Language Generation from Brain Recordings

Ziyi Ye
Tsinghua University
2024.3.12

Introduction

- **Application of Brain-Computer Interface (BCI)**
  - Instruction decoding [NeuraLink 2021]
  - Emotion recognition [Edgar 2020]
  - Sematic decoding
    - Visual information reconstruction [Takagi 2023]
    - Language information reconstruction [Makin 2020]

Fig: Neuralink's monkey use BCI to play games [Cooney 2021]

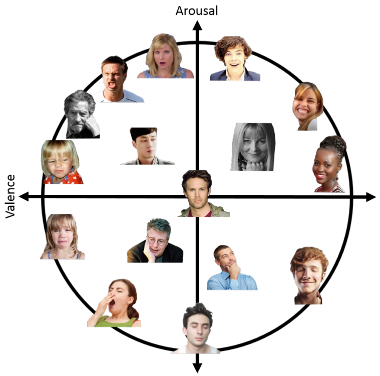


Fig: Emotion recognition [Edgar 2020]

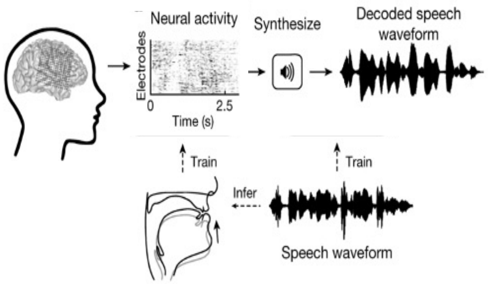
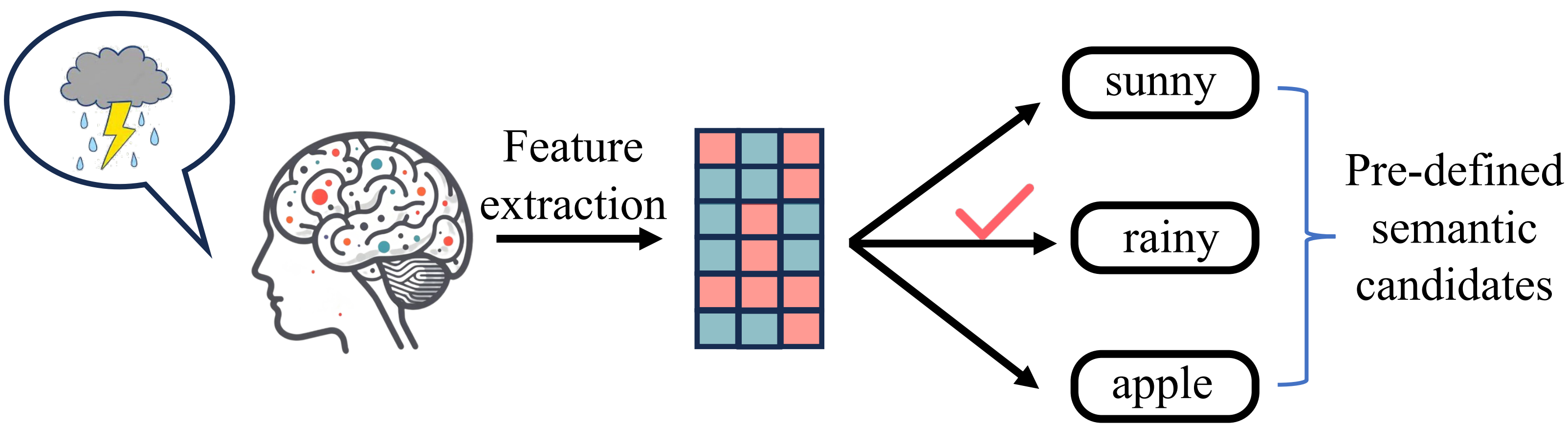


Fig: Speech decoding [Makin 2020]

Background

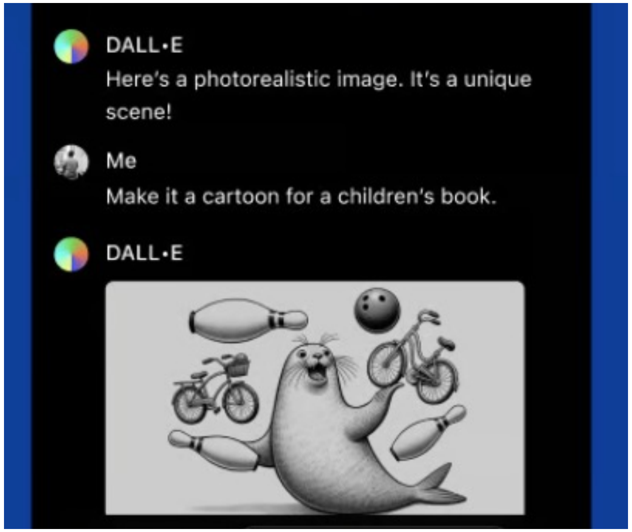
- **Existing language BCIs**
  - Pre-defining a series of semantic candidates
  - Limitations
    - A limited number of semantic candidates (usually 2-50)
    - High task dependency



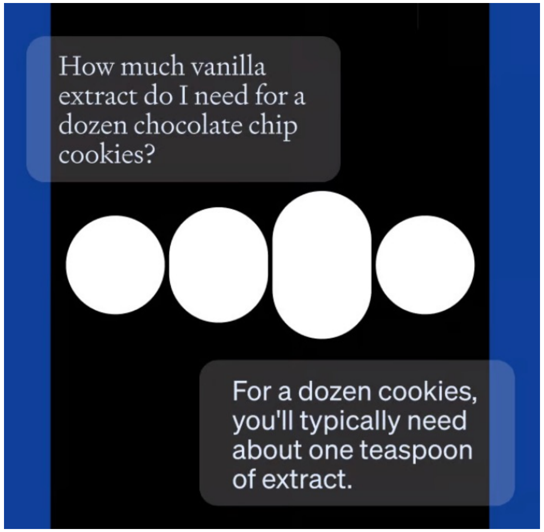
Fig： Language BCIs by pre-definition and post-hoc selection/classification

Background

- **Emergence of generative language models (LMs)**
  - Reconstructing mental language is difficult
  - The LM might be able to provide **contextual knowledge**



Fig： ChatGPT + DALL-E



Fig： ChatGPT + speech synthesis

Background

- **A language BCI with generative model** [Tang 2023]
  - Pre-generation with post-hoc selection
  - **Limitations**
    - Brain information is not involved in the language generation phase
    - Still use a limited amount of candidates

