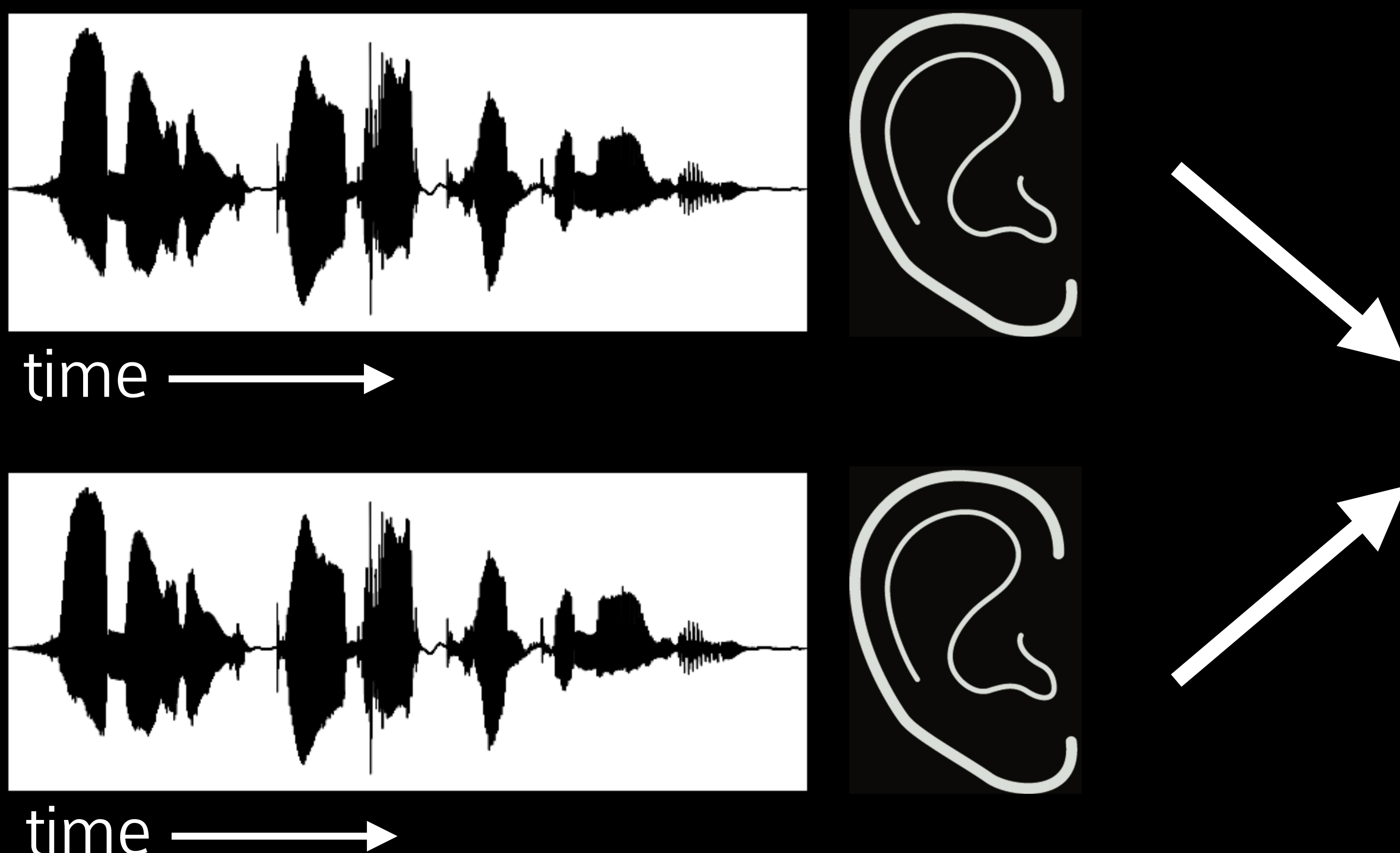
Invariant & hierarchical computation in human auditory cortex

Alex Kell

2018.07.17 :: CSGF Program Review

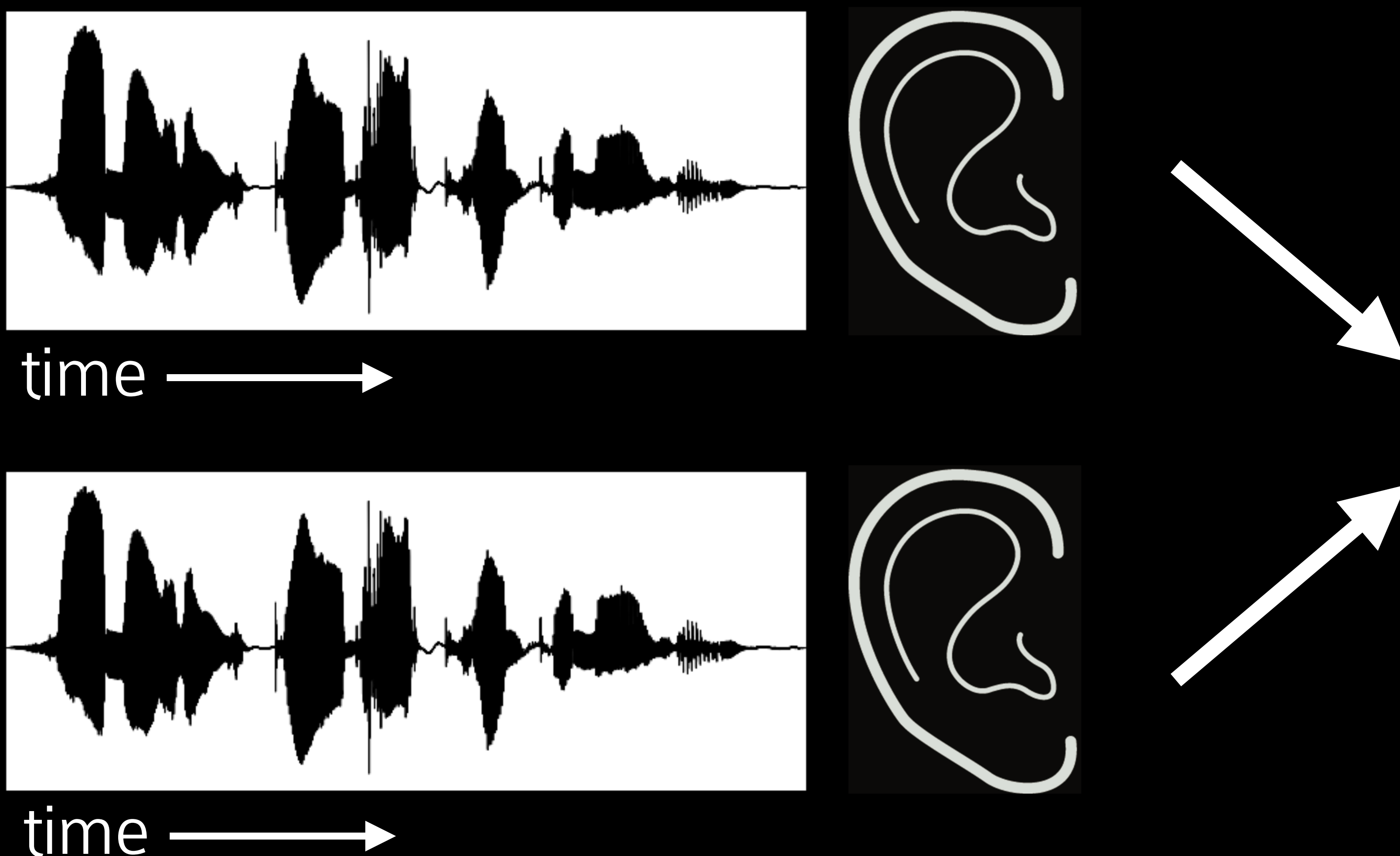

time →


time →



What was said?
Who said it?
How did they feel when they said it?

What caused the sound?
Where?

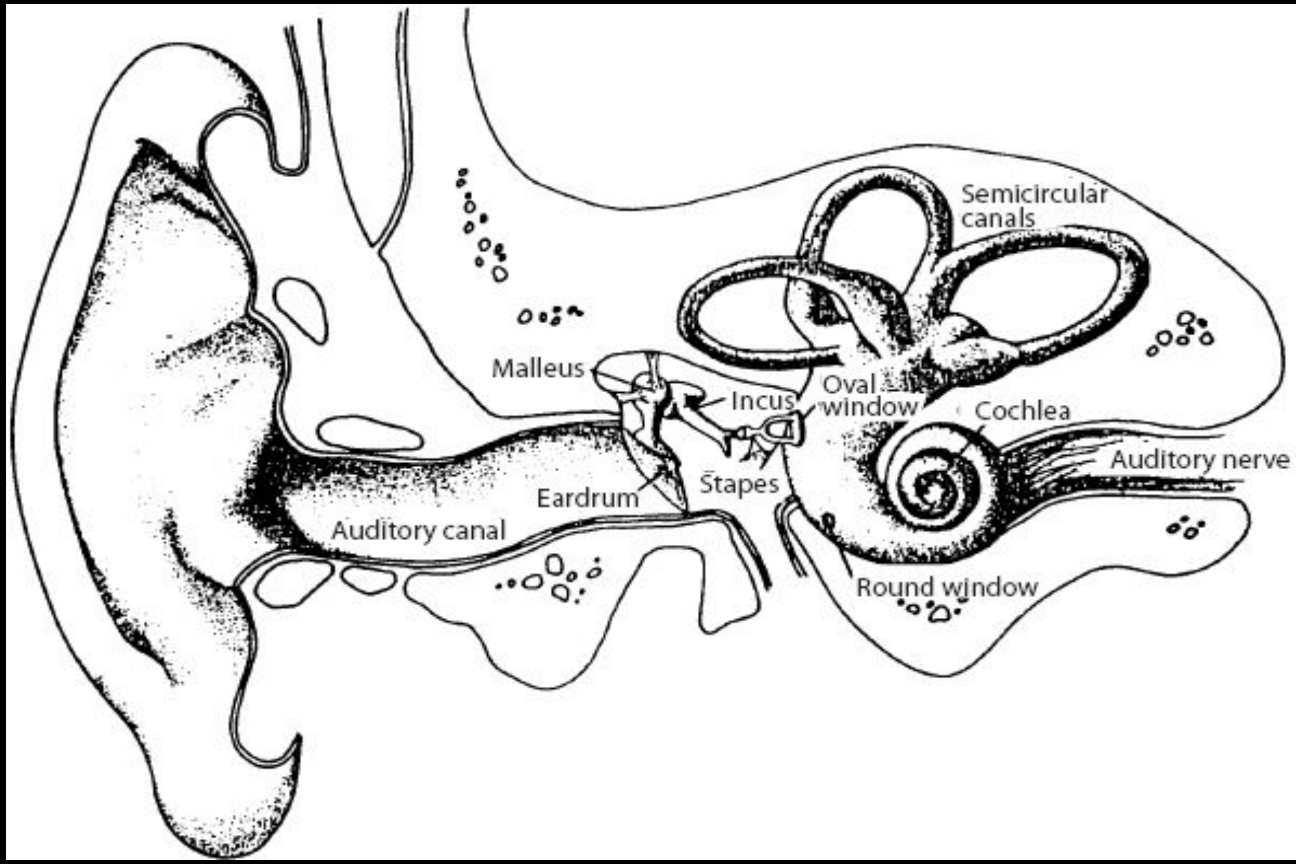


What was said?
Who said it?
How did they feel when they said it?

