



SPINT: Spatial Permutation-Invariant Neural Transformer for Consistent Intracortical Motor Decoding

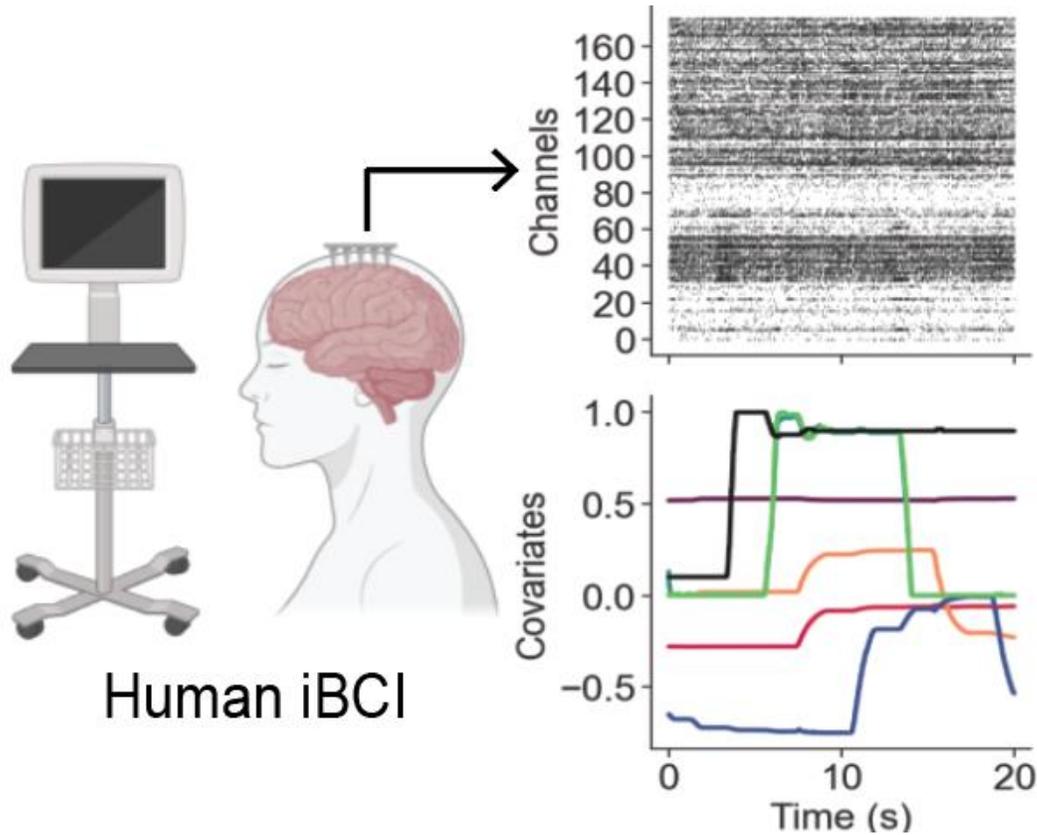
Presenter: Dr. Hao Fang

Aug 16, 2025

(Submitted to NeurIPS25, under review)

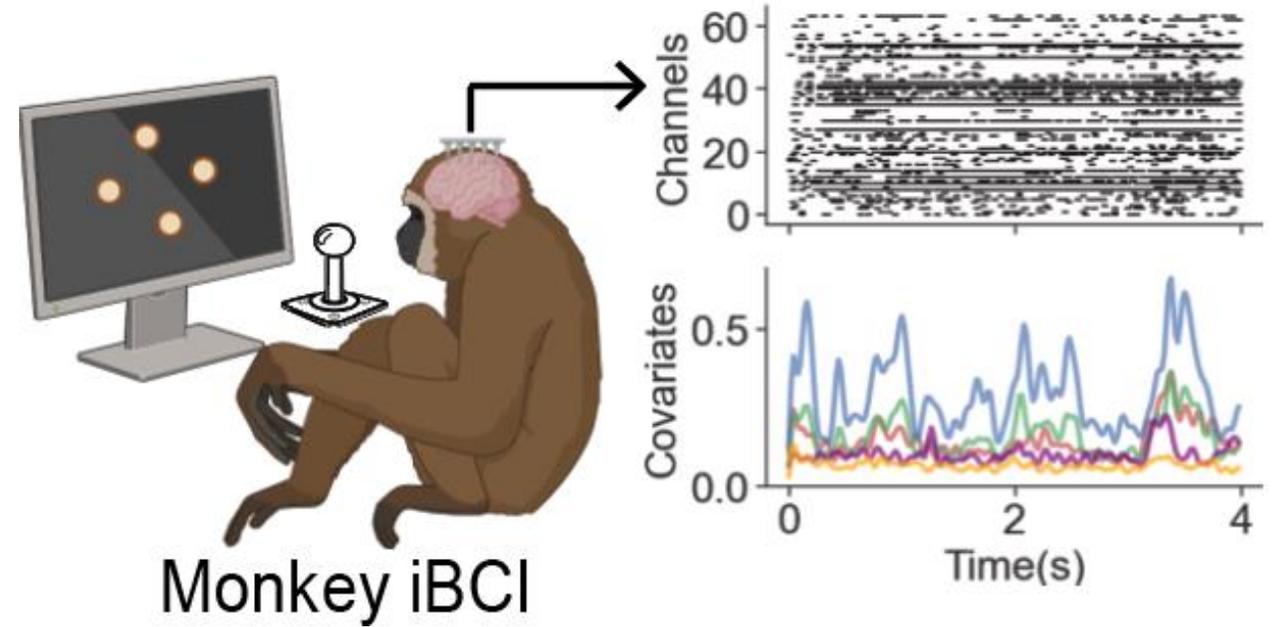


- Intracortical brain-computer interfaces (iBCIs)



Example: decode attempted human intentions^{1,2,3}

[1] Fan et al., *NeurIPS*, 2023;
[2] Flesher et al., *Science*, 2021;
[3] Willett et al., *Nature*, 2021