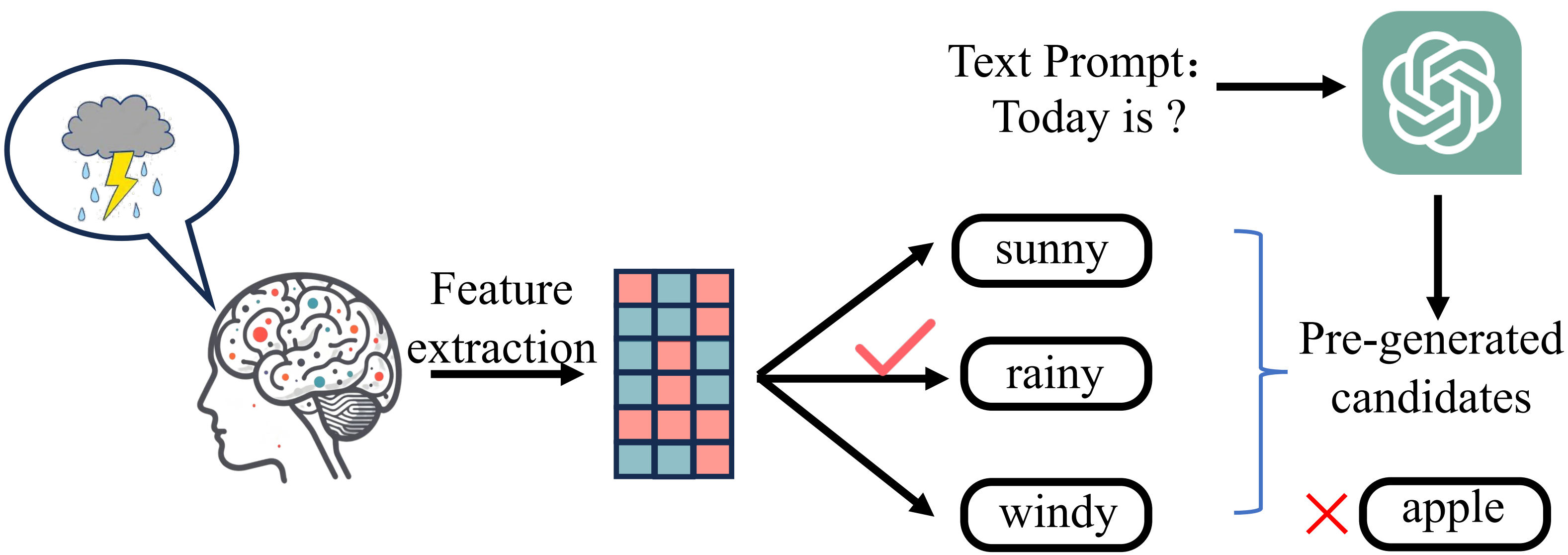


Fig: Language BCIs by pre-generation and post-hoc selection

Background

- **Language in LM and language in the Brain**
  - Brain and LM might have similarities in language processing

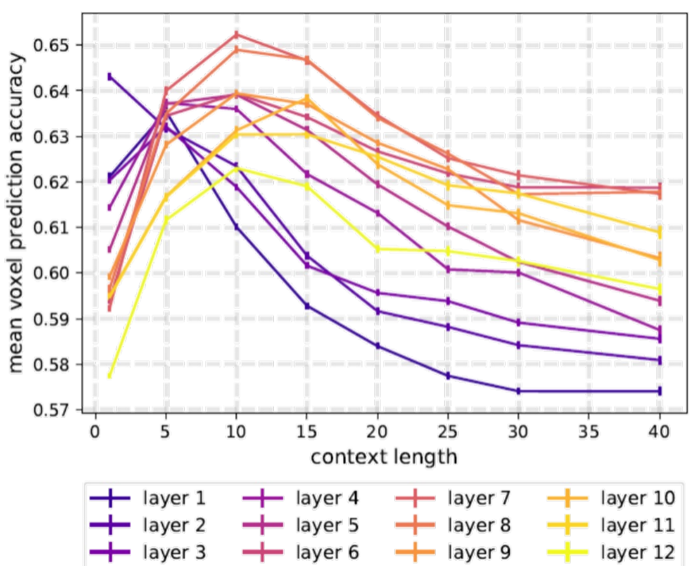


Fig: The **representation** in different layers of the language model have **similarities** to the human brain. [Mariya 2019]

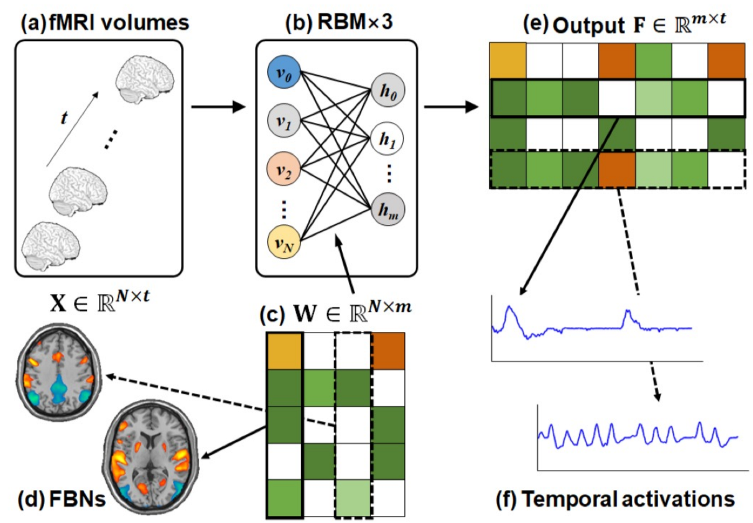


Fig: The **physical neurons in the brain** exhibit synchrony in activation with the **neurons in language models.** [Liu 2023]

Background

- **Is the similarity more pronounced in larger models?**
  - **Scaling laws** when mapping brain representations to computational representations

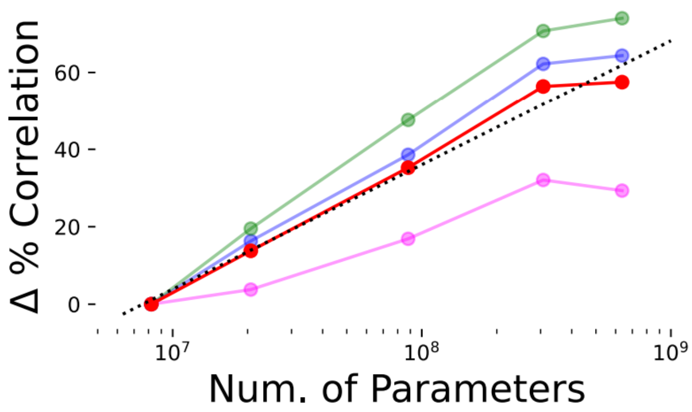


Fig: **Larger correlations** in audio model **with a larger parameter size**. [Anntonello 2023]

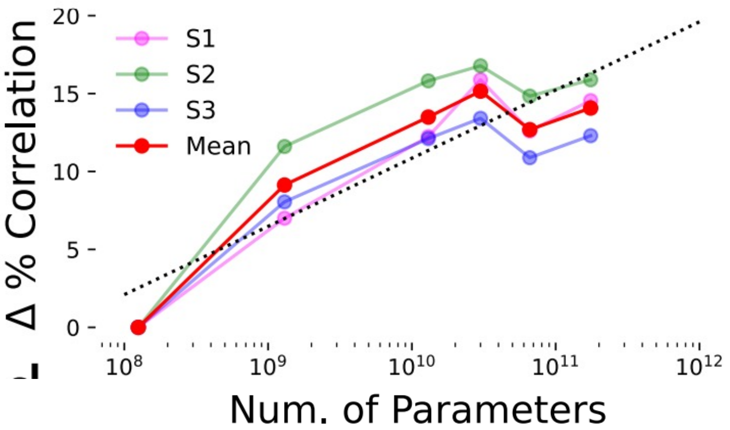


Fig: **Larger correlations** in language model **with a larger parameter size**. [Anntonello 2023]

