Exploring Abstract Pattern Representation in the Brain and Non-symbolic Neural Networks

Enes Avcu[1], Skyla Lynch[1], and David Gow[1,2]
eavcu@mgh.harvard.edu

Massachusetts General Hospital/Harvard Medical School[1], Salem State University[2]

Overview

- Human infants easily discover and generalize identity relationships in temporal sequences that are challenging for variable-free recurrent network models.
- This gap has been interpreted as evidence that human performance depends on operations over variables that models cannot capture.
- We investigated the role of task and pretraining in network model performance and relate those models to human MR-constrained simultaneous MEG/EEG neural data.
- The results suggest that pretraining on syllables improves network performance, and that pretrained machine representations are highly correlated with human neural data.

Background

- Generativity, the capacity to create and comprehend novel forms, is a defining feature of both language and human cognition. One perspective in the literature is that achieving proficiency in generativity implies the ability to contemplate abstract relationships between stimuli (Marcus, 2001: Berent 2013).

Behavior, Models and Brains

- Marcus et al. (1999) demonstrated that infants, but not simple recurrent models (SRNs, Elman, 1990) easily learn to discriminate novel syllable sequences based on abstract syllable repetition patterns.

- A series of modelling studies examined the role of pretraining on recurrent model acquisition of repetition-based rules:
  - Seidenberg and Elman (1999a,b) extensively pre-trained an SRN with syllables, enabling the SRN to recognize identity relationship between them.
  - Altmann's (2002) pretrained a model with 10,000 sentences wherein the model predicts the subsequent word without considering syllables or phonemes.

- Prior knowledge is a valid assumption! In fact, in accordance with the findings of Hart and Risley (2003), it is estimated that children are exposed to approximately 4 million words by the age of 7 months.

- Gow et al. (2022) demonstrated that ROI-level localized M/EEG data can be decoded using SVMs to discriminate abstract repetition patterns consistent with either abstract variables or token-level abstract representations.

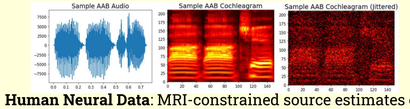
Questions

- Is pretraining a necessary precondition for model learning and perhaps a reasonable assumption for infants?
- Do variable-free network models discover the same kinds of representations that brains discover to produce generalization of abstract syllable repetition patterns?

Acknowledgements: This work was supported by National Institute on Deafness and Other Communication Disorders (NIDCD) grant R01DC015455 (P.I. D.G). We would like to thank Kevin Brown for helpful comments and John Rho for helping us run the early versions of the models.

Methods

- **Stimuli**: 23 unique CV syllables were recorded (250 ms) and concatenated to form 720 training (with 10 different jittered tokens) and 126 test unique trisyllabic nonce words consistent with 3 repetition patterns: AAB (e.g., *babapo*), ABB (e.g., *bapopo*), and ABA (e.g., *bapoba*).



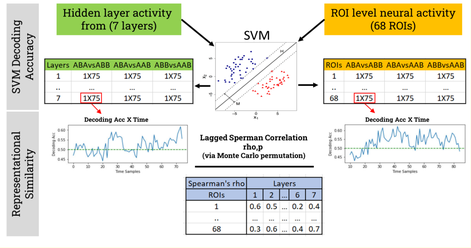
- **Human Neural Data**: MRI-constrained source estimates of combined MEG/EEG data collected from 14 adults performing an auditory artificial grammar task (from Gow et al., 2022).

- **Models:** Models trained on the same stimuli with data augmentation (10 jittered versions of each token).
  - **Pretraining**: One layer LSTM with 512 nodes trained to identify 23 syllables using one hot vectors.
  - **Training Tasks**: Four separate 7-layer LSTM models (two including a pretrained layer) with dropout (0.85) trained independently on the same training data with different labels (output vectors):
    - **Pattern learners** w/ and w/o pretraining trained on random vectors representing three patterns
    - **Word learners** w/ and w/o pretraining trained on vectors representing each word based on syllable position

