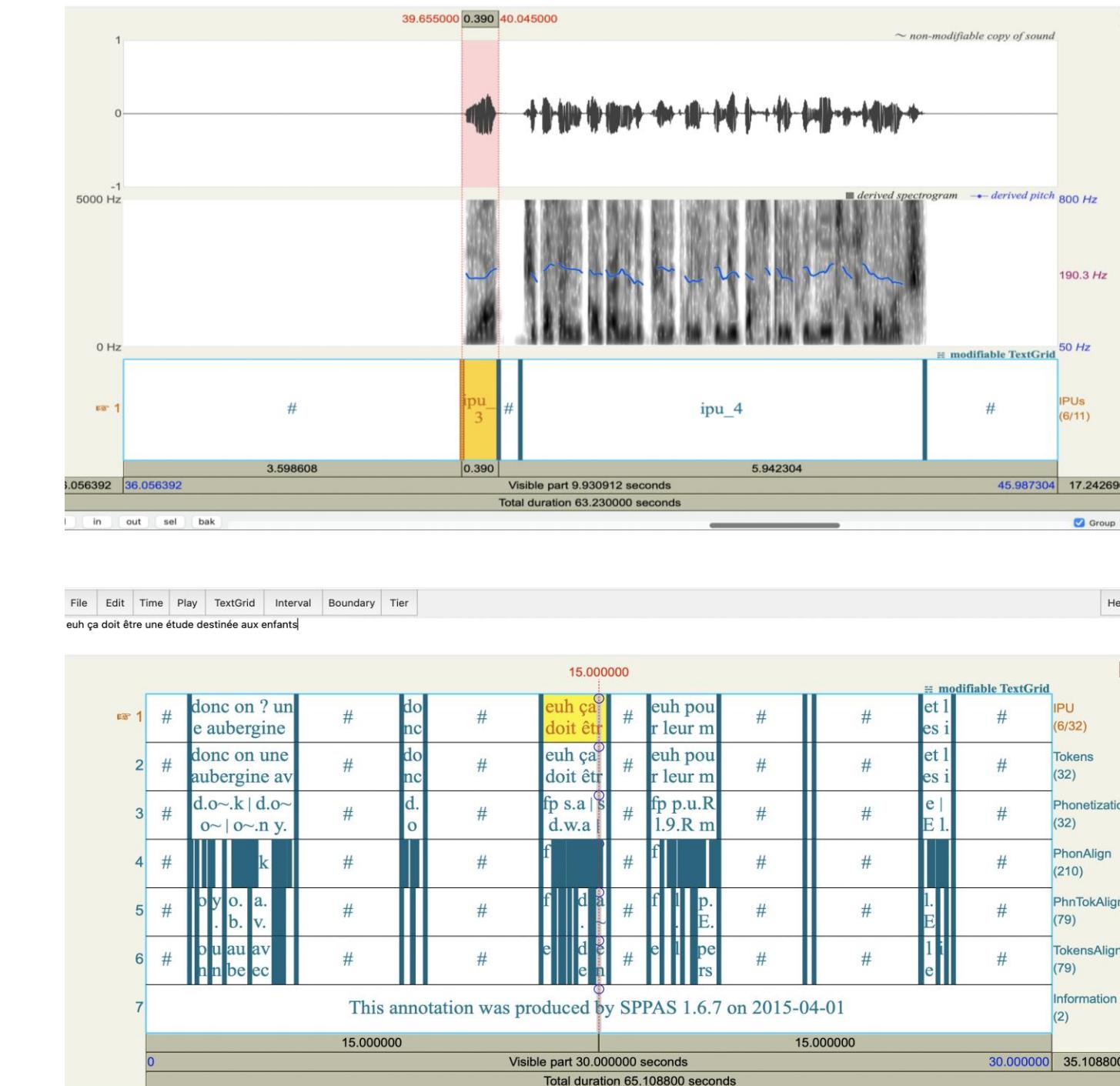
Are these responses localized within canonical ToM or broader “social cognition” brain regions?



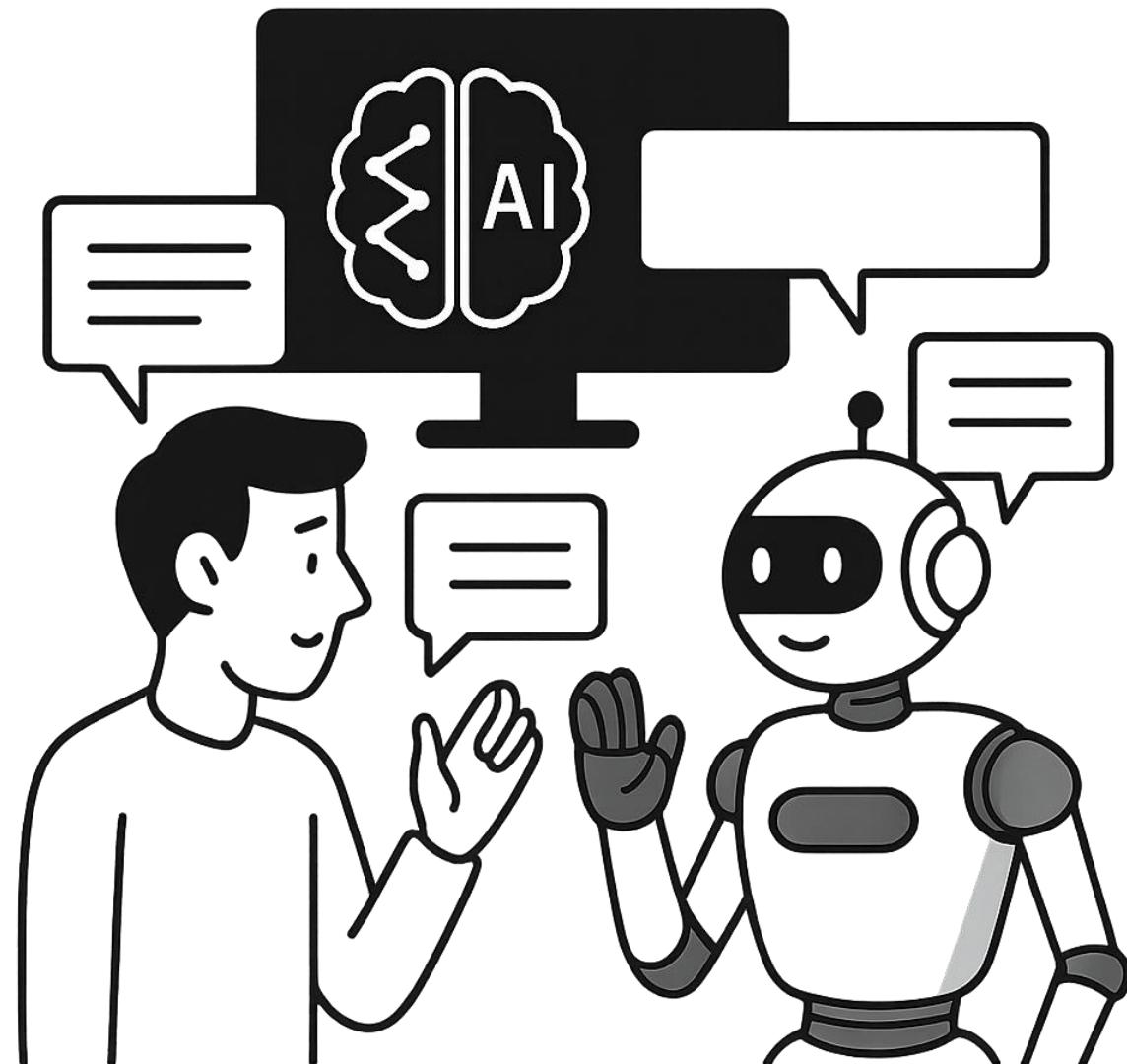
Building the ToM Tagger – The Data

Development of a methodology to leverage an LLM to identify ToM exchanges within transcripts of authentic conversations

Modality of data	Extracted variables per conversation
Neural data (fMRI)	Begin and end of conversations from logfiles
	BOLD signal for group and time series analysis
Speech and transcribed data	Number of IPUUs and tokens
	Time series and mean duration of conversational features: IPUUs,



Building the ToM Tagger



What is the ToM Tagger

- A process that analyses and tags conversational text for **Theory of Mind** expressions using an **LLM (GPT-4o)**

How does it work?

- **Loads** a designed prompt from a text file
- **Sends** the conversation data (.csv) to GPT for automatic annotation
- **Tags** the data based on GPT's response
- **Saves** the updated conversation into a new CSV file

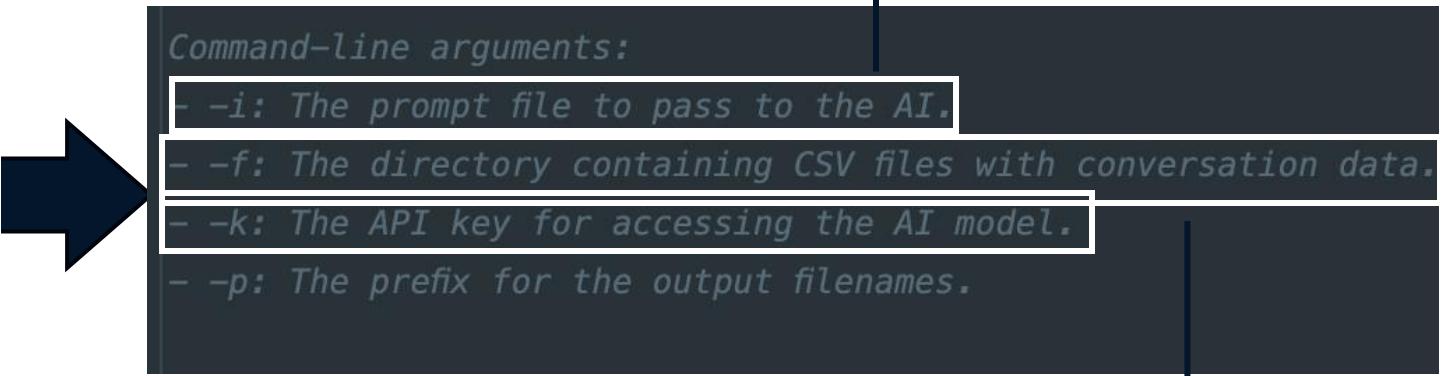
Tagging Procedure – Input

INPUT FILE (conversation Sub (1-25)-Sess (1-4)-Trial (1-6)-Conv (H/R) → csv format)

	0	1	2	3
0	1.90494	9.44998	Interlocutor	ah ça marche pas pour la ... pour la les légumes tu sais ce que tu me disais ... parce que euh ... la pomme c'est un fruit
1	10.91501	11.76998	Participant	ah ben oui c'est vrai
2	11.59501	15.52998	Interlocutor	et et on a la trouve dans la dans la ... on la trouve en \$Spiderman\$
3	16.16499	16.66998	Participant	et ouais
4	16.95002	18.02998	Interlocutor	ou en \$Deadpool\$ je sais pas
5	19.49002	19.82998	Participant	ouais
6	25.83501	27.70998	Interlocutor	peut-être que les premiers ils protègent les deuxièmes
7	28.97341	33.20998	Participant	ben je sais pas parce que y a des *
8	33.47501	37.50998	Interlocutor	ouais ... donc ils ont besoin d'être protégés
9	39.06	43.08998	Participant	euh ... ou bien euh ... je ... ouais
10	44.55501	46.42998	Interlocutor	ou alors ils sont tristes parce qu'on les oublie
11	48.38998	49.18998	Participant	ouais aussi



prompt_fr.txt



Input_file_test_1_fr.csv
Input_file_test_2_fr.csv
Input_file_test_3_fr.csv
Input_file_test_4_fr.csv



