



HAL
open science

MEG Encoding using Neural Language & Speech Models and Shared Context Semantics in Listening Stories

Subba Reddy Oota, Nathan Trouvain, Gaël Jobard, Frédéric Alexandre,
Xavier Hinaut

► **To cite this version:**

Subba Reddy Oota, Nathan Trouvain, Gaël Jobard, Frédéric Alexandre, Xavier Hinaut. MEG Encoding using Neural Language & Speech Models and Shared Context Semantics in Listening Stories. SNL 2023 - Society for the Neurobiology of Language, Oct 2023, Marseille, France. . hal-04401769

HAL Id: hal-04401769

<https://hal.science/hal-04401769v1>

Submitted on 17 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Subba Reddy Oota^{1,2,3} Nathan Trouvain^{1,2,3} Gael Jobard³ Frédéric Alexandre^{1,2,3} Xavier Hinaut^{1,2,3}
¹INRIA Bordeaux ²LaBRI, UMR 5800 ³IMN, UMR 5293, Université de Bordeaux, France

Summary

- Language models (LMs) predict both text-evoked and speech-evoked brain activity to an impressive degree irrespective of choice of modality.
- Speech-based models (SMs), which would be expected to predict speech-evoked brain activity better, provided they model language processing in the brain well.
- These findings poses several questions:
 - Are speech models (SMs) outperform language models (LMs) during speech-evoked brain activity?
 - Do both LMs and SMs combine a new word representation with previous context?
 - What type of information is shared between LMs and SMs that results in improved brain predictivity?
- We investigate these questions using both LMs (GPT2) and SMs (Wav2vec2.0, HuBERT and Data2vec), and observe how these models predict MEG brain recordings acquired while participants listened to the naturalistic stories.
- Language models outperform speech models irrespective of speech-evoked brain activity.
- Both type of models still behind the estimated noise-ceiling performance.
- Like LMs, previous context is important in predicting MEG recordings for SMs

Language Models (LMs) / Speech Models (SMs) are trained to predict missing words/speech

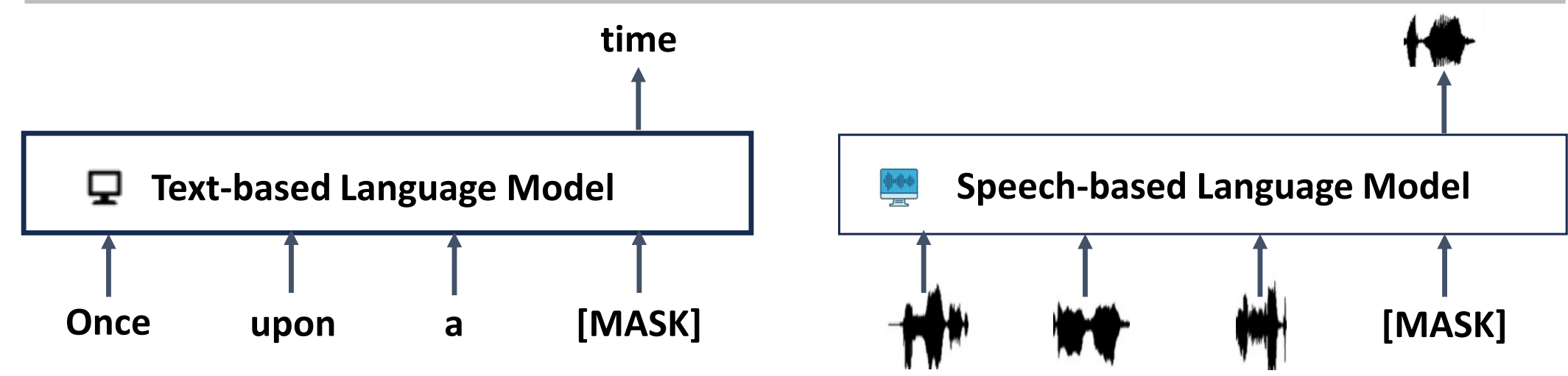


Figure 1: Language Models (LMs) and Speech Models (SMs) have achieved impressive performance across many benchmarks

LMs/SMs also predict brain activity evoked by complex language (e.g. listening a story) to an impressive degree