Motivation

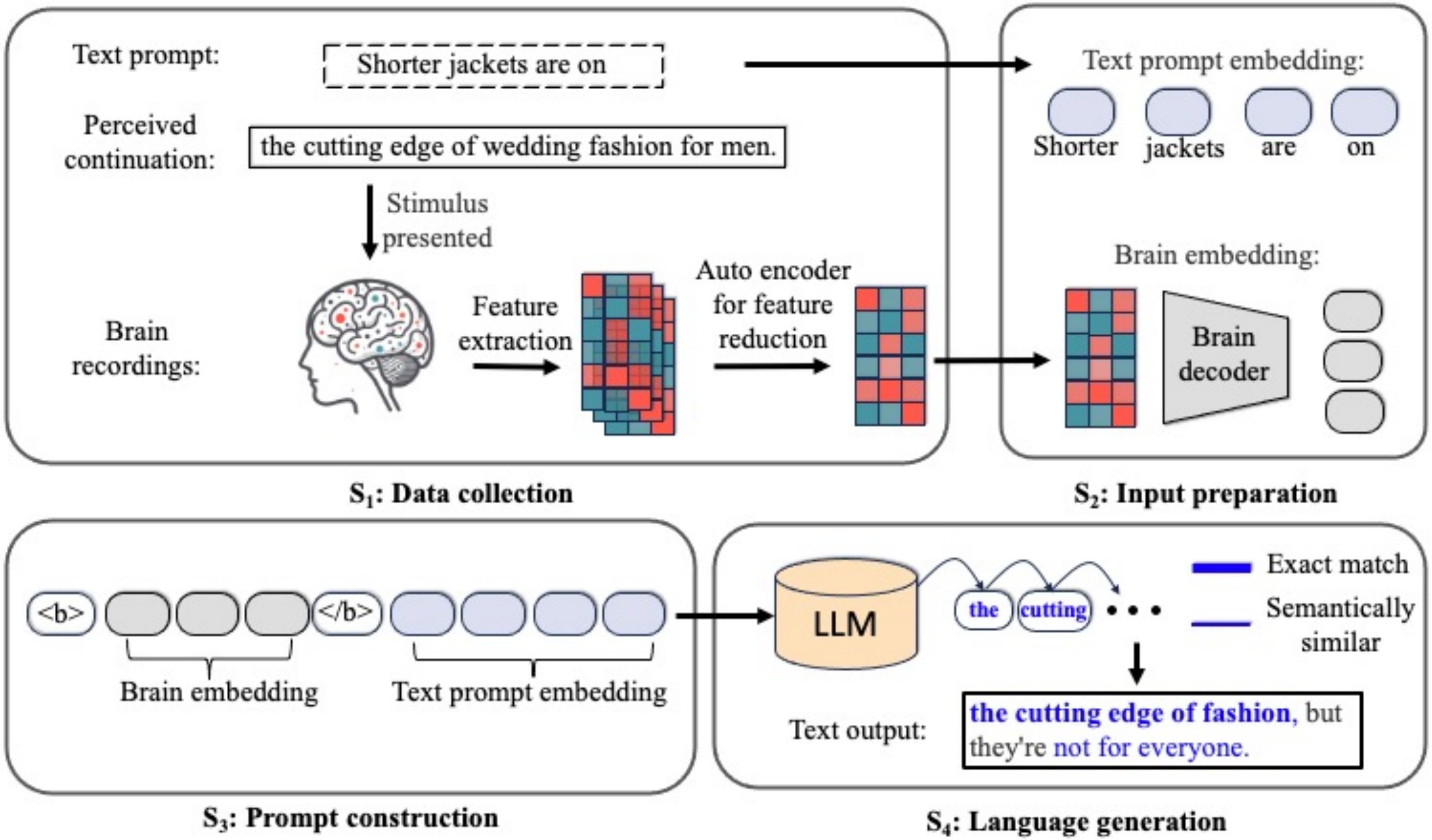
- Designing BCIs with **direct language generation feature**

- **Limitations of existing work:**
  - Classification-based setting
  - Limited candidate set and limited performance
  - Ignoring the potential relationship between brain and LLM