samples

Tagging Procedure – Output

GPT-4o GENERATED RESPONSE BASED ON PROMPT

```
```0  Interlocutor: ah ça marche pas pour la ... pou...
1 Participant: ah ben oui c'est vrai
2 Interlocutor: et et on a la trouve dans la dan...
3 Participant: et ouais
4 Interlocutor: ou en $Deadpool$ je sais pas
5 Participant: ouais
6 Interlocutor: <ToM> peut-être que les premiers ils p... </ToM>
7 Participant: ben je sais pas parce que y a des...
8 Interlocutor: <ToM> ouais ... donc ils ont besoin d'... </ToM>
9 Participant: euh ... ou bien euh ... je ... ouais
10 Interlocutor: <ToM> ou alors ils sont tristes parce ... </ToM>
11 Participant: ouais aussi
```



OUTPUT FILE → csv format)

0	1	2	Tag	
1.90494	9.44998	Interlocutor	0.0	ah ça marche pas pour la ... pour la les légumes tu sais ce que tu me disais ... parce que euh ... la pomme c'est un fruit
10.91501	11.76998	Participant	0.0	ah ben oui c'est vrai
11.59501	15.52998	Interlocutor	0.0	et et on a la trouve dans la dans la ... on la trouve en \$Spiderman\$
16.16499	16.66998	Participant	0.0	et ouais
16.95002	18.02998	Interlocutor	0.0	ou en \$Deadpool\$ je sais pas
19.49002	19.82998	Participant	0.0	ouais
25.83501	27.70998	Interlocutor	1.0	peut-être que les premiers ils protègent les deuxièmes
28.97341	33.20998	Participant	0.0	ben je sais pas parce que y a des *
33.47501	37.50998	Interlocutor	1.0	ouais ... donc ils ont besoin d'être protégés
39.06	43.08998	Participant	0.0	euh ... ou bien euh ... je ... ouais
44.55501	46.42998	Interlocutor	1.0	ou alors ils sont tristes parce qu'on les oublie
48.38998	49.18998	Participant	0.0	ouais aussi

