Computational pressures behind the development of parallel dorsal and ventral stream lexica

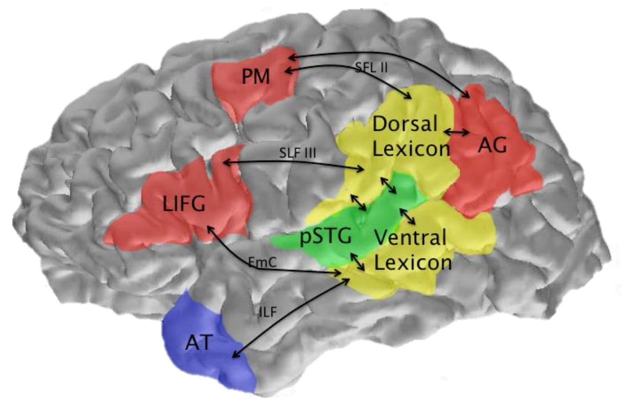
Enes Avcu[1], Michael Hwang[2], Kevin Brown[3], and David Gow[1,4]
eavcu@mgh.harvard.edu
Massachusetts General Hospital/Harvard Medical School[1], Harvard College[2], Oregon State University[3], Salem State University[4]

Overview

- Words play a pivotal role in almost every aspect of language processing.
- There is evidence for the existence of parallel lexica serving the dorsal and ventral processing streams.
- We use ANNs to examine what role computational pressures play in the emergence of this parallel architecture.

A Dual Lexicon System

- Gow's (2012) dual lexicon model synthesizes evidence from aphasia, behavioral and neural results to identify two wordform areas that mediate the mapping between acoustic-phonetic input and processing in the dorsal and ventral speech streams identified by Hickok and Poeppel (2007).
  - The **dorsal lexicon**, located in the supramarginal gyrus (SMG), mediates the mapping between acoustic-phonetics and articulation.
  - The **ventral lexicon**, located in the posterior middle temporal gyrus (pMTG), mediates the mapping between acoustic-phonetics and semantic/syntactic processing.



A Computational Hypothesis for the Division

- Distributed feature-based lexical representations in these areas act as hidden nodes to facilitate mappings.
- **Hypothesis**: This separation arose in part because of fundamental differences in the computational requirements of these mappings.
  - The **mapping between sound and articulation**, though complex, is largely systematic and temporally continuous.
  - The **mapping between sound and syntactic/semantic information**, though partially systematic at the level of productive morphology, is largely arbitrary and dependent on identifying larger temporal units.

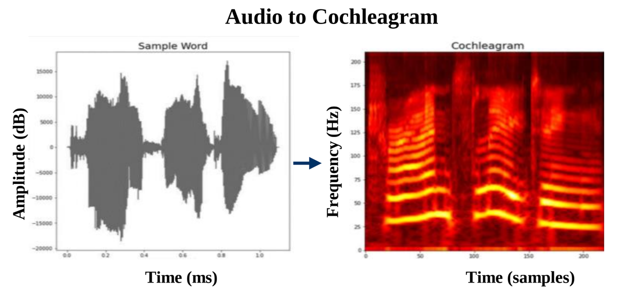
Acknowledgements: This work was supported by National Institute on Deafness and Other Communication Disorders (NIDCD) grant R01DC015455 (P.I. D.G) and NSF BCS-PAC 2043903 (P.I. K.B). We would like to thank Jim Magnuson for helpful comments, Alison Xin for helping us create the lexicon, and Seppo Ahlfors and Skyla Lynch for thoughtful comments and feedback on the work.

Predictions

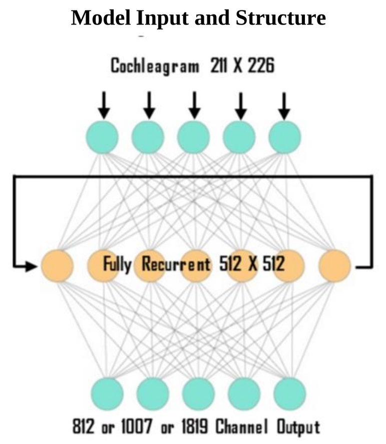
- ANNs trained on either dorsal or ventral mappings should produce features that support individual word identification because **both representations discover unique features**.
  - Features from ANNs trained on dorsal mappings should have an **advantage for phonological categorization** but not semantic/syntactic categorization.
  - Features from ANNs trained on ventral mappings should have an **advantage for semantic/syntactic categorization** but not phonological categorization.
  - Features from ANNs trained on both dorsal and ventral mappings **should NOT have any advantage for semantic/syntactic categorization or phonological categorization.**