- **Can representation in the brain and in the LLM be jointly modeled?**

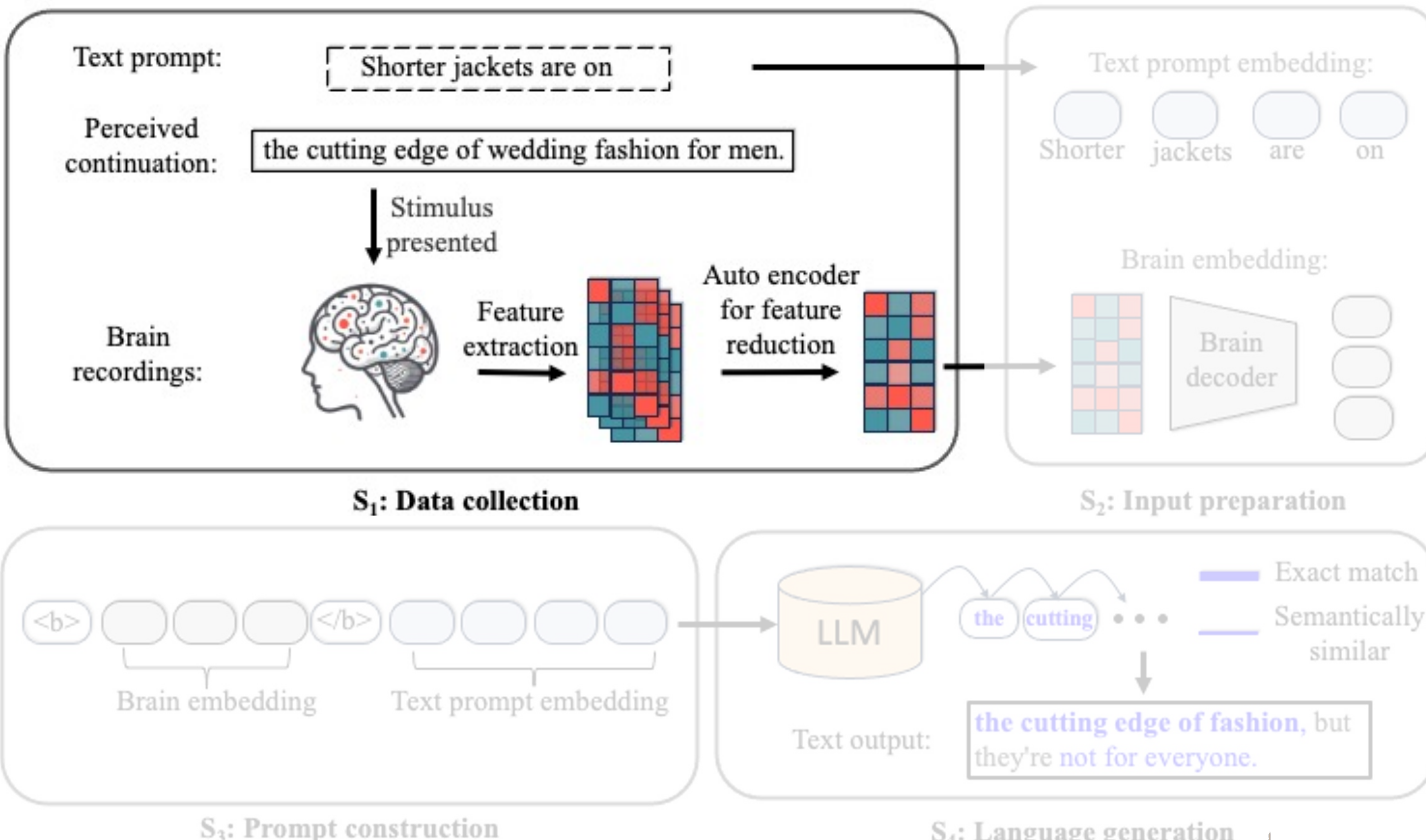
Method

- Language generation by jointly modeling of brain and the LLM (**BrainLLM**)



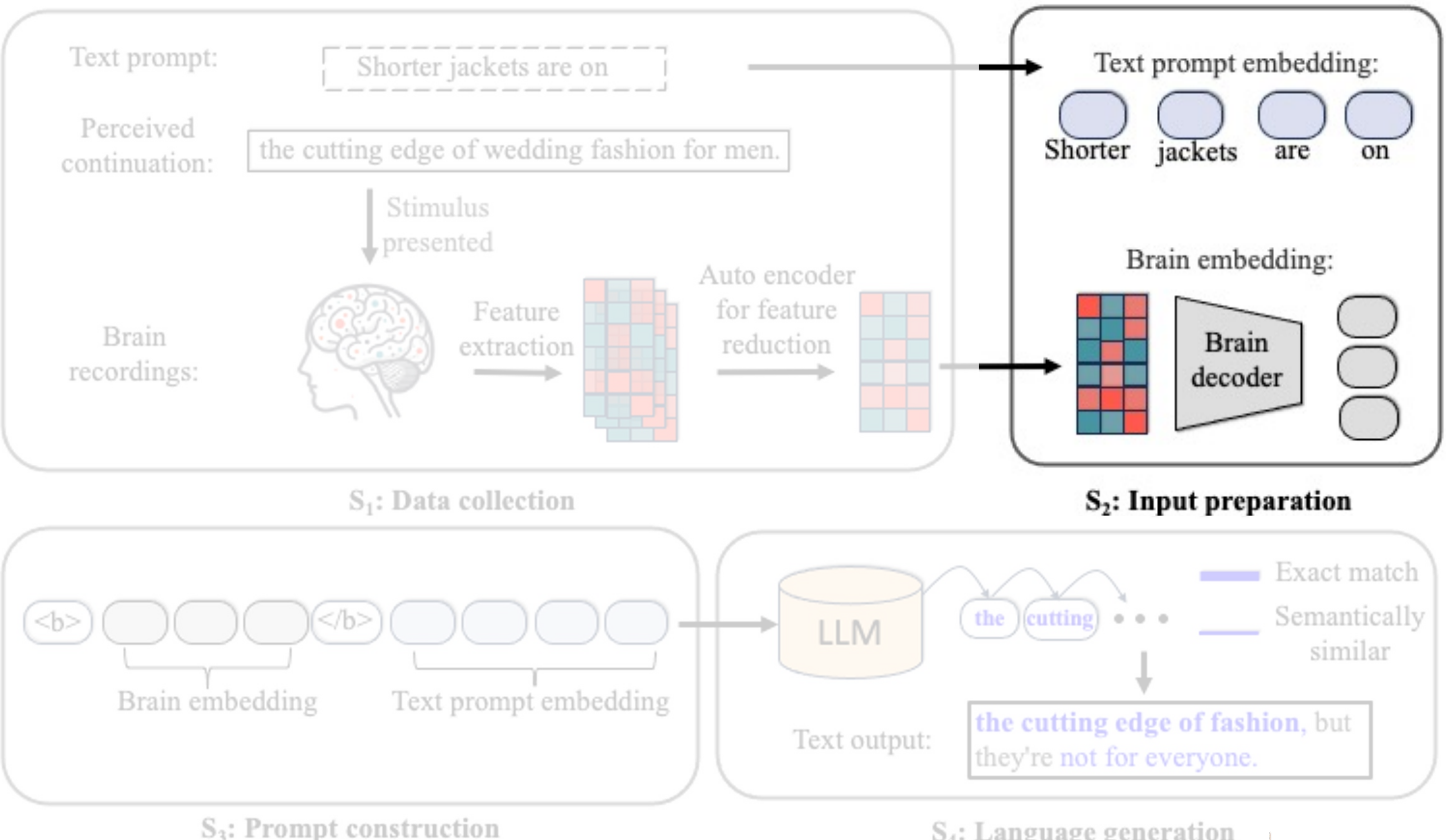
Method

- Language generation by jointly modeling of brain and the LLM (**BrainLLM**)