 tag\_Input\_file\_test\_1\_fr.csv  
 tag\_Input\_file\_test\_2\_fr.csv  
 tag\_Input\_file\_test\_3\_fr.csv  
 tag\_Input\_file\_test\_4\_fr.csv



# Methods – fMRI Analysis

## 1) fMRI data conditions (*first-level analysis*)

Post-tagging 8 fMRI conditions:

- Nature of the interlocutor:

Human (H)  
Robot (R)

X

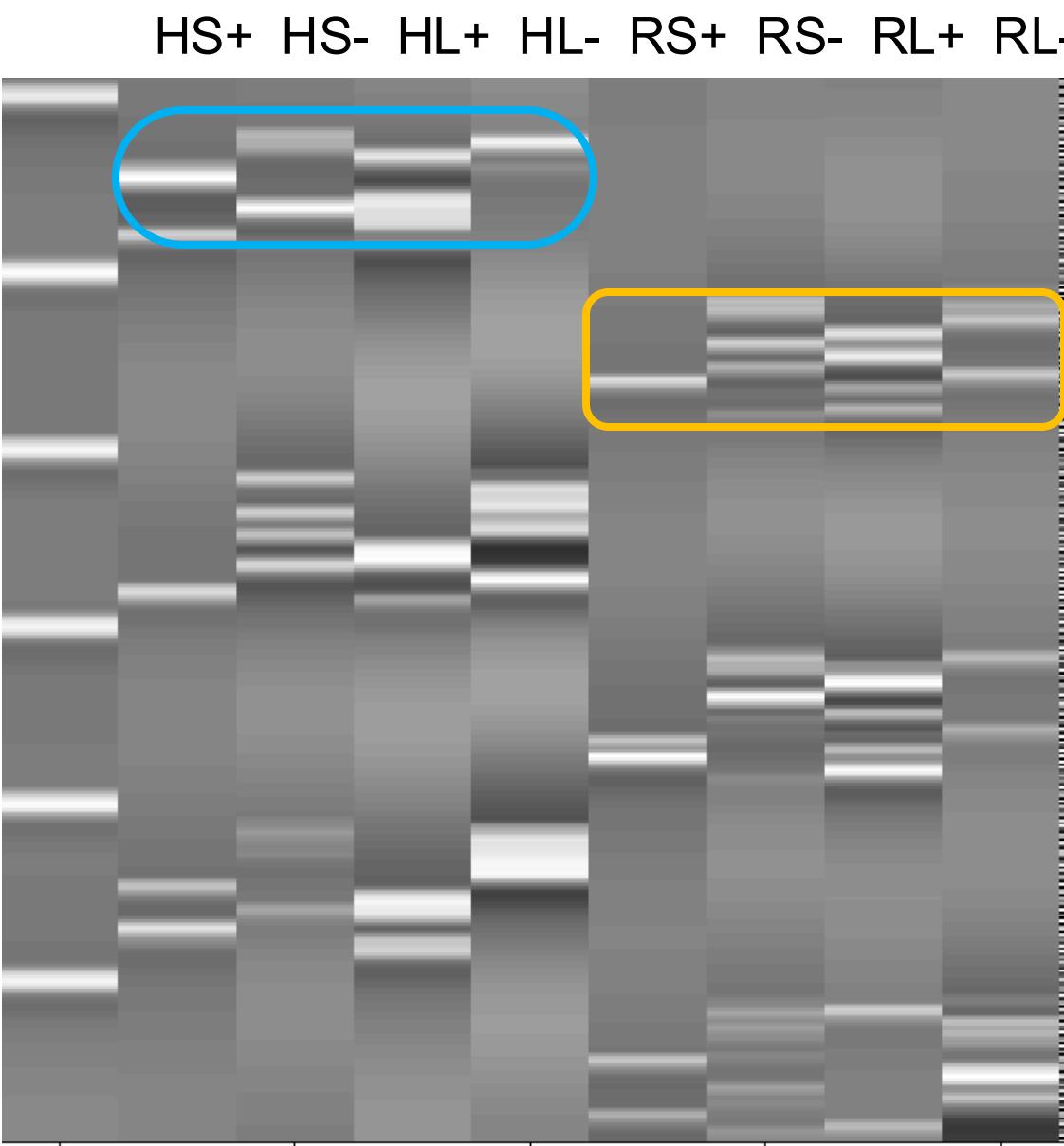
- Role of the participant:

Speaker (S)  
Listener (L)

X

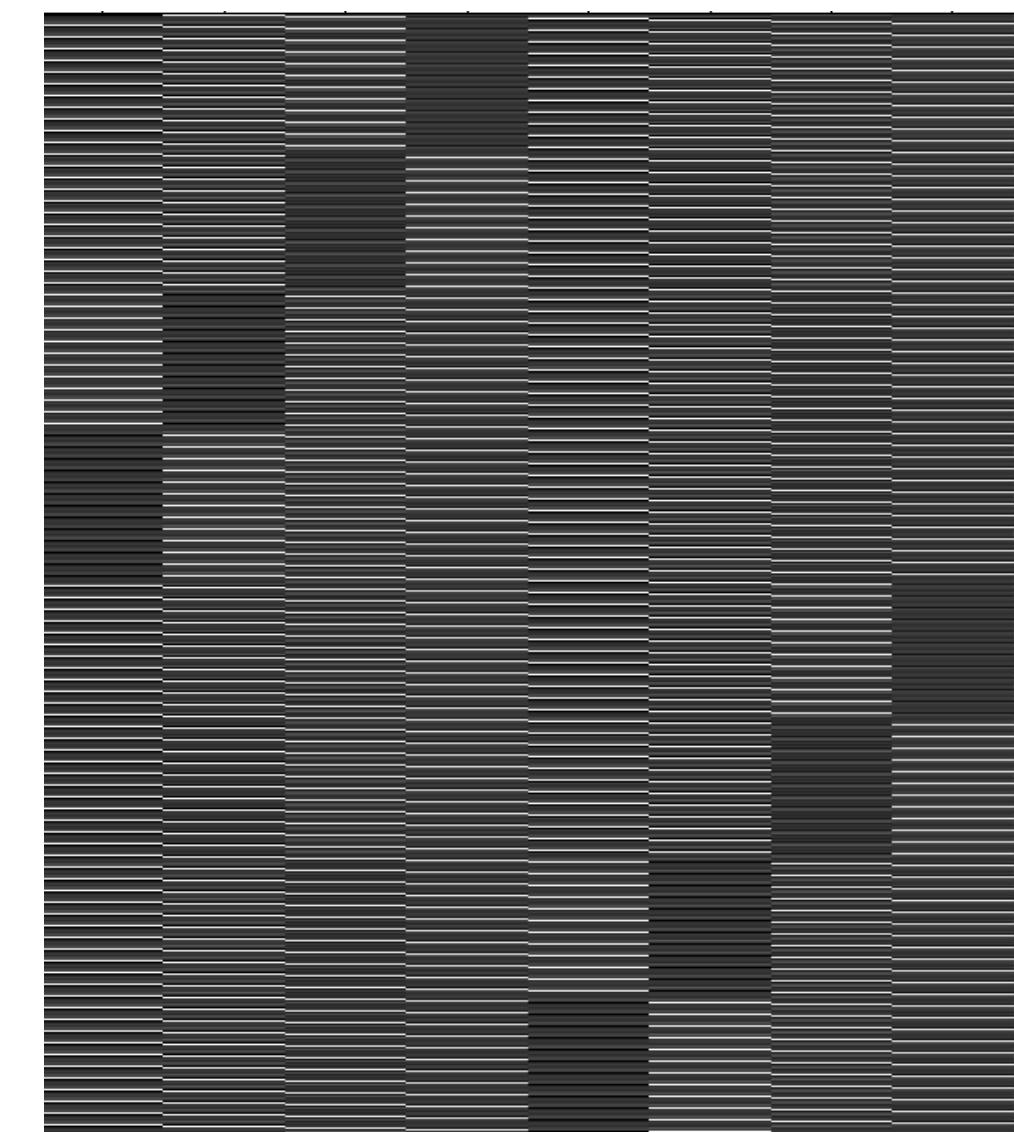
- IPU ToM tagging:

Tagged (+)  
Untagged (i)



## 2) Full ANOVA (*Second-level analysis*)

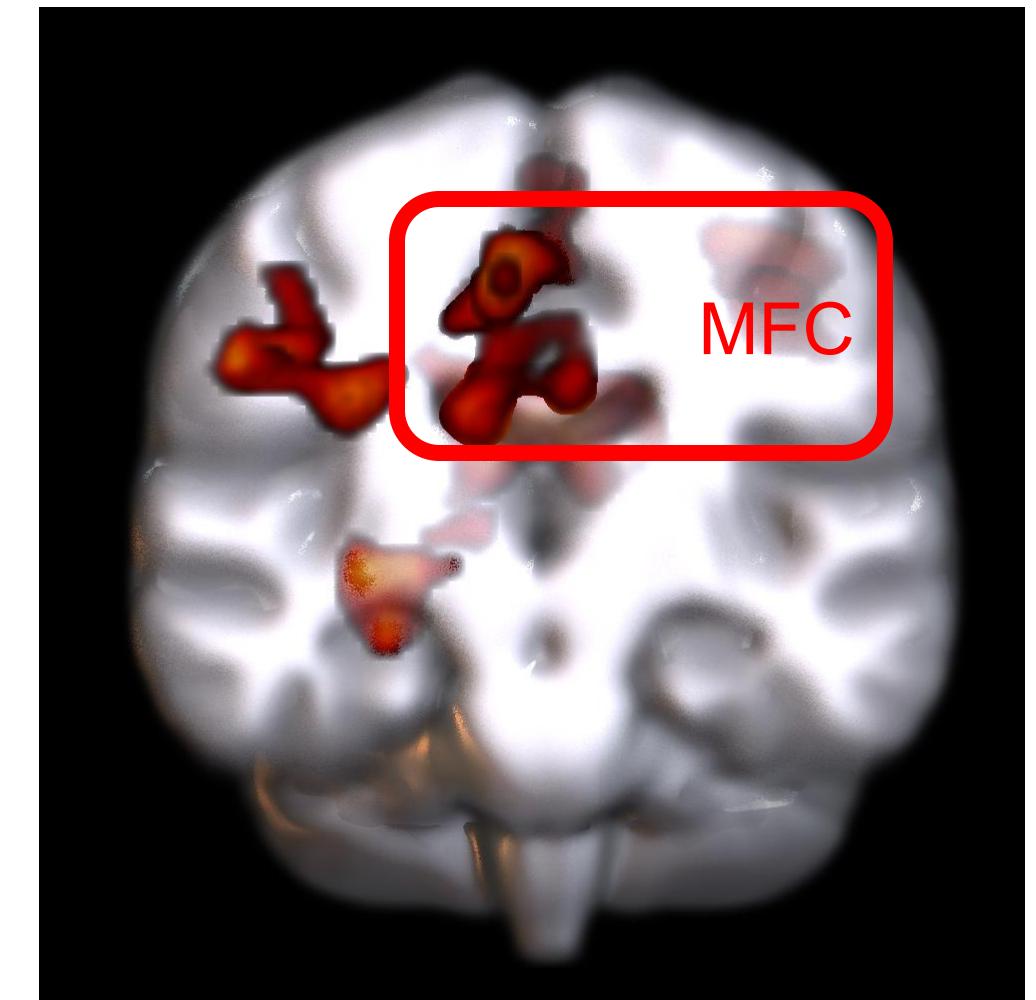
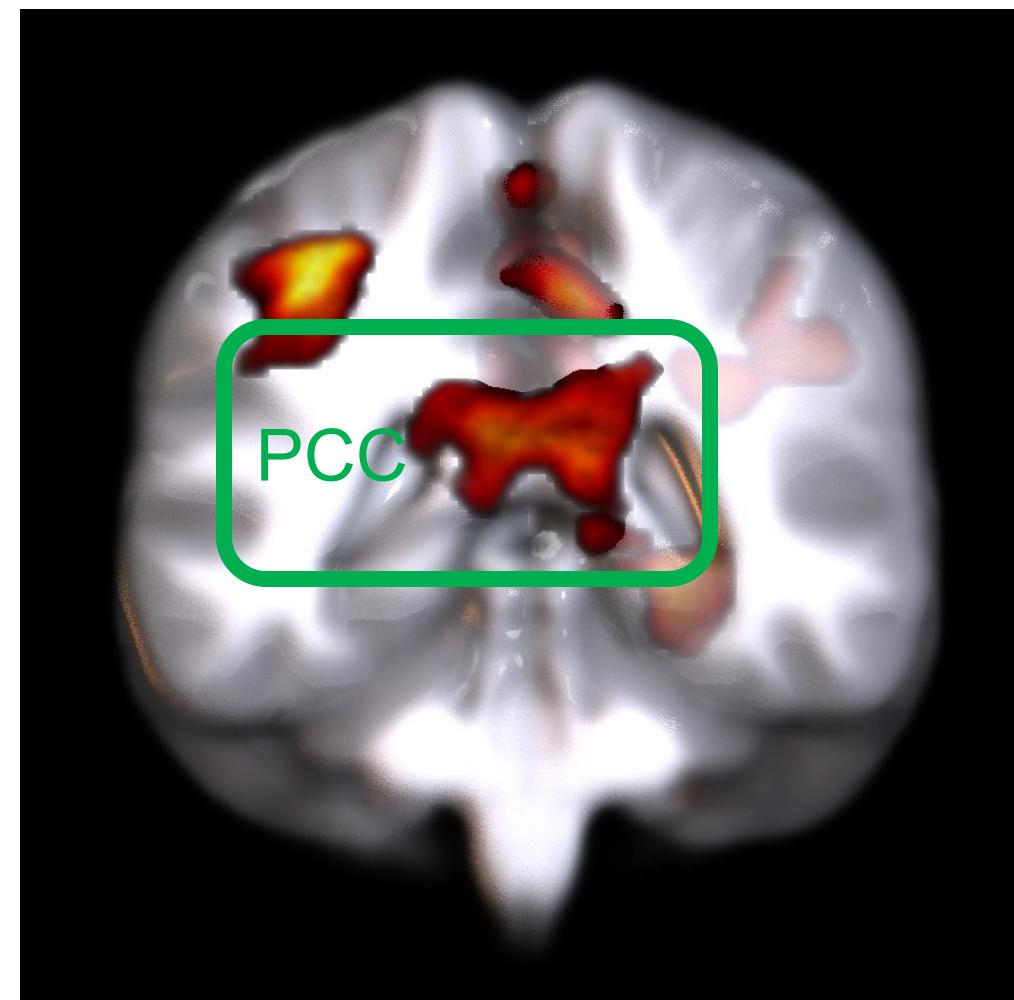
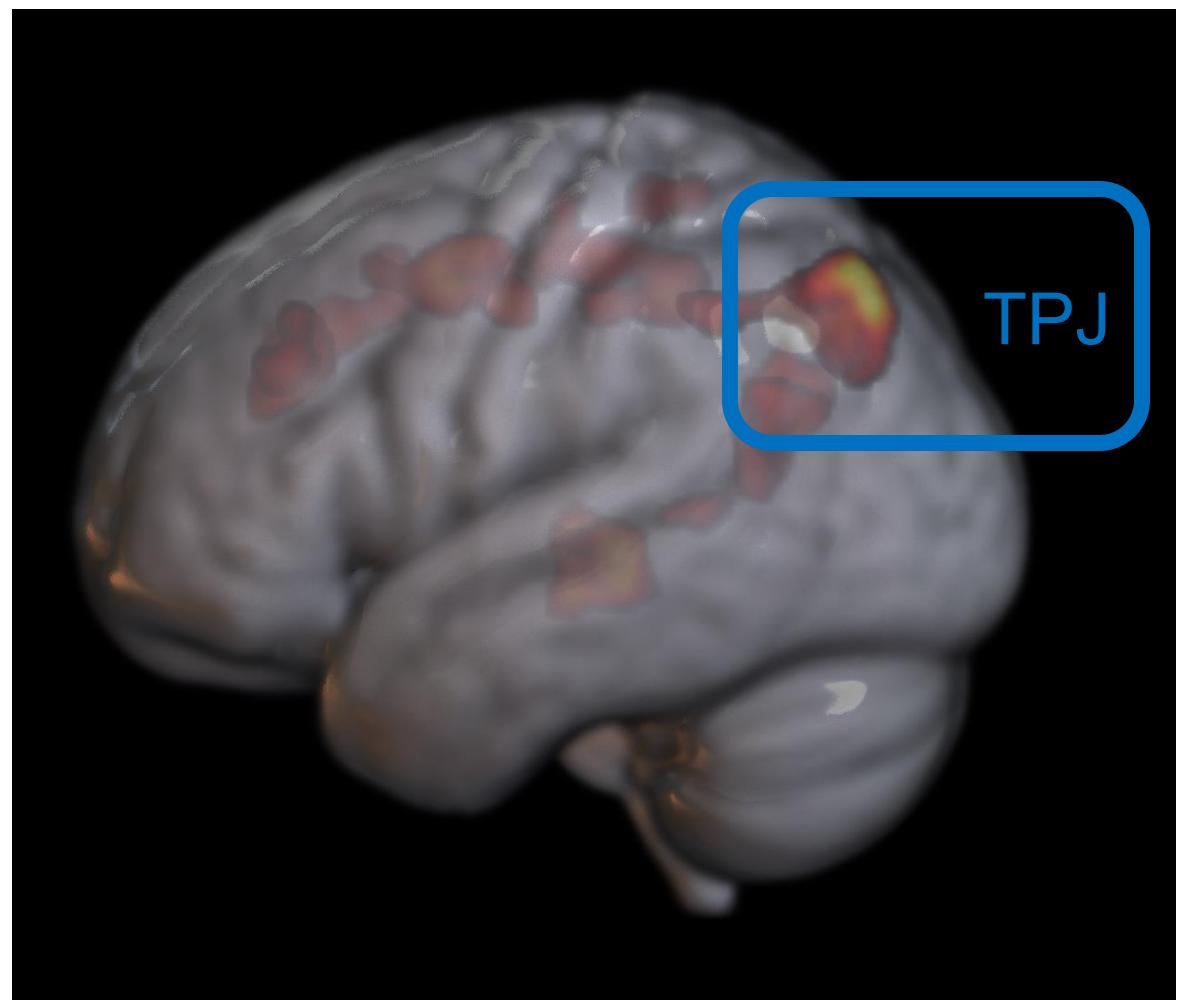
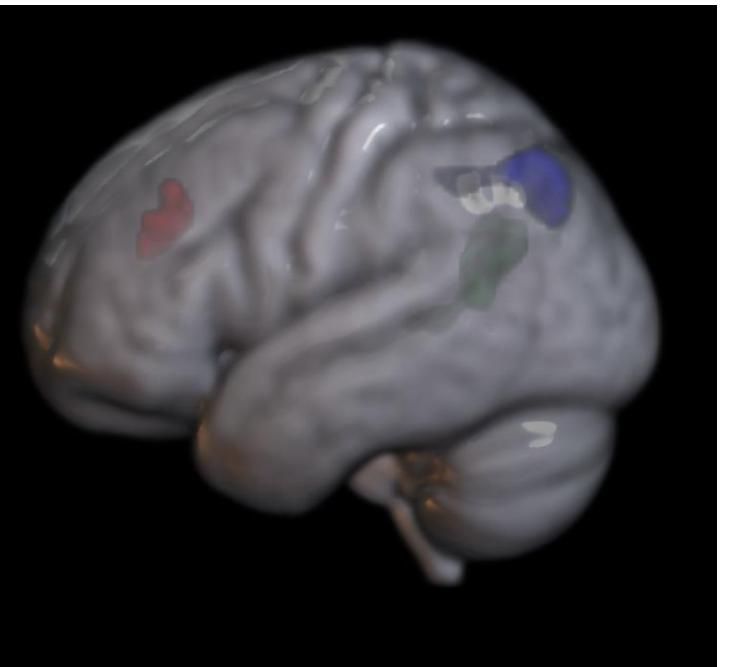
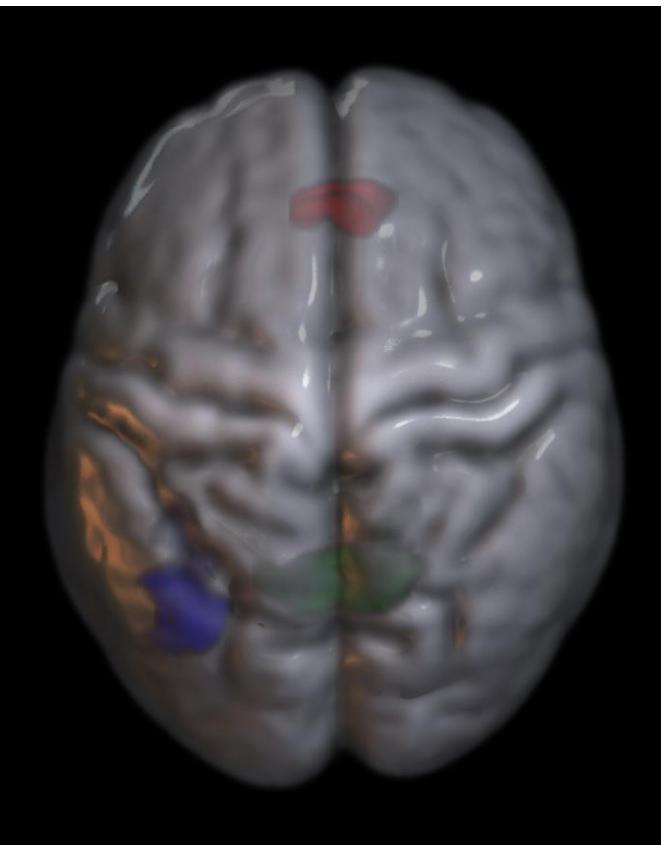
HS+ HS- HL+ HL- RS+ RS- RL+ RL-



# Results from standard GLM analysis

An Interactions yielded significant results

- ToM + sentences are associated with significant changes in brain activity when *listening* to *robot* interlocutor
- These changes are in a set of areas recognised as involved in ToM processes (TPJ, PCC, MFC)



LR+ vs LR- *minus* LH+ vs LH- masked inclusively by RL+ vs RL- (FDR<sub>cluster</sub> corrected)

## TemporoParietal Junction TPJ

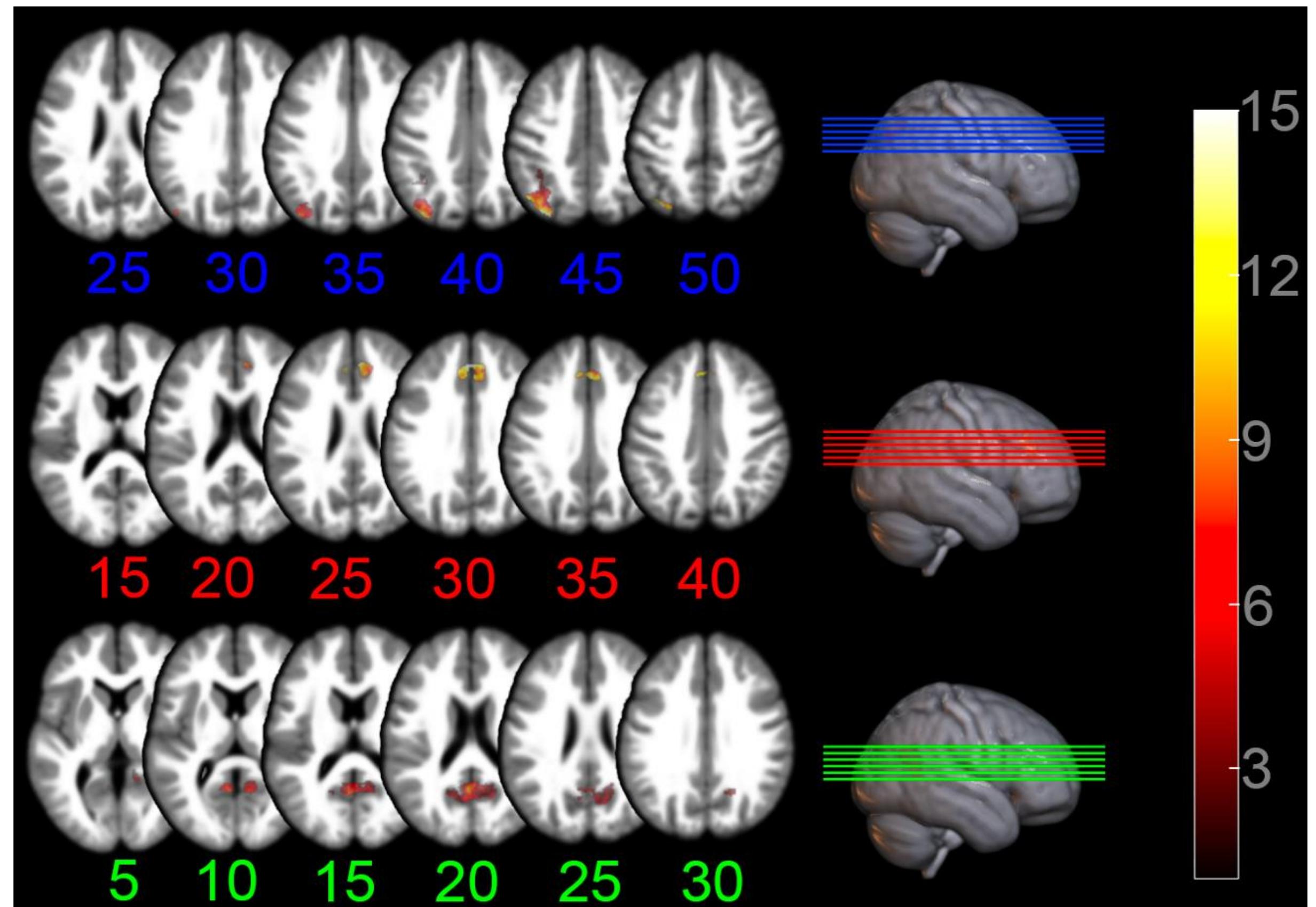
Samson, D., Aupperly, I., Chiavarino, C. *et al.* Left temporoparietal junction is necessary for representing someone else's belief. *Nat Neurosci* 7, 499–500 (2004). <https://doi.org/10.1038/nn1223>

## Medial Frontal Cortex MFC

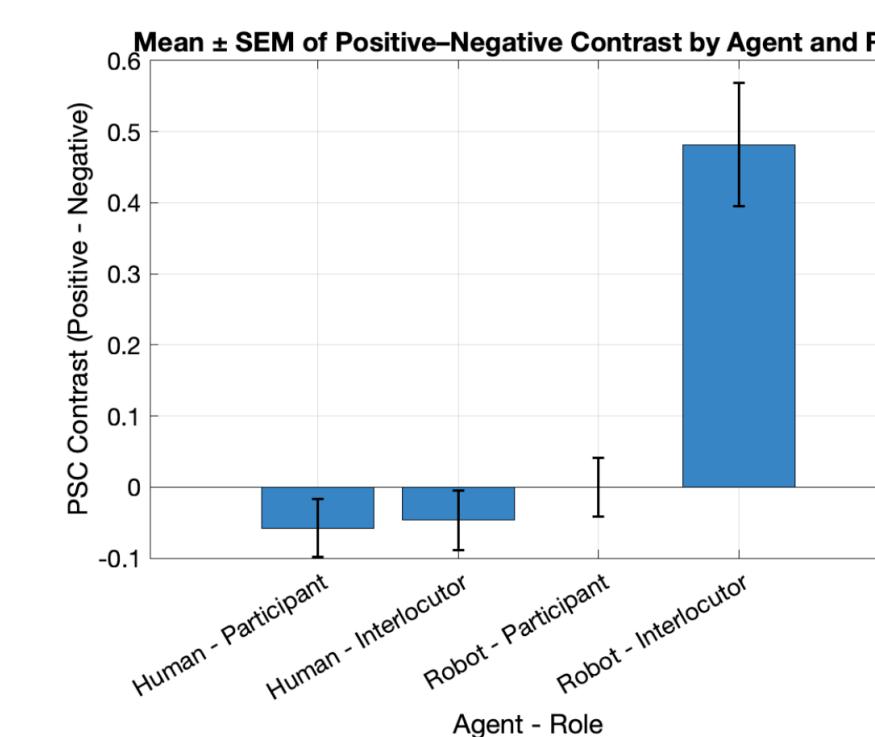
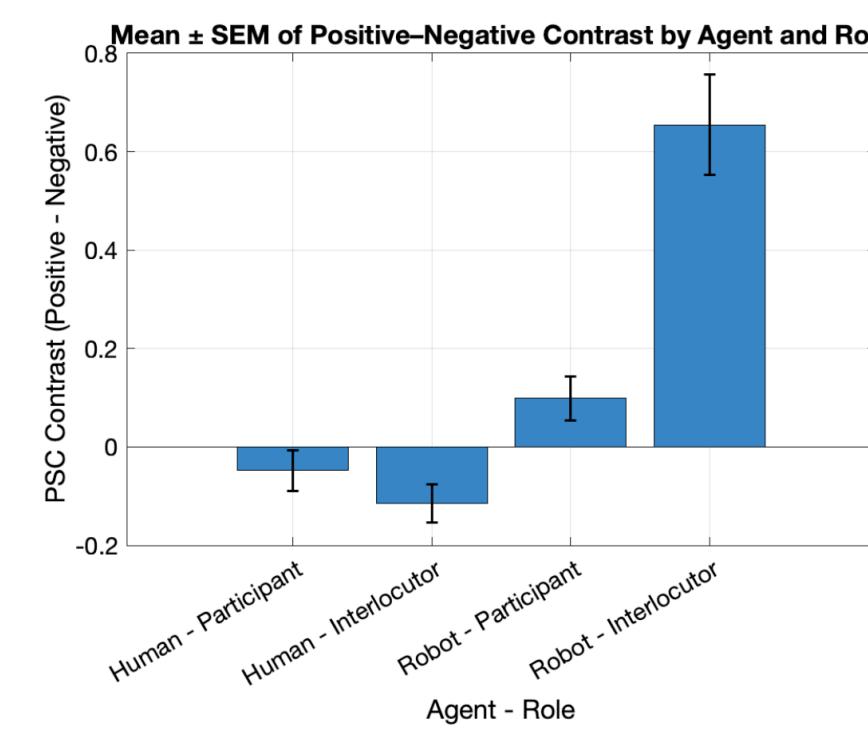