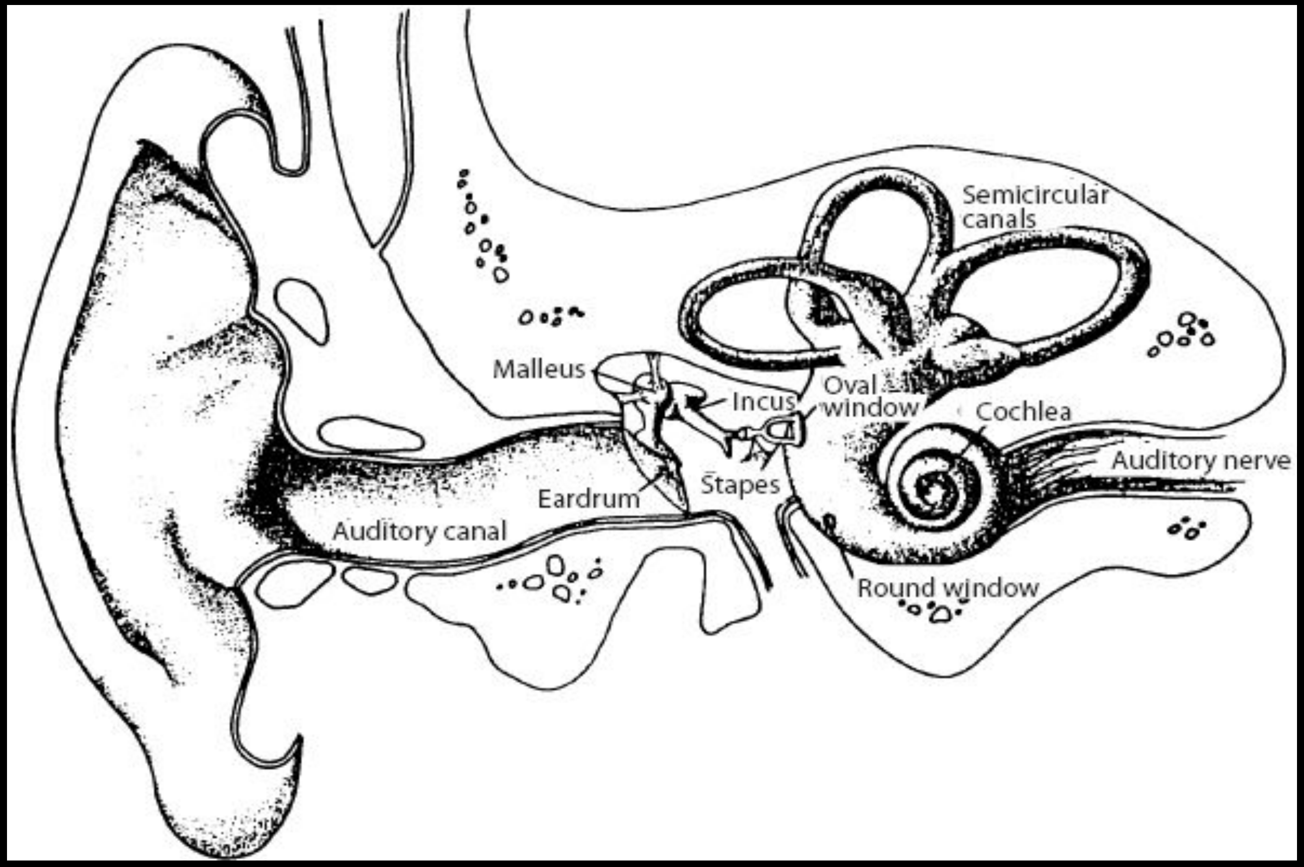
What caused the sound?
Where?

How does the brain extract behaviorally relevant information from these waveforms?

Peripheral auditory system: well characterized...



Peripheral auditory system: well characterized...



... but auditory cortex is poorly understood. (Particularly in humans.)

TODAY:

Basic questions about functional organization of human auditory cortex.

TODAY:

Is there a hierarchical organization?

TODAY:

Is there a hierarchical organization?

If so, how many stages?

TODAY:

Is there a hierarchical organization?

If so, how many stages?

What do different stages do?

TODAY:

Is there a hierarchical organization?

If so, how many stages?

What do different stages do?

Use modeling to generate specific hypotheses in a principled manner

Neuron

Article

A Task-Optimized Neural Network Replicates Human Auditory Behavior, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy

Highlights

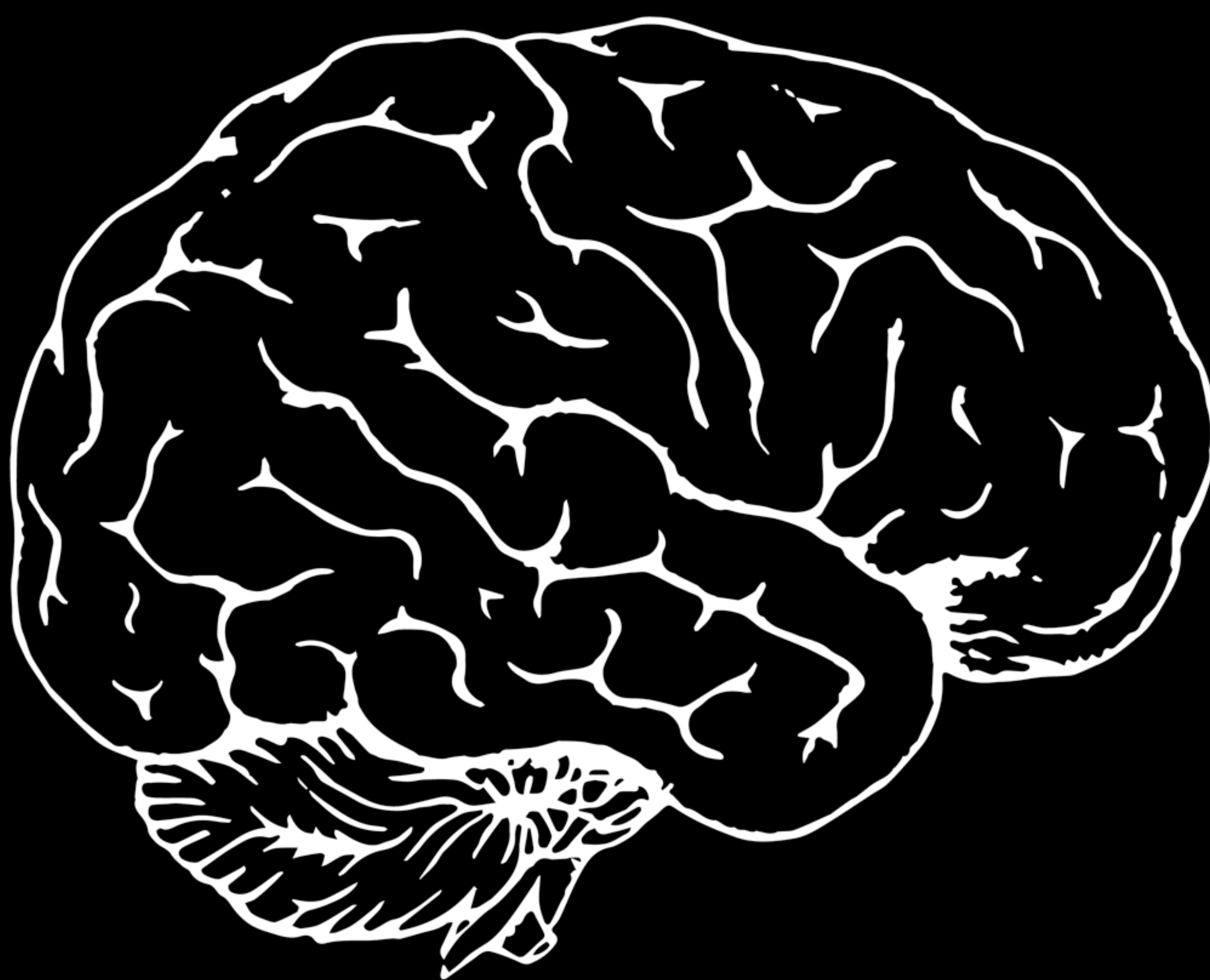
- A deep neural network optimized for speech and music tasks performed as well as human listeners
- The optimization produced separate music and speech pathways after a shared front end

Authors

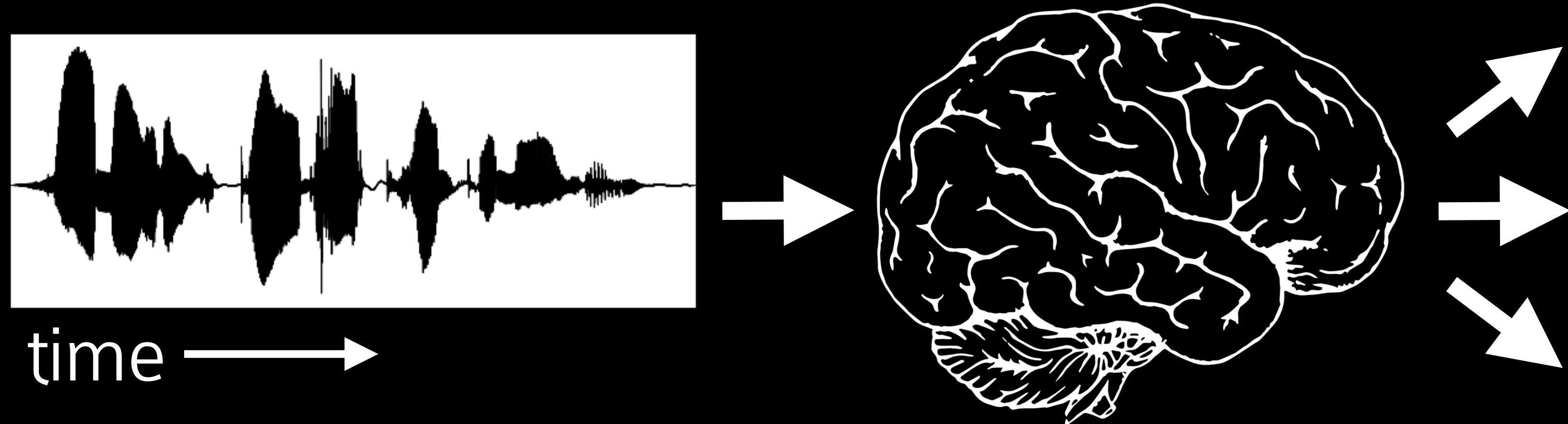
Alexander J.E. Kell, Daniel L.K. Yamins,
Erica N. Shook,
Sam V. Norman-Haignere,
Josh H. McDermott

Work with:
Dan Yamins, Erica Shook, Sam Norman-Haignere,
and Josh McDermott

How to build better models of auditory cortex?



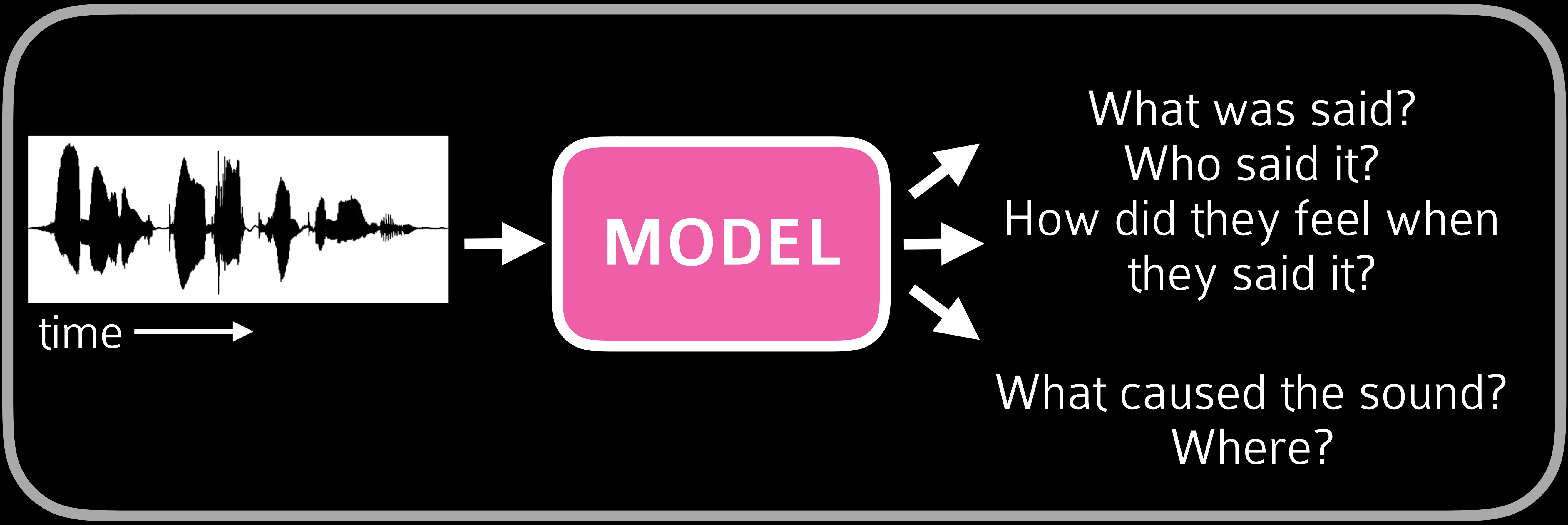
How to build better models of auditory cortex?



What was said?
Who said it?
How did they feel when they said it?

