MAX PLANCK INSTITUTE
FOR SOFTWARE SYSTEMS

SIC Saarland Informatics Campus

Brain-tuning Improves Generalizability and Efficiency of Brain Alignment in Speech Models

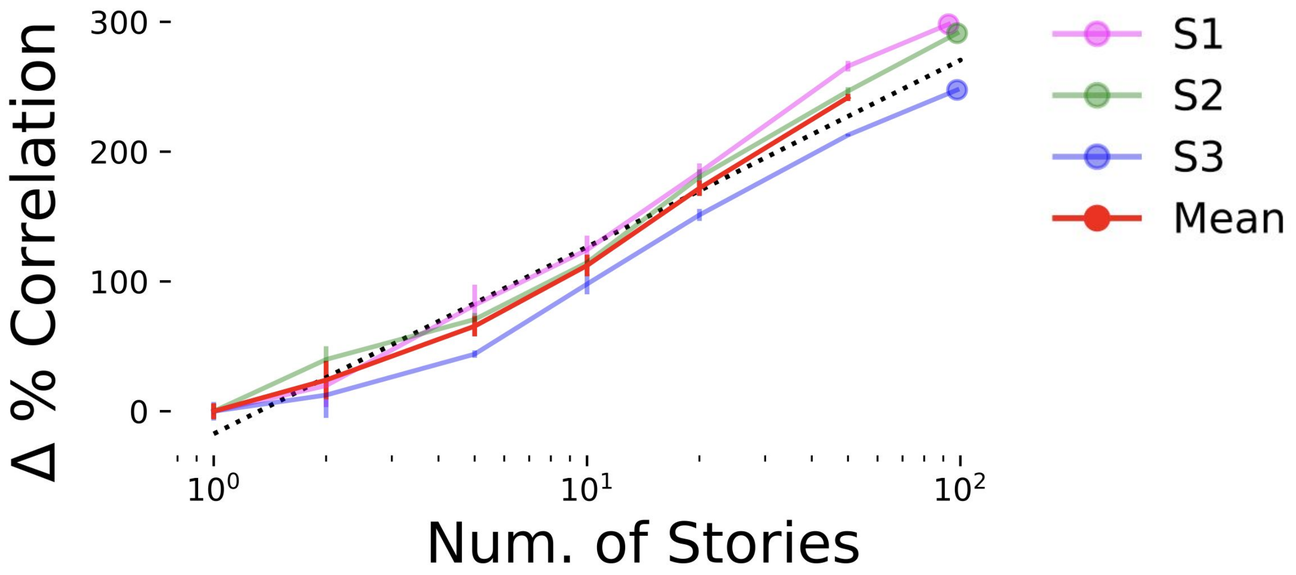
Omer Moussa, Mariya Toneva

https://neurips.cc/virtual/2025/loc/san-diego/poster/119924

Studying the brain with encoding models

Encoding models with fMRI data are participant-specific

Estimating brain alignment via brain encoding models is data hungry [1]



Brain encoding models lack efficiency and generalisability

[1]Antonello et al., 2024

Brain-tuning of speech models

Brain-tuning: inducing brain relevant **bias** directly into the model by **fine-tuning** with brain fMRI data.

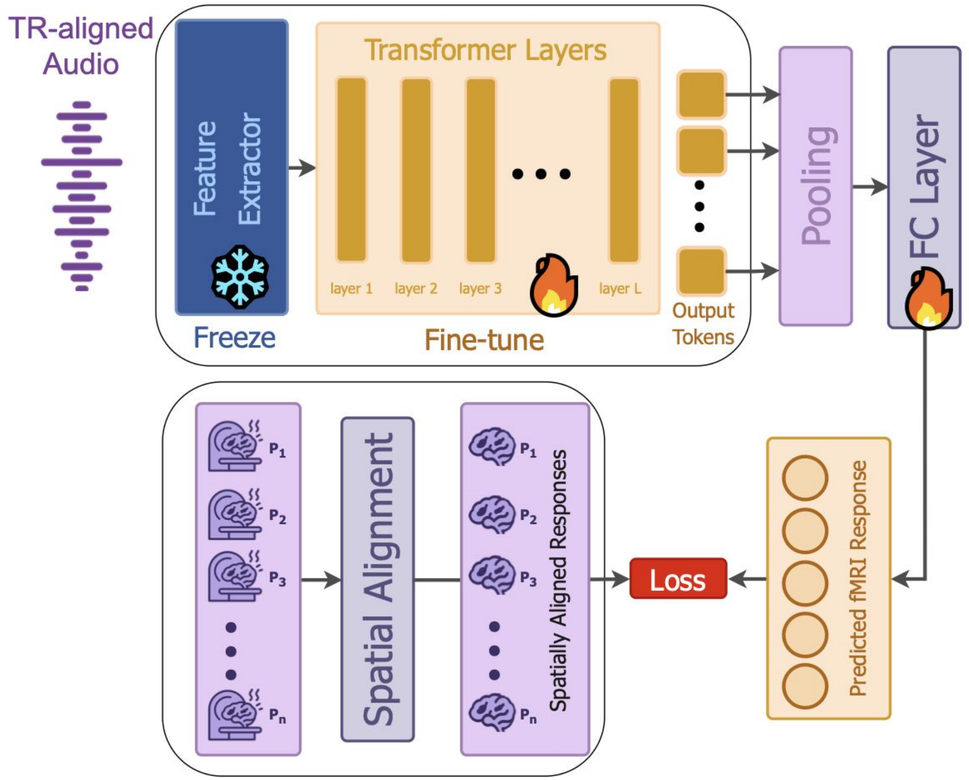
Brain-tuning with multiple participants makes speech model representations more efficient and generalisable for encoding models.