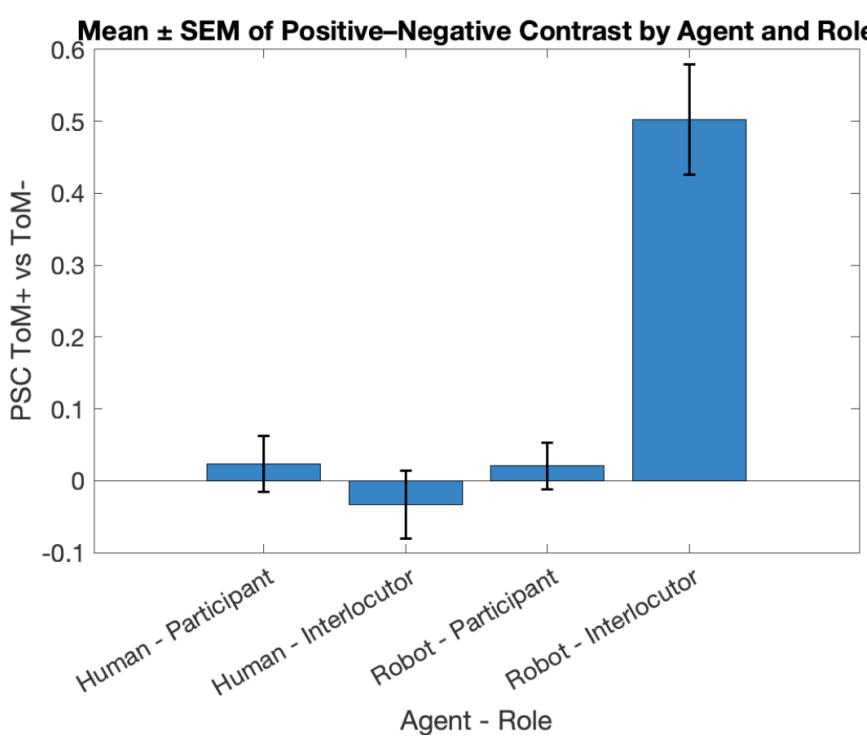
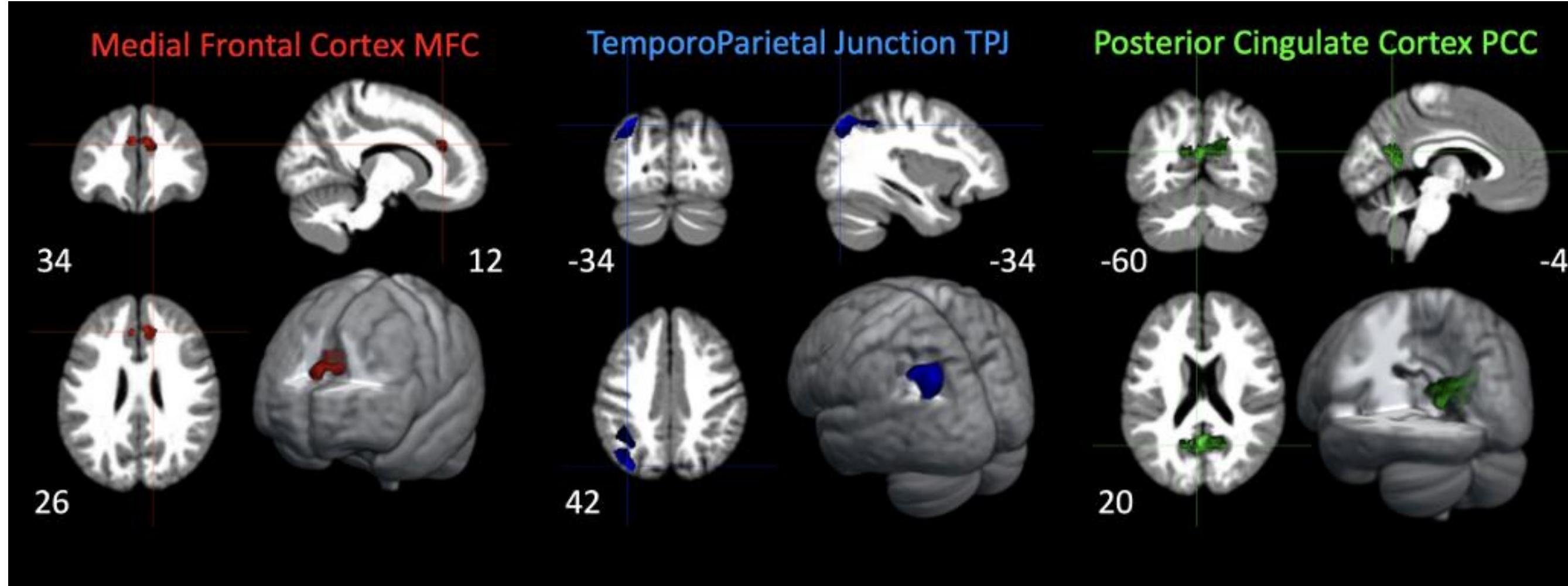
Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci*. 2006 Apr;7(4):268-77. doi: 10.1038/nrn1884. PMID: 16552413.

## Posterior Cingulate Cortex PCC

Spatola, N., Chaminade, T. Precuneus brain response changes differently during human–robot and human–human dyadic social interaction. *Sci Rep* 12, 14794 (2022). <https://doi.org/10.1038/s41598-022-14207-9>



# Results from Regions of Interest

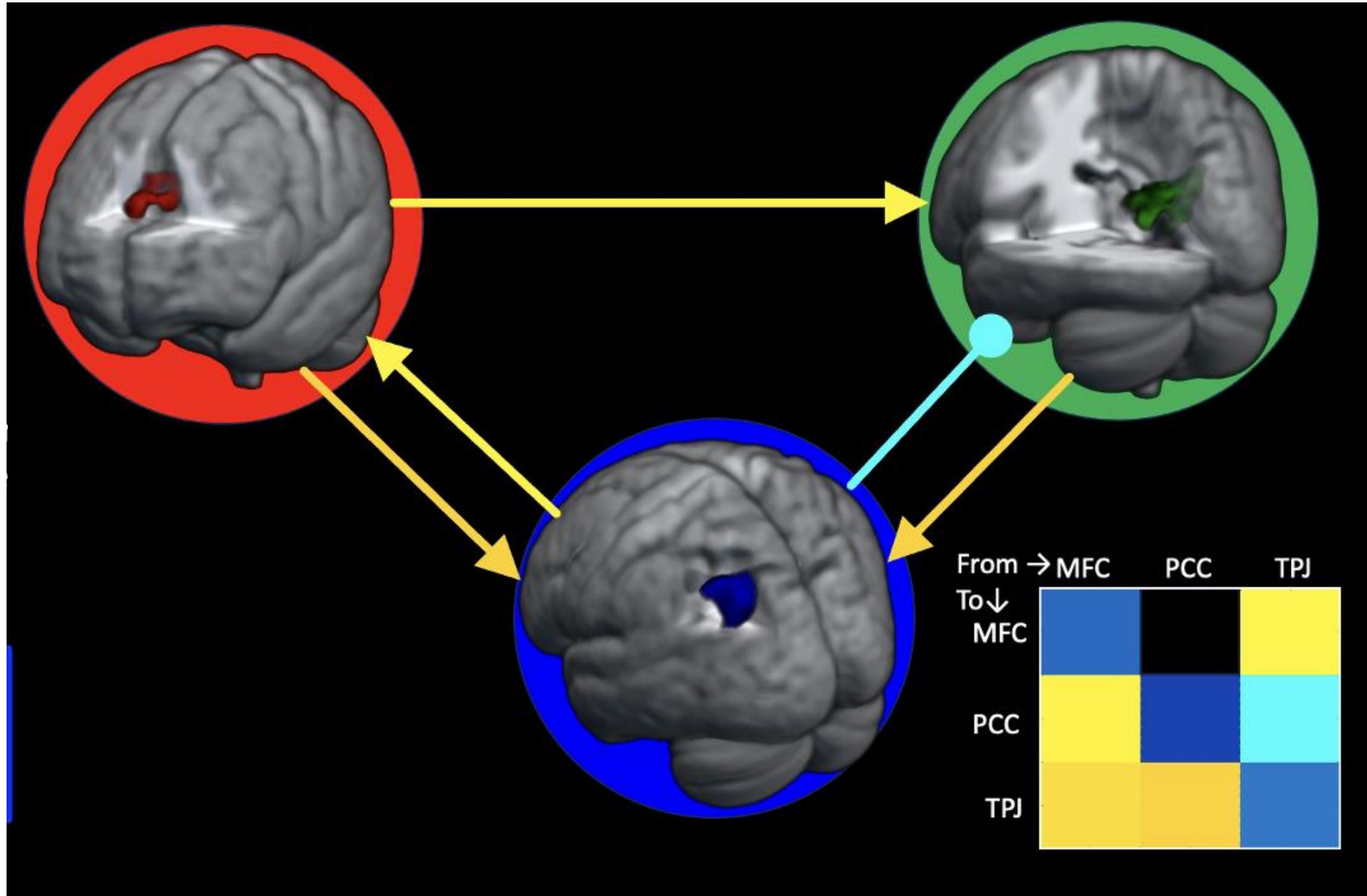


PSC was extracted at the individual level and explored statistically:

- Agent x Role by ToM significant in all ROIs
- Response to ToM is increased in the condition Listening to Robot
- Human Responses not affected by tagging

# Effective Connectivity with DCM

1) Static ROI-to-ROI connections for this experimental paradigm (Matrix A)

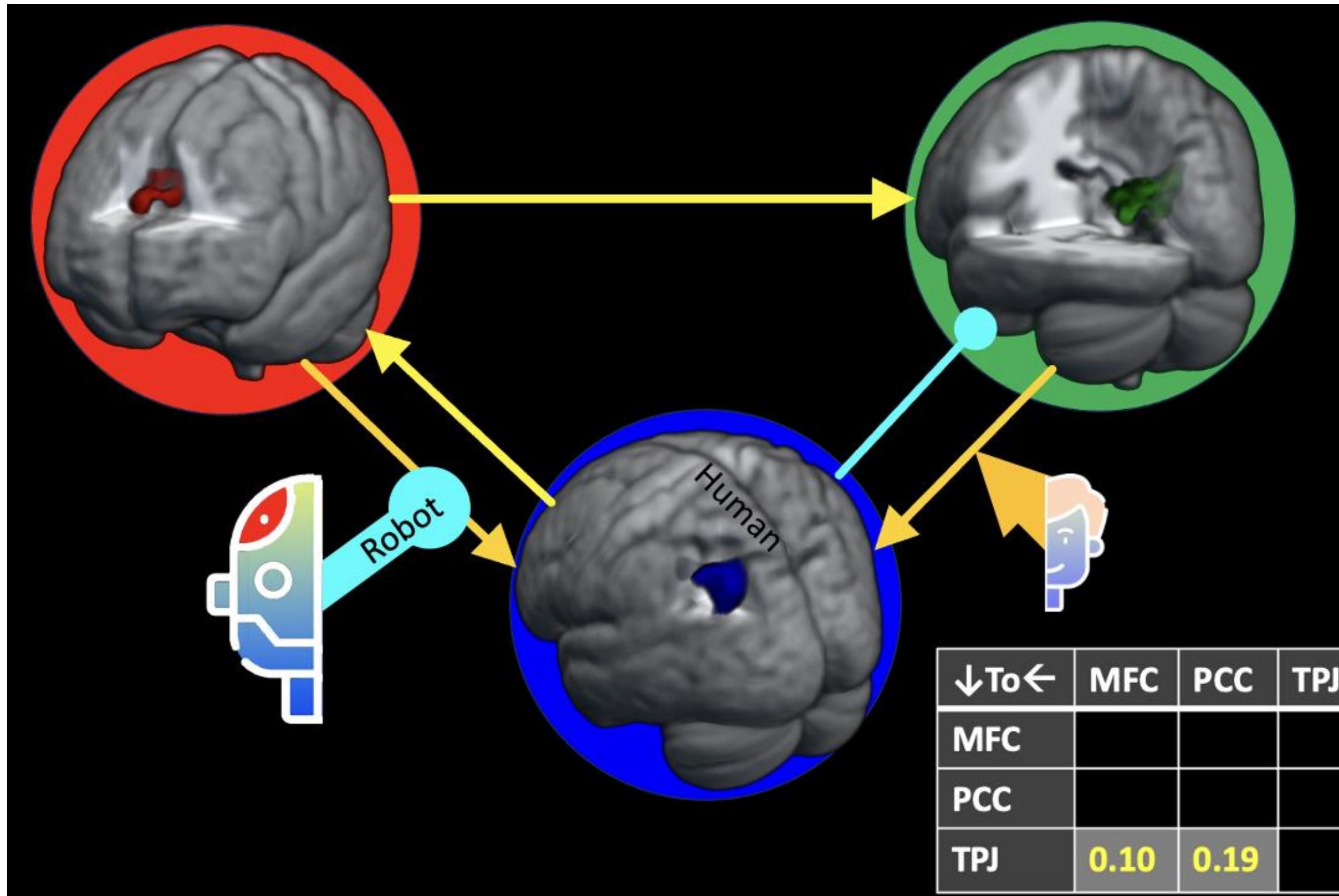


- PCC → MFC connection not significant
- Only TPJ is fully connected in the network

**The left TPJ has a central role in the natural conversation mentalizing network**

# Effective Connectivity with DCM

2) How the Agent (Robot<sub>+vs-</sub> & Human<sub>+vs-</sub>) affects the fixed connections



PCC → TPJ: Human	0.81	(+0.07)
: Robot	NC	
MFC → TPJ: Human	NC	
: Robot	0.10	(-0.12)

The nature of the agent affects connections to the left Temporo Parietal Junction, confirming its central function

# Conclusions

- No significant main effect, in all conditions

## Speaking:

- No significant ToM effects, regardless of interlocutor.