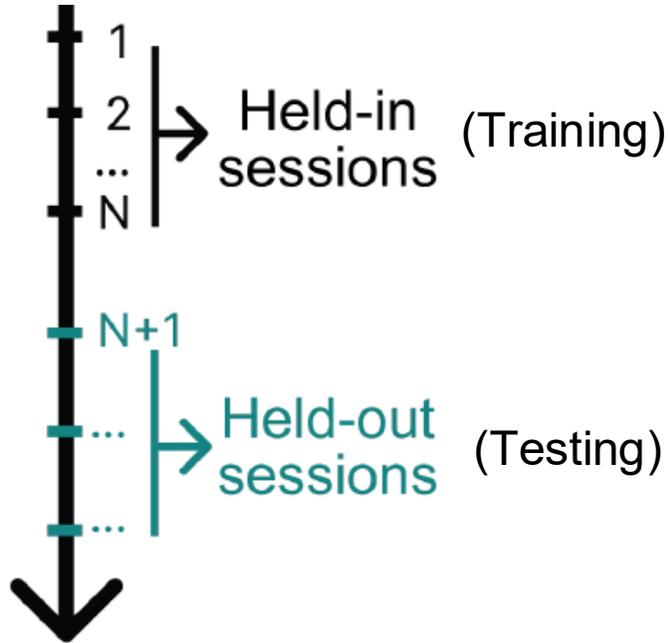
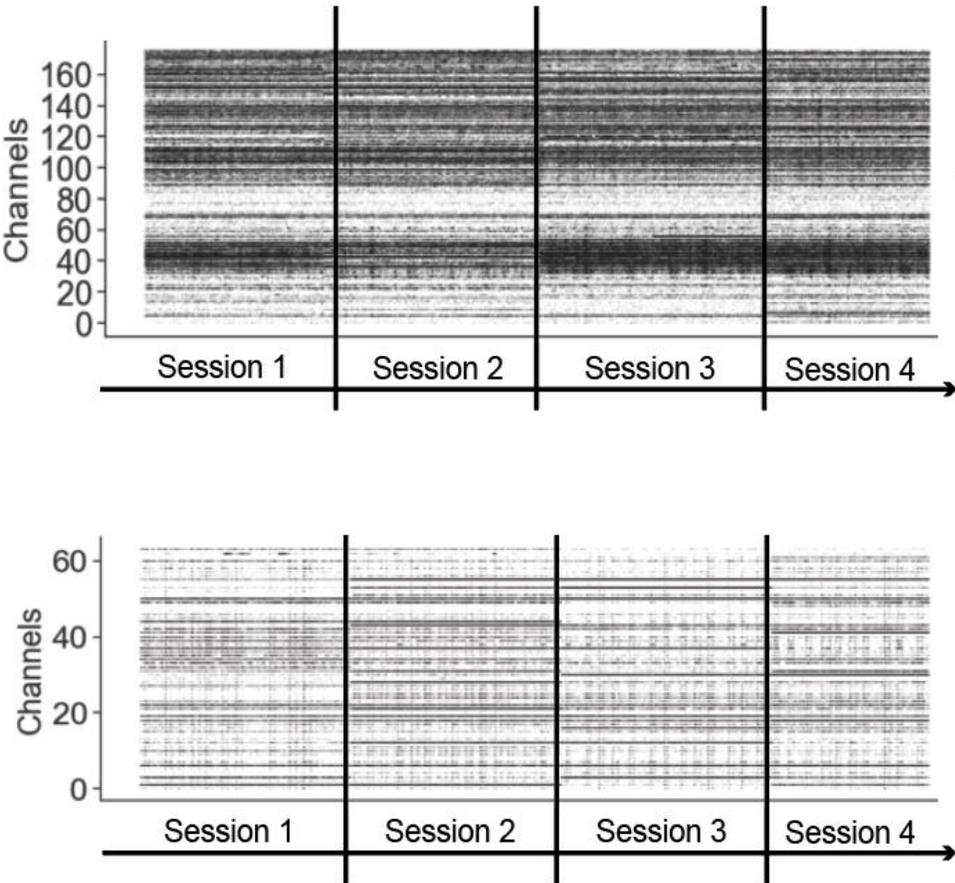
Example: decode cursor/joystick movement^{4,5}

[1] Nason et al., *Neuron*, 2021;
[2] Rouse et al., *Cell Report*, 2019;

Background

- Nonstationarity degrades the decoder's performance across sessions

Nonstationary neural activity over sessions



Methods:

- Zero-shot (ZS)
- Few-shot unsupervised^{1,2} (FSU)
- Few-shot supervised² (FSS)

Usually, FS methods will use finetuning.

