On Length Bias in EEG-to-Text Decoding

How the length of a sentence inflates brain-decoding scores



Iakovos Tenedios



Yashar Moshfeghi

NeuraSearch Laboratory, University of Strathclyde, Glasgow, UK

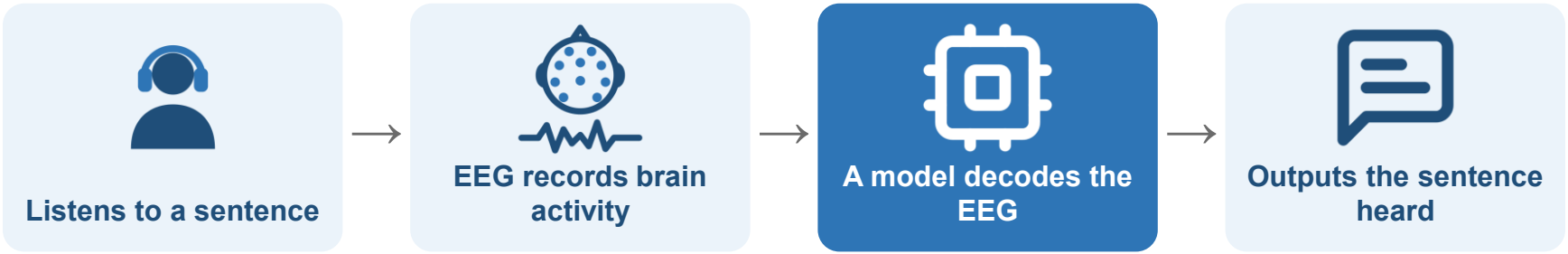
NeurIPS 2026 · Paris Poster Hall, France · December 9, 17-19:00

NEURAL INFORMATION PROCESSING SYSTEMS

NeuraSearch Laboratory

University of Strathclyde Glasgow

EEG-to-text decoding aims to turn brain activity into sentences, but do current scores come from the brain?



What

Decode the sentence a person heard from their EEG alone

Why

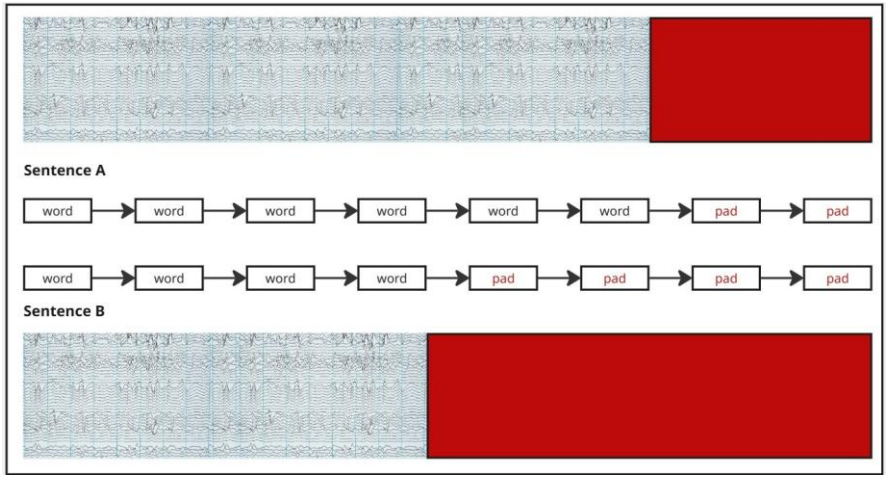
A route to communication for people who cannot speak, such as people with ALS

The problem now

Models report high scores, which however are not completely derived from brain activity

York et al. (2014), J. Neurology; decoder example: BrainECHO, Li et al. (2025), Findings of ACL

The state of the art is inflated by length cues: up to 88% of its score survives with no brain signal



EEG padded with zeros (red) to one fixed length

- Padding varies per sentence, so the amount of padding reveals which sentence was heard
- We replace the real EEG with noise and keep the padding

Our test: length-matched noise

Real EEG | pad → Gaussian noise | pad

25–88%

of the state of the art's score survives with no brain signal, across 3 datasets

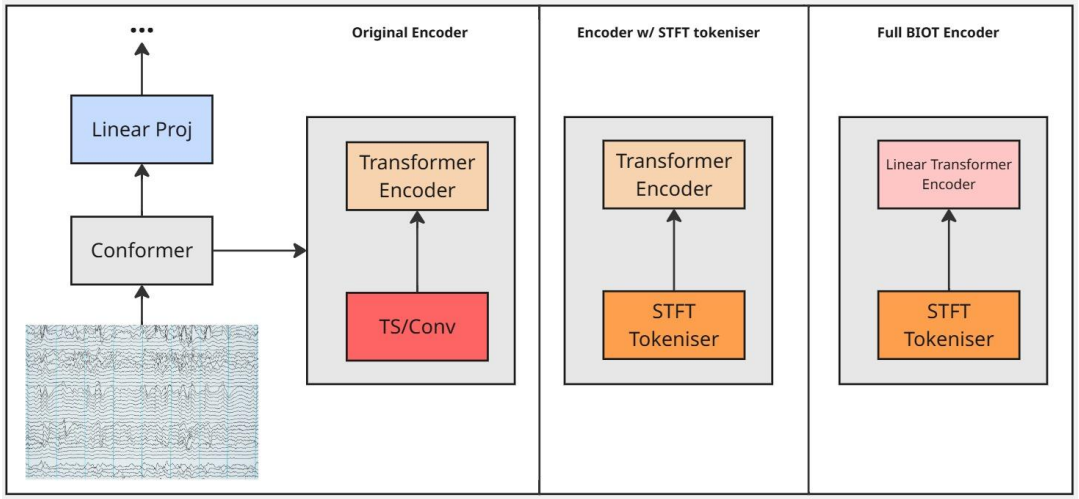
State of the art: BrainECHO, Li et al. (2025). Data: Brennan & Hale (2019); Alexander et al. (2017); NeuraSearch (ours)