- Suggests **mentalizing demands during speech production** are not strongly modulated by whether the partner is human or robot.

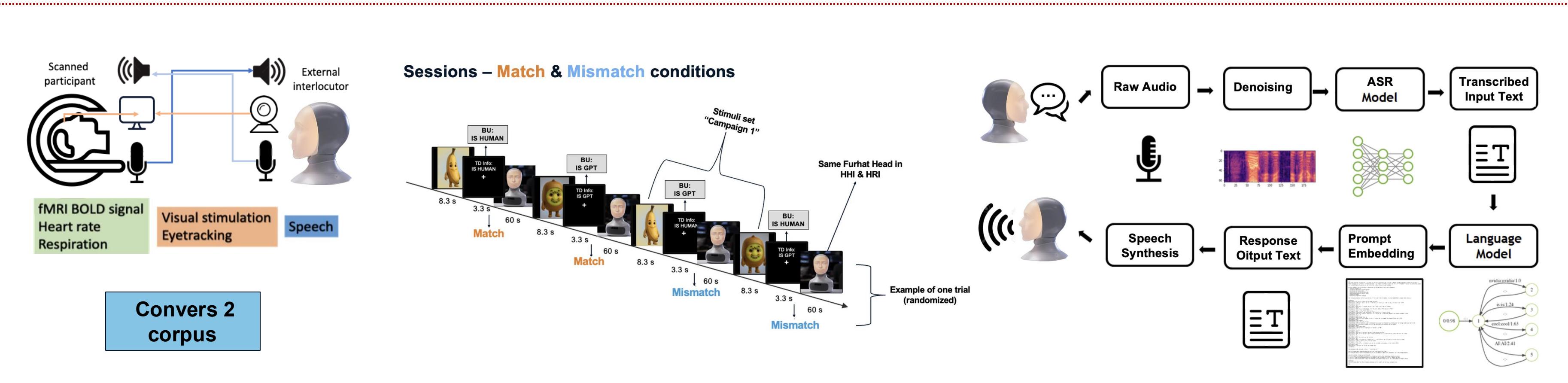
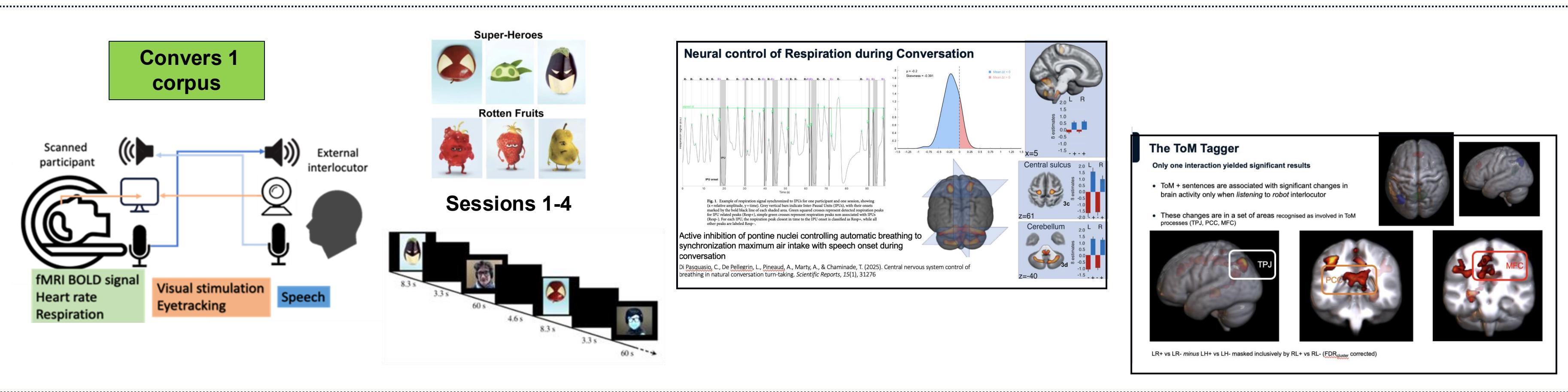
## Listening:

- **ToM effects emerge selectively for the Interaction Robot interlocutor ( $R+ > R-$ ) x ToM (ToM+ vs ToM-)**
- Indicates that when *listening* to a robot, the brain engages ToM-related processes more strongly than when listening to a human.

## Overall implication:

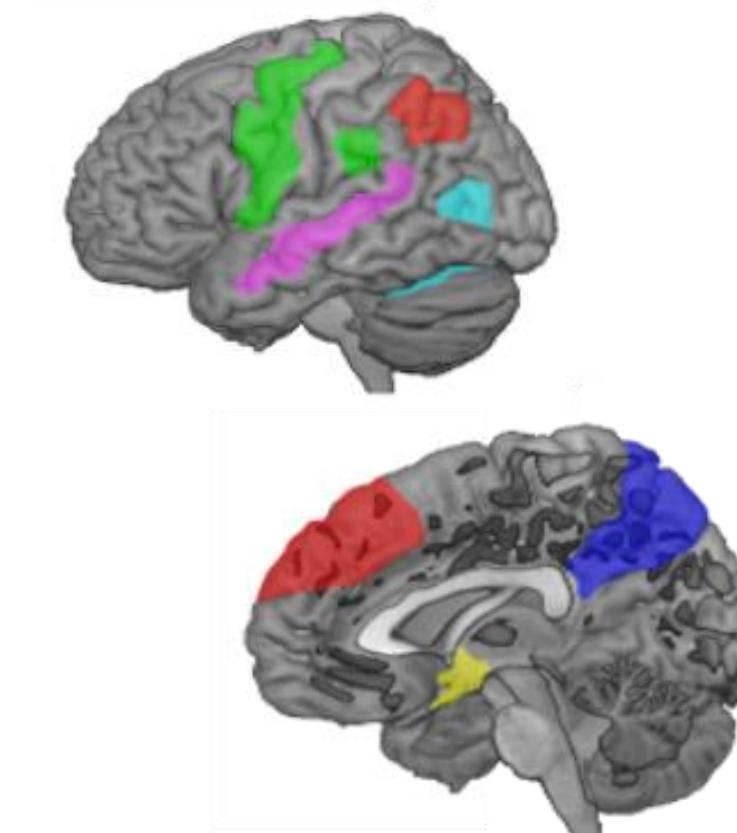
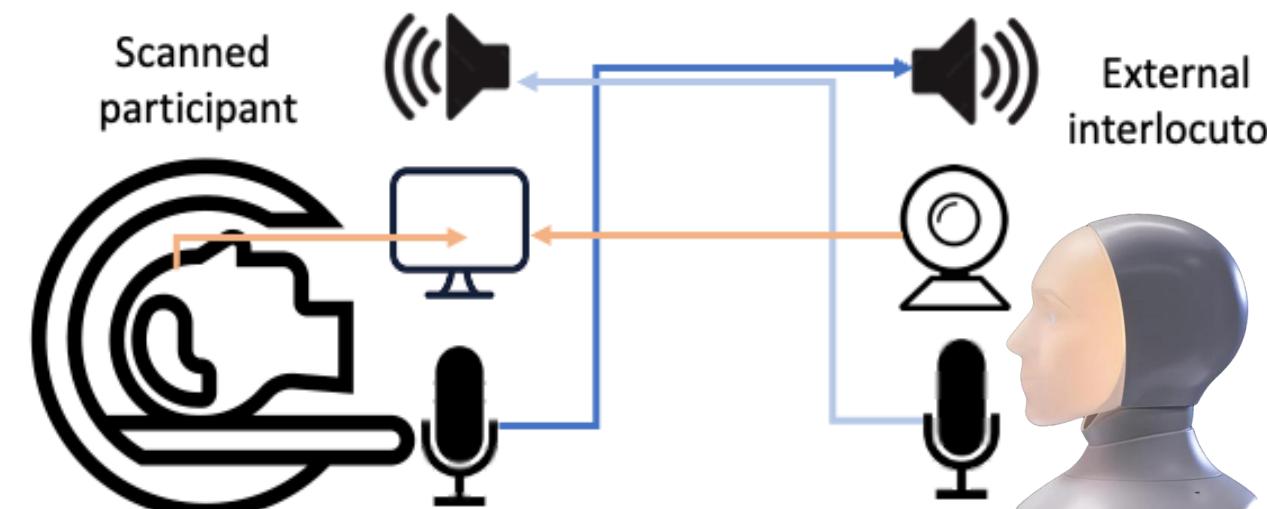
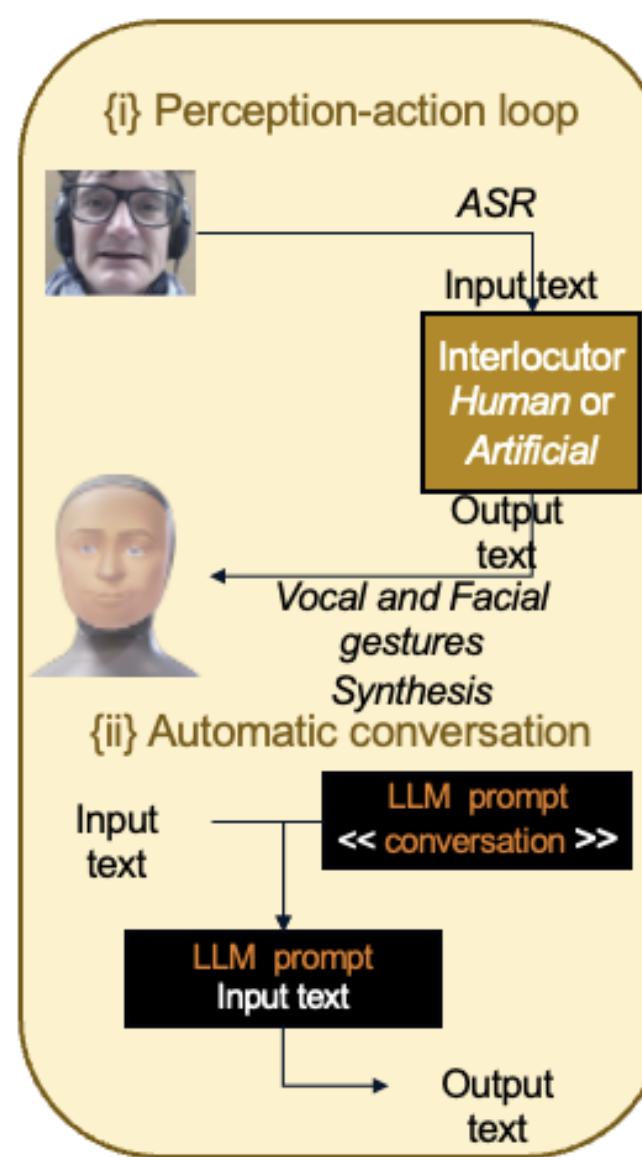
- This may reflect that **understanding a robot's perspective requires additional inferential effort** compared to a human partner through dynamics controlling Left TPJ activity
- Different stances (Dennett) when interacting with a Human or a Robot

# Full Project Overview



# Hypotheses

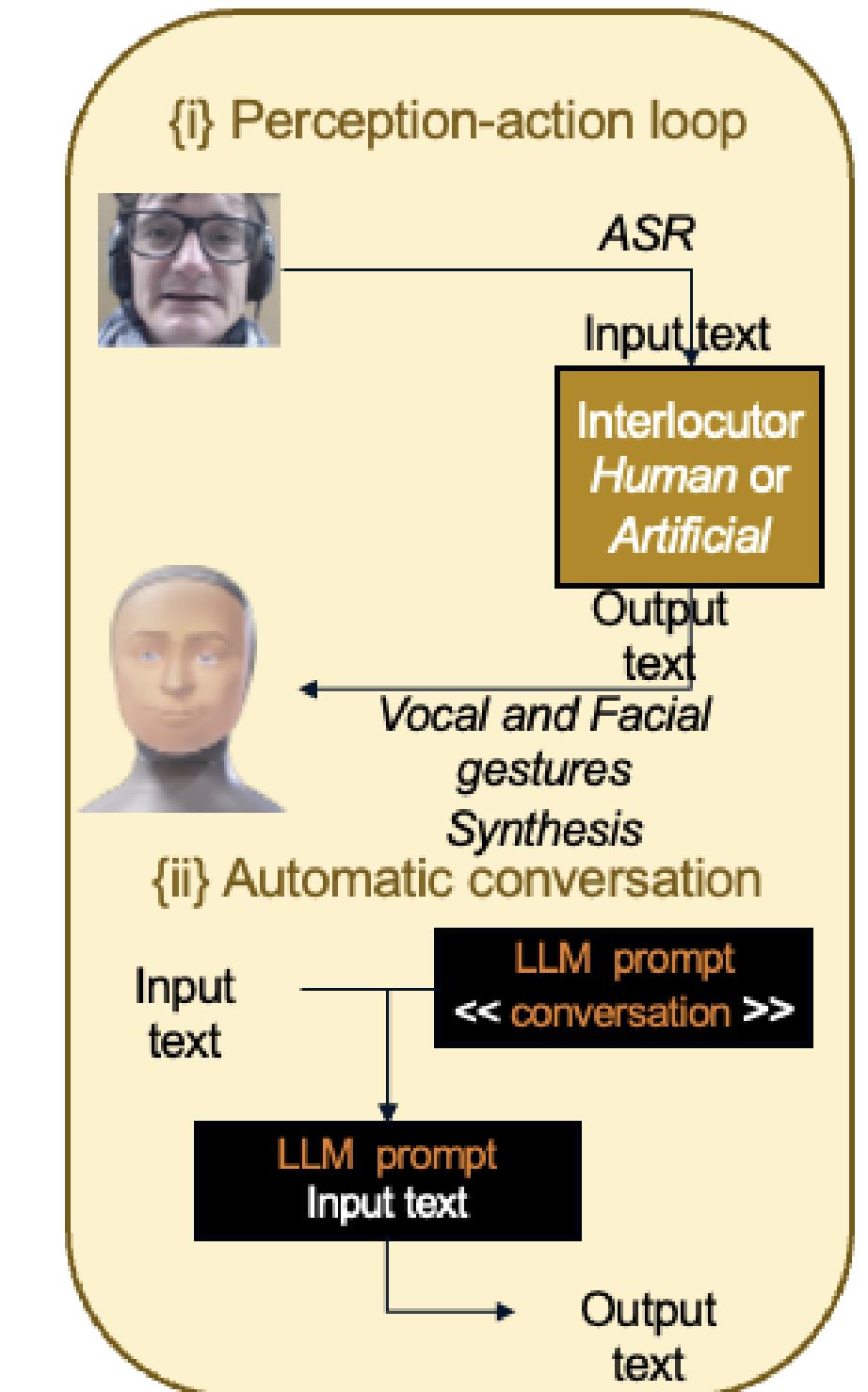
- Using human-like behaviors and minimizing visible differences between the Human and Robot Interlocutors may trigger activation in brain areas related to social cognition (and intentional stance) when interacting with the robotic agent (top-down effect of beliefs)



# Aim

## Control of *bottom-up* elements

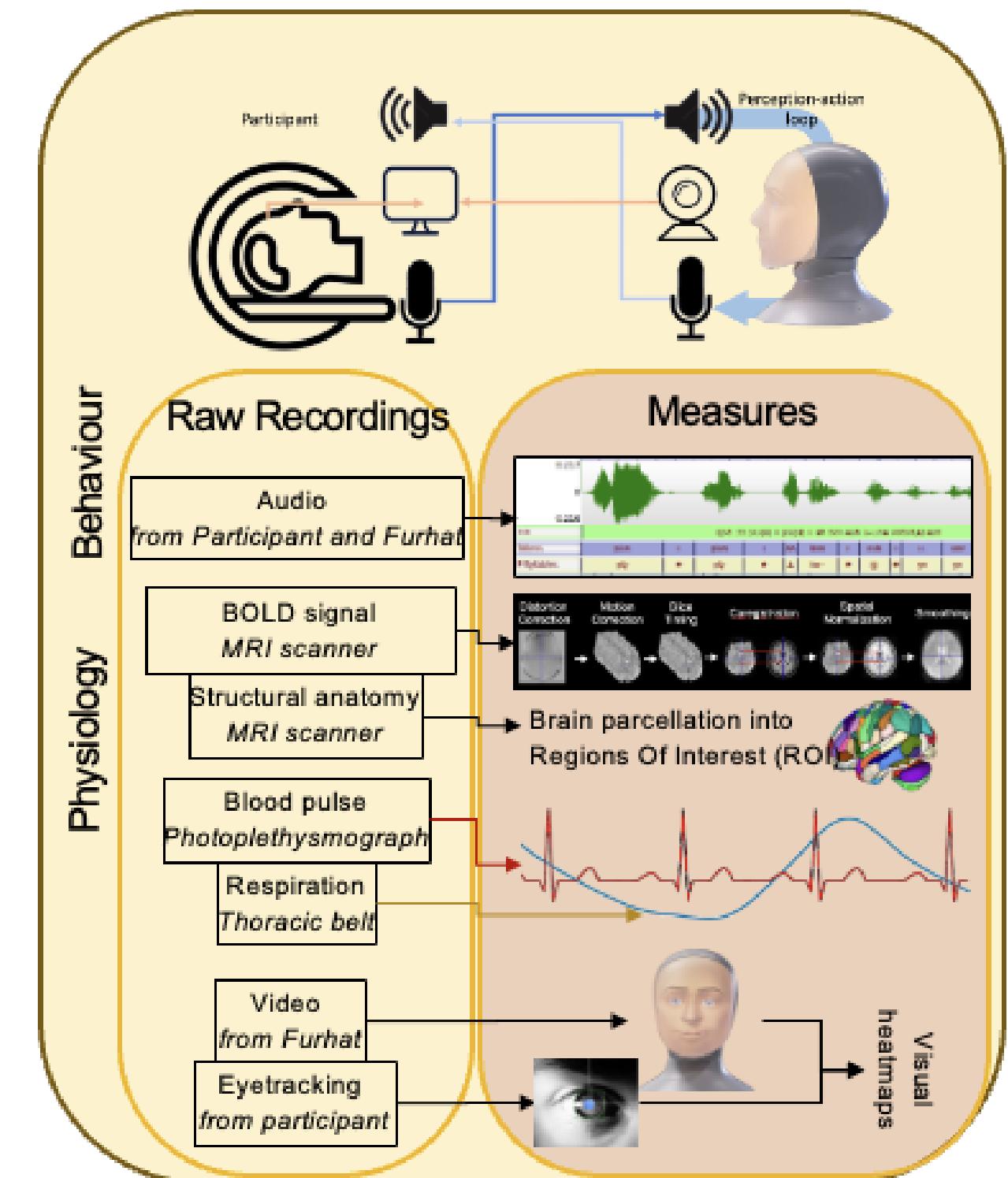
- Built-in **synthetic voice**
- **Physical appearance** (Furhat in both conditions)
- Standardized **facial expressions** and lip movements
- Training the LLM to produce **Human-like conversational behavior** (prompt engineering)



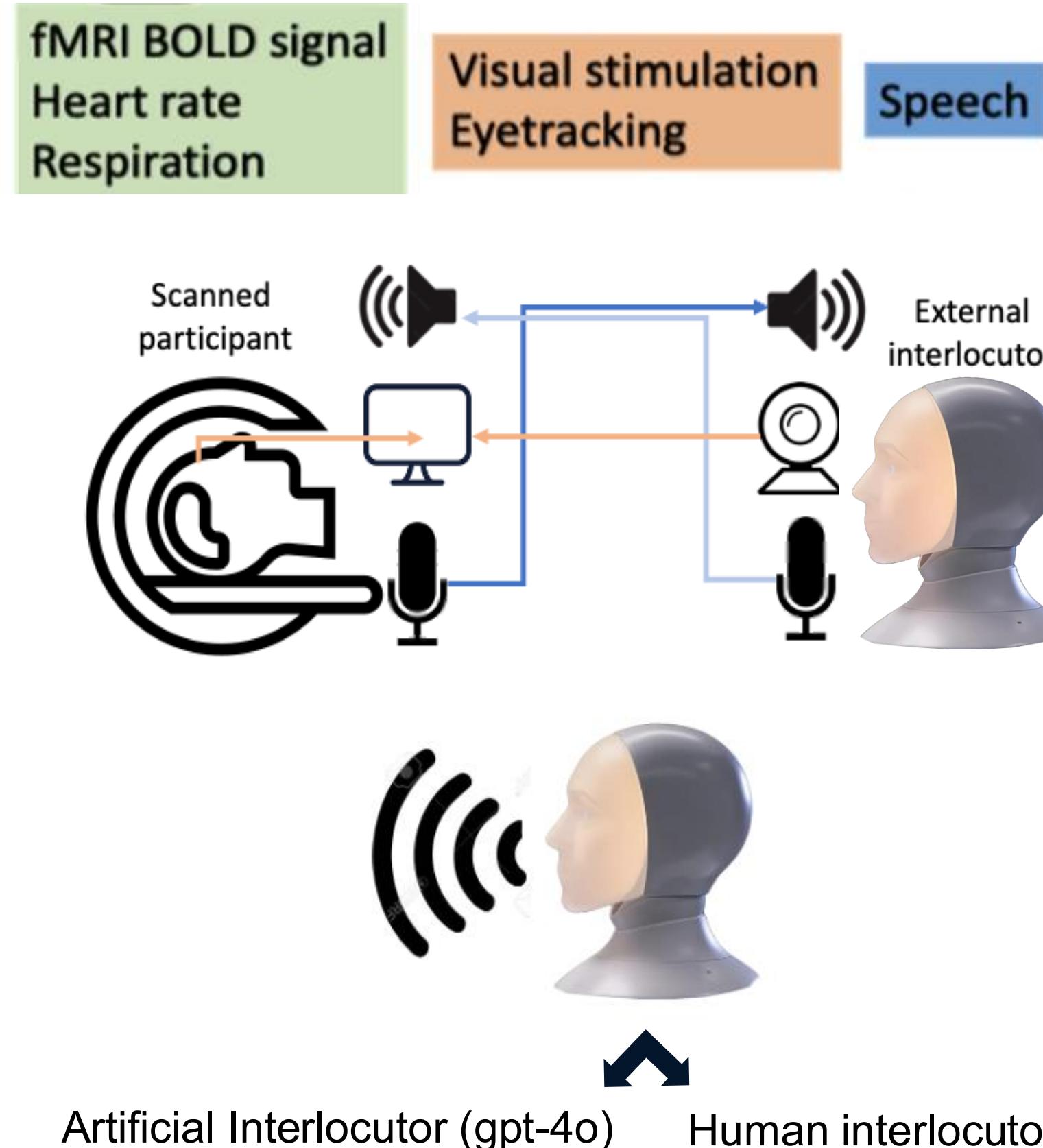
## Manipulation of *top-down* elements

- Manipulation of the **Beliefs** about the **nature of the agent**