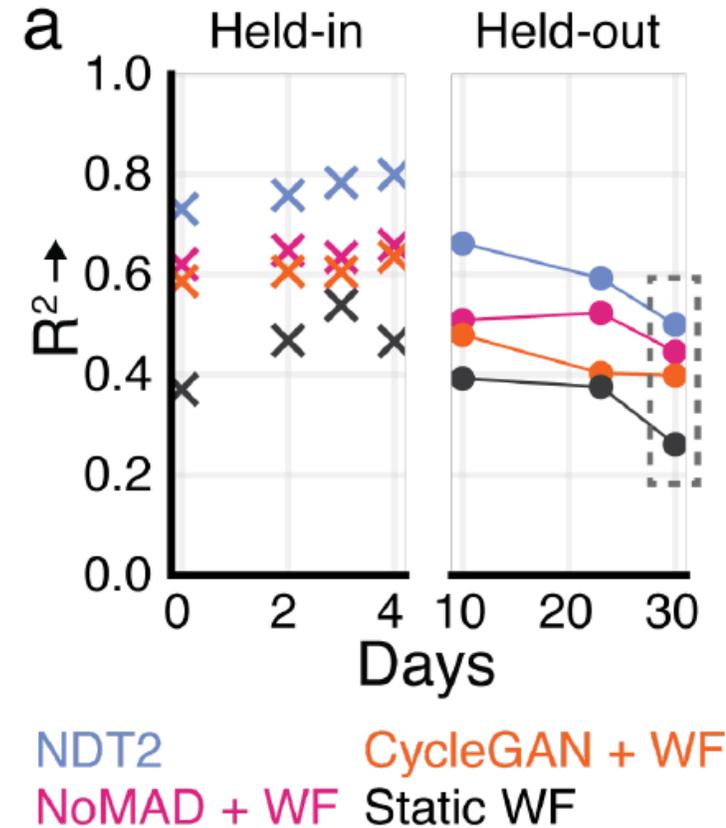
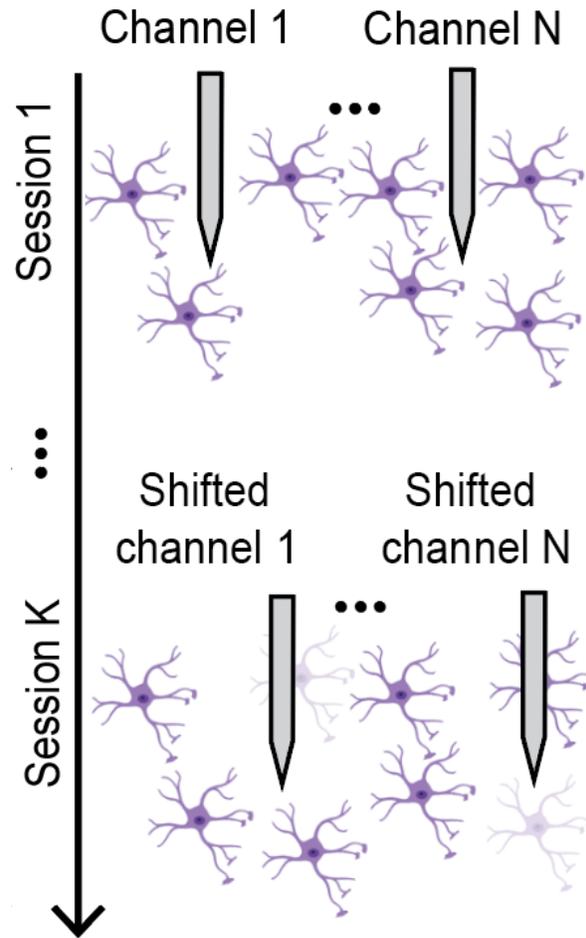


Figure a is from Karpowicz et al., NeurIPS, 2024

- [1] Ma et al., *elife*, 2023;
- [2] Karpowicz et al., *bioRxiv*, 2022;
- [3] Ye et al., *NeurIPS*, 2023;

- Possible situations causing nonstationarity across sessions

1. Neuron disappears
2. Neuron appears
3. Channel location shifts



Question: how can we model these situations when we build/train our decoder?

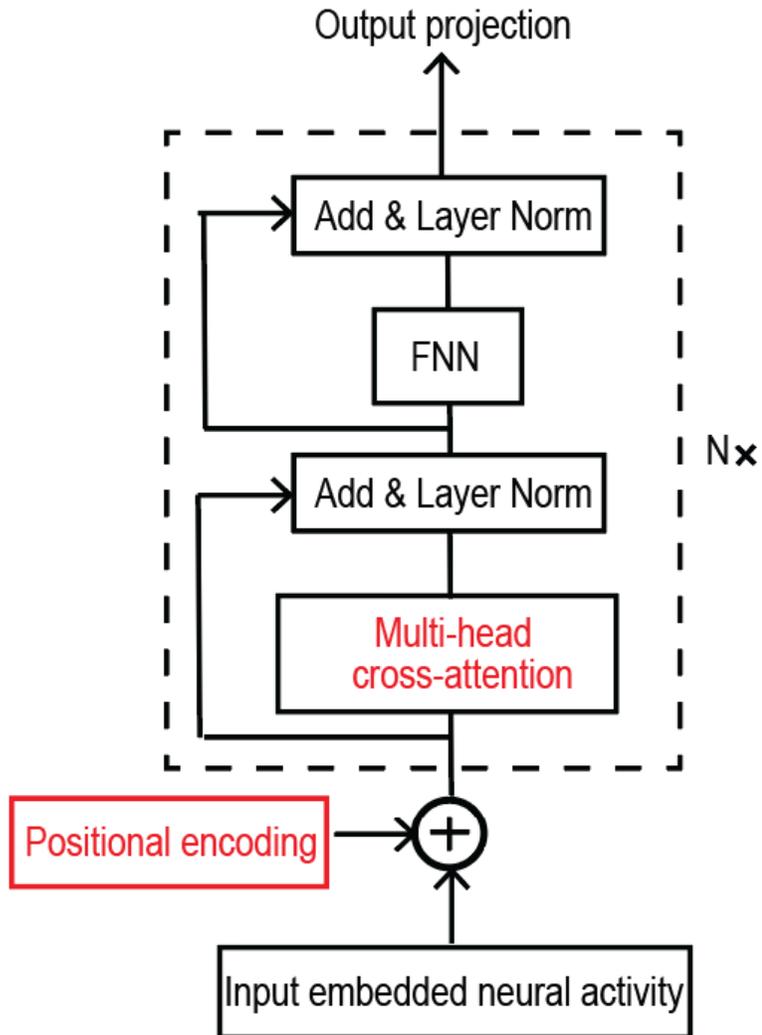
Hint: **Rethink** the drawbacks of standard architecture of transformer model.

Solution: leverage the **few-shot data for better inductive information** of cross-session nonstationarity.