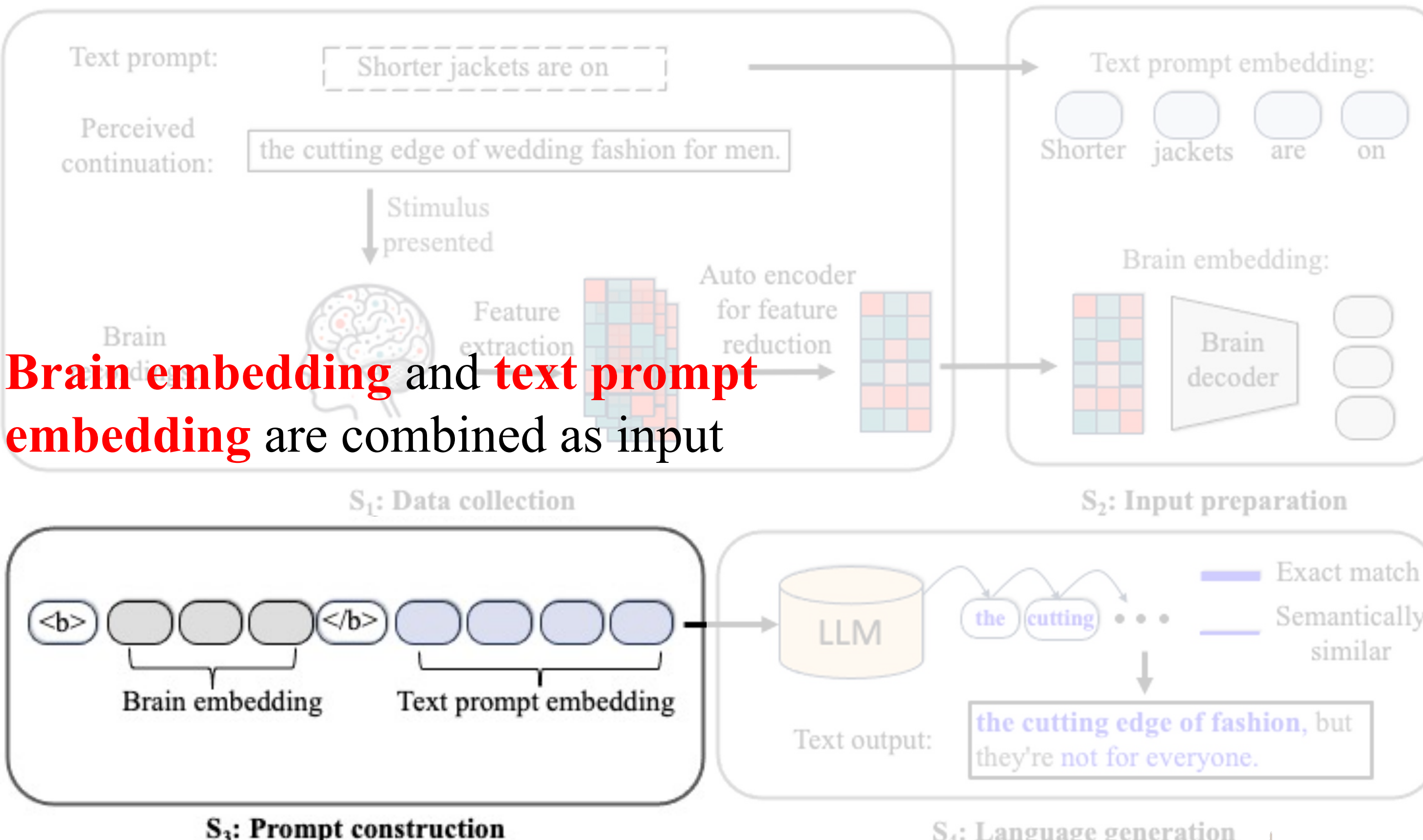
Method

- Language generation by jointly modeling of brain and the LLM (**BrainLLM**)



Method

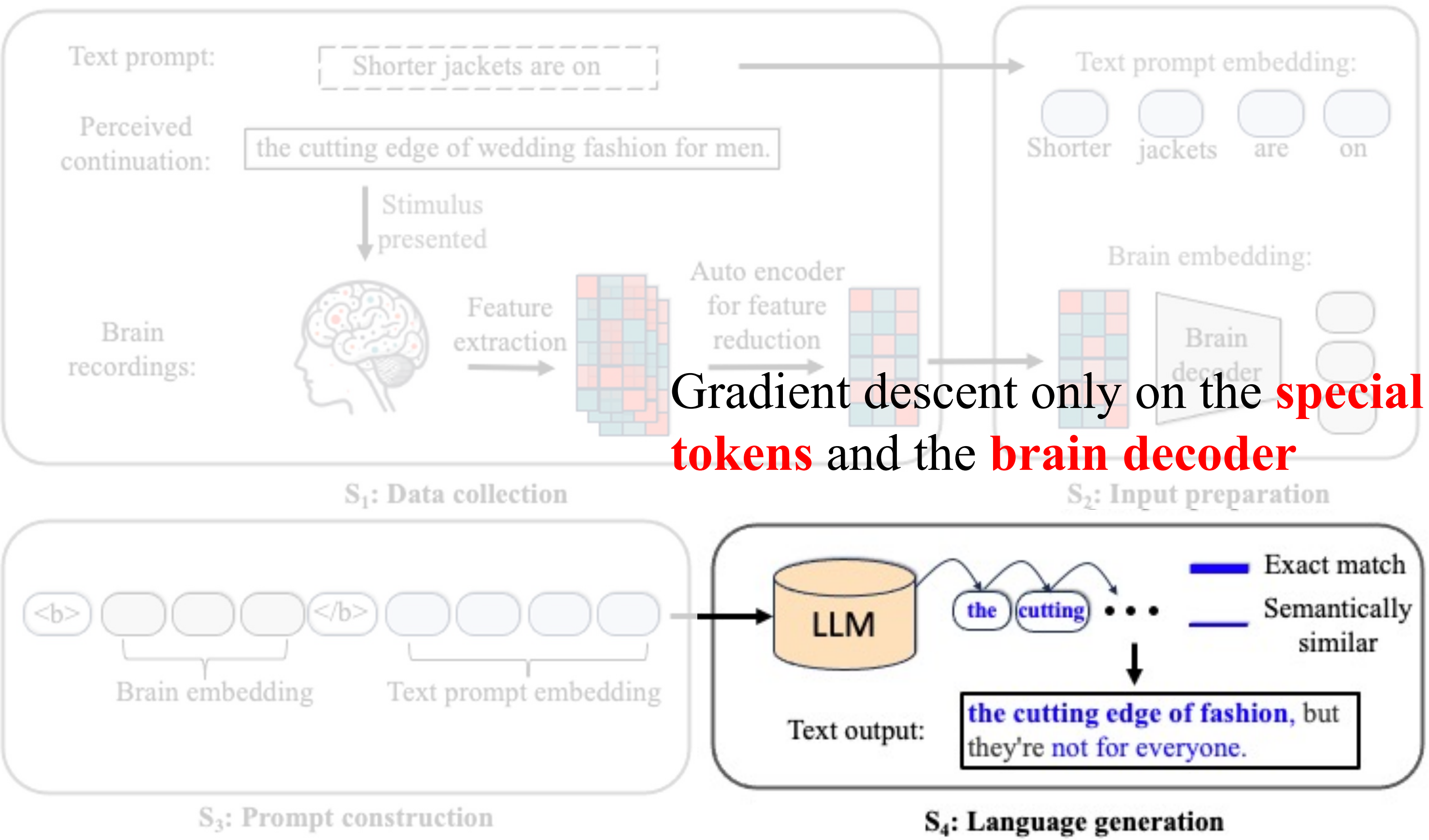
- Language generation by jointly modeling of brain and the LLM (**BrainLLM**)