Methods

- **Training Data**: 883 (631 multi-morphemic) English word audio files with 10 different speakers, making a total of 8830 total training items. We used cochleagrams of each sound file as the input to the network (Feather et. al., 2019; Kell et al., 2018).

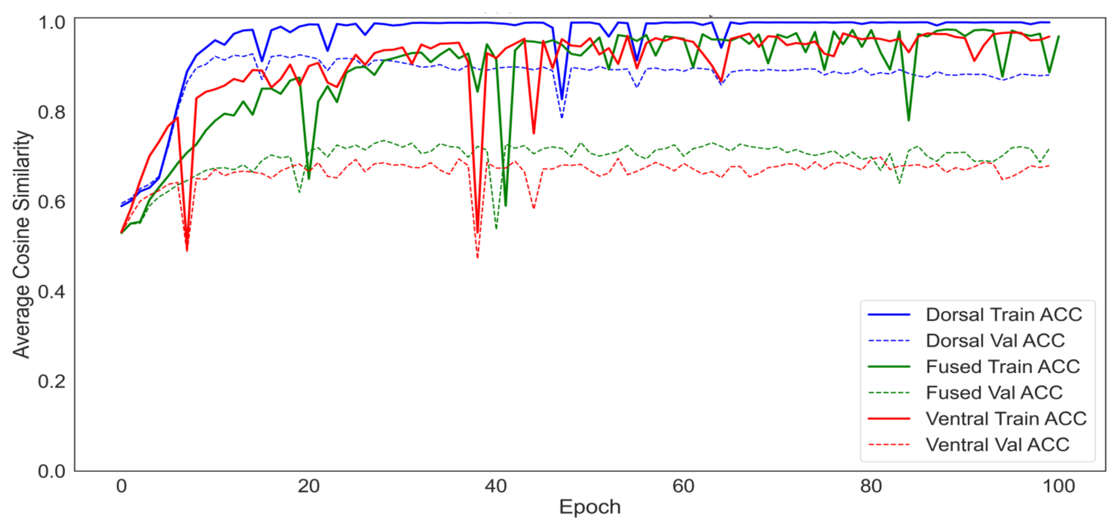


- **Training Tasks:** Three separate LSTM models trained independently on the same training data with different labels (output vectors):
  - A dorsal model trained on vectors representing articulatory features,
  - A ventral model trained on vectors derived from lexical co-occurrence statistics,
  - A fused model trained on the combined vectors of both the dorsal and ventral models,