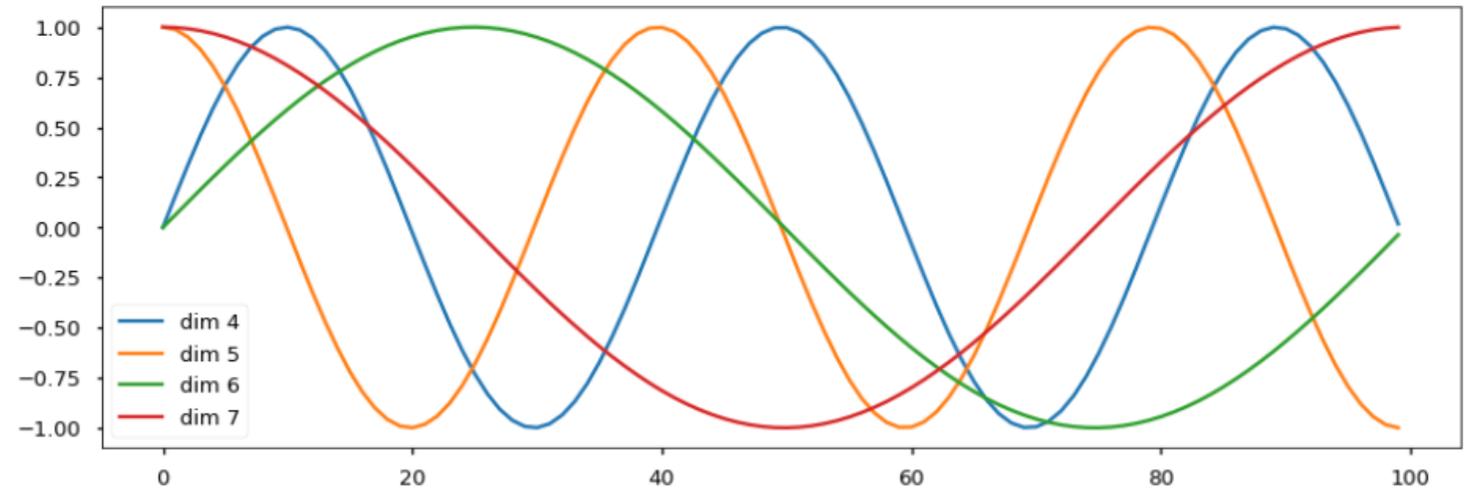
Standard transformer model



- Standard transformer model



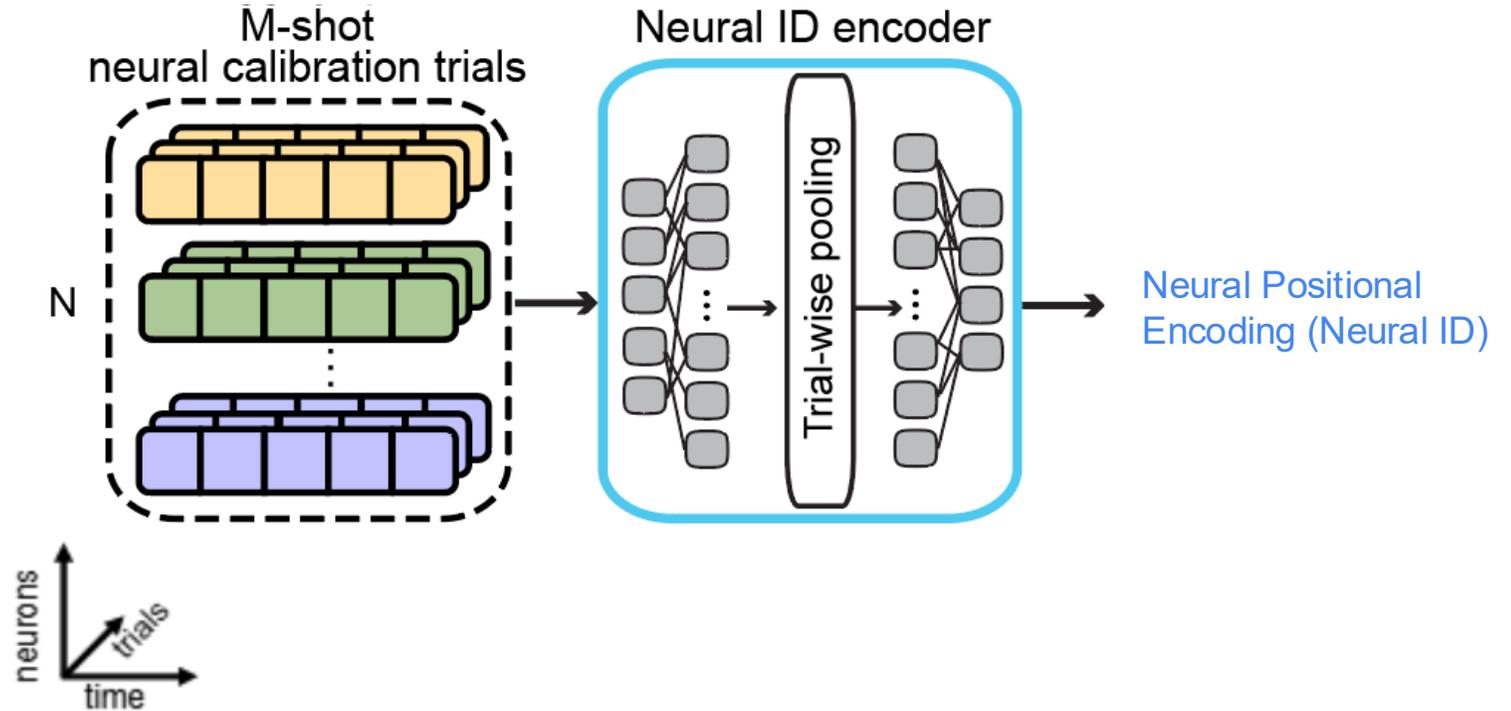
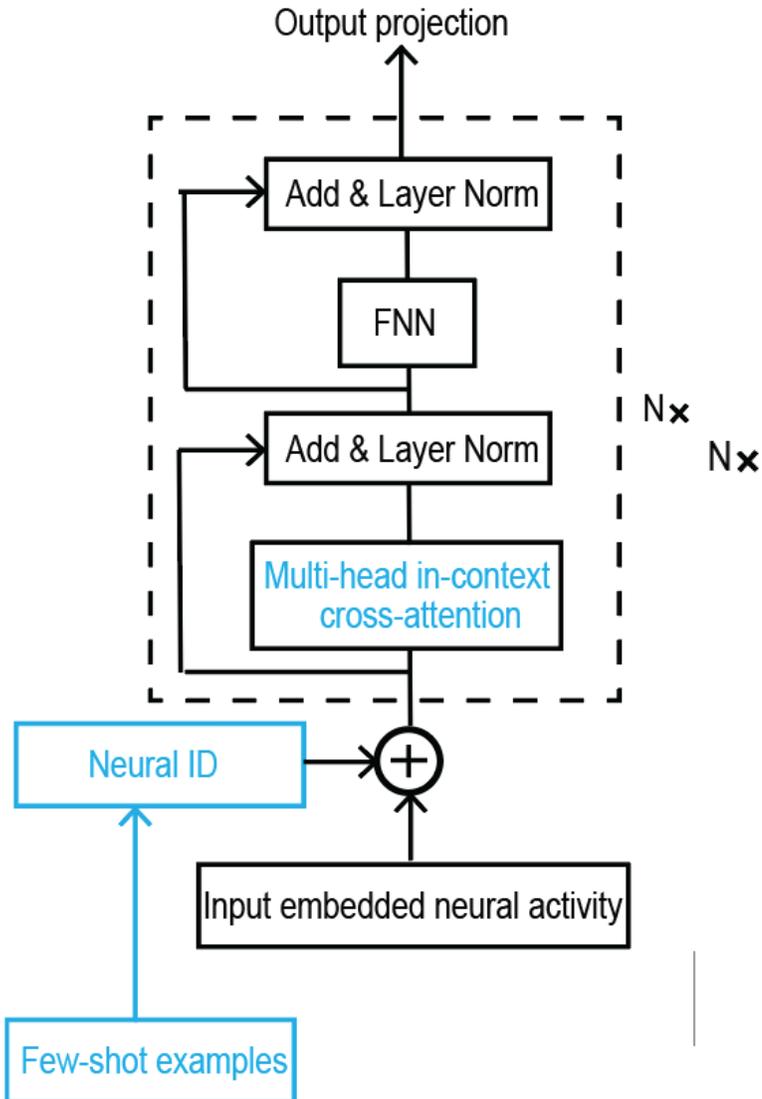
Positional encoding: Fixed sinusoid signal for different dimensions



