BrainBERT: Self-supervised representation learning for intracranial recordings

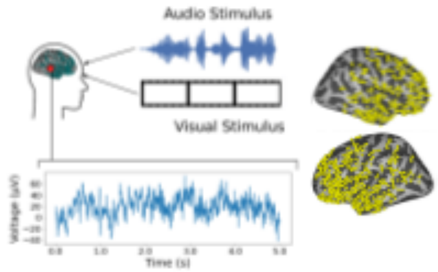
Christopher Wang, Vighnesh Subramaniam, Adam Uri Yaari, Gabriel Kreiman, Boris Katz, Ignacio Cases, Andrei Barbu

Summary: Self-supervised embeddings improve neural decoding across a variety of tasks

Motivation: Linear decoders on neural signal have poor performance, but an established interpretation. Deep NNs have better performance but an unclear interpretation.

We propose BrainBERT: task-agnostic, pre-trained embeddings of neural signal for downstream use in linear decoding

Data: 10 subjects, 21 movies, 1,249 sEEG electrodes (yellow), 4,551 electrode-hours

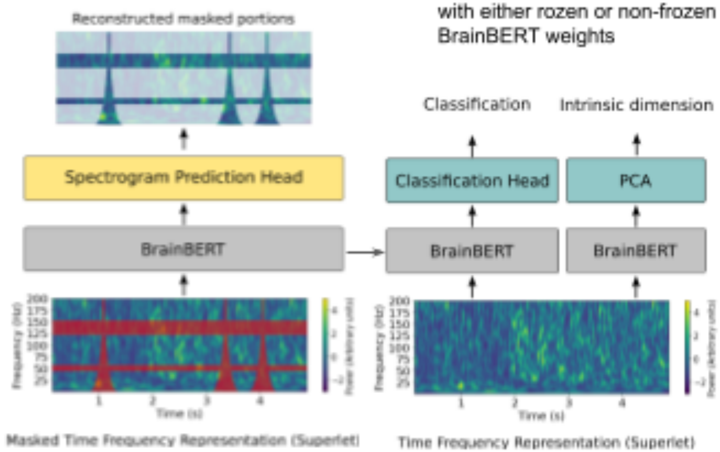


Voltage at one electrode

Model inputs: Time frequency representation of neural signal, which is either:
- Short time Fourier transform (STFT)
- Superlets (a composite of Morlet wavelet transforms)

Pretraining → Fine-tuning

Fine-tuning can either occur with either rozen or non-frozen BrainBERT weights



For superlets, we use an adaptive mask to reflect the variable trade-off between time and frequency resolution

Pretraing loss: L1 Reconstruction loss + "content aware" term

$$\mathcal{L} = \sum_{y_{input}} |y_{input} - y_{pred}| + \alpha \sum_{y_{input}>1} |y_{input} - y_{pred}|$$

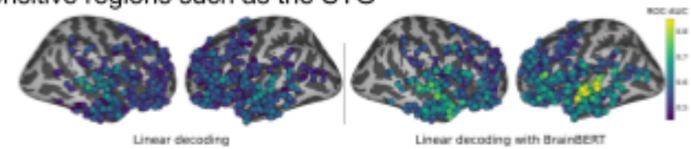
Classification tasks: Sentence onset vs. baseline, Word onset vs. baseline, Volume (high vs. low), Pitch (high vs. low)

Fine-tuning results

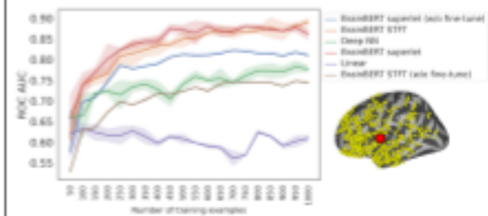
	Sentence onset	Speech/Non-speech	Pitch	Volume	Task Avg.
Linear (.25s, time domain)	.54 ± .04	.52 ± .03	.48 ± .09	.54 ± .09	.52 ± .07
Linear (5s, time domain)	.63 ± .04	.58 ± .06	.58 ± .07	.56 ± .19	.59 ± .11
Linear (.25s, STFT)	.60 ± .04	.53 ± .04	.51 ± .06	.52 ± .06	.54 ± .06
Linear (.25s, superlet)	.59 ± .03	.53 ± .03	.52 ± .06	.53 ± .08	.54 ± .06
Deep NN (5s, 5 FF layers)	.72 ± .10	.67 ± .08	.57 ± .06	.54 ± .11	.63 ± .12
BrainBERT (STFT)	**.82 ± .07**	**.93 ± .03**	**.75 ± .03**	.83 ± .09	**.83 ± .09**
BrainBERT (superlet)	.78 ± .08	.86 ± .06	.62 ± .05	.70 ± .10	.74 ± .12

Results continued

BrainBERT improves linear decoding, particularly in language sensitive regions such as the STG

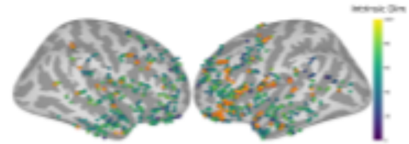


Results shown for the sentence decoding task and superlet inputs



Data efficiency: With fine-tuning, BrainBERT only needs 1/5th as many examples to outperform the deep NN baseline

Intrinsic dimension: For each electrode, we can ask what is the dimension of the space on which the embeddings lie?



Highest ID regions: supramarginal gyrus, lateral obitofrontal cortex, amygdala

Future work:
How does the intrinsic dimensionality vary across time?
How should we combine embeddings across locations?
Can we use 24 hour recordings to uncover the mechanisms underlying sleep in a data-driven way?

Model weights and source code at https://github.com/czlwang/BrainBERT/