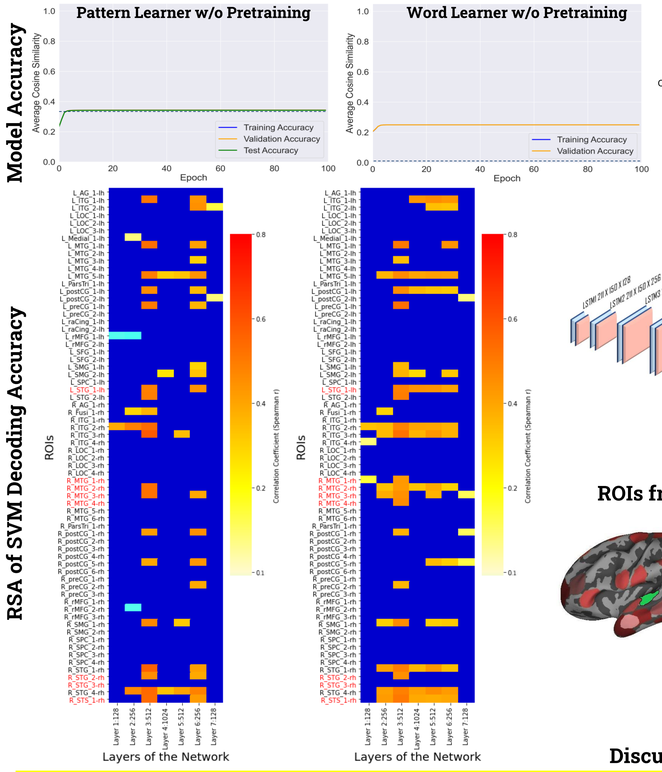
- **Testing**: Cosine Similarity between target observed patterns was calculated for trained tokens (training acc.), reserved tokens of trained syllable patterns (validation acc.) and novel tokens (test acc.).

Decoding and RSA

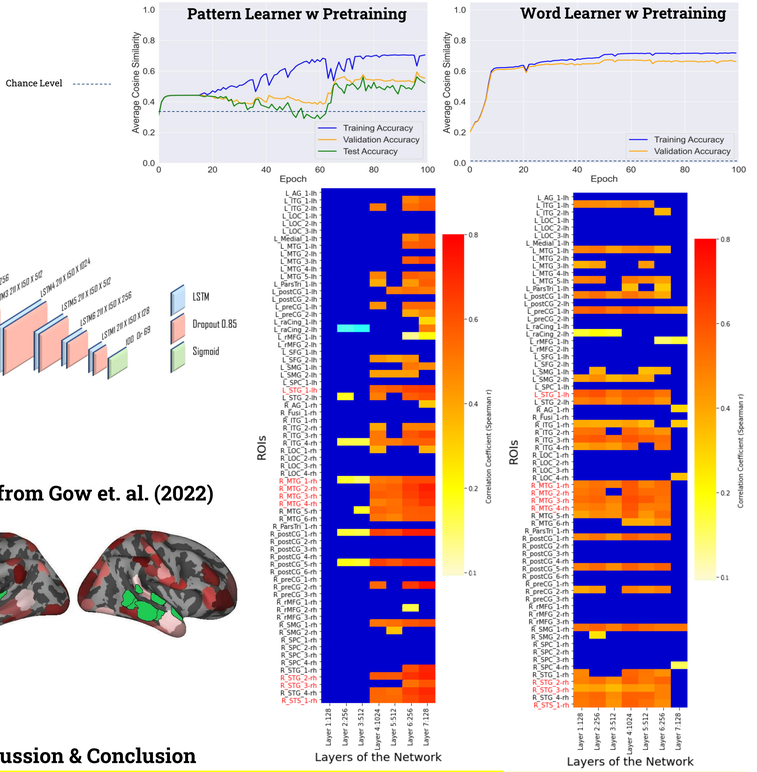
- RSA is looking at the correlation between decoding accuracy based on SVMs applied to ROI activation vectors in the brain (8 MNE measures per ROI per timepoint) and SVMs applied to activation vectors defined over each of the 7 layers in the model.



W/O PRETRAINING



W PRETRAINING



ROIs from Gow et. al. (2022)

Discussion & Conclusion

- We explored how task specificity and pretraining impact the performance of network models, drawing connections between these models and human neural data obtained through MR-constrained simultaneous MEG/EEG.
  - Simulation results showed that both word learner and pattern learner models benefited from pretraining which suggests that the seven-month-old infants tested by Marcus et al. might have benefited from prior exposure to speech sounds as well.
  - RSA comparing patterns of regional brain activity to patterns of hidden node activation showed significant correlations between the decoder ROIs from Gow et al. (2022) and pretrained model representations. Strong correlations were observed between the early layers of the word model and the decoder ROIs, in contrast to the late layers in the pattern model. These results suggest that adult performance benefits representations that can develop from either pattern learning or word learning.
- To sum up, our results suggest that associative mechanisms operating over discoverable representations capturing abstract stimulus properties account for a critical example of human cognitive generativity highlighting the crucial significance of generative AI models in simulating and understanding cognitive generativity within the realms of human learning and representation.