Brain-tuning with multiple participants

Spatial Alignment: Method to project responses into a common space (e.g., FreeSurfer)

Binding by stimulus: for each shared stimulus, present responses consecutively

- Use LoRA Updates
- One FC layer to predict responses
- Fine-tune FC layer + LoRA params



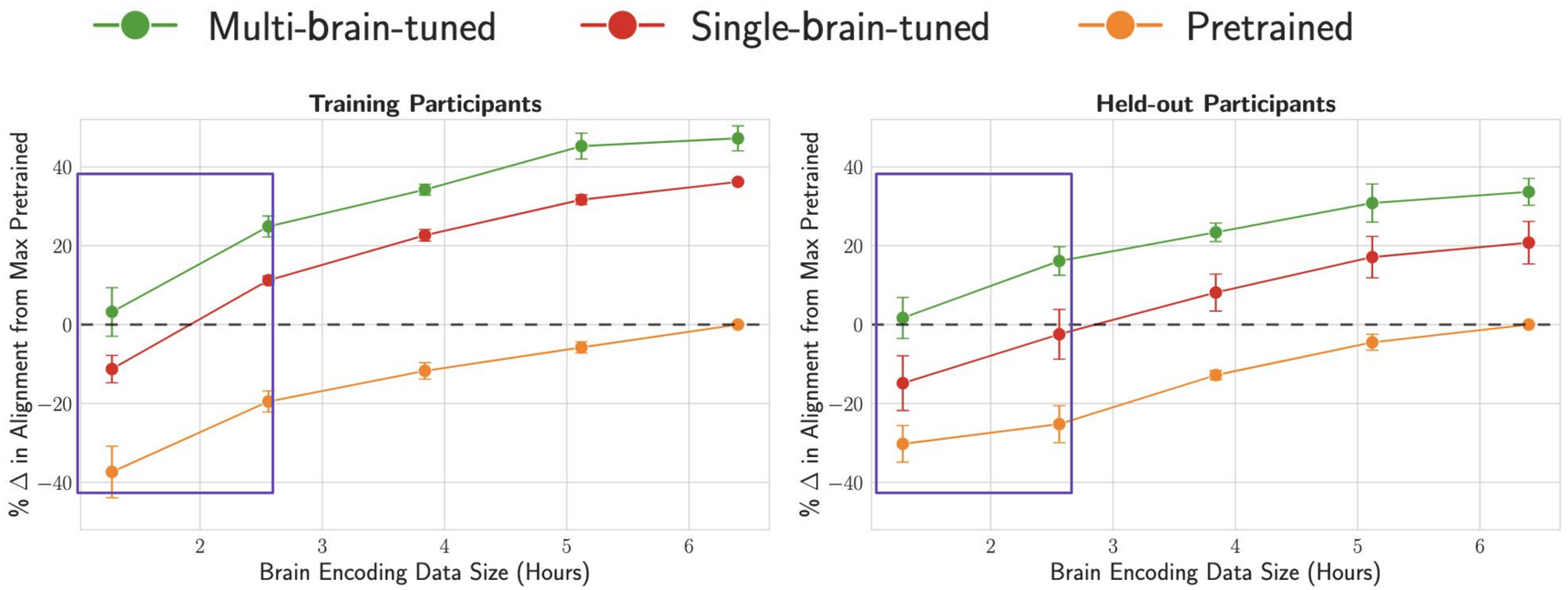
Setup: one FC layer + FreeSurfer (Lang regions) + Binding by stimulus + L2 Loss

Evaluating Multi-brain-tuning efficiency

How much data does it take for to perform the same (same brain alignment) as the pretrained one?