Problems: Fail to capture the in-context information across session

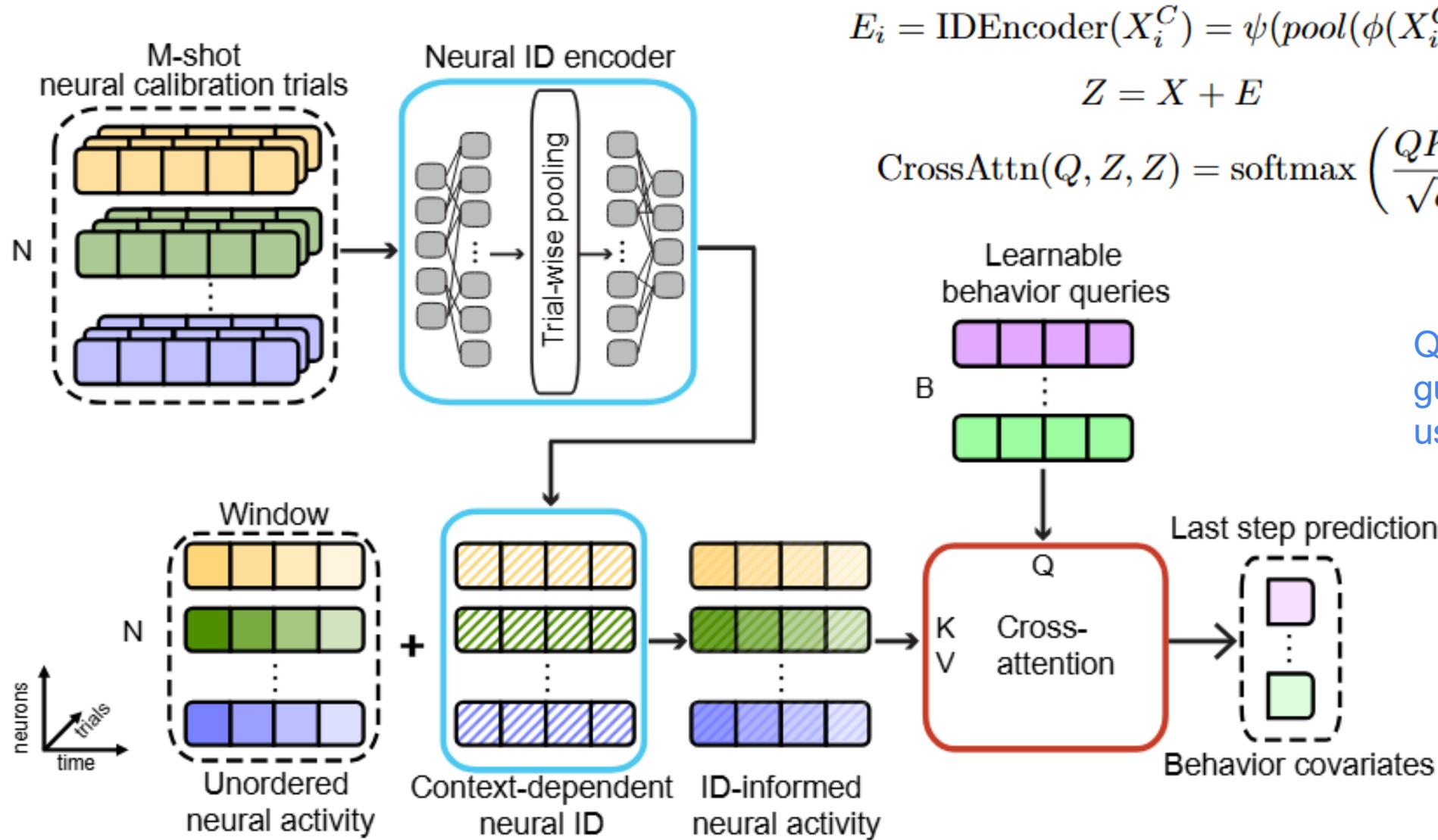
Transformer model from standard to new

- Standard transformer model
- Neural ID encoder design



1. The Neural ID serves as **context-dependent inductive bias** to improve the cross-session performance.
2. **No finetune/retrain is required.** Only a few-shot test examples at test phase.

Complete transformer model diagram



$$E_i = \text{IDEncoder}(X_i^C) = \psi(\text{pool}(\phi(X_i^C))) \quad (1)$$

$$Z = X + E \quad (2)$$

$$\text{CrossAttn}(Q, Z, Z) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

Question: any mathematical guarantees/benefits by using Neural ID encoder?

Permutation invariant property

Proposition 1. *Cross-attention with identity-informed neural activity (Equation 3) is invariant to the permutation of neural units, i.e.,*

$$\text{CrossAttn}(Q, Z, Z) = \text{CrossAttn}(Q, P_R Z, P_R Z), \quad (4)$$

where P_R is the row permutation matrix. (See proof in Appendix).

$$\text{CrossAttn}(Q, Z, Z) = \text{softmax} \left(\frac{QK^\top}{\sqrt{d_k}} \right) V \quad (3)$$

It follows that:

$$Z' = X' + E' = P_R X + P_R E = P_R (X + E) = P_R Z$$

In other words, Z is equivariant to the permutation of neural units.

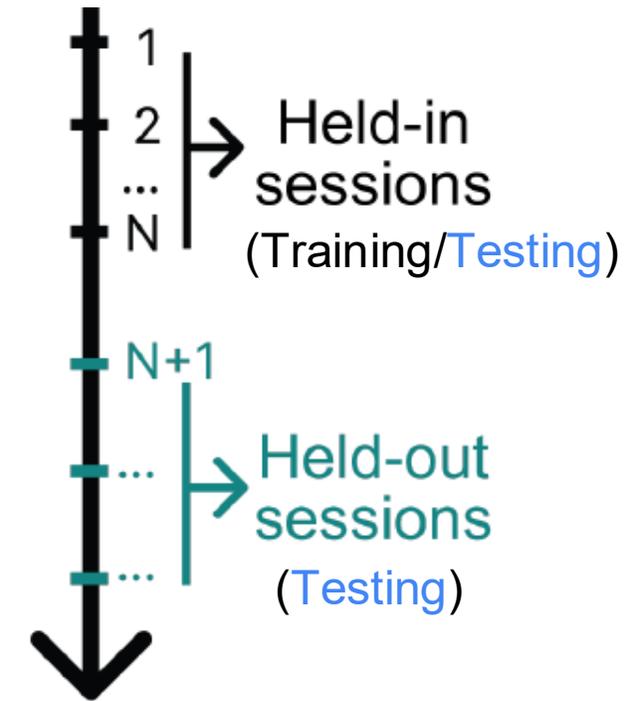