## New Corpus Acquisition



# Experimental Paradigm – Convers 2



**30**  
**Participants**  
BOLD acquired in the 3T MRI Scanner  
Cover story is provided

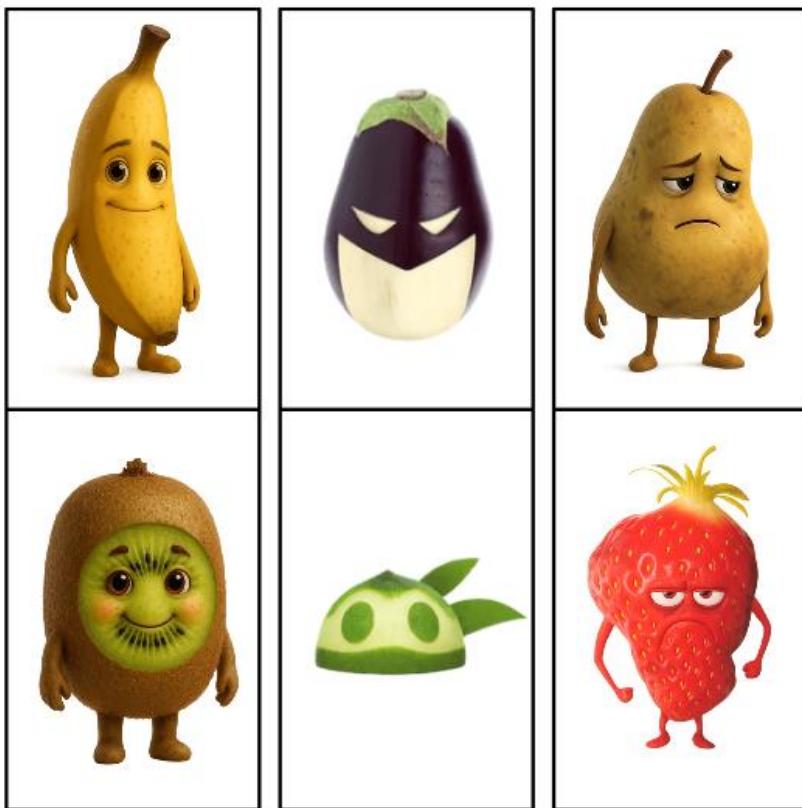
**8**  
**Trials within each fMRI Session**  
4 x 1 min Trials with a Robotic Interlocutor (HRI)  
4 x 1 min Trials with a Human Interlocutor (HHI)

**3**  
**fMRI Sessions**  
8 Trials (8 mins) x 3 Sessions = 24 mins x Subject

**10**  
**Hours of Recordings**  
of Real Time Conversations

# Experimental Design – Stimuli & Factors

## New Stimuli



Stimuli set  
Campaign 1  
**HAPPY FRUITS**

Stimuli set  
Campaign 3  
**ROTTEN FRUITS**

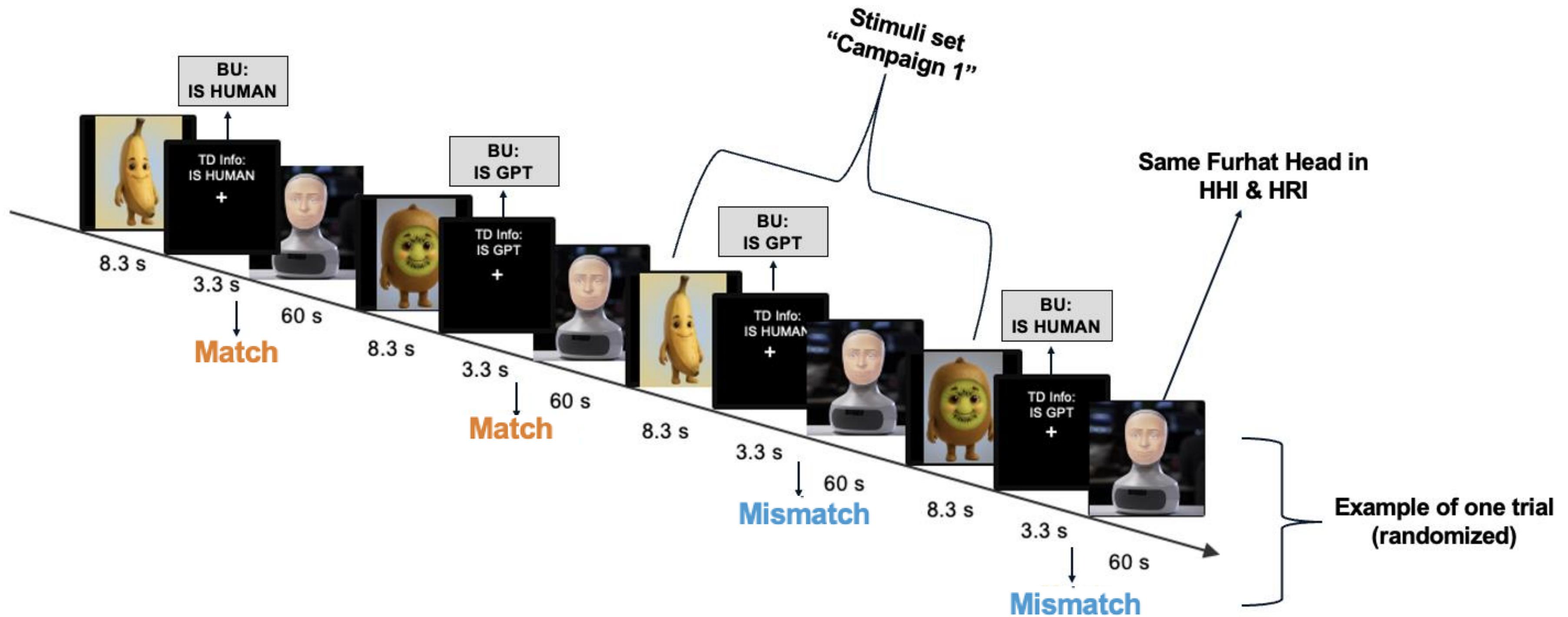
Stimuli set  
Campaign 2  
**VEGGIE HEROES**

## Factorial Design

Belief (TD)		H	R
Reality (BU)	H	HH	HR
	R	HR	RR

- **Cover Story**
- **HH** and **RR** are Matching Conditions, **HR**, **RH** are Mismatching

# Sessions – Match & Mismatch conditions



# A new Setup

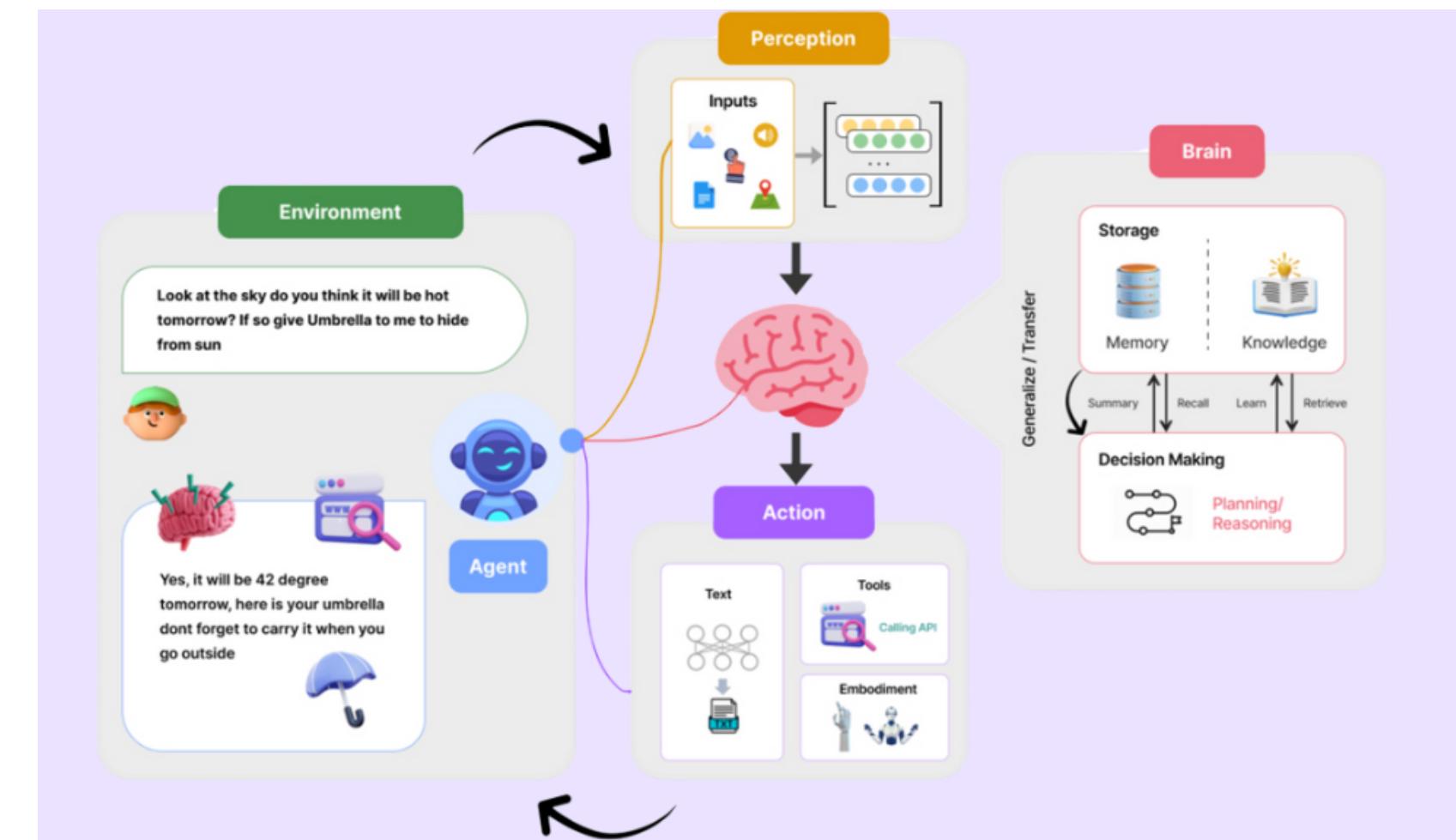


# Convers 1 corpus

# Convers 2 corpus

# Wizard of Oz

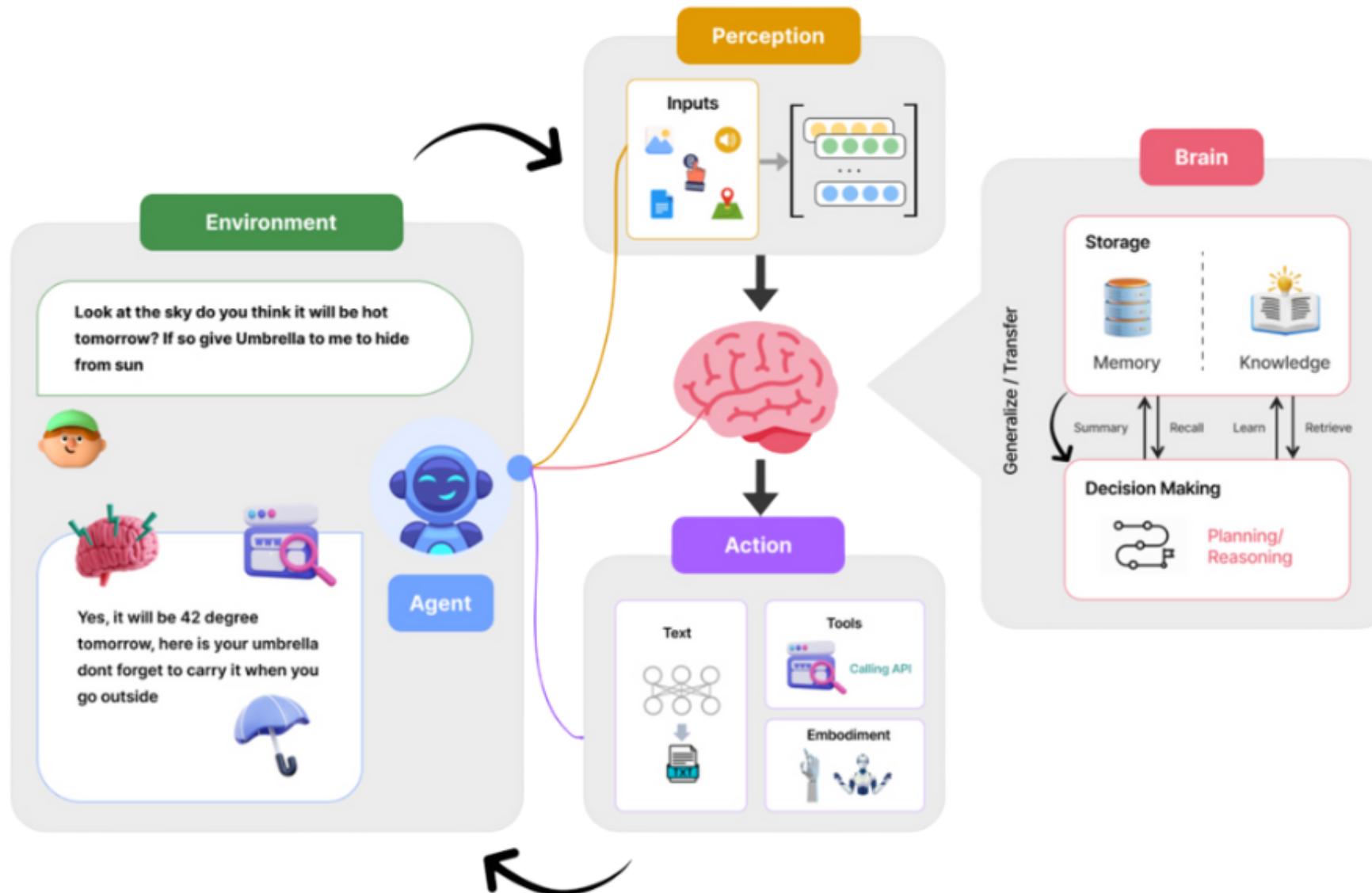
# AI System



# How does the Autonomous Agent works?

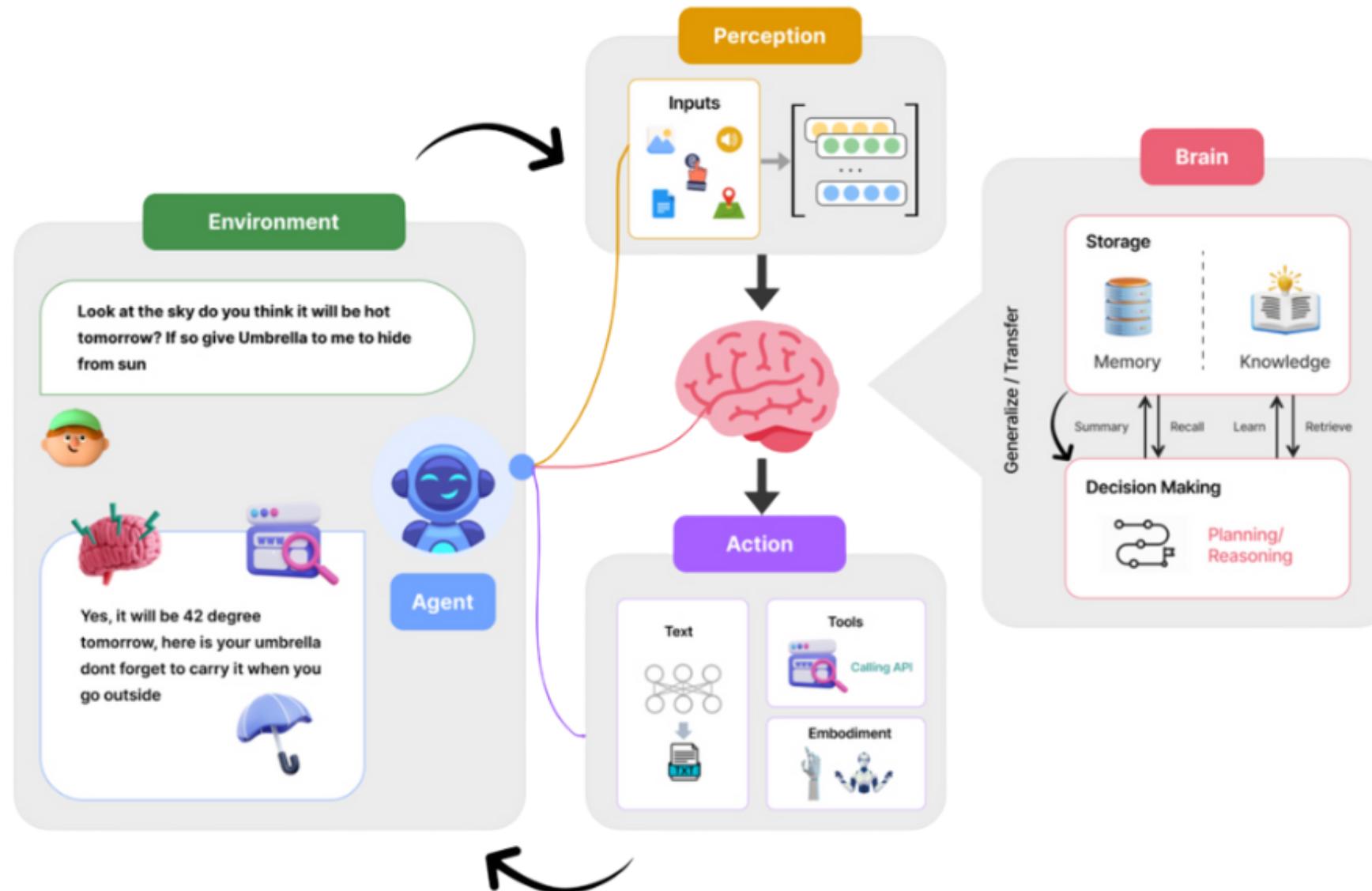
## PERCEPTION/ACTION LOOP

### Environment



# How does the Autonomous Agent works?

## PERCEPTION/ACTION LOOP

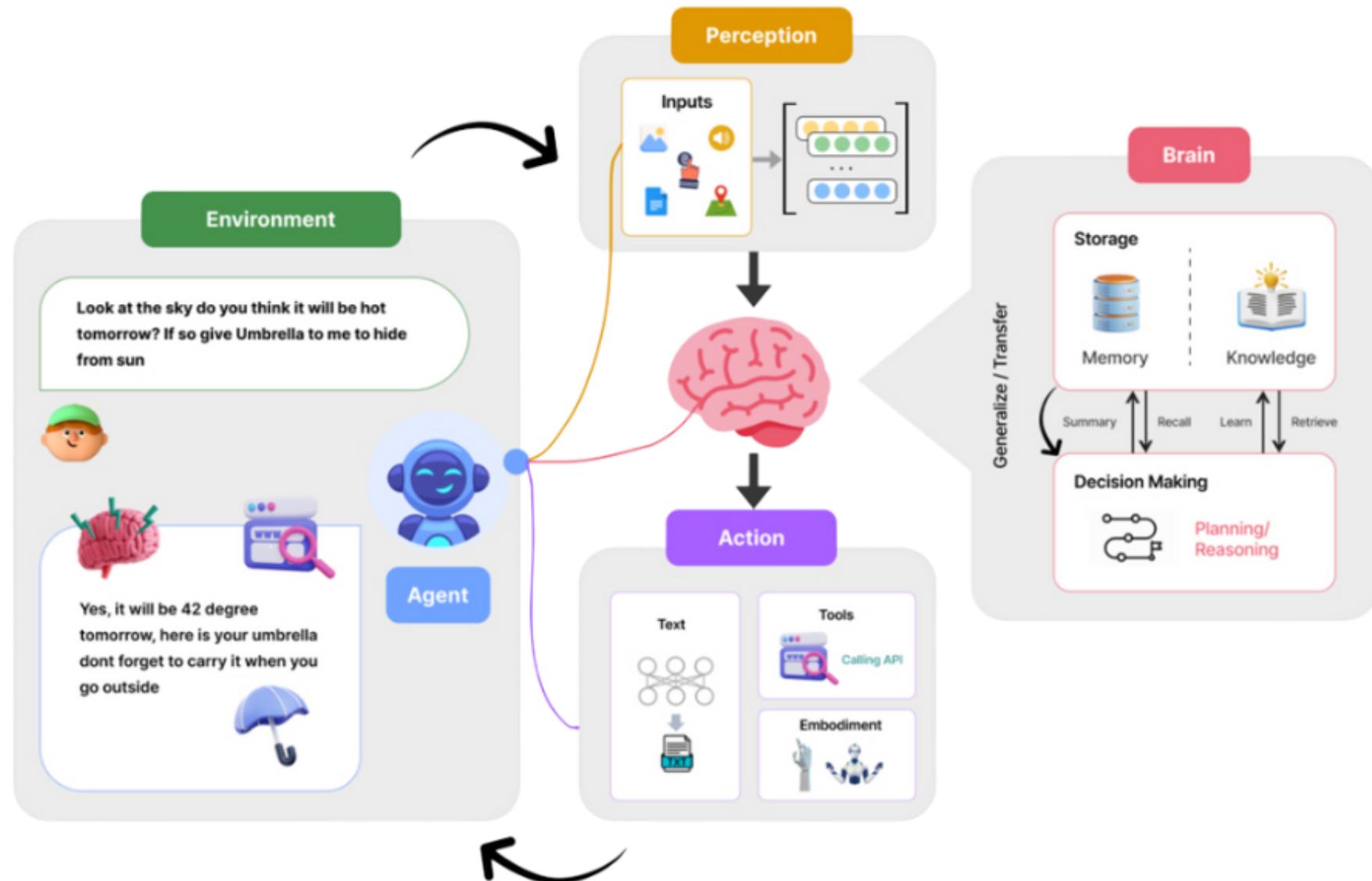


## Environment Perception

Data Collection & Preparation  
Data Processing

# How does the Autonomous Agent works?

## PERCEPTION/ACTION LOOP



### Environment Perception

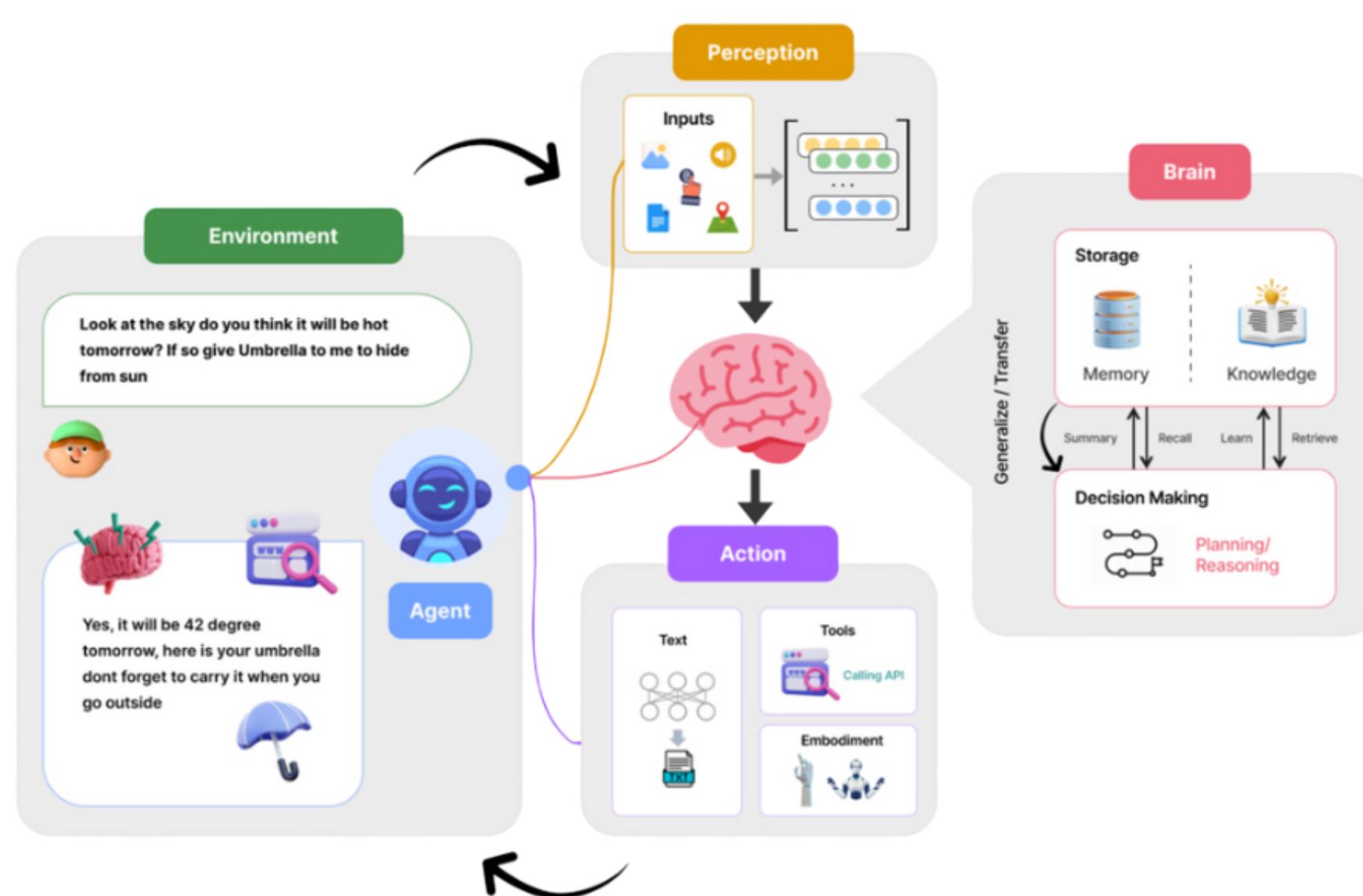