Performance is measured as pct change from pretrained with all fMRI data

Multi-brain-tuning reduces data needed by 5-fold



Evaluating Multi-brain-tuning generalization

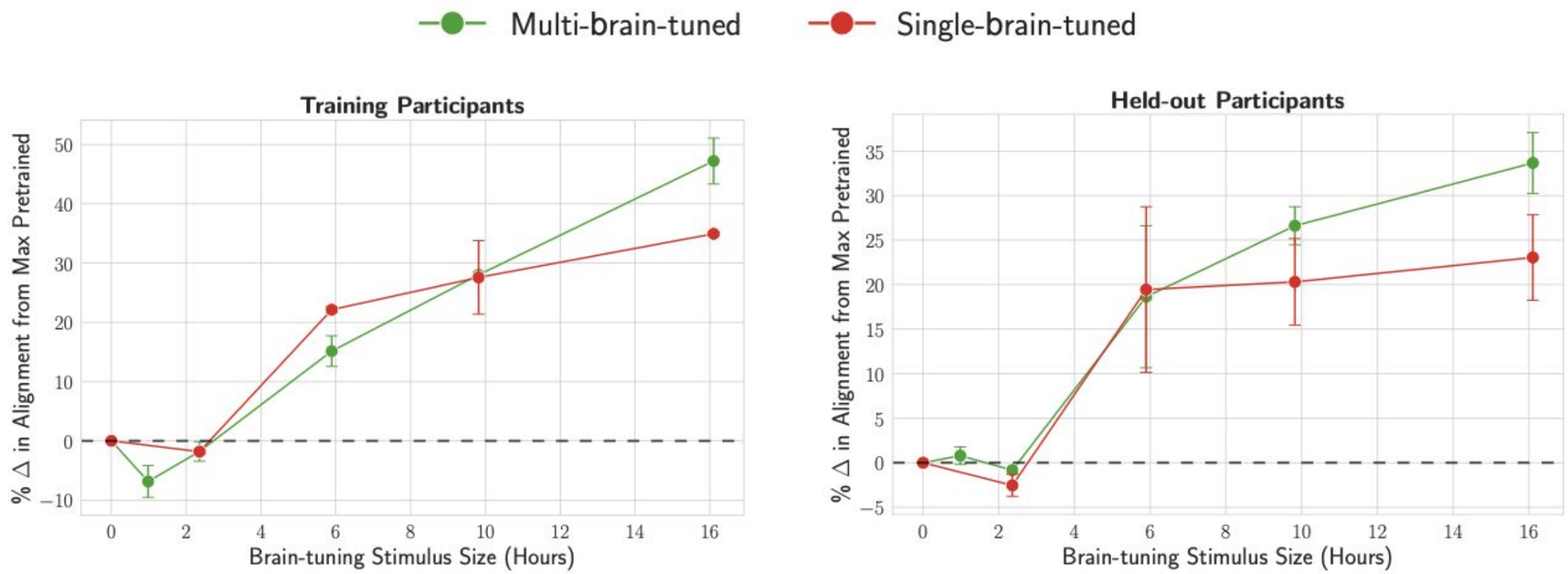
(1) How does improvement scale with data?

(2) Can improvement generalize to other datasets?

