

COMPUTATIONAL SIMILARITIES BETWEEN VISUAL AND AUDITORY CORTEX STUDIED WITH CONVOLUTIONAL NEURAL NETWORKS & FMRI



Alexander Kell*, Daniel Yamins*, Sam Norman-Haignere, Darren Seibert, Ha Hong, James DiCarlo, Josh McDermott

alexkell@mit.edu
AK & DY contributed equally

MOTIVATION

Visual and auditory cortex underlie impressively robust recognition
But operate on distinct classes of signals

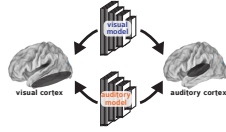
TO WHAT EXTENT ARE SIMILAR COMPUTATIONS USED ACROSS SENSORY MODALITIES?

Recent work: Computational models **optimized to perform visual tasks**^{1,2}
--> Predict visual cortical responses well
--> Recapitulate known properties of functional organization

Perhaps we can use these kinds of models to examine potential similarities across modalities

HERE:

1. Build models to perform visual and auditory tasks
2. Predict AUDITORY cortical responses from AUDITORY model
3. Predict AUDITORY cortical responses from VISUAL model



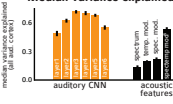
Future: Predict visual cortical responses

PREDICTING AUDITORY CORTICAL RESPONSES WITH AUDITORY MODEL

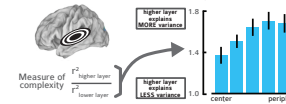
Auditory CNN:

1. Predicts auditory cortical responses better than alternatives
2. Clarifies known properties of auditory cortical functional organization

Across all auditory cortex: Median variance explained

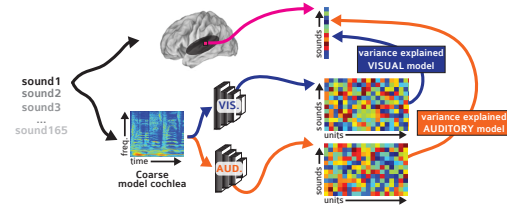


Higher layers better predict more peripheral auditory cortex

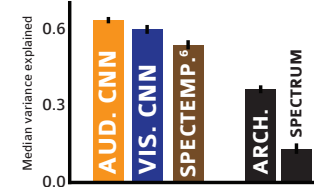


PREDICTING AUDITORY CORTICAL RESPONSES WITH VISUAL MODEL

Predict each voxel using cross-validated regularized linear regression



Median variance explained across all of auditory cortex



BUILD MODELS, COLLECT DATA

Model class: convolutional neural networks (CNNs)^{1,4,5}



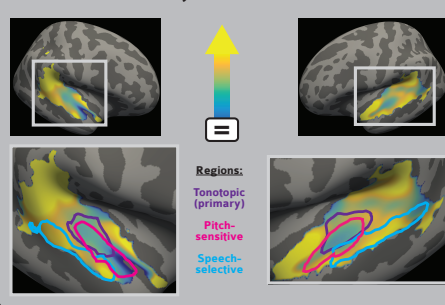
Visual task: Invariant object recognition
Trained on ImageNet?
Similar to Yamins, Hong, et al¹

Auditory task: Invariant word recognition
excitatory speech + heavy, complex backgrounds (e.g., music, speech babble, auditory scenes)

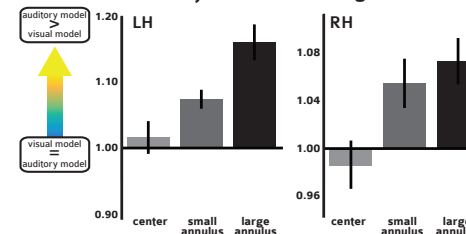
Auditory fMRI data: Responses to 165 natural sounds

- person screaming
- car windows
- airplane taking off
- vacuo
- printer printing
- rap song
- whistling
- cat purring
- shuffling cards
- Italian speech
- walking on leaves
- door knocking

Each voxel: Ratio variance explained (auditory cnn/visual cnn)



Ratio variance explained: Auditory cortex's radial organization



TAKEAWAYS

Novel method for **quantitatively** measuring computational similarities across **distinct modalities** & **neural populations**.

Model of visual cortex captures non-trivial aspects of responses in **auditory cortex**.

Equally predictive in central auditory cortex...
... less predictive peripherally.

Consistent with possibility:

Across modalities: similar computations in early sensory cortex...
...but subsequent processing more modality-specific.

FUTURE DIRECTIONS

Test whether holds in other sensory modality.
-- predict visual cortical fMRI responses with auditory model

Predict electrophysiology data.
-- voxel responses: coarse -- worry: fMRI "easy" to predict?

Begin to understand computational similarities.
-- examine where CNN representations diverge
-- synthesize images/sounds from different layers^{7,8}

THANKS



DOE Computational Science Graduate Fellowship to AK
NSF Graduate Research Fellowship to SNH
McDonnell Scholar Award to JHM
VSS Travel Award to AK
NEC Award to JHM

fMRI data: Nancy Kanwisher
fMRI protocol help: Atsushi Takahashi

REFS

- [1] Yamins, Hong, 8 co., PNAS, 2014; [2] Seibert & co. (submitted); [3] Hinton & co., 2012 IEEE Signal Processing Magazine; [4] Krizhevsky & co., Neural Information Processing Systems (NIPS), 2012; [5] Ding & co., International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2013; [6] Chi, Ru, & Shamma, Journal of Acoustical Society of America (JASA), 2005; [7] Mahendran & Veladi, arXiv 2014; [8] Zeller & Fergus, ECCV, 2014