Cross-attention performed on Z' then becomes:

$$\begin{aligned} \text{CrossAttn}(Q, Z', Z') &= \text{CrossAttn}(Q, P_R Z, P_R Z) \\ &= \text{softmax} \left(\frac{QW_K^\top Z'^\top P_R^\top}{\sqrt{d_k}} \right) P_R Z \\ &= \text{softmax} \left(\frac{QW_K^\top Z'^\top P_C}{\sqrt{d_k}} \right) P_R Z \\ &= \text{softmax} \left(\frac{QW_K^\top Z'^\top}{\sqrt{d_k}} \right) P_C P_R Z \\ &= \text{softmax} \left(\frac{QW_K^\top Z'^\top}{\sqrt{d_k}} \right) Z \\ &= \text{CrossAttn}(Q, Z, Z) \end{aligned}$$

We use the property of row/column permutation matrix, i.e., $P_C = P_R^\top$ and $P_R P_C = I$

Experimental results

- Two evaluations



Methods:

- Oracle (OR: ideal upper bound)
- Zero-shot (ZS)
- Few-shot unsupervised (FSU)
- Few-shot supervised (FSS)

- SOTA** in held-in sessions

	Class	M1	M2	H1
Wiener Filter (WF)	OR	0.54 ± 0.01	0.27 ± 0.02	0.24 ± 0.02
RNN	OR	0.75 ± 0.03	0.59 ± 0.07	0.51 ± 0.09
NDT2 Multi [1]	OR	0.77 ± 0.03	0.62 ± 0.03	0.68 ± 0.05
NDT2 Multi [1]	FSS	0.77 ± 0.03	0.63 ± 0.03	0.62 ± 0.04
WF	ZS	0.46 ± 0.06	0.15 ± 0.07	0.20 ± 0.04
RNN	ZS	0.52 ± 0.15	0.20 ± 0.29	0.31 ± 0.13
CycleGAN + WF [2]	FSU	0.61 ± 0.02	0.32 ± 0.03	0.15 ± 0.04
NoMAD + WF [3]	FSU	0.64 ± 0.01	0.35 ± 0.05	0.21 ± 0.06
SPINT (Ours)	GF-FSU	0.77 ± 0.02	0.59 ± 0.01	0.47 ± 0.06