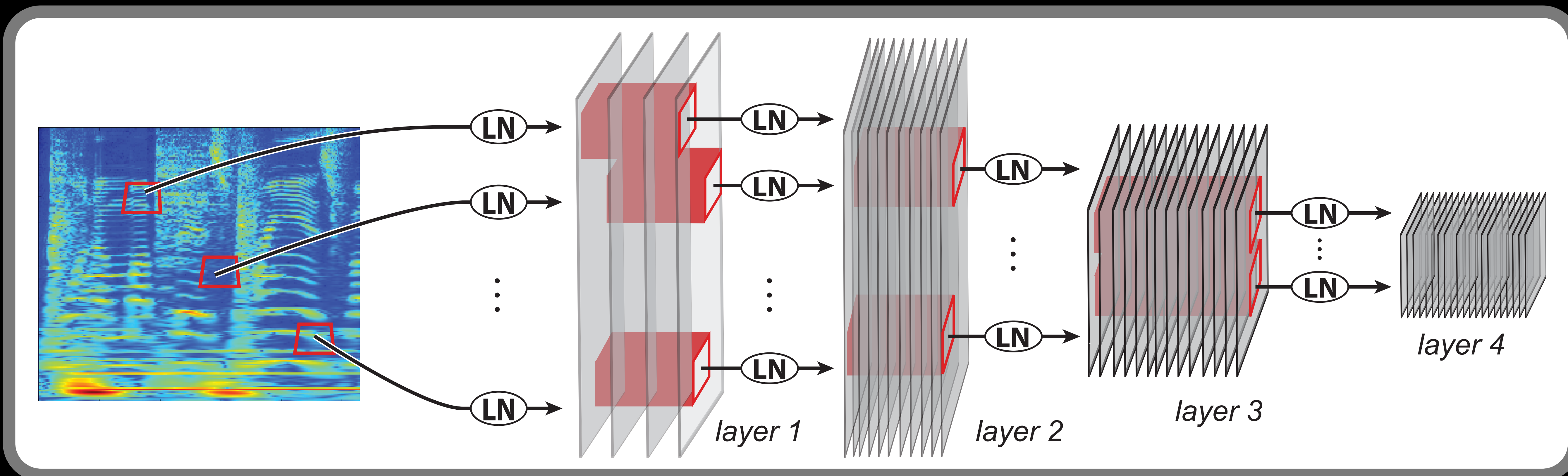
What caused the sound?
Where?

How to build better models of auditory cortex?



Recent machine learning advances: Deep learning

Recent machine learning advances: Deep learning



Hierarchical convolutional neural networks (CNNs)

(Fukushima, 1980; Lecun et al., 1989; Krizhevsky et al., 2012; Yamins, Hong, et al., 2014; etc.)

KEY HYPOTHESIS:

A model optimized to perform real-world auditory tasks may converge to brain-like computations

KEY HYPOTHESIS:

A model optimized to perform real-world auditory tasks may converge to brain-like computations

Approach pioneered in the visual cortex

(Yamins, Hong, et al. 2014; Cadieu et al. 2014; Hong, Yamins, et al. 2016)

<u>KEY HYPOTHESIS:</u>

A model optimized to perform real-world auditory tasks may converge to brain-like computations

Approach pioneered in the visual cortex

(Yamins, Hong, et al. 2014; Cadieu et al. 2014; Hong, Yamins, et al. 2016)

Potentially:
Particularly useful in auditory cortex

Unsatisfying aspects of deep learning as a neuroscience model

Unsatisfying aspects of deep learning as a neuroscience model

- Unrealistic amount of (supervised) training data.
- Unrealistic learning rule (backprop).
- Discriminative models (rather than generative).
- etc.

Unsatisfying aspects of deep learning as a neuroscience model

- Unrealistic amount of (supervised) training data.

- Unrealistic learning rule (backprop).

- Discriminative models (rather than generative).

- etc.

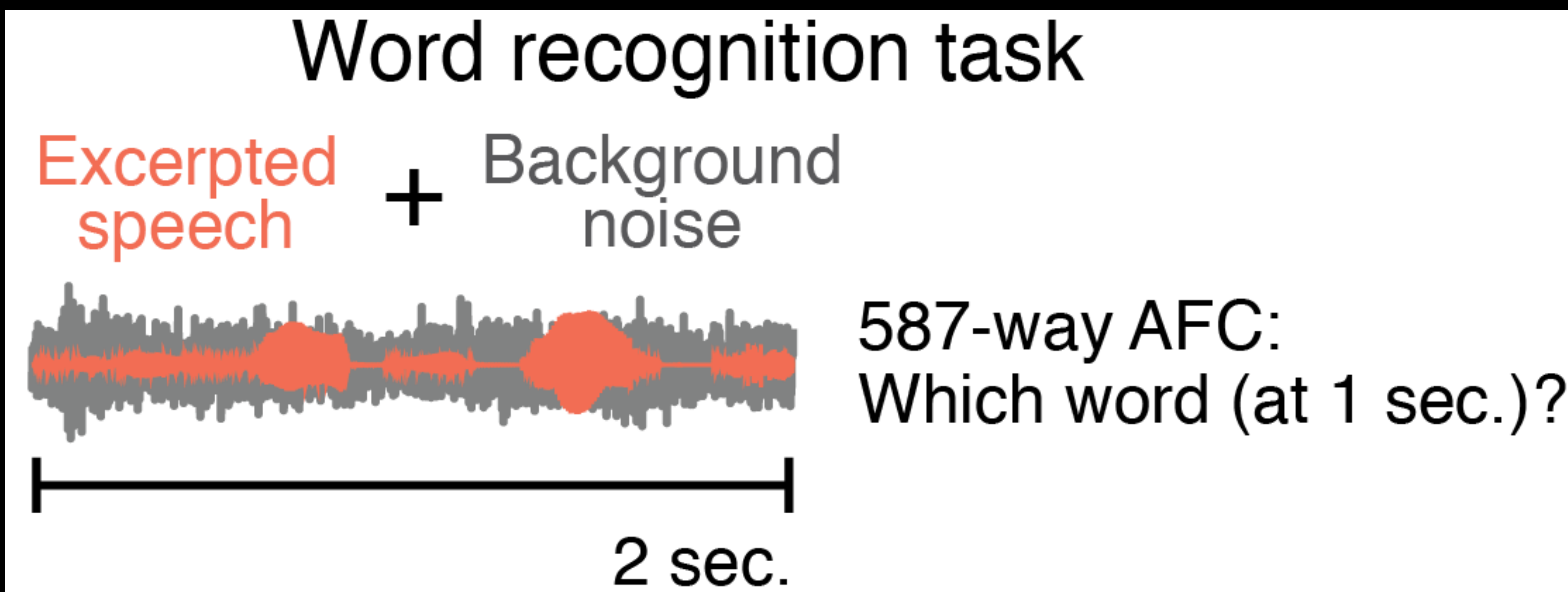
Network optimization: Real-world tasks

Network optimization: Real-world tasks

... that have large labelled datasets.

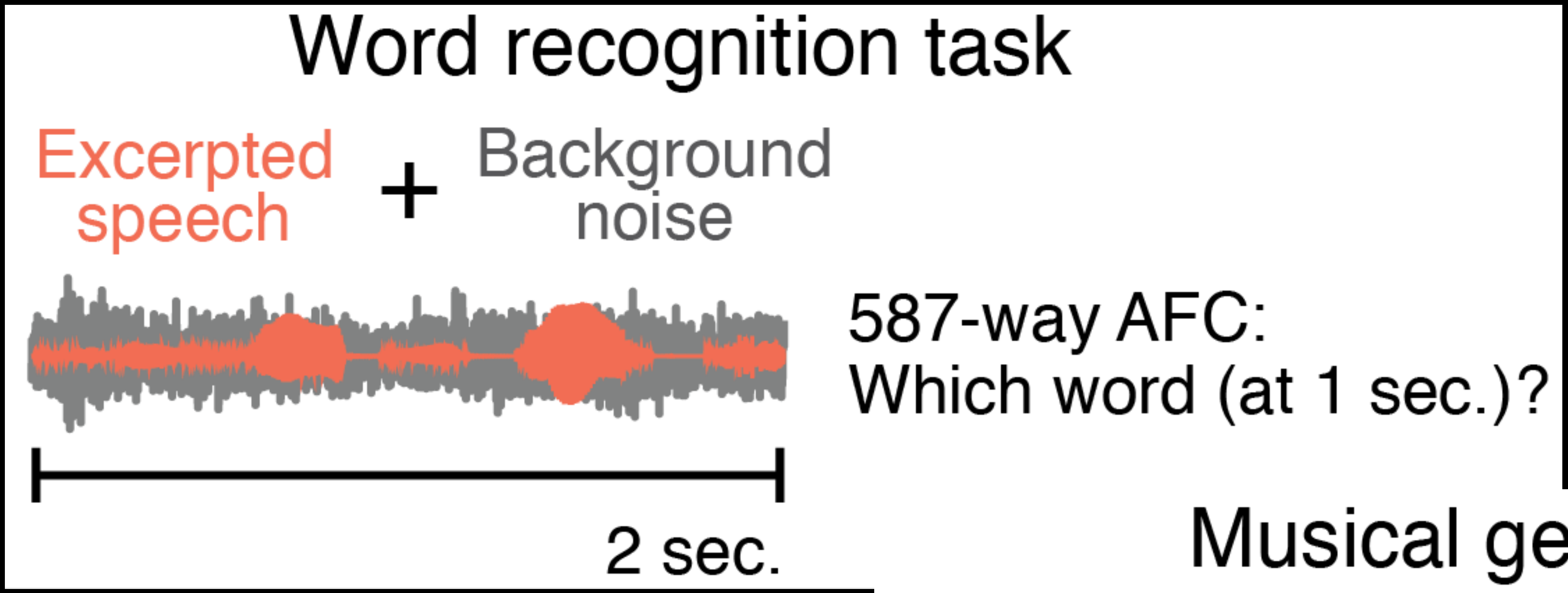
Network optimization: Real-world tasks

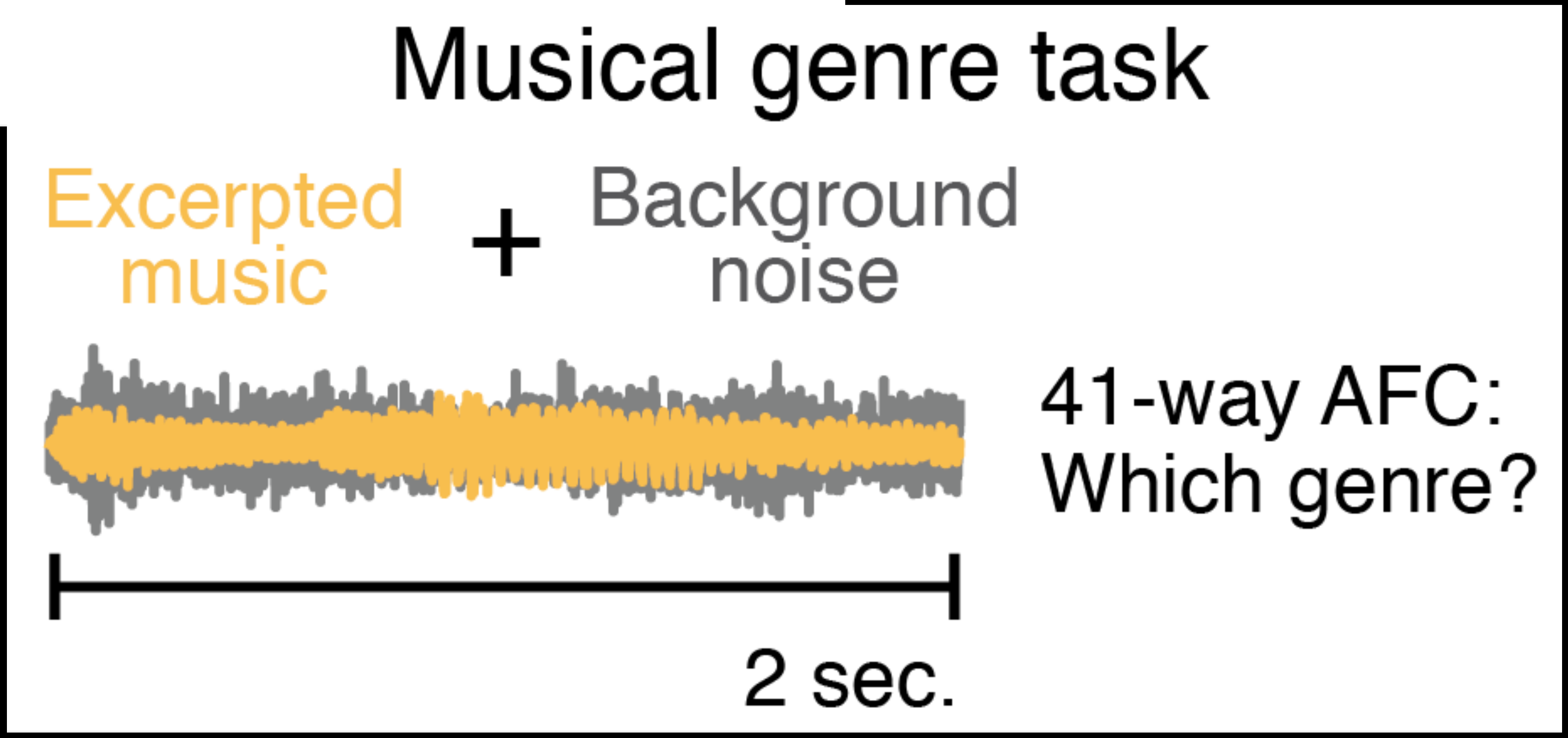
… that have large labelled datasets.