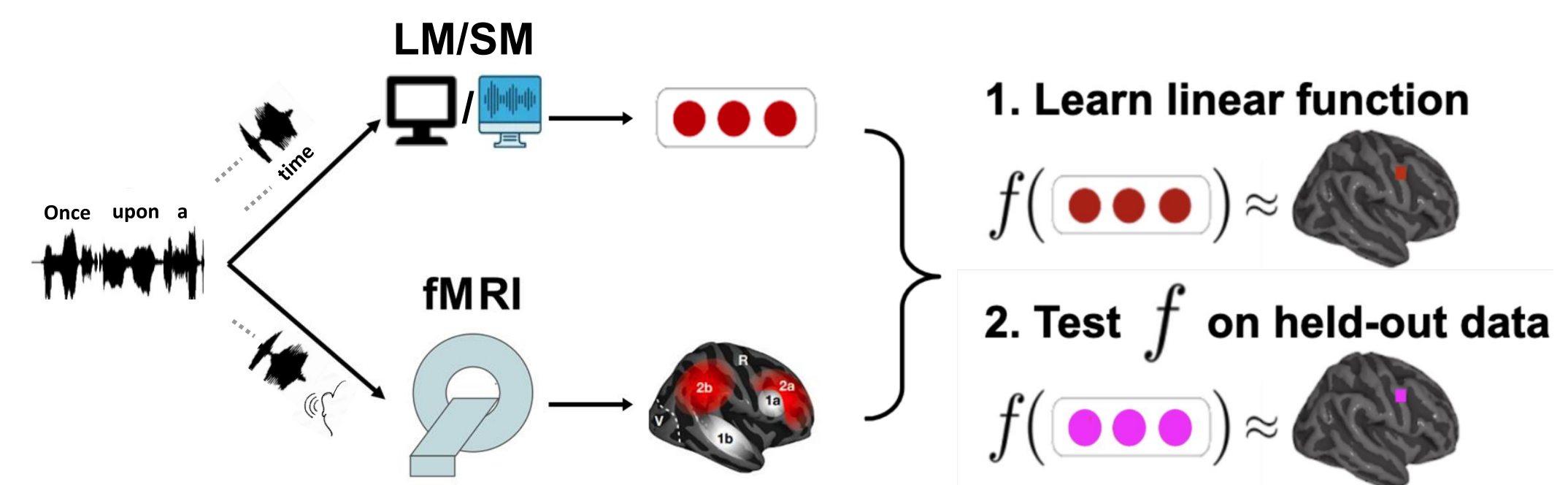
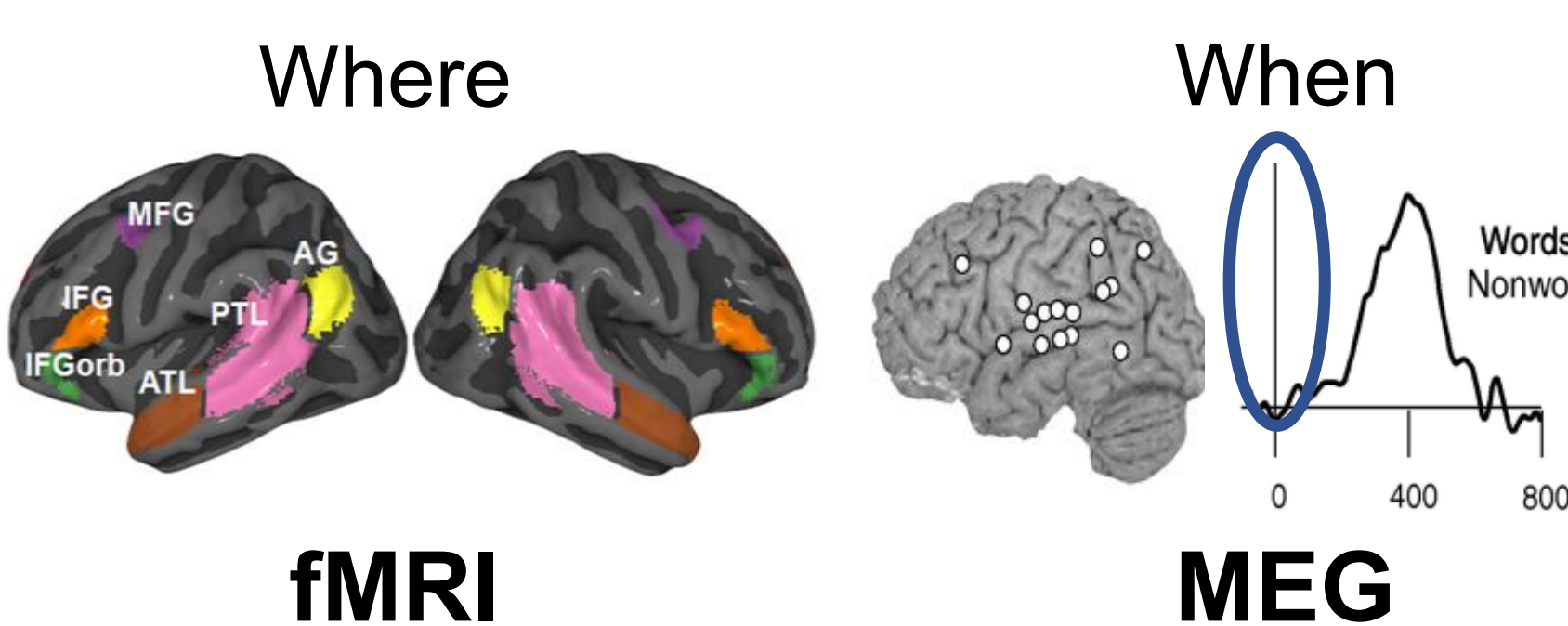