NEURAL INFORMATION PROCESSING SYSTEMS

NeuraSearch Laboratory

University of Strathclyde Glasgow

Our fix: an STFT tokeniser in the EEG encoder makes the padding invisible to the model

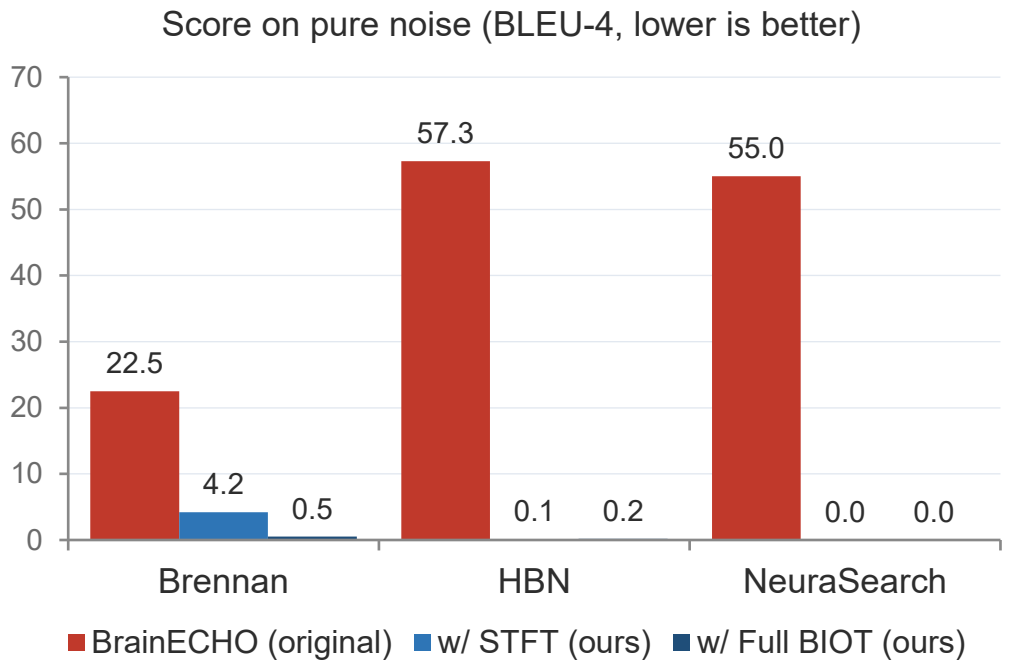


Left: original BrainECHO. **Centre, right:** our two variants

- STFT turns short windows of EEG into their frequency content
- A window of pure padding becomes exactly zero, so padding carries no signal
- We found this by taking apart BIOT, the only classifier robust to length

BrainECHO: Li et al. (2025); BIOT: Yang et al. (2023), NeurIPS

Our variants remove the length shortcut, but models appear to perform sentence retrieval



Shortcut removed

Score on noise falls from up to 57 to about 0

Cost

Lower scores on real EEG (Brennan: 91 → 46)

Retrieval

Classifying a sentence and retrieving its text from its label, nearly matches generation (89 vs 91)

BLEU-4: word overlap with the true sentence (0–100). Full results in the paper.

NEURAL INFORMATION PROCESSING SYSTEMS

NeuraSearch Laboratory

University of Strathclyde Glasgow

Conclusion: length bias is a measurable flaw in EEG-to-text evaluation, and it can be removed

What we showed

1. Up to 88% of state-of-the-art scores come from sentence length, not the brain
2. STFT tokenisation removes the shortcut, and our two variants bring the fix to text generation
3. Current models largely retrieve sentences from a fixed set rather than generate

Future directions

- Locate remaining bias in the rest of the generation pipeline
- Test on sentences never seen in training
- Recover accuracy with better length-invariant encoders

Take-home: if using sentence in EEG-to-text, report how you handle length bias; with a fixed sentence set, also report a retrieval baseline (Cls2Gen)

NEURAL INFORMATION PROCESSING SYSTEMS

NeuraSearch Laboratory

University of Strathclyde Glasgow

Thank you

iakovos.tenedios@strath.ac.uk
yashar.moshfeghi@strath.ac.uk

Acknowledgements: This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/Y009800/1], through funding from Responsible AI UK (KP0011).

Paper

LinkedIn

X (Twitter)

NEURAL INFORMATION PROCESSING SYSTEMS

NeuraSearch Laboratory

University of Strathclyde Glasgow