Network optimization: Real-world tasks

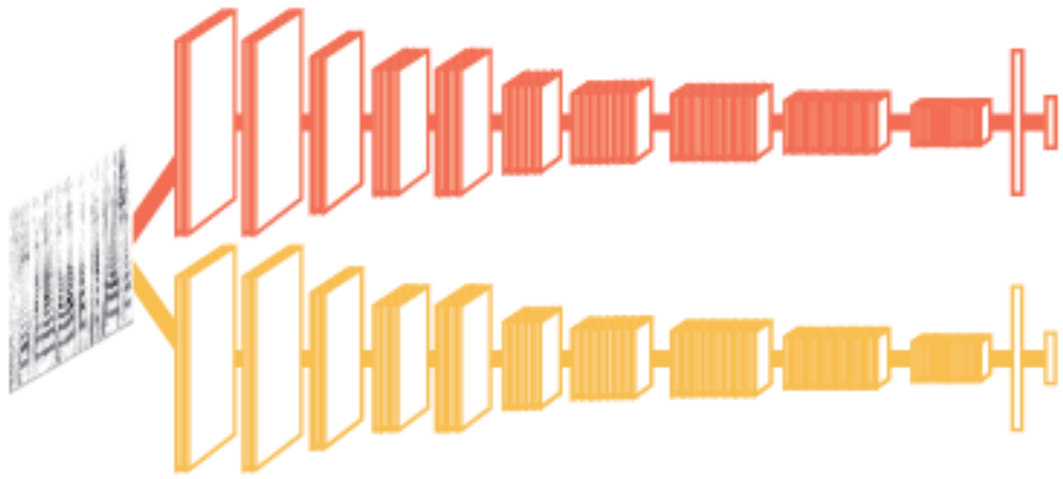
... that have large labelled datasets.



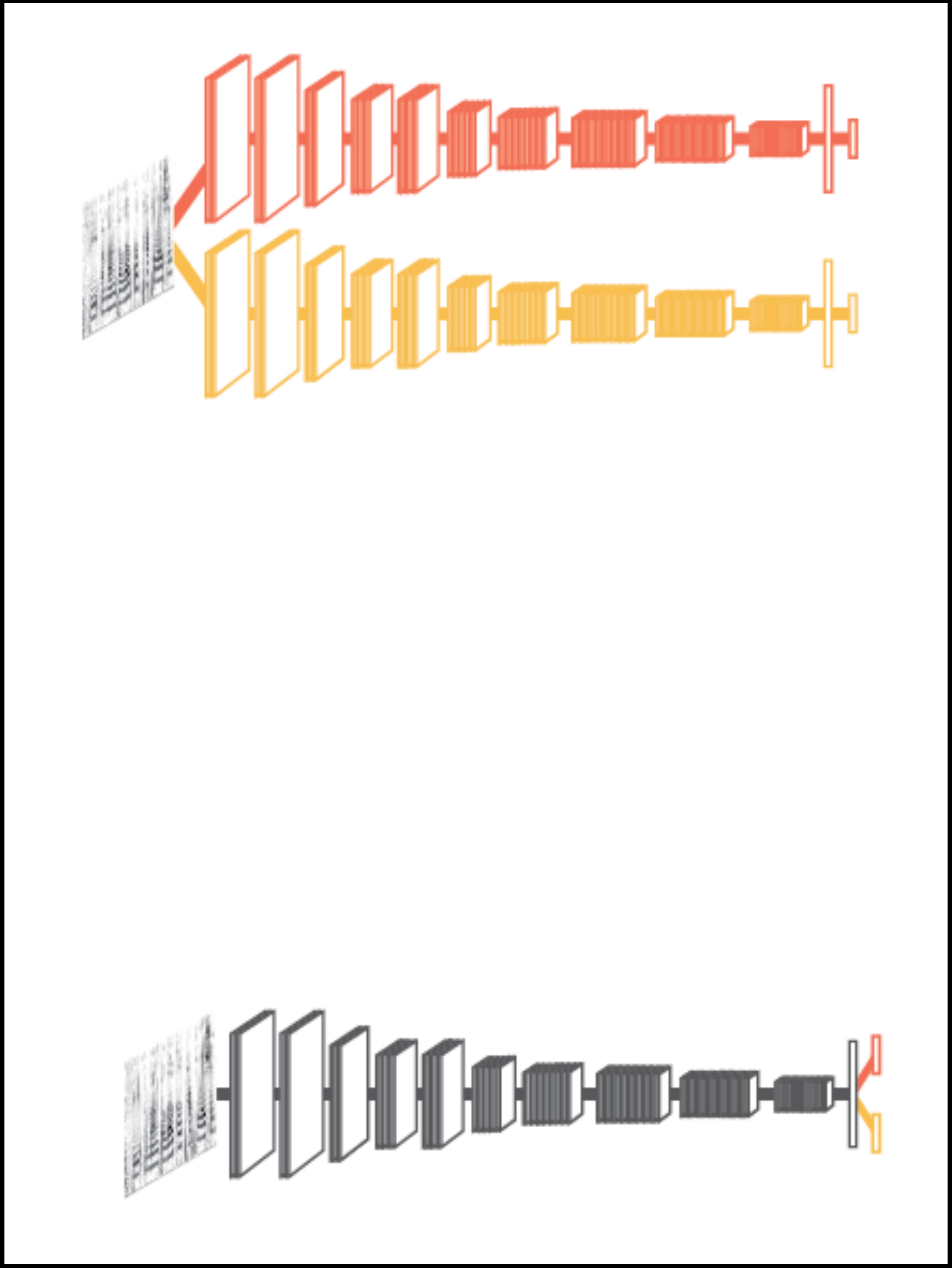


Network optimization: Architecture search

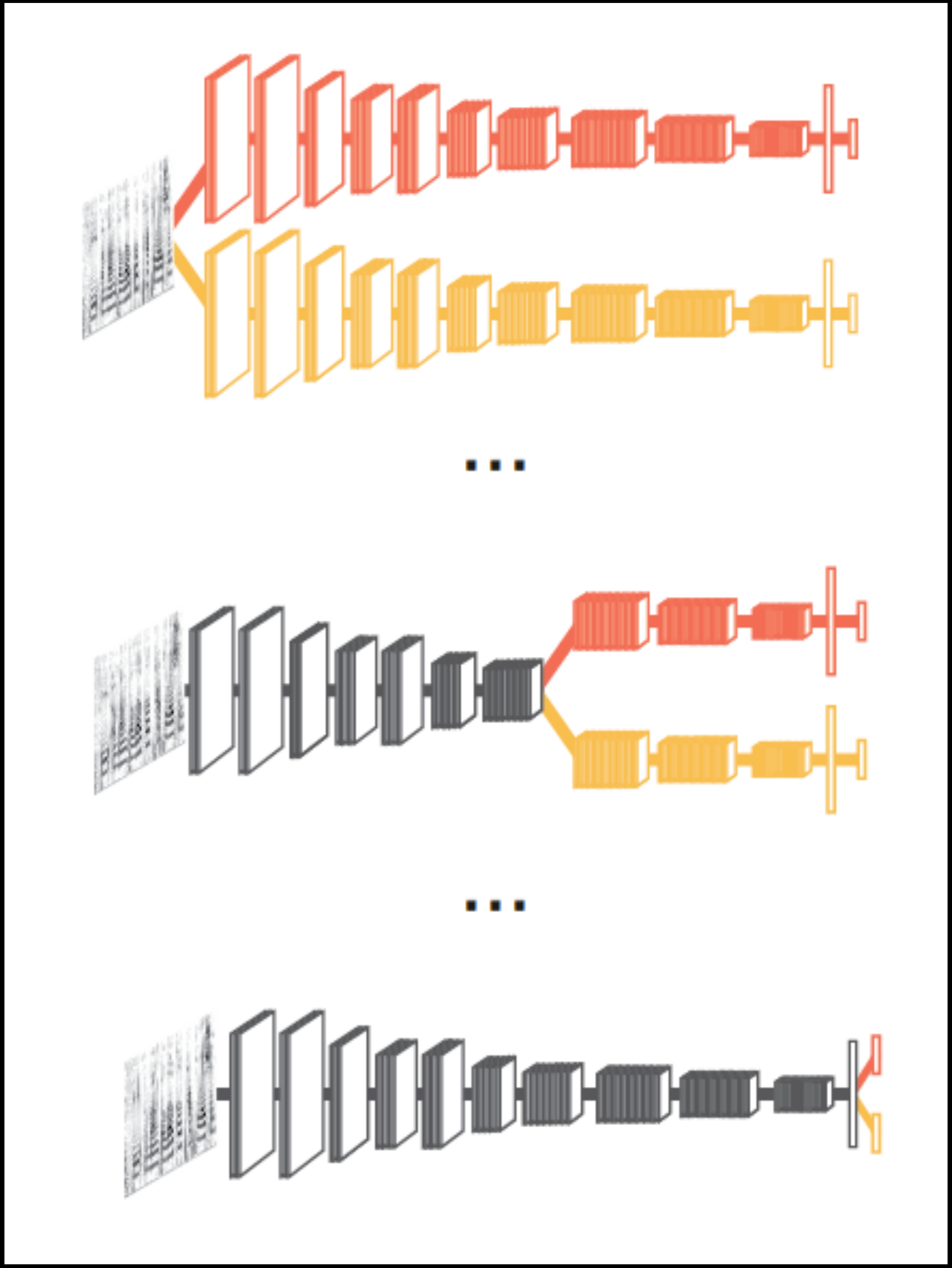
Network optimization: Architecture search