Brain embedding and **text prompt embedding** are combined as input

Method

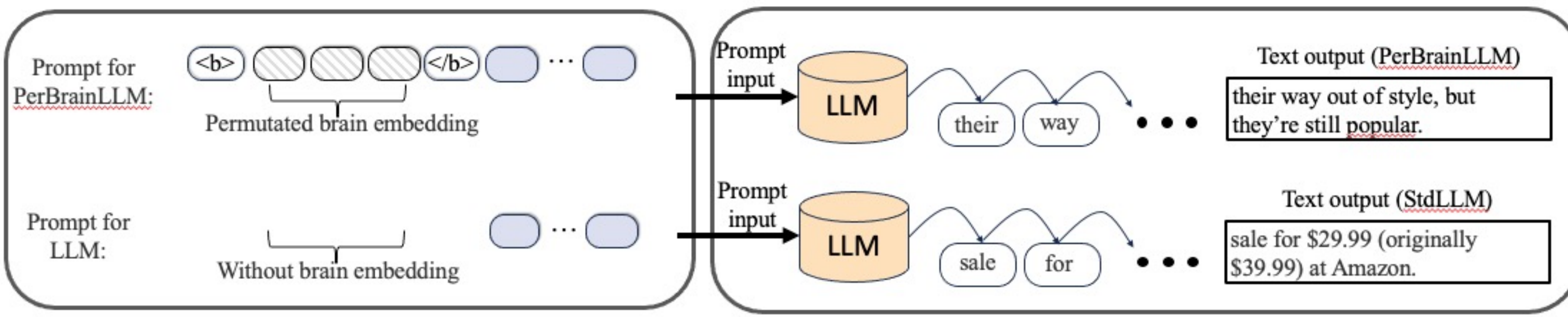
- Language generation by jointly modeling of brain and the LLM (**BrainLLM**)



Gradient descent only on the **special tokens** and the **brain decoder**

Method

- Language generation by jointly modeling of brain and the large language model (**BrainLLM**)
- Control models:
  - ***PerBrainLLM***: BrainLLM with brain input randomly sampled
  - ***StdLLM***: the standard LLM with only text input



S_3': Prompt construction for control models

S_4': Language generation for control models

Evaluation

- Evaluation protocols:
  - **Pairwise accuracy:**
    - comparing the likelihood of generating the perceived continuation
    - i.e., $Pairwise\ ACC = \begin{cases} 1, & if\ P_{BrainLLM} > P_{PerBrainLLM} \\ 0, & \text{else} \end{cases}$
  - **Language similarity metrics:**
    - Bleu, WER, Rouge, perplexity/surprise
  - **Human evaluation:**
    - pairwise preference judgment from human annotators

Results

- Case study：

Text input in BrainLLM

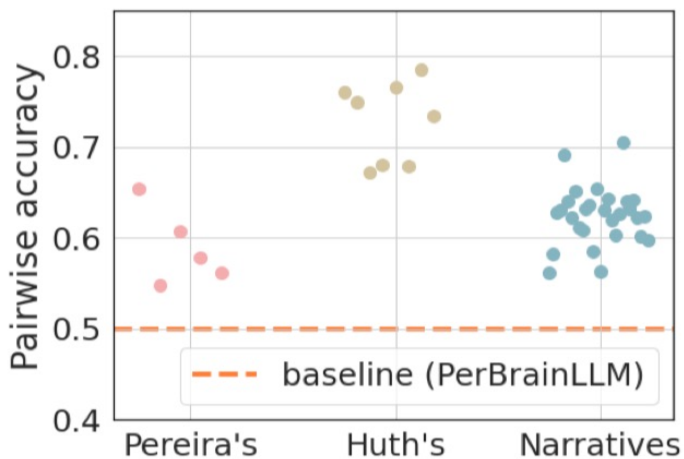
Brain response to perceived continuation as BrainLLM input

Text prompt	**Perceived continuation**
Shorter jackets are on	the cutting edge of wedding fashion for men.

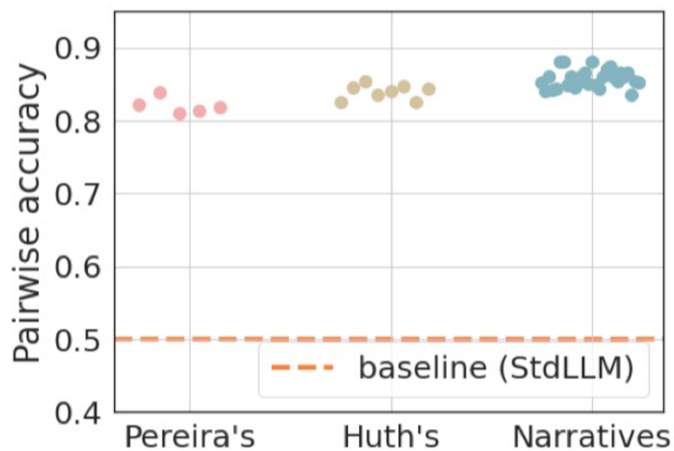
BrainLLM	**PerBrainLLM**	**StdLLM**
the cutting edge of fashion, but they're not for everyone.	their way out **of** style, but they're still popular.	sale for \$29.99 (originally \$39.99) at Amazon.

Results

- Pairwise accuracy：
  - BrainLLM outperforms PerBrainLLM and StdLLM
  - PerBrainLLM is a stronger control than StdLLM
    - PerBrainLLM contains brain prompt that make the LLM generate content more aligned with the distribution of tokens in the training set



(a) BrainLLM vs. PerBrainLLM

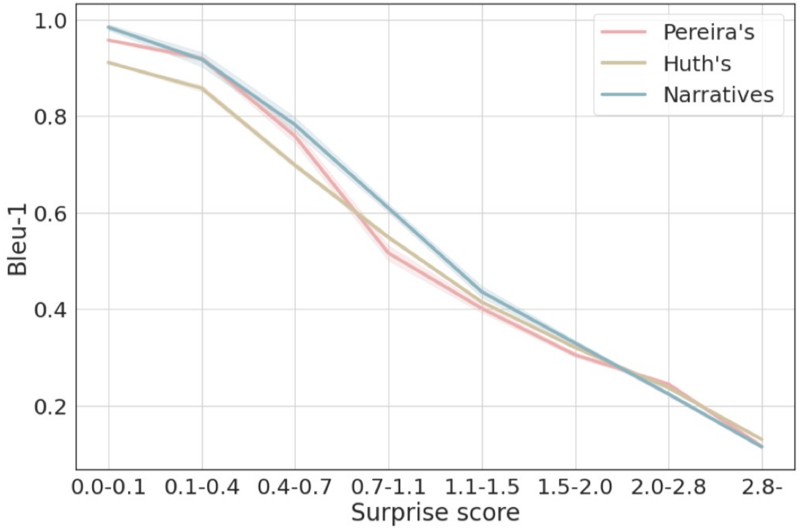


(b) BrainLLM vs. StdLLM

Results

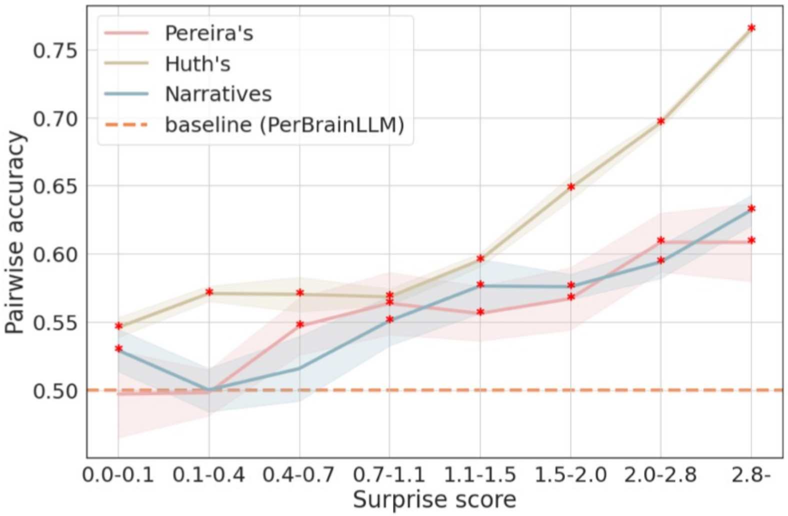
- Analysis regarding surprise score：

Higher surprise, worse performance



Fig：PerBrainLLM's performance w.r.t. different surprise

Higher surprise, BrainLLM gains more when compared to PerBrainLLM



Fig：Pairwise accuracy of BrainLLM v.s. PerBrainLLM in terms of different surprise