- SOTA** in held-out sessions **without finetuning**

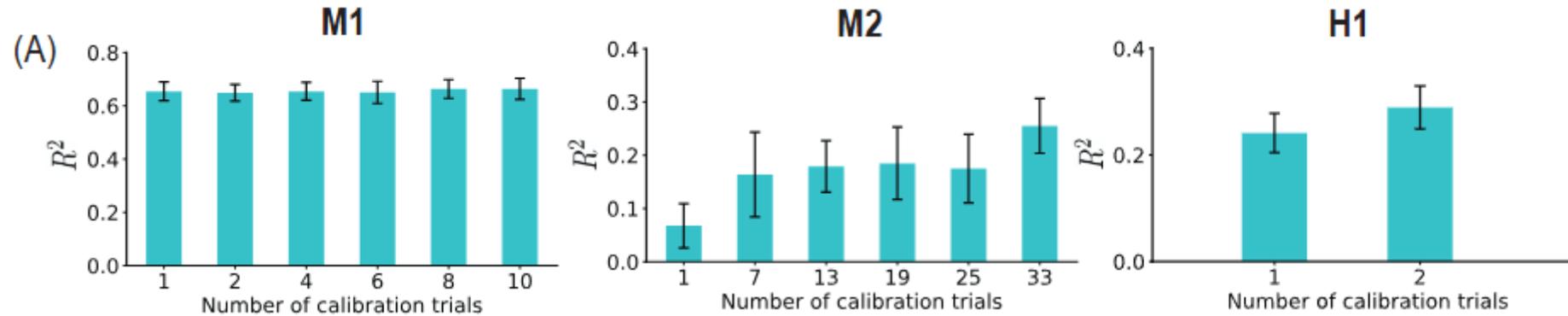
	Class	M1	M2	H1
Wiener Filter (WF)	OR	0.53 ± 0.04	0.26 ± 0.03	0.21 ± 0.04
RNN	OR	0.75 ± 0.05	0.56 ± 0.04	0.44 ± 0.13
NDT2 Multi [14]	OR	0.78 ± 0.04	0.58 ± 0.04	0.63 ± 0.08
NDT2 Multi [14]	FSS	0.59 ± 0.07	0.43 ± 0.08	0.52 ± 0.04
WF	ZS	0.34 ± 0.06	0.06 ± 0.04	0.16 ± 0.03
RNN	ZS	-0.60 ± 0.45	-0.07 ± 0.23	0.09 ± 0.18
CycleGAN + WF [23]	FSU	0.43 ± 0.04	0.22 ± 0.06	0.12 ± 0.06
NoMAD + WF [19]	FSU	0.49 ± 0.03	0.20 ± 0.10	0.13 ± 0.10
SPINT (Ours)	GF-FSU	0.66 ± 0.07	0.26 ± 0.13	0.29 ± 0.15

Ablation study and sweep analysis

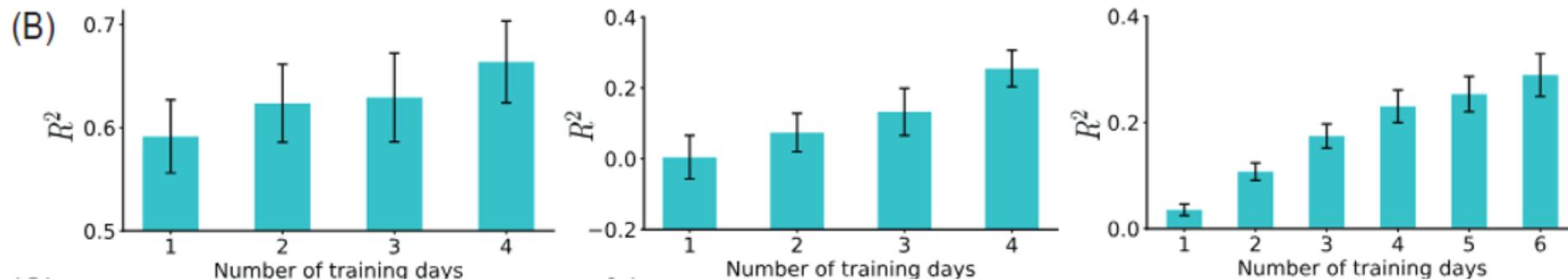
- Held-out sessions

	M1	M2	H1
Without PE	0.50 ± 0.02	-0.09 ± 0.09	0.09 ± 0.06
Absolute PE	0.46 ± 0.01	-0.06 ± 0.09	0.09 ± 0.09
Context-dependent ID	0.66 ± 0.07	0.26 ± 0.13	0.29 ± 0.15

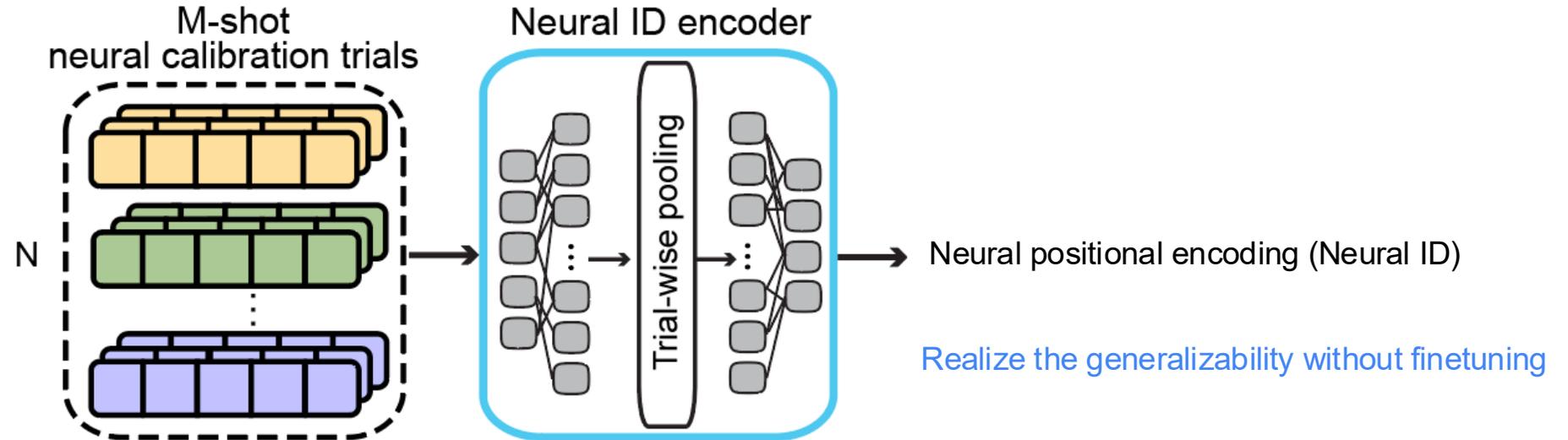
- Investigation of different numbers of calibration trials (A)