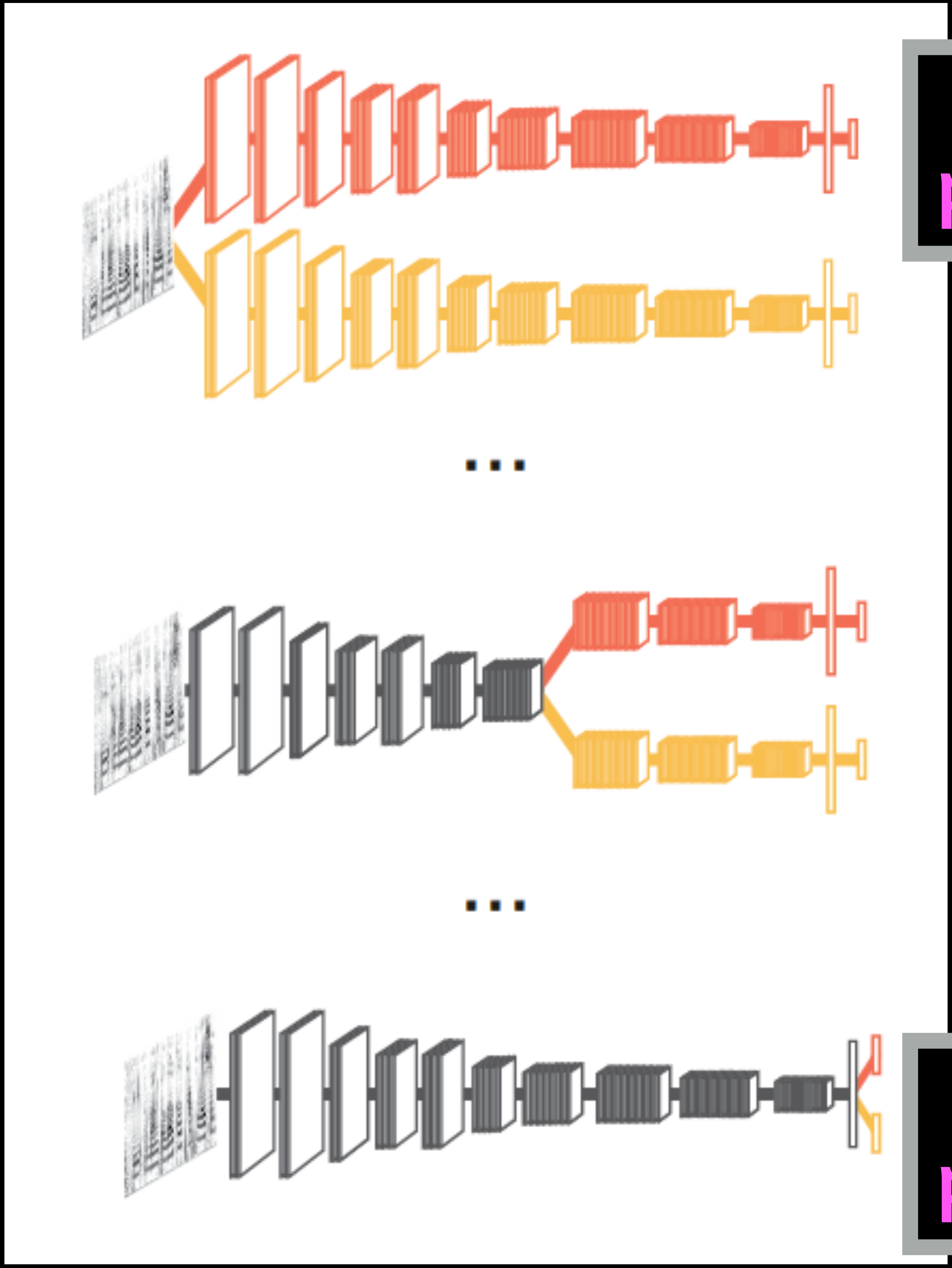
Network optimization: Architecture search



Network optimization: Architecture search



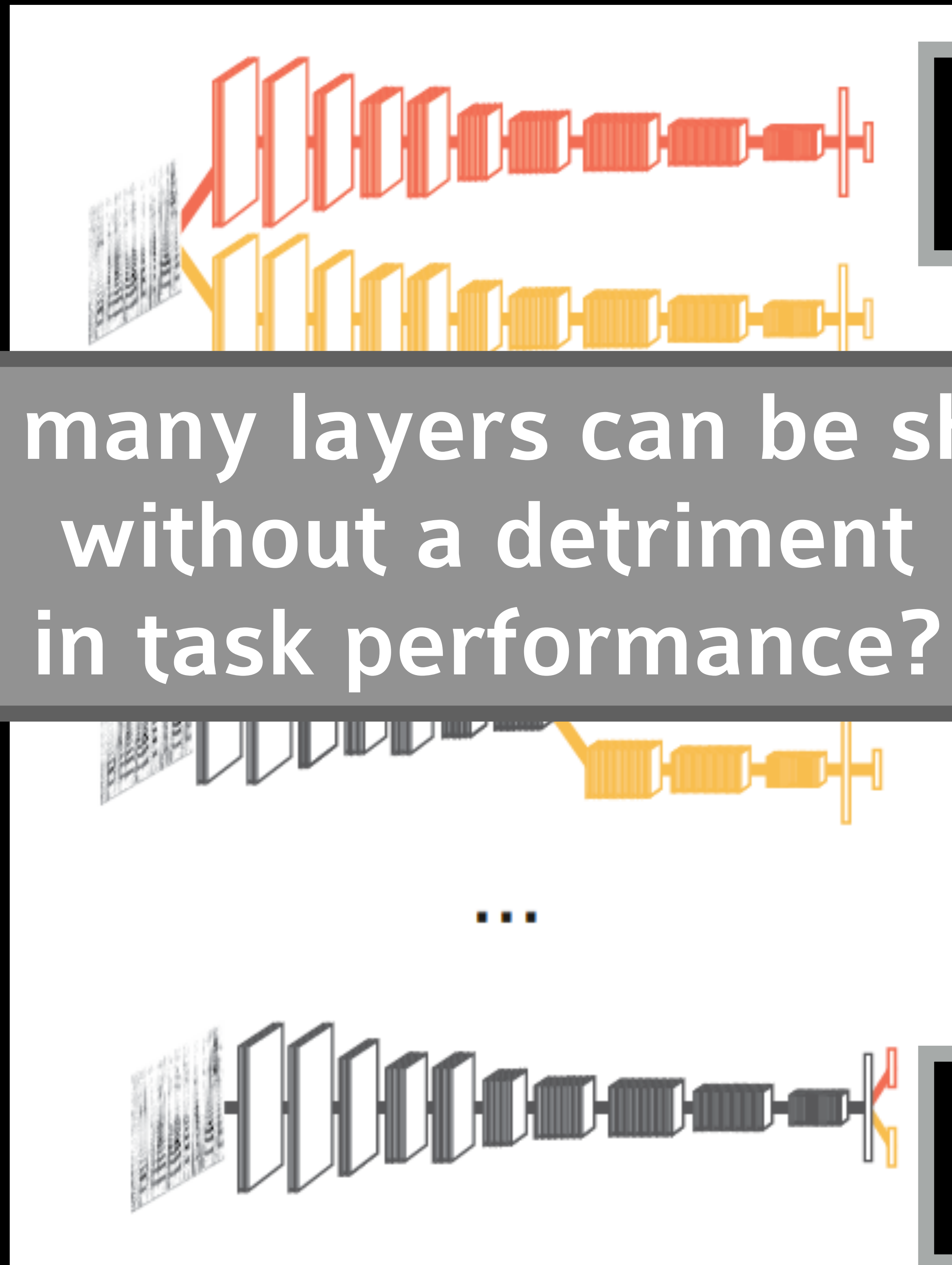
Network optimization: Architecture search



More parameters

Fewer parameters

Network optimization: Architecture search

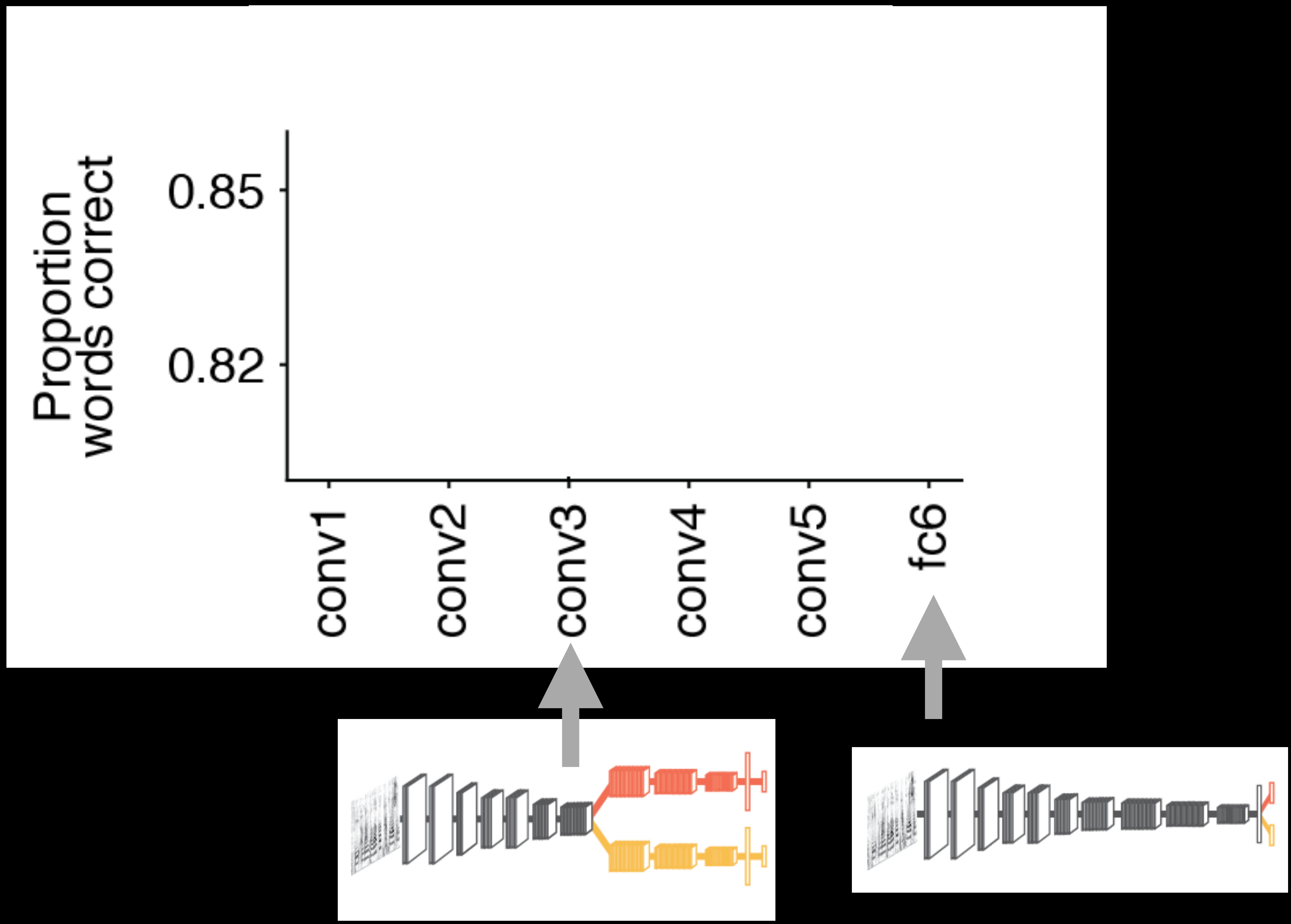


More parameters

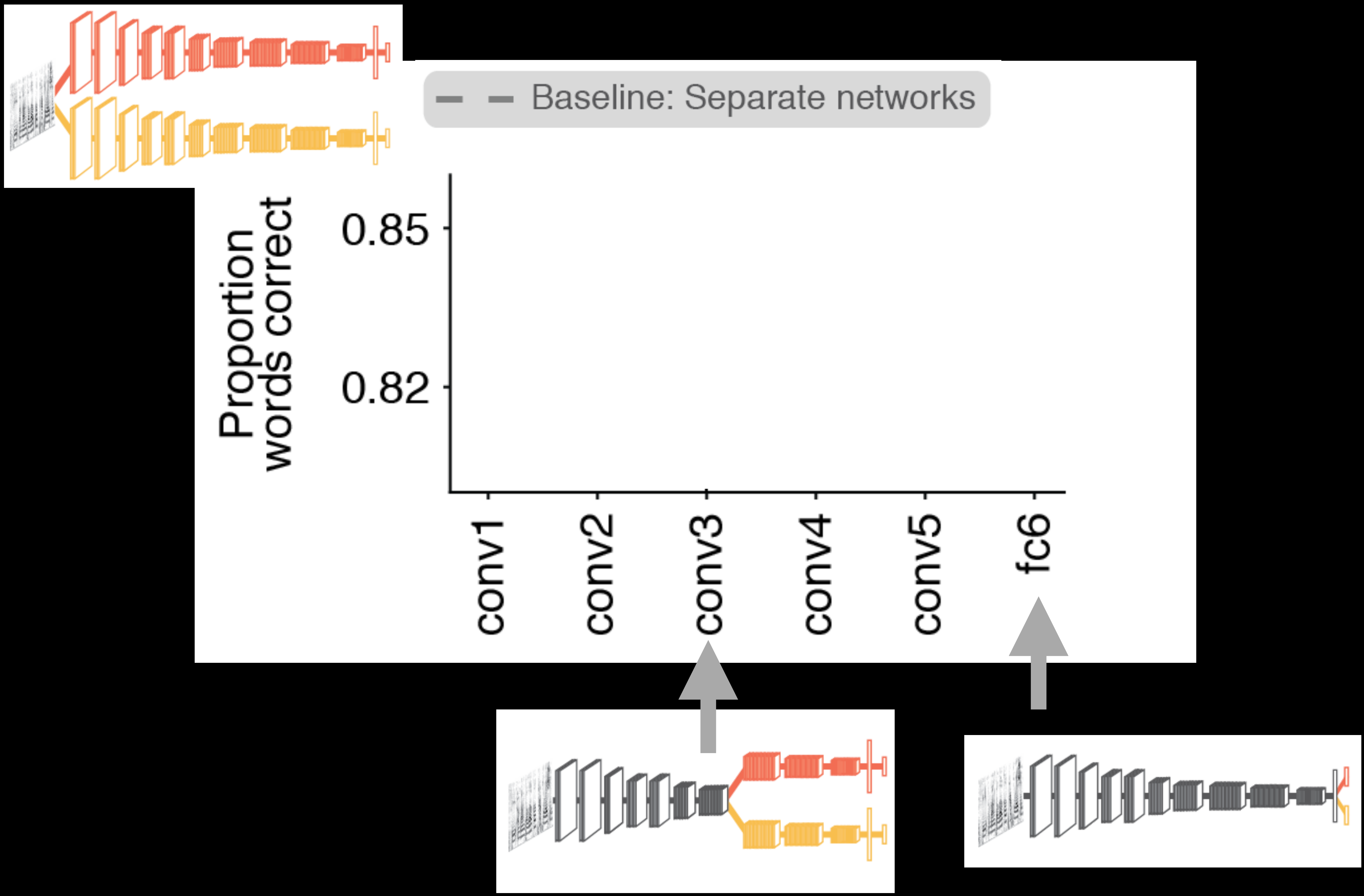
How many layers can be shared without a detriment in task performance?