(1) Multi-brain-tuning scales well with sufficient data

Worse than Single-brain-tuned when data is very small

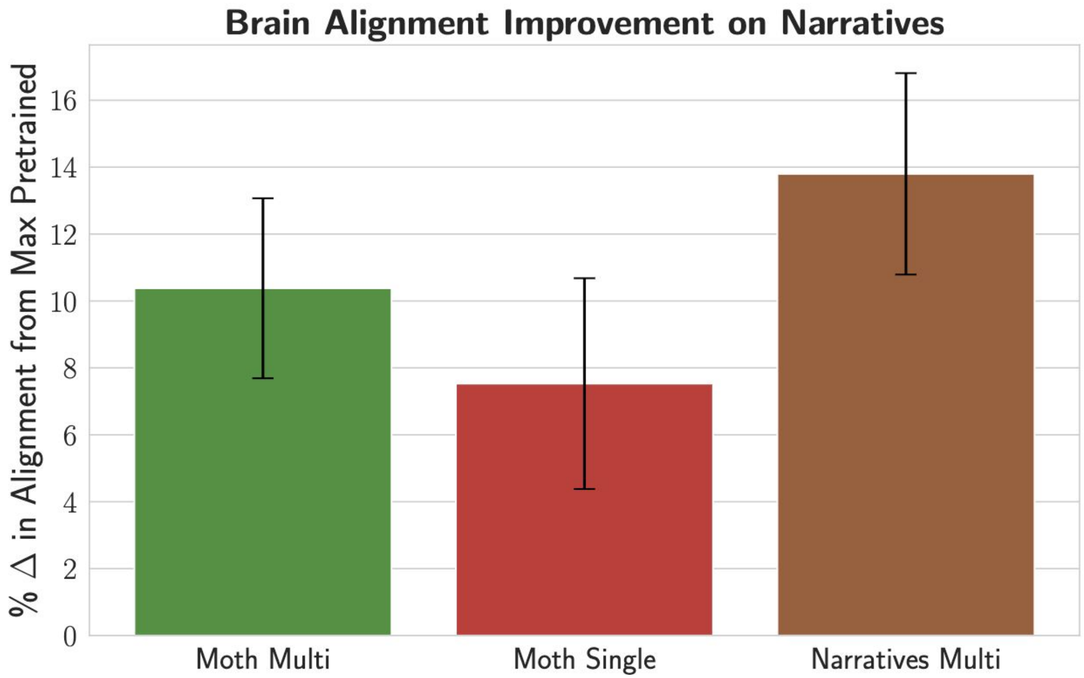
upward trend with more data



(2) Multi-Brain-tuned improves alignment on narratives

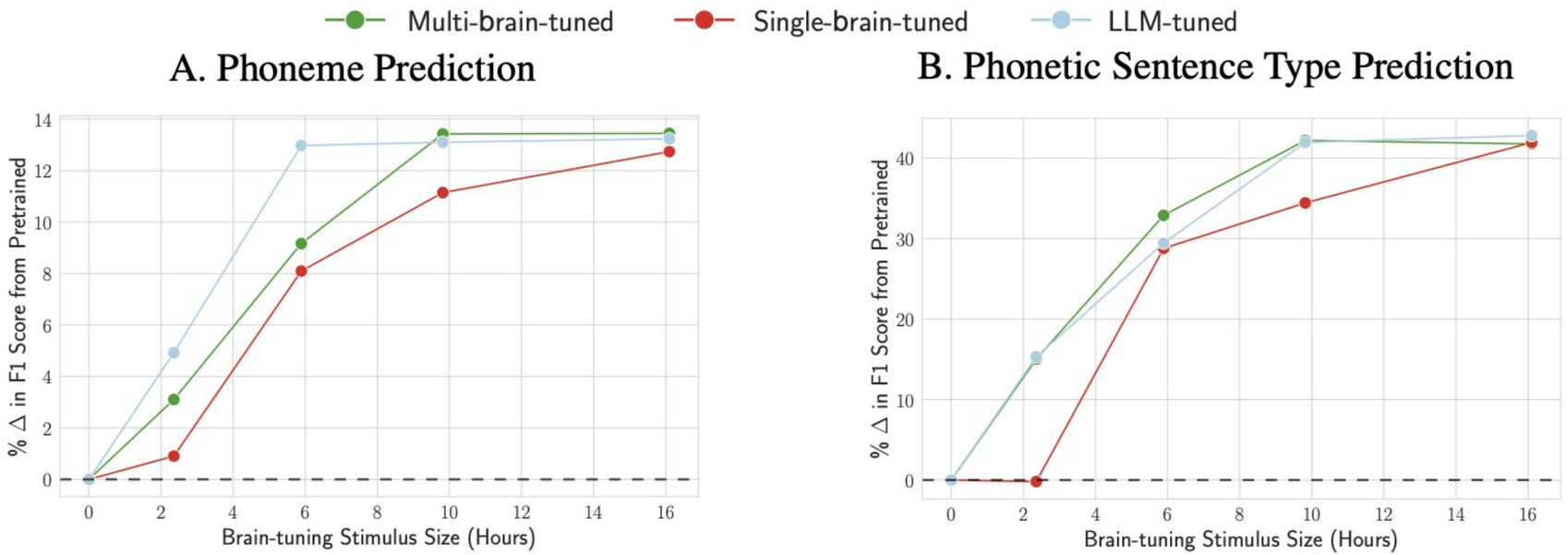
16 participants and only a 56min stimulus, compare with brain-tuned on the same participants

We observe improvement even with many participants and limited per-participant data



Evaluating downstream performance

To ensure downstream utility of the model is not affected



With more data, reaches same performance as LLM-tuned (fine-tuned with Llama2 7B representations)

Conclusion and future work

✓ **Multi-brain-tuning leads to better generalization and efficiency of brain alignment models**

✓ **Improves downstream performance**

Future work: extending to bigger models, more diverse datasets, and multiple modalities

Thank you!

Questions?

Poster: Wed 3 Dec 4:30 p.m. — 7:30 p.m.

Feel Free to reach us at:

omoussa@mpi-sws.org
mtoneva@mpi-sws.org

https://neurips.cc/virtual/2025/loc/san-diego/poster/119924

References

- Scaling laws for language encoding models in fMRI: https://arxiv.org/pdf/2305.11863
- Scaling laws for decoding images from brain activity: https://arxiv.org/pdf/2501.15322
- Semantic language decoding across participants and stimulus modalities: https://www.sciencedirect.com/science/article/pii/S096098222500054 5
- BrainWavLM: Fine-tuning Speech Representations with Brain Responses to Language: https://arxiv.org/pdf/2502.08866v1