- Investigation of data scalability: training days (B)



- The first work towards future plug-and-play robust iBCI decoder.

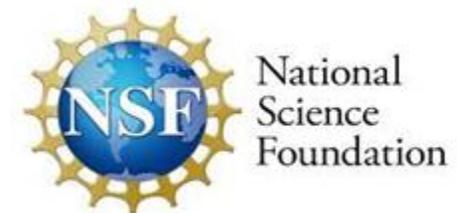


- I prefer to call it **semi zero-shot** to distinguish with prior few-shot model finetuning/retraining methods.
- Our model supports causal and real-time decoding (*in-silico* simulation) but is needed to be evaluated *in vivo*.
- Extend to more complicated decoding task, e.g., communication or speech

W

NeuroAI

Q & A



- Our novel positional embedding consistently improve the results.

1. Held-in sessions

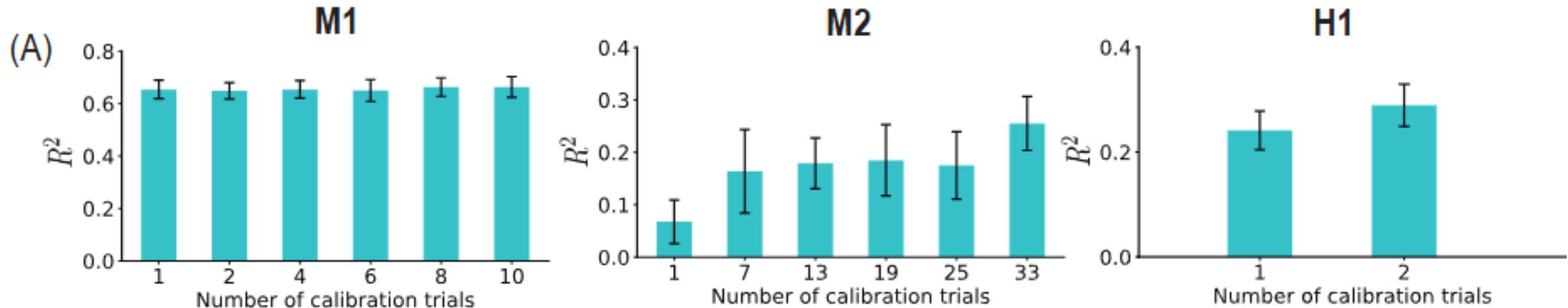
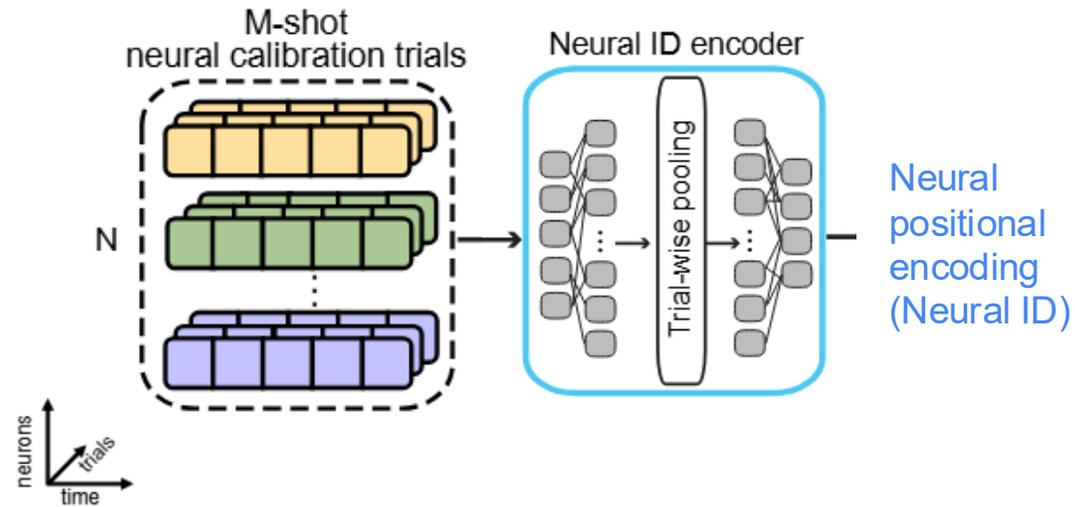
	M1	M2	H1
Without PE	0.59 ± 0.03	0.09 ± 0.04	0.17 ± 0.03
Absolute PE	0.60 ± 0.03	0.13 ± 0.04	0.18 ± 0.05
Context-dependent ID	0.77 ± 0.02	0.59 ± 0.01	0.47 ± 0.06

2. Held-out sessions

	M1	M2	H1
Without PE	0.50 ± 0.02	-0.09 ± 0.09	0.09 ± 0.06
Absolute PE	0.46 ± 0.01	-0.06 ± 0.09	0.09 ± 0.09
Context-dependent ID	0.66 ± 0.07	0.26 ± 0.13	0.29 ± 0.15

More sweeping analysis

- Investigation of different numbers of calibration trials (A)



More sweeping analysis

- Investigation of model scalability: training days (B) and population size (C)