Fewer parameters

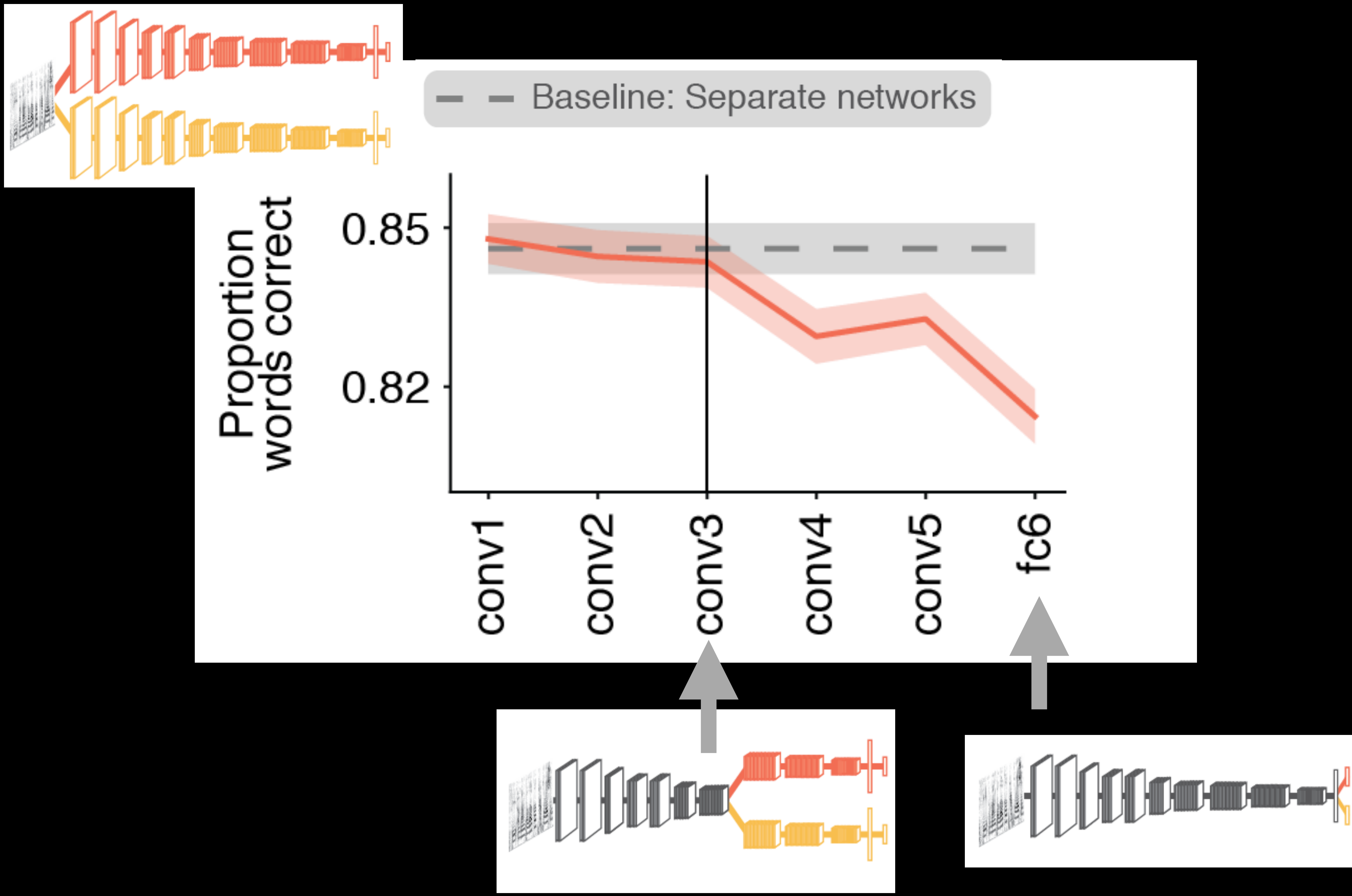
Network optimization: Architecture search



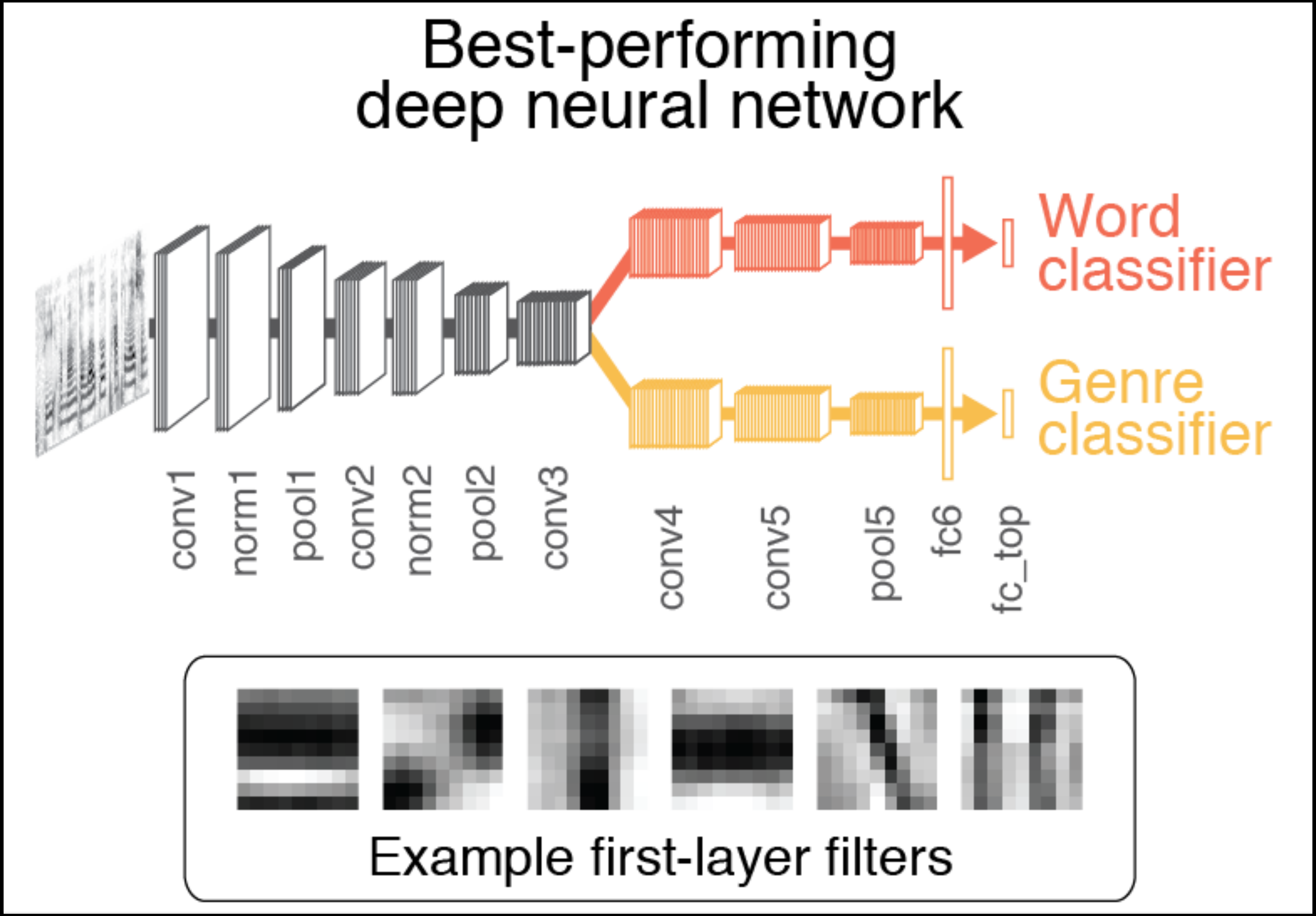
Network optimization: Architecture search