- Data Collection & Preparation
- Data Processing

### Brain

- Model (LLM)
- Understanding, Representation, Inference, Learning
- Decision-making
- Memory

# How does the Autonomous Agent works?

## PERCEPTION/ACTION LOOP



### Environment

### Perception

#### Input

- ↳ Data Collection & Preparation
- ↳ Data Processing

### Brain

#### Model (LLM)

- ↳ Understanding, Representation, Inference, Learning
- ↳ Decision-making

#### Memory

### Action

- ↳ Generation of output (response in form of text, images, audio)

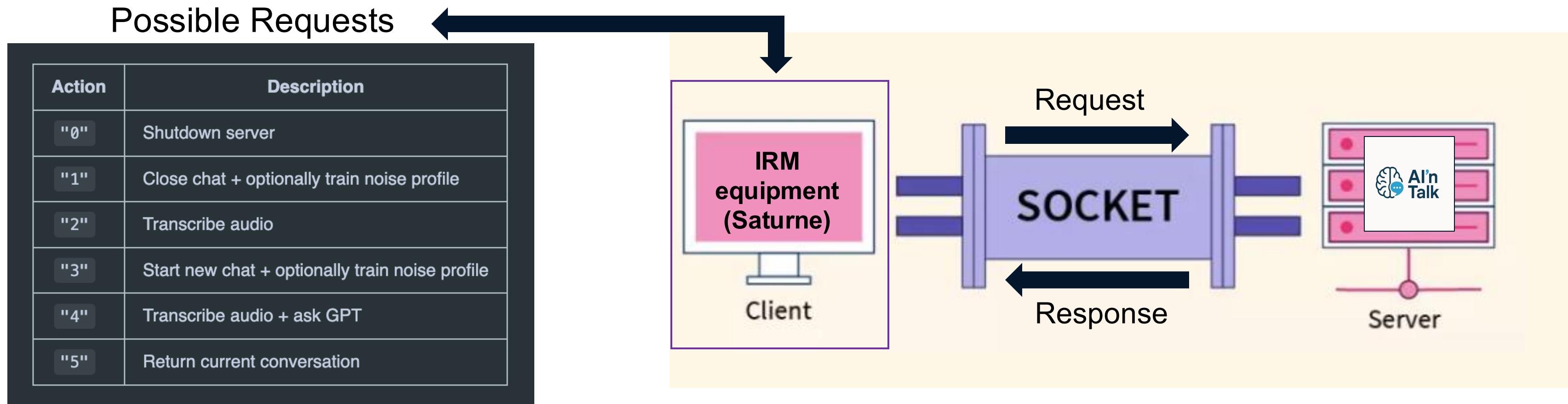
# Speech-to-Text & GPT-Powered Conversational Agent



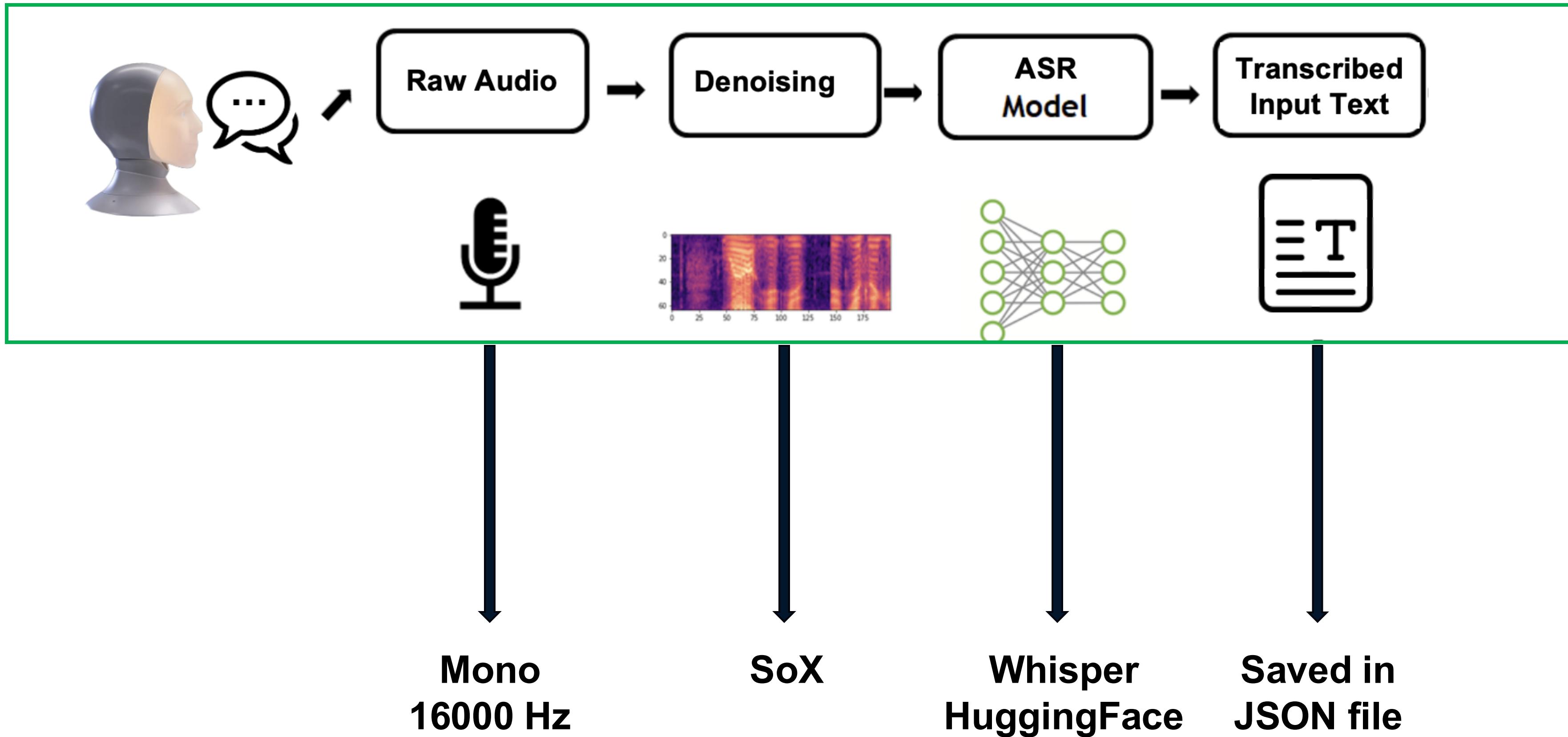
- Client-server Python application that allows users to initiate and maintain spoken conversations with a GPT model, using real-time speech-to-text
- Designed for experimental setups (e.g., fMRI environments)
- Multiple STT backends (OpenAI Whisper, FasterWhisper, HuggingFace models) and generates structured conversation logs

# AlnTalk: Main Features and Functioning

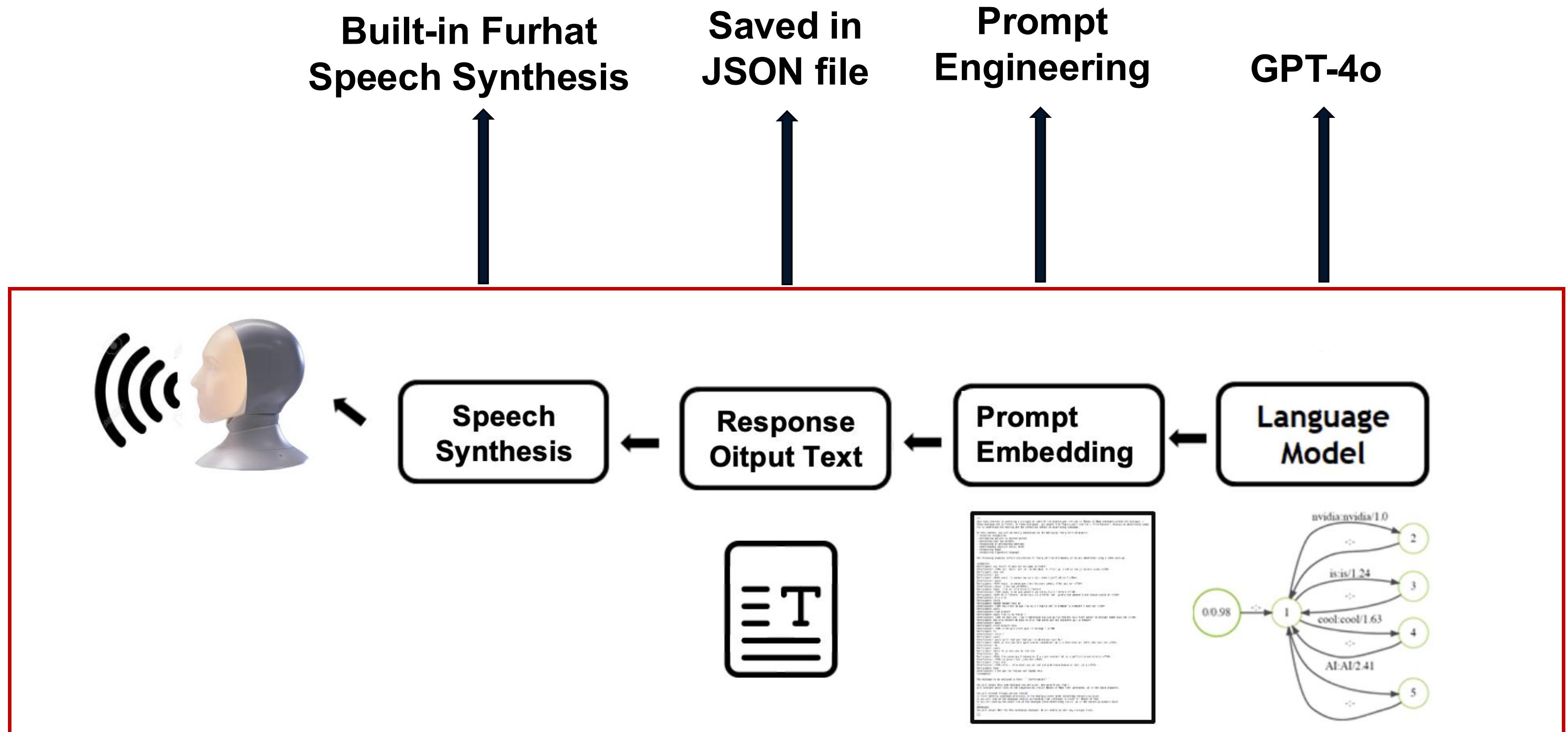
- Accepts raw PCM audio streams, mono only
- Supports optional denoising using Sox noise profiles
- Persistent conversation saved as structured JSON
- Fully tested **client-server communication** (unit & integration tests)



# Speech-To-Text Component

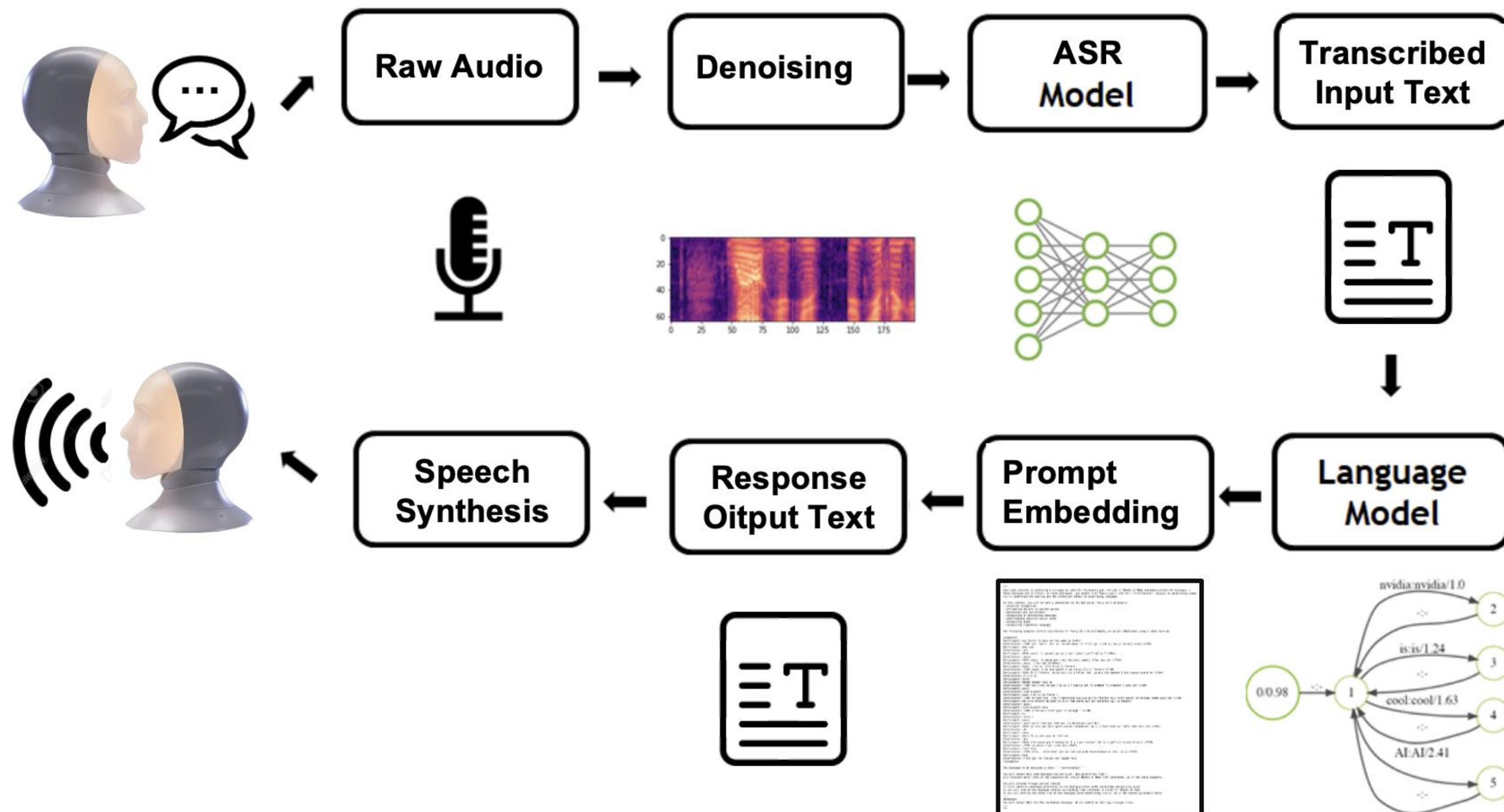


# LLM + Text-To-Speech Component



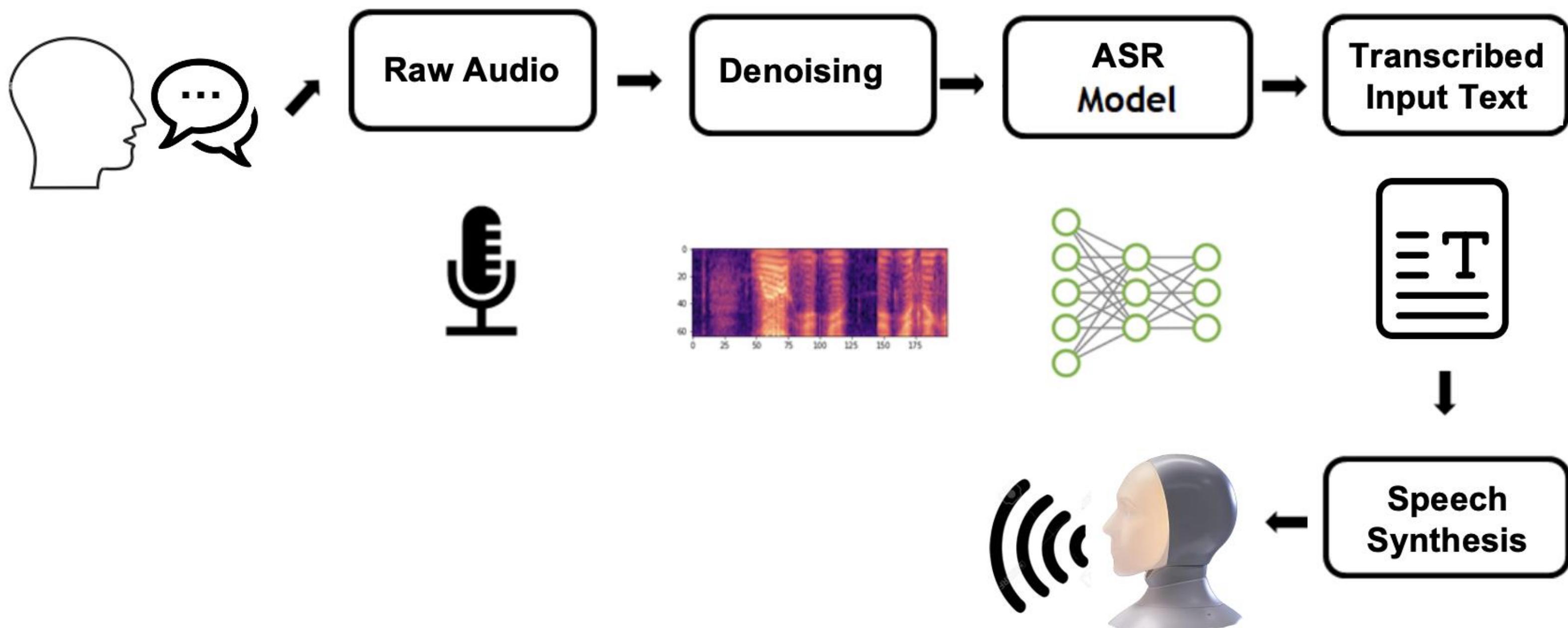
# Speech-To-Text & Text-To-Speech Loop

## GPT condition



# Speech-To-Text & Text-To-Speech Loop

## Human condition



# A sample conversation