Figure 2: Brain alignment of a LM/SM

With MEG we can analyze sub-word time course

- MEG recording data at very fast temporal resolution
- So, we can look at sub-word process
- fMRI recording data at very high-spatial resolution



Brain encoding schema

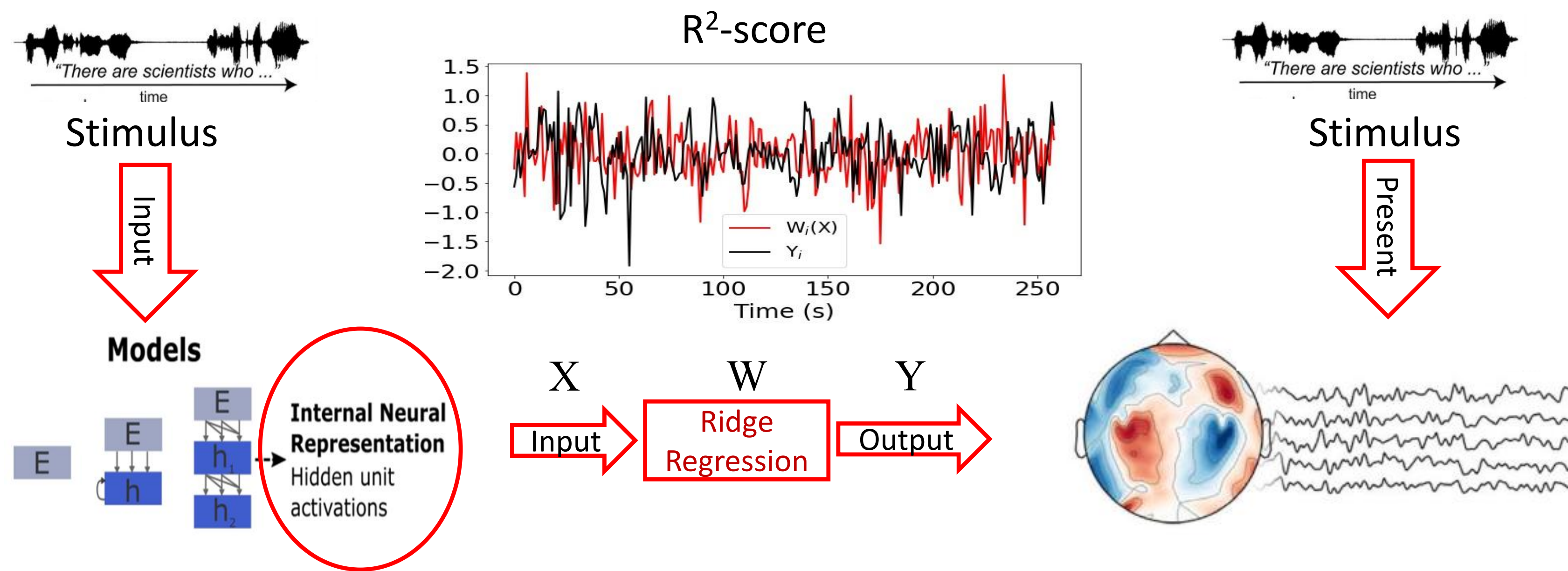
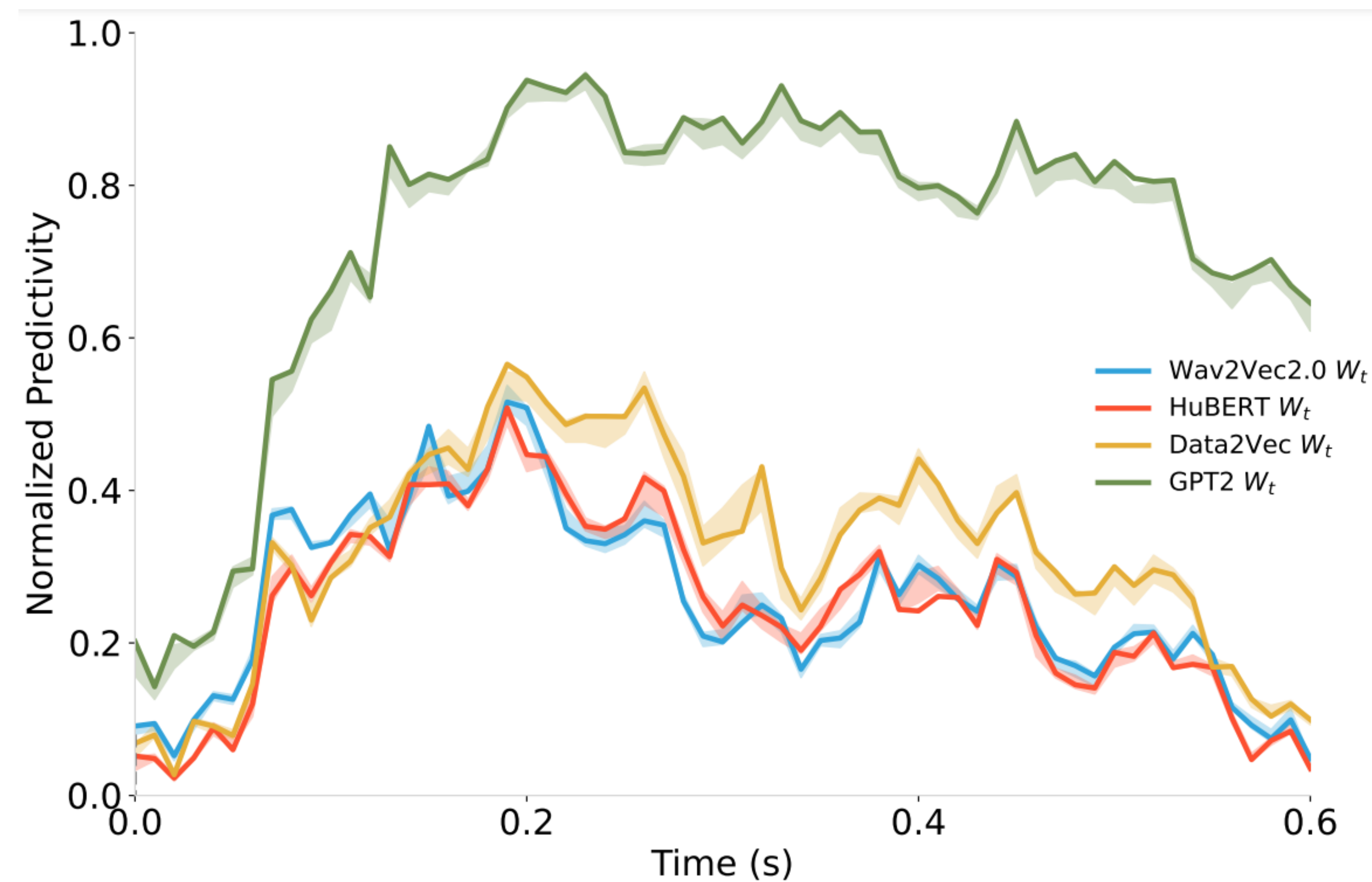


Figure 3: Approach to directly test for the alignment between a LM/SM and MEG brain recordings.

Encoding performance of LM and SMs



Brain alignment – 4-fold Cross-Validation + Ridge regression

MEG-MASC Dataset

- MEG recordings
- 27 participants
- 2 hours of story listening
- 2 repeated sessions
- structural MRIs
- audio, phonetic and word annotations
- standardised BIDS structure

(1) Naturalistic story listening dataset

- 4 natural stories (11,002 words) x 768 → LM/SM representations
- 800ms signal window around word onset: 200ms before (baseline correction), 600ms after
- 208 MEG sensors x 81 time points → MEG predictions (same dimensions as actual brain activity)
- 18 participants
- 3 Speech language models
 - Wav2vec2.0
 - HuBERT
 - Data2Vec
- 1 Text language Model (GPT2)