Network optimization: Architecture search

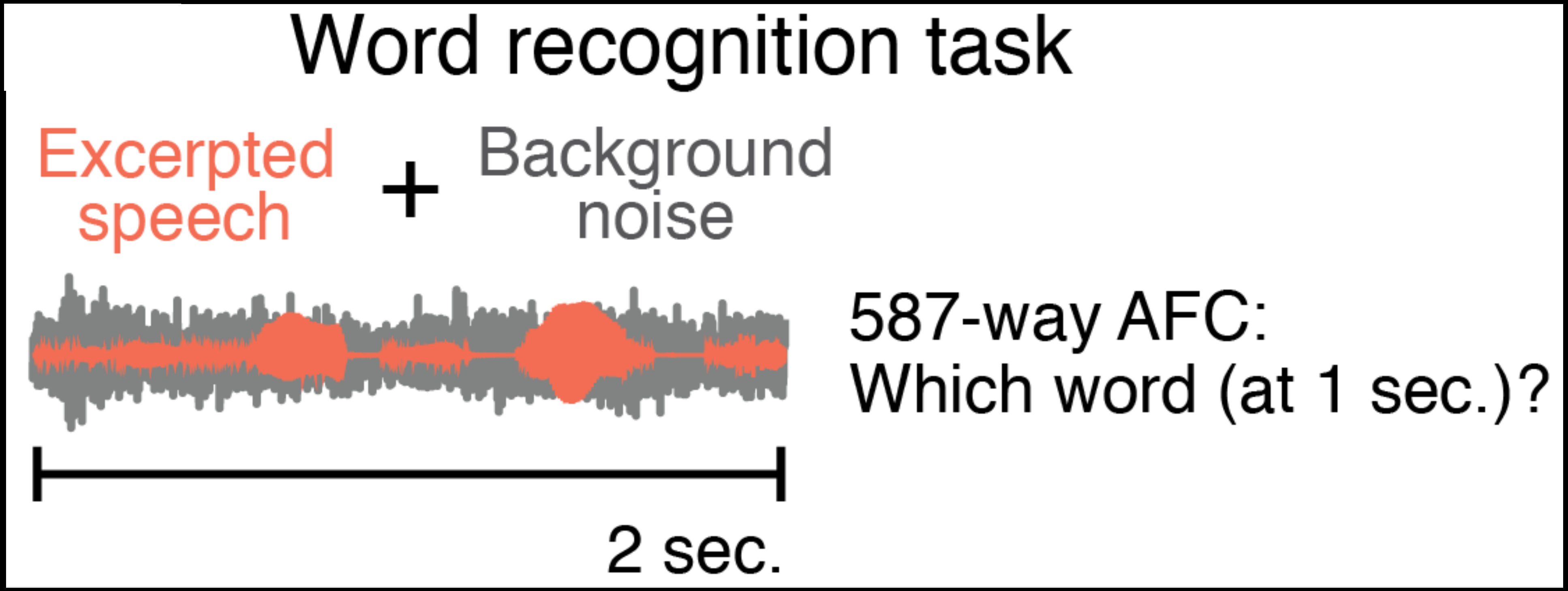


Network optimization: Resulting network

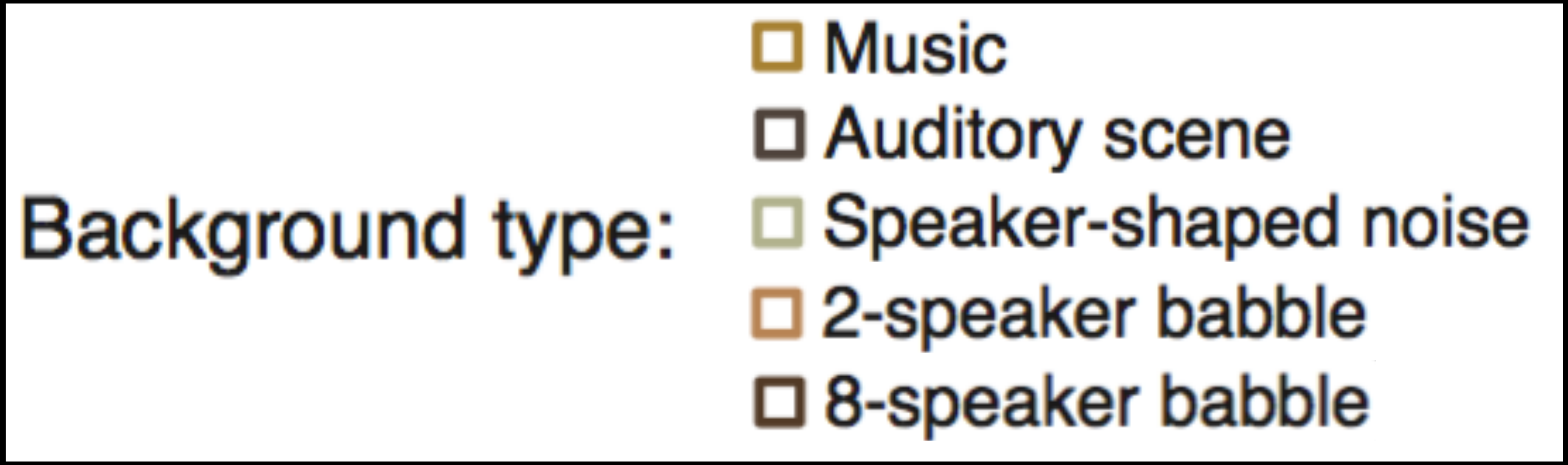


Comparing human & model behavior

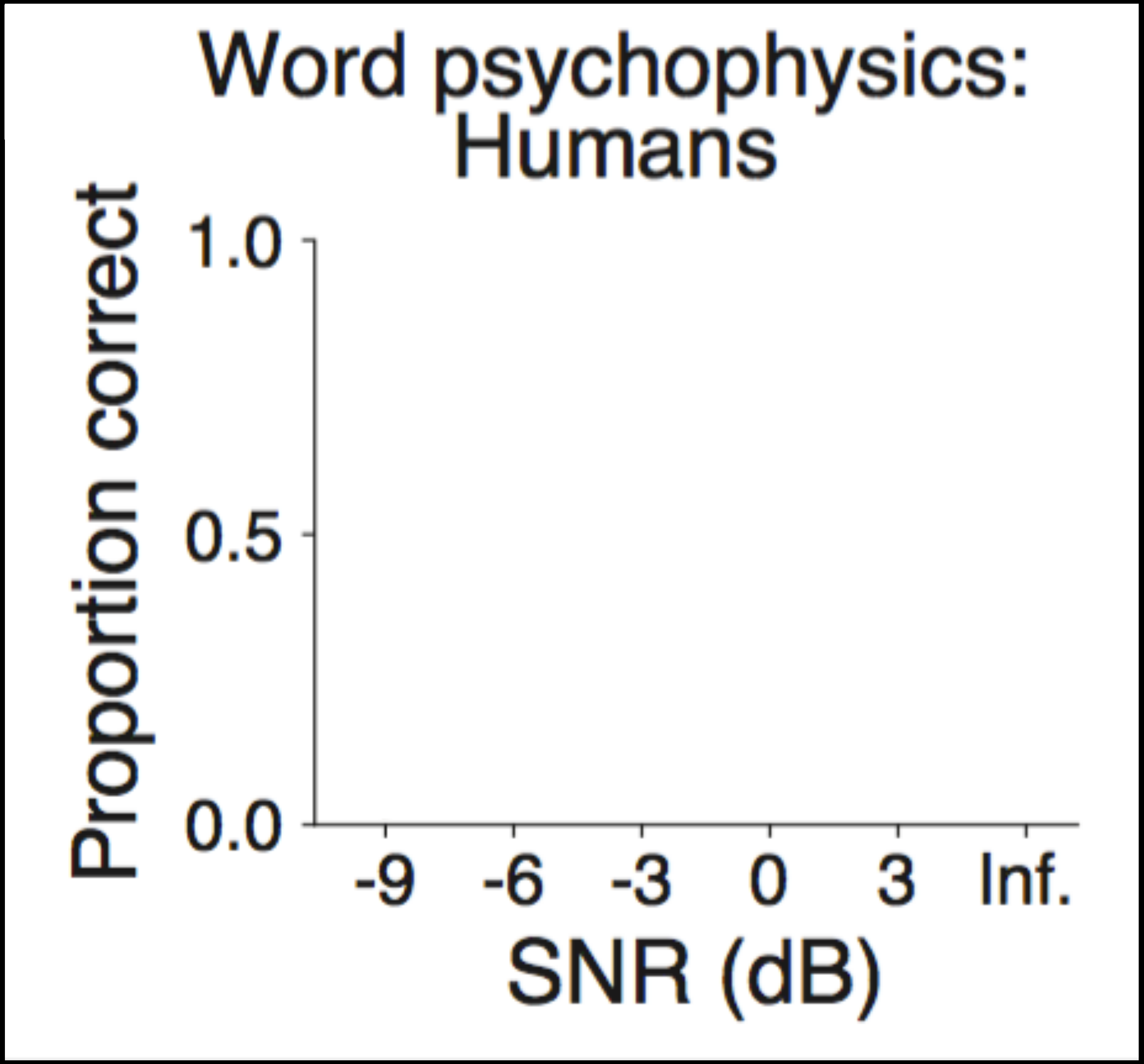
Comparing human & model behavior

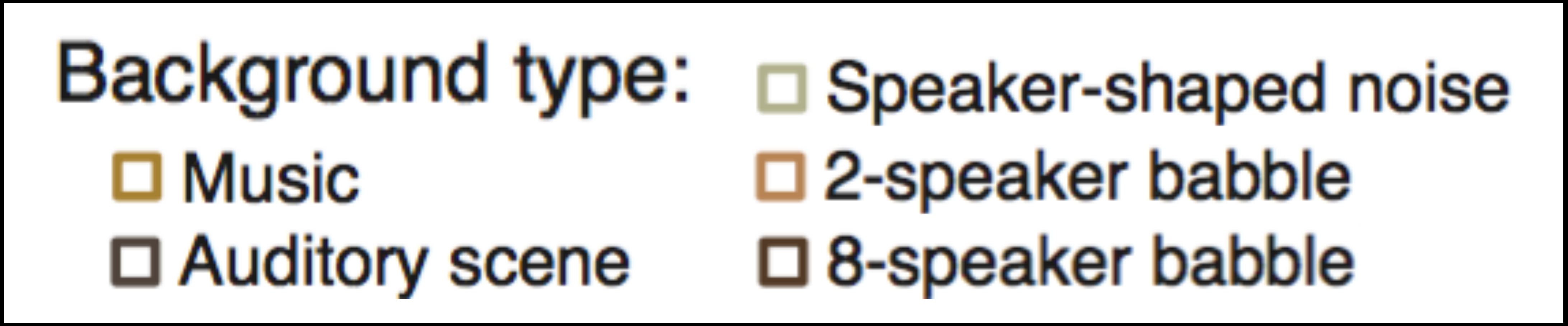


26 conditions:
5 background types x 5 signal-to-noise ratios (SNRs)
+ noiseless

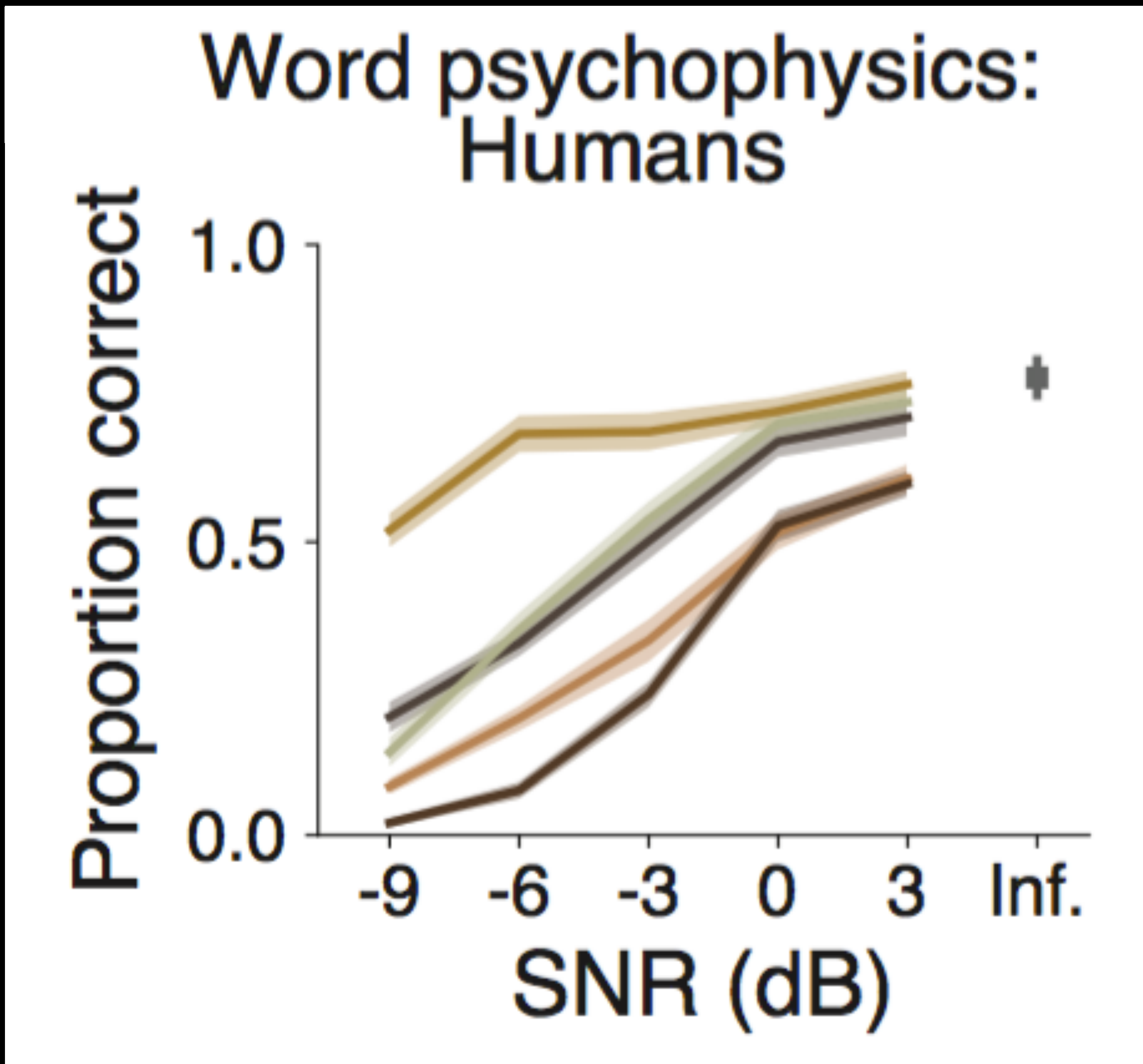


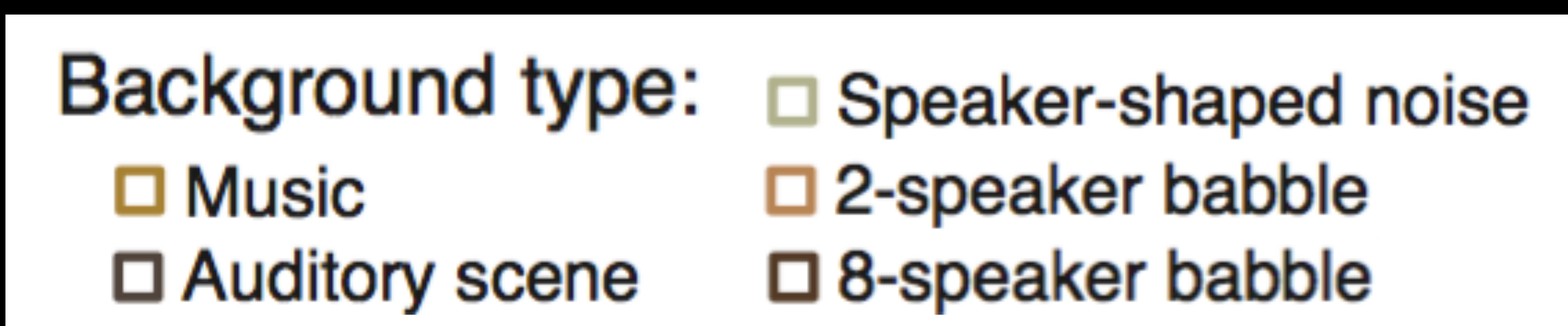
CNN & human psychophysics



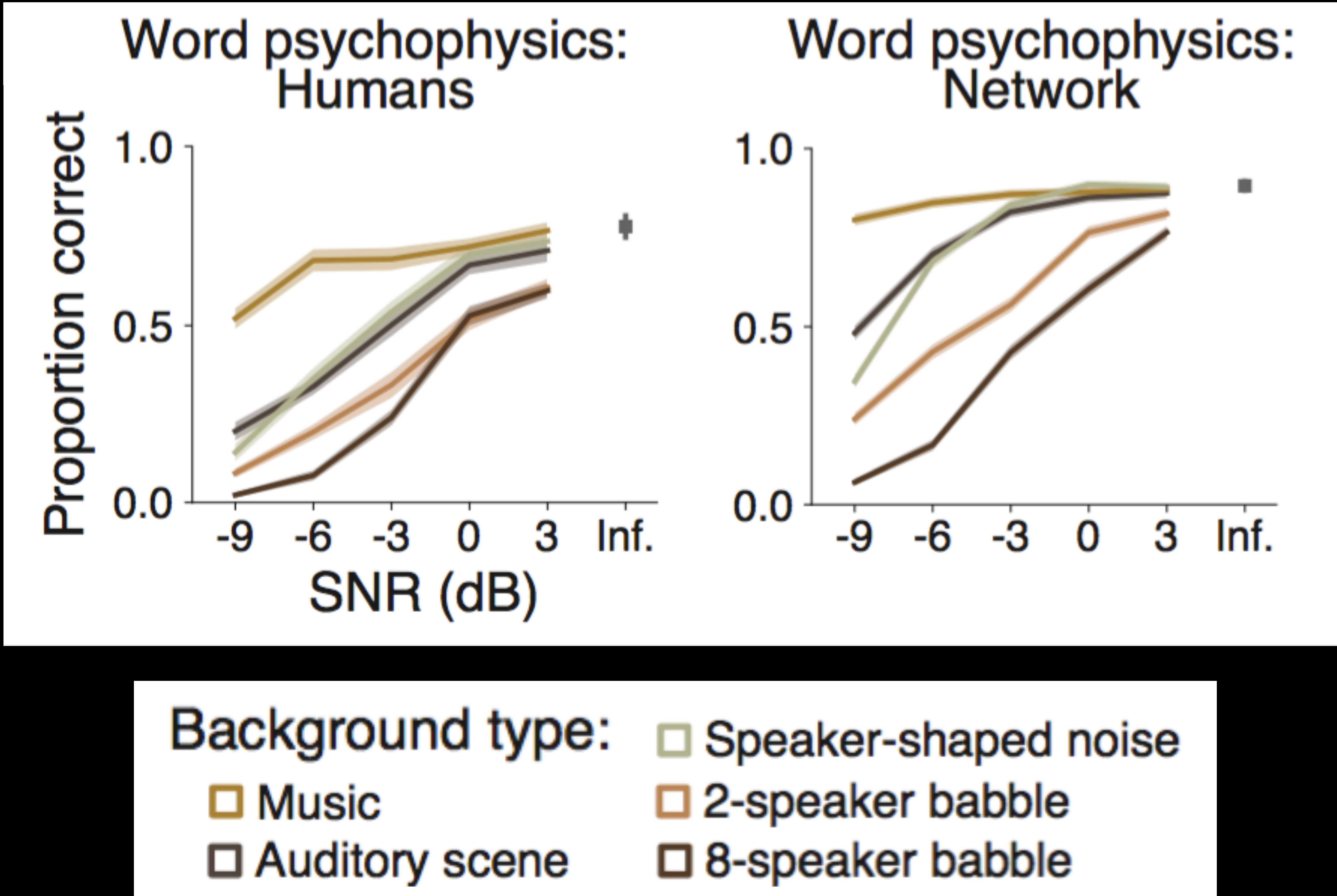


CNN & human psychophysics





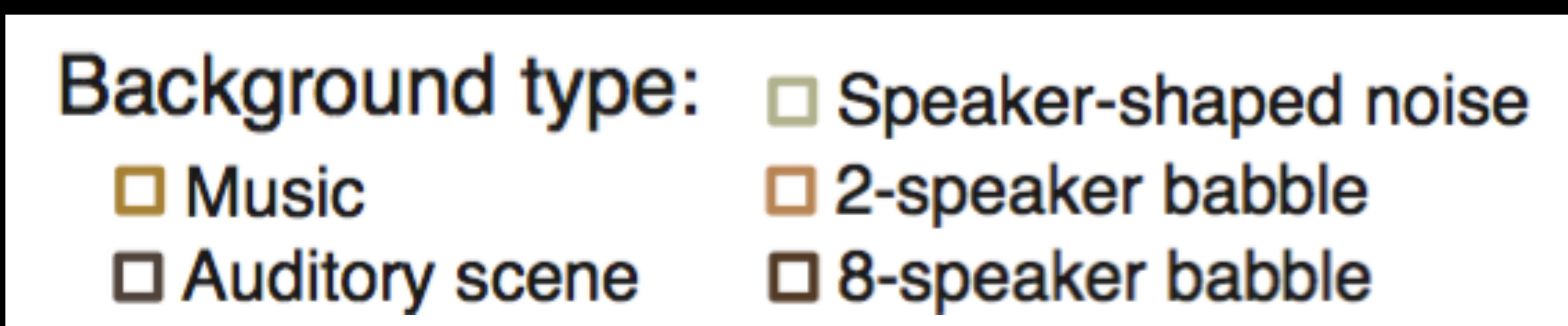
CNN & human psychophysics



CNN & human psychophysics

NOTE:

CNN optimized ONLY for task performance
NOT optimized to behave similarly to humans

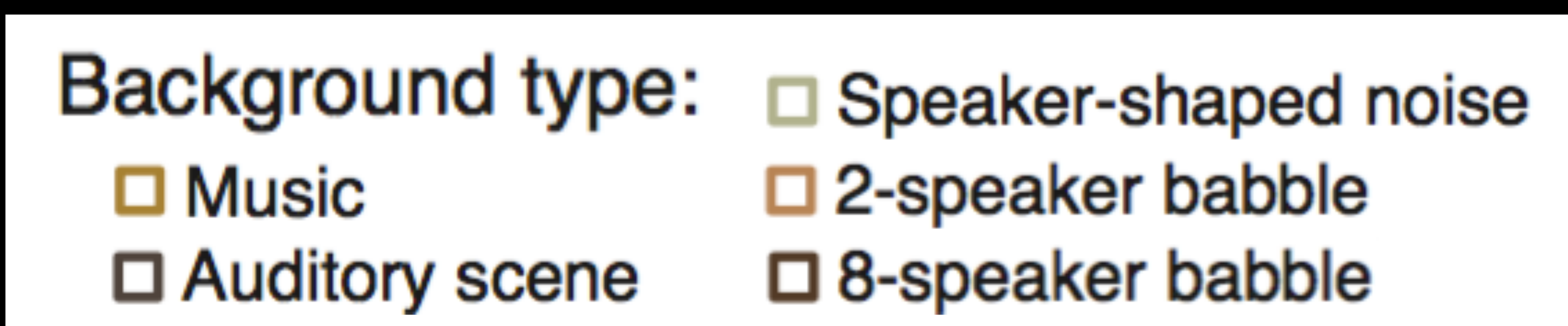


CNN & human psychophysics

NOTE:

CNN optimized ONLY for task performance
NOT optimized to behave similarly to humans

POTENTIAL REASONS FOR SIMILARITY:



CNN & human psychophysics

NOTE:

CNN optimized ONLY for task performance
NOT optimized to behave similarly to humans

POTENTIAL REASONS FOR SIMILARITY:

1. Both network & humans near optimal?

Background type:
- Music
- Auditory scene
- Speaker-shaped noise