Results

- Analysis regarding length of text prompts:

Shorter text prompts, worse performance

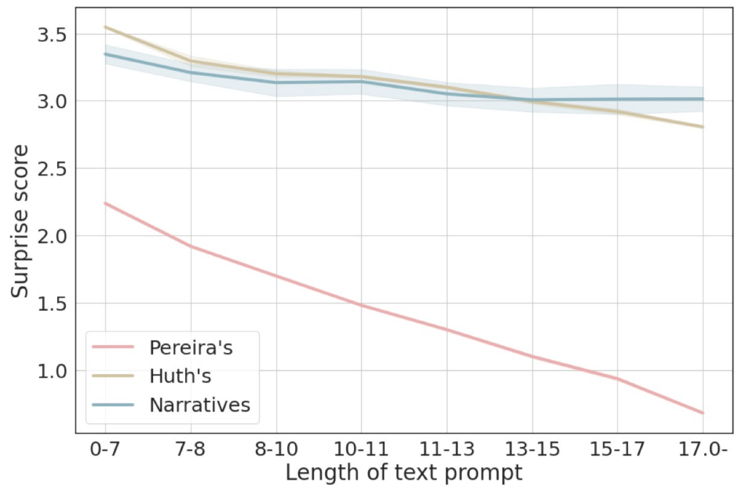


Fig: Surprise w.r.t. length of text prompts

Shorter text prompts, more performance gain with BrainLLM

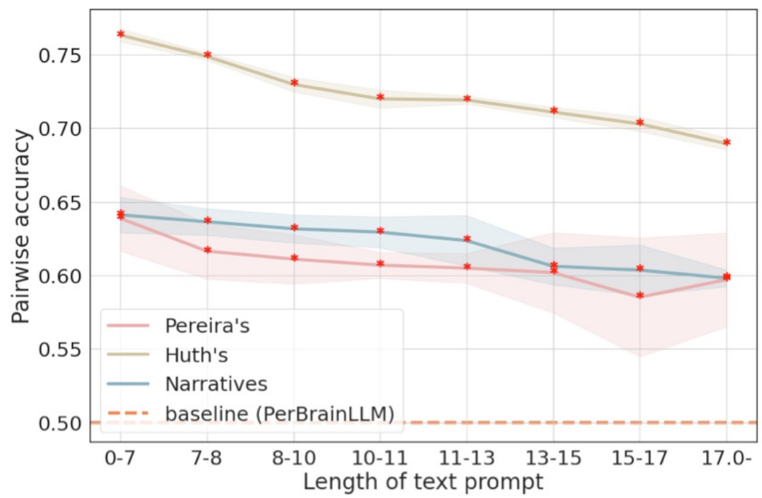


Fig: Pairwise accuracy of BrainLLM v.s. PerBrainLLM w.r.t. length of text prompts

Results

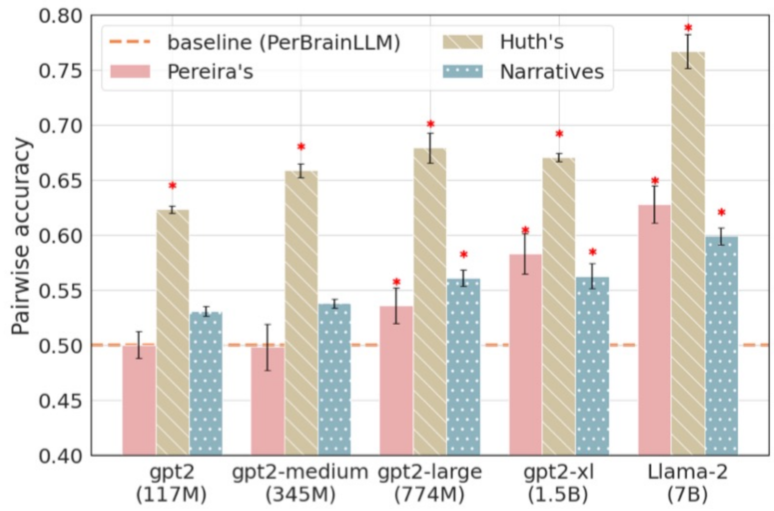
- Analysis regarding the parameter size of LLM：

LLM with more parameters yields better performance

BrainLLM gains even more when using LLM with more parameters!

LLM backbone	Model	BLEU-1(↑)	ROUGE-1(↑)
Llama-2 (7B)	StdLLM	0.2415*	0.2133*
	PerBrainLLM	0.3249*	0.2875*
	BrainLLM	**0.3333**	**0.2987**
GPT-2-xl (1.5B)	PerBrainLLM	0.2772	0.234
	BrainLLM	0.2814*	0.2378*
GPT-2-large (774M)	PerBrainLLM	0.2605*	0.213*
	BrainLLM	0.2655	0.2182
GPT-2-medium (345M)	PerBrainLLM	0.2100	0.1649*
	BrainLLM	0.2118	0.1672
GPT-2 (117M)	PerBrainLLM	0.1866	0.1456
	BrainLLM	0.1846	0.1445

Fig：Language generation performance in Pereira's dataset with different number of LLM parameters