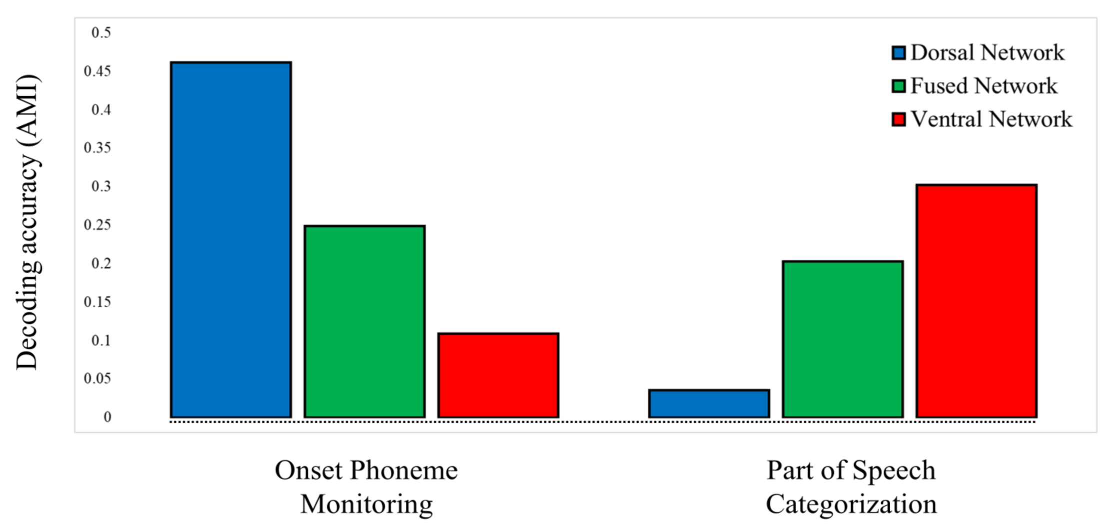


- **Testing:** Cosine similarity as the accuracy metric of training.
  - **Generalization Tasks**: Extracted hidden layer activation patterns for 8830 words from all 3 models and categorized into 7 (articulatory task), and 9 (semantic/syntactic task) classes, respectively. We used unsupervised learning (Hierarchical Clustering) to quantify decoding accuracy of activations of each network.
  - **Hidden Unit Selectivity Analyses:** Created two Selectivity Indices (SIs) to measure the degree to which hidden units of all 3 networks encode information related to phonemes (PSI) and morphemes (MSI).
  - **Error Analysis:** Examined the kinds of errors models make when asked to identify individual words to see whether they pattern with dissociations in aphasia (dorsal: phonological errors as in reproduction conduction aphasia, and ventral: semantic errors as in transcortical sensory aphasia)