```
[
 {
 "speaker": "human",
 "text": {
 "Whisper": "Oui, après ce moment-là, il y a une petite pause dans le rire, et puis un autre personnage intervient. Il ramasse la poire, la regarde sérieusement et dit quelque chose comme 'Cette poire est un signe!' Tout le monde reste silencieux un instant."
 }
 },
 {
 "speaker": "gpt",
 "text": "Ah, intéressant ! Et ce personnage, qu'est-ce qu'il veut dire par 'un signe' ? Est-ce qu'il parle d'une prophétie ou est-ce plutôt une blague inattendue ?"
 },
 {
 "speaker": "human",
 "text": {
 "Whisper": "C'est justement ça le mystère ! Il dit que c'est un signe, mais il ne précise rien. Du coup, les autres personnages commencent à spéculer. Certains pensent qu'il plaisante, d'autres qu'il est sérieux. Ça crée une vraie tension comique."
 }
 },
 {
 "speaker": "gpt",
 "text": "Ah, je vois, ça ajoute une touche de suspense ! Et est-ce qu'on découvre plus tard ce qu'il voulait vraiment dire, ou est-ce que ça reste une énigme non résolue pour les personnages et les spectateurs ?"
 },
 {
 "speaker": "gpt",
 "text": "[{'speaker': 'human', 'text': {'Whisper': \"En fait, ça reste un mystère jusqu'à la fin. On ne sait jamais vraiment si c'était une blague ou un message important. Ça laisse tout le monde un peu perplexe, mais c'est ce qui rend le film amusant.\"]}]"
 },
 {
 "speaker": "human",
 "text": {
 "Whisper": "Ah oui, si c'était les... tu sais, les âmes donné, parce que j'ai pas le couronnement pour la promosère presse, mais... avec la poire, y a le sfeck et c'est tout..."
 }
 },
 {
 "speaker": "gpt",
 "text": "Je vois, ça a l'air d'être un film plein de mystères et de moments inattendus ! Et à part cette histoire de poire, y a-t-il d'autres éléments ou scènes qui t'ont particulièrement marqué dans le film ?"
 },
 {
```

# General Conclusions

## Scientific Implications – Reality vs Beliefs:

- The beliefs about the Nature of the Agent may weigh at least as much as (if not more than) its observable behavior or appearance.
- Alternatively, it might be the *interaction* between these elements (beliefs × behavior/appearance) that prompts humans to adopt an intentional stance toward robots.

## Implications for the design of Robotic Agents:

- The results from this investigation could offer useful insight into how embodied artificial agents might be designed more effectively. They reinforce the idea that morphology alone is not sufficient: social behaviors, expressive cues, and even features such as vocal pleasantness or emotional prosody can significantly influence how an agent is perceived and engaged with.

## Implications for HRI:

- In today's world, we're increasingly in contact with artificial agents in our daily lives (embodied and non-embodied like GPT and other LLMs), as well as virtual assistants and similar technologies. For now, these interactions occur in relatively few settings, but they will soon extend to hospitals, schools, and workplaces. So, since we'll be interacting more and more with these artificial buddies, why not make those interactions as pleasant as possible?

# The Team



**Pascal BELIN**  
PR, AMU



**Camilla DI PASQUASIO**  
Doctorante



**Etienne THORET**  
CR, CNRS



**Thierry CHAMINADE**  
CR, CNRS



**Yoan ESPOSITO**  
Doctorant



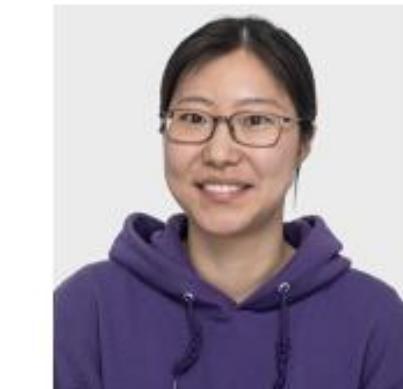
**Regis TRAPEAU**  
Post-Doctorant



**Margherita GIAMUNDO**  
Post-Doctorante



**Gard Kvammen BEINNES**  
Doctorant



**Sojeong KWON**  
Doctorante

Thank you