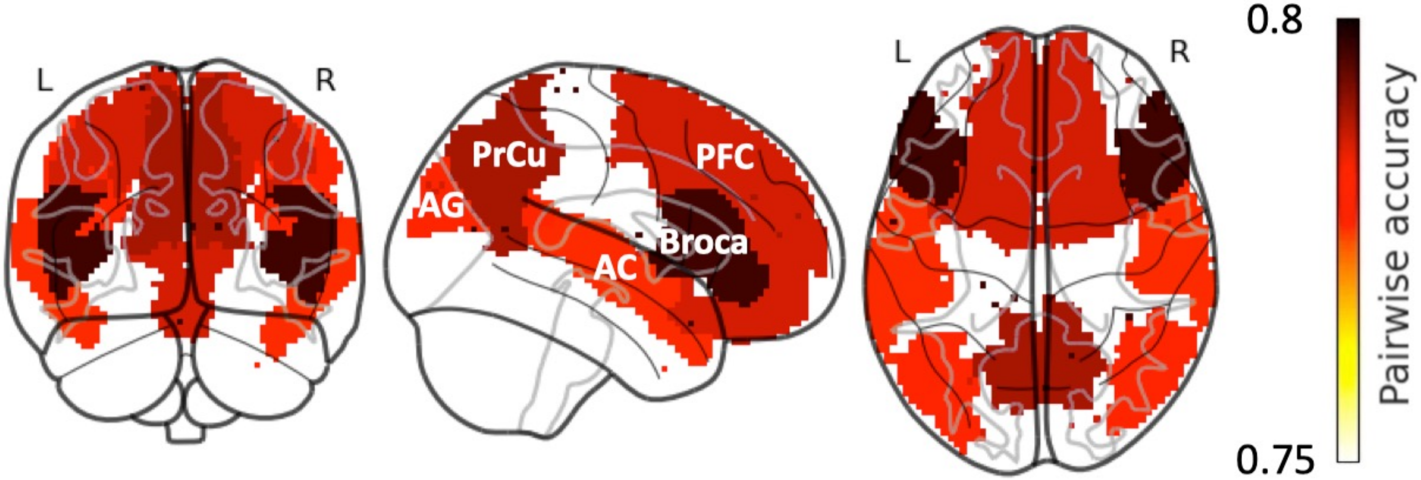


Fig：Pairwise accruacy of BrainLLM vs PerBrainLLM

Results

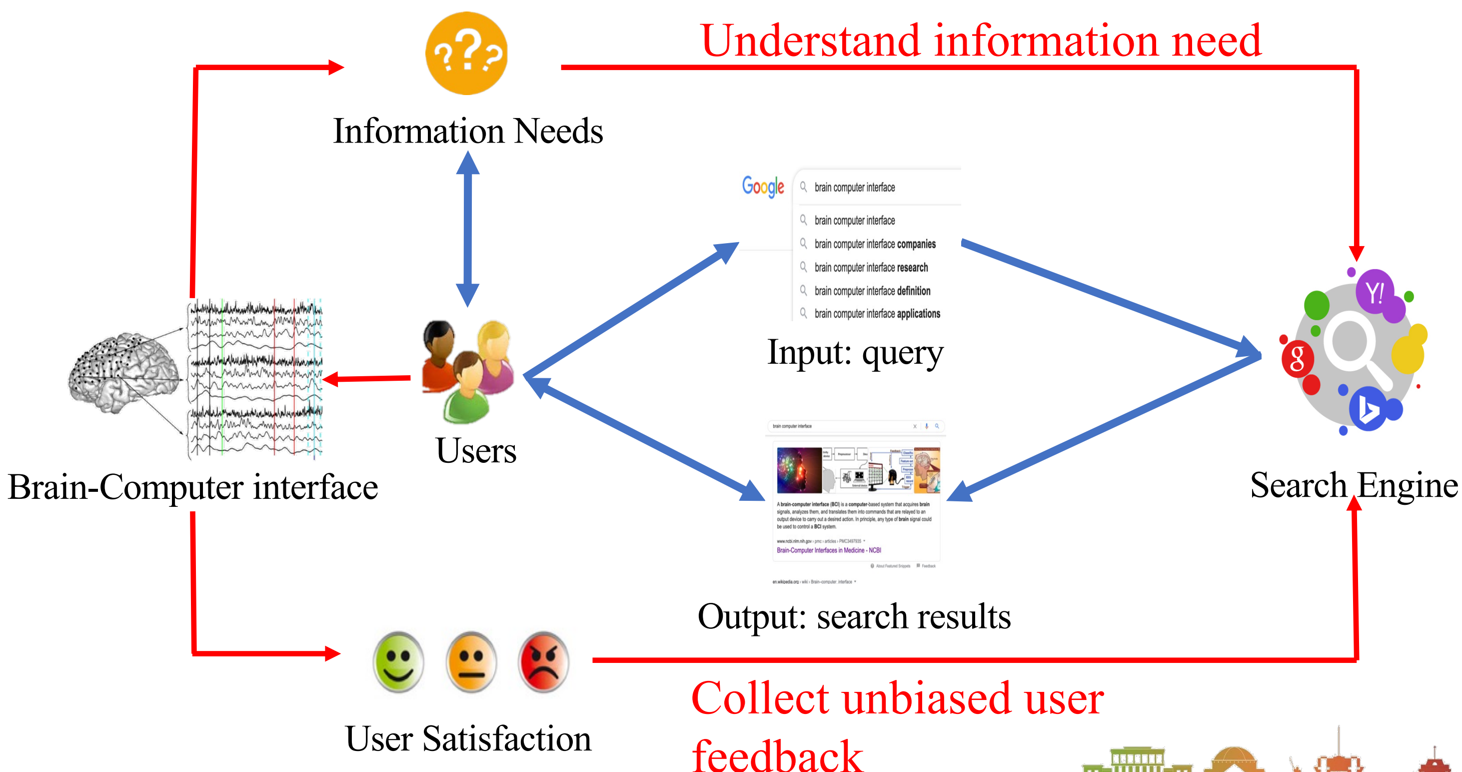
- Analysis regarding region of interests (ROIs)：
  - Broca: language production and grammar processing
  - PrCu: language memory, and language consciousness
  - PFC: decision-making
  - AC: auditory information processing
  - AG: semantic and phonological processing

Semantics encoded in human brain might be overlapping



Future application: BCI for Search

- **How can BCI help search**
  - Query augmentation via decoding information need from brain
  - Feedback modeling by decoding



Future application: more

- **BrainLLM for language BCIs**
  - Language construction without pre-generation
  - Integration with BCIs that utilize motor representations

- **Neurolinguistic research**
  - Quantification ability on the generation likelihood of textual content
  - E.g., no longer need manipulation for neurolinguistic experimental design

- **Personalized LLM**
  - Content deemed surprising by LLMs could potentially be corrected by individual's brain recordings

Ethics

- **Reconstruct language from the human brain**
  - Challenging the deeply ingrained notion of the mind as a private sanctuary
  - Currently at a very early stage

- **Direct language generation feature**
  - Without human-controlled pre-definition step
  - May decode contents that participants may wish to keep private

- **What should we do?**
  - Processing and remove privacy content from the output
  - Training a safe brain decoder
  - Reviewing the output by the participant

Reference

- [Cooney 2021] Moving Forward with Brain Machine Interfaces.
- [Makin 2021] Machine translation of cortical activity to text with an encoder-decoder framework. *Nature neuroscience.*
- [Edgar 2020] EEG-based BCI emotion recognition: A survey. *Sensors*.
- [Takagi 2023] High-resolution image reconstruction with latent diffusion models from human brain activity. *CVPR 2023.*
- [Mariya 2019] Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain). *Neurips 2023.*
- [Liu 2023] Coupling Artificial Neurons in BERT and Biological Neurons in the Human Brain. *AAAI 2023.*
- [Tang 2023] Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature neuroscience.*
- [Antonello 2023] Scaling laws for language encoding models in fMRI. *Neurips 2023.*

Thanks for your listening!