(2) Brain Alignment: Dataset Curation

3 stories in training & 1 story in testing

(3) 4-fold Cross-Validation

Layer-wise Performance of SM: Data2vec

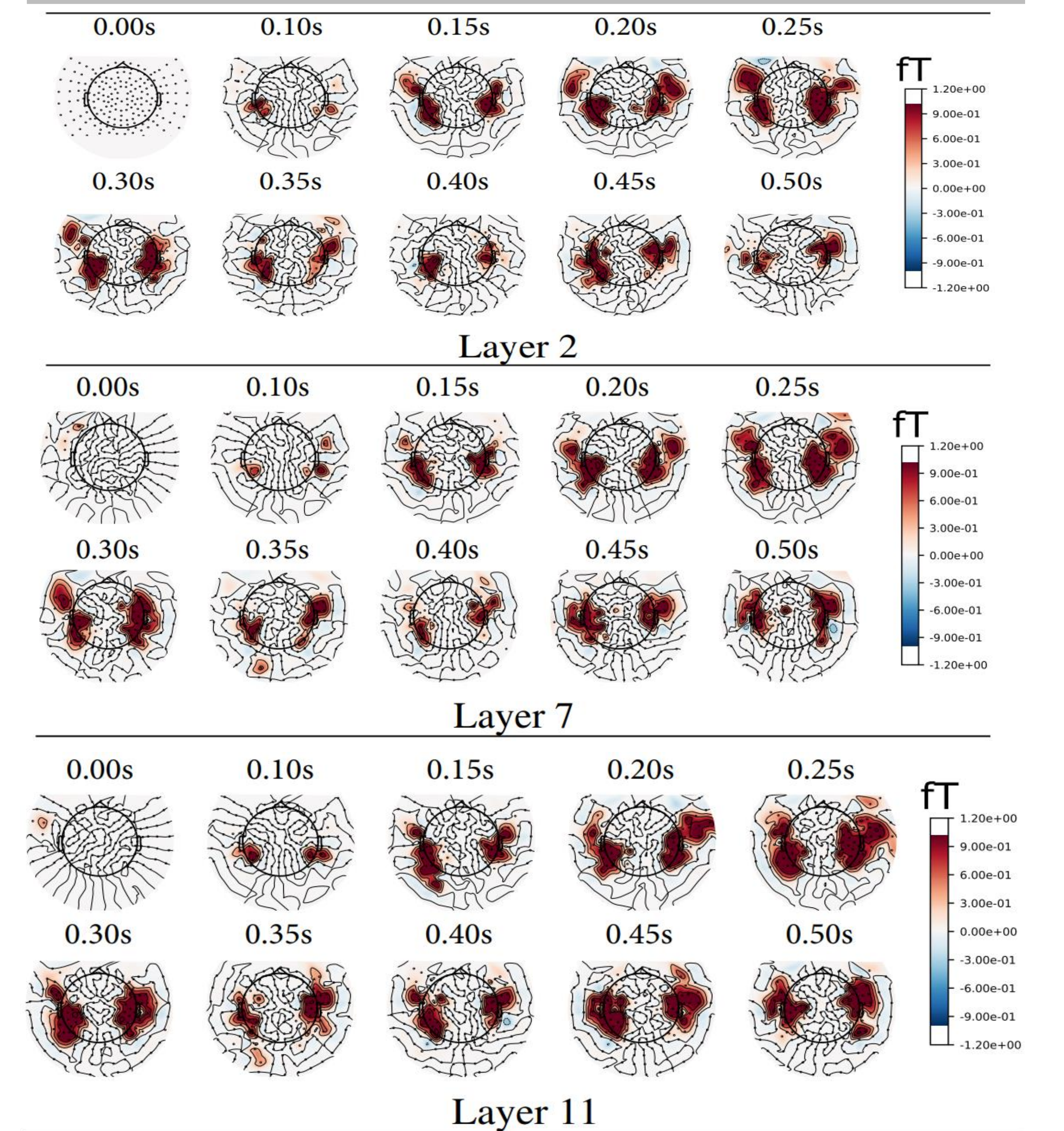


Figure 5: Topo map for Average Normalized predictivity by Data2Vec (across subjects) for MEG activity. Word onset is at 0ms. Layer 2 (top), Layer 7 (middle), and Layer 11 (bottom).

References

- Toneva et al. 2019. Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain).
- Gwilliams et al. 2022. Meg-masc: a high-quality magneto-encephalography dataset for evaluating natural speech processing.
- Oota et al. 2023. Meg encoding using word context semantics in listening stories.

Acknowledgements

1 We thank Inria for the PhD funding and ANR DeepPool project.