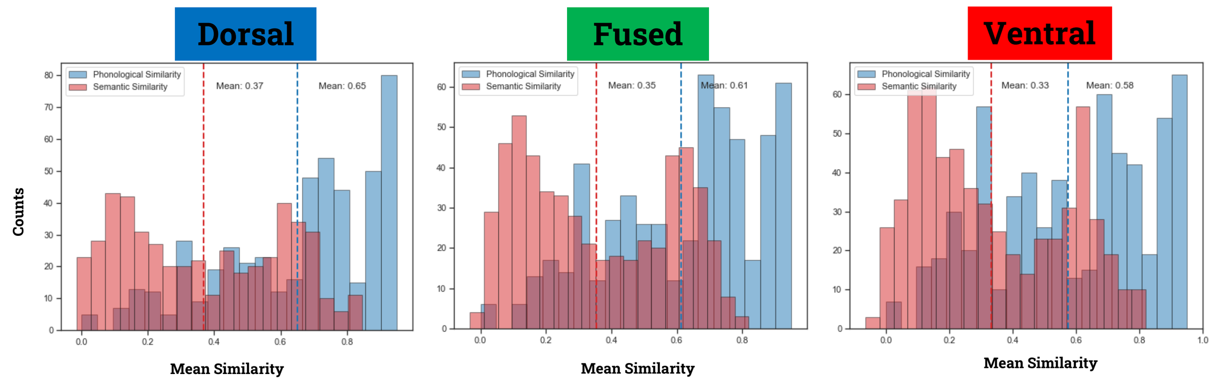
Training and Validation Accuracy



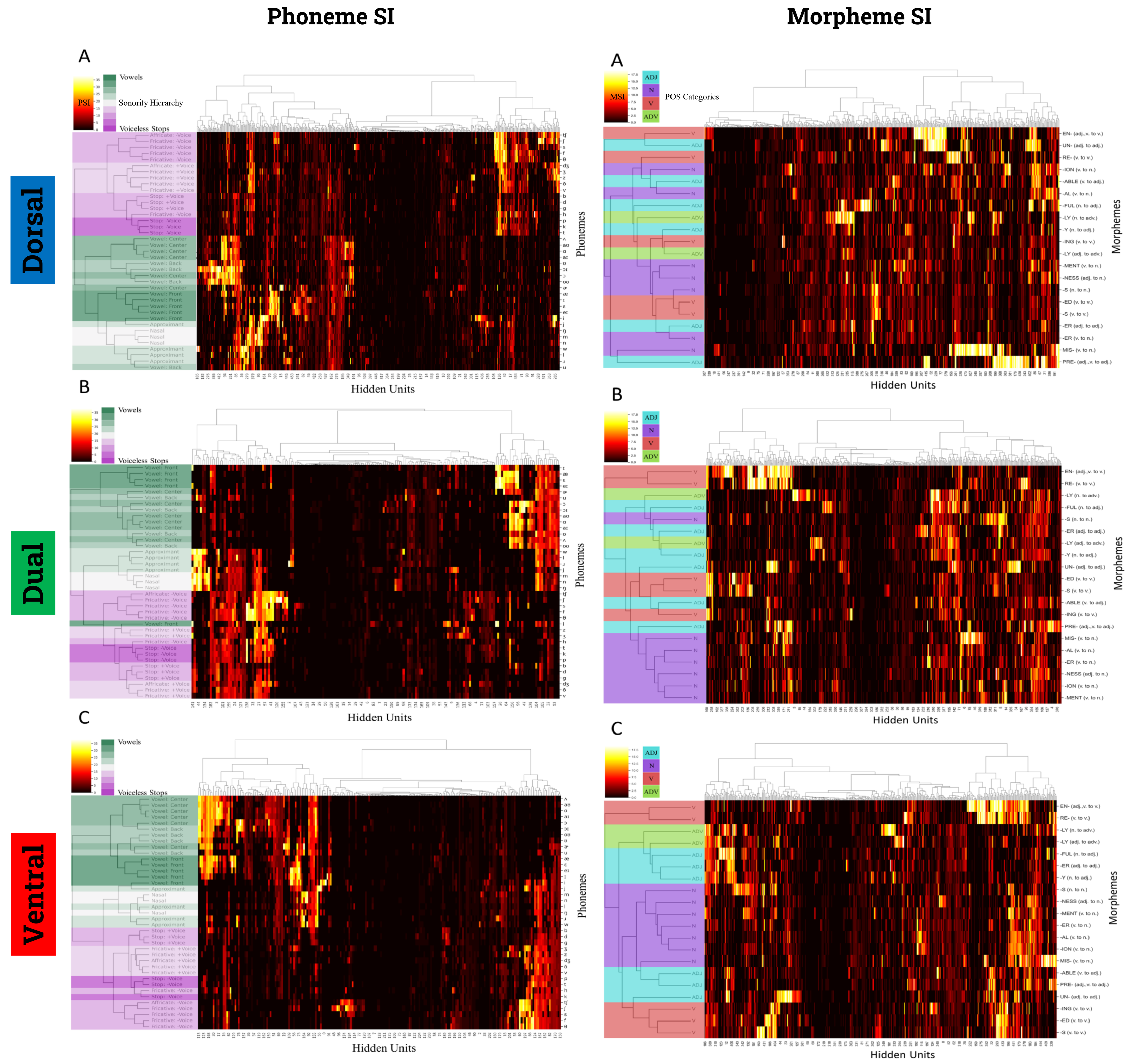
Decoding Results



Error Analysis



Hidden Unit Selectivity



Discussion & Conclusion

- Training the networks on differently structured wordform representations produced different featural representations and these emergent representations supported different patterns of performance on secondary tasks
- The results on the generalization tasks and the hidden unit SIs show that
  - (i) the dorsal features were more successful on an articulatory task,
  - (ii) the ventral features were more successful on a semantic/syntactic task,
  - (iii) the fused features were less successful on task optimization,
- Representations optimized for one task would not be transferable to the other task
- Development of parallel lexica in the dorsal and ventral pathways arose from computational pressures for optimizing the primary mapping functions that support lexically organized processes in the dorsal and ventral processing streams

References

Gow, D. W. (2012). The cortical organization of lexical knowledge: A dual lexicon model of spoken language processing. Brain and language, 121(3), 273-288.
Hickok, G., Poeppel, D. (2007). The cortical organization of speech processing. Nat Rev Neurosci 8, 393–402.
Feather, J., Durango, A., Gonzalez, R., & McDermott, J. (2019). Metamers of neural networks reveal divergence from human perceptual systems. Advances in Neural Information Processing Systems, 32.
Kell, A. J., Yamins, D. L., Shook, E. N., Norman-Haignere, S. V., & McDermott, J. H. (2018). A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. Neuron, 98(3), 630-644. doi: 10.1016/j.neuron.2018.03.044.