- 2-speaker babble
- 8-speaker babble

CNN & human psychophysics

NOTE:

CNN optimized ONLY for task performance
NOT optimized to behave similarly to humans

POTENTIAL REASONS FOR SIMILARITY:

1. Both network & humans near optimal?
2. Algorithmic similarities between net & humans?

Background type:
- Music
- Auditory scene
- Speaker-shaped noise
- 2-speaker babble
- 8-speaker babble

Using this model to predict cortical responses to natural sounds

Using this model to predict cortical responses to natural sounds

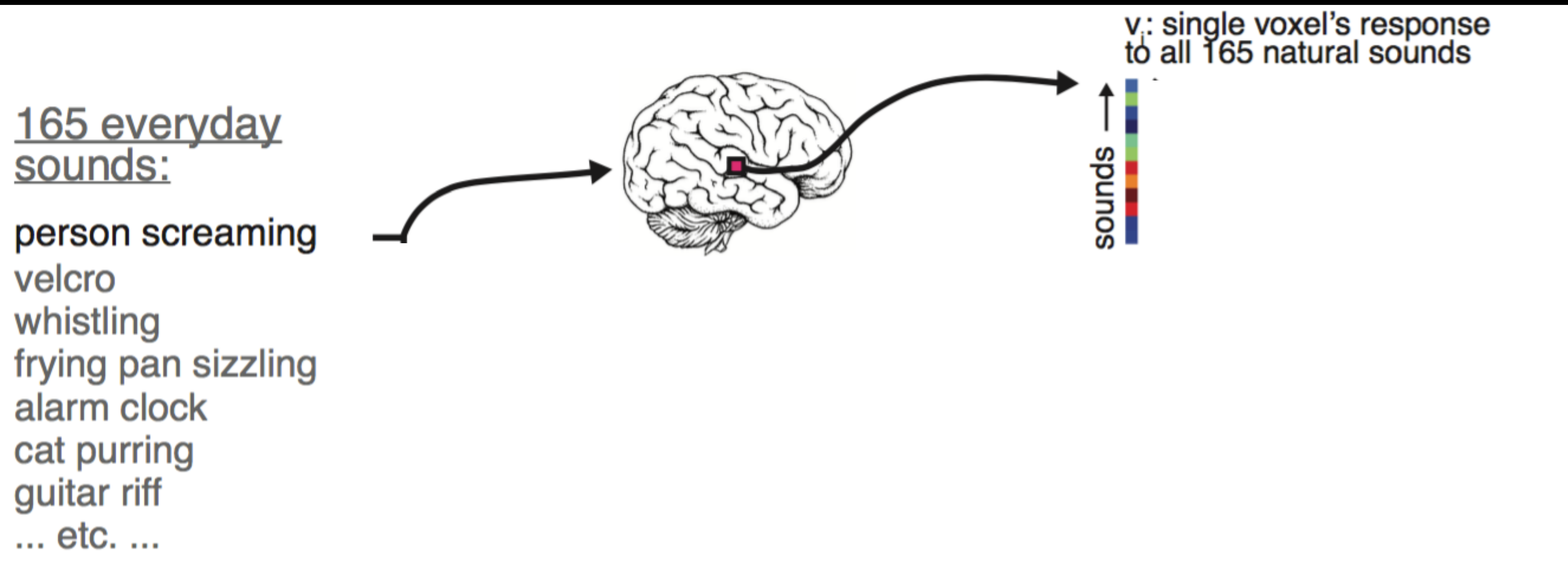
Measure fMRI responses to 165 natural sounds*

person screaming
man speaking
flushing toilet
pouring liquid
tooth-brushing
woman speaking
car accelerating
biting and chewing
laughing
typing
car engine starting
running water
breathing
keys jangling
dishes clanking
...

road traffic
zipper
cellphone vibrating
water dripping
scratching
car windows
telephone ringing
chopping food
telephone dialing
girl speaking
car horn
writing
computer startup sound
background speech
songbird
...

guitar
coughing
crumpling paper
siren
splashing water
computer speech
alarm clock
walking with heels
vacuum
wind
boy speaking
chair rolling
rock song
door knocking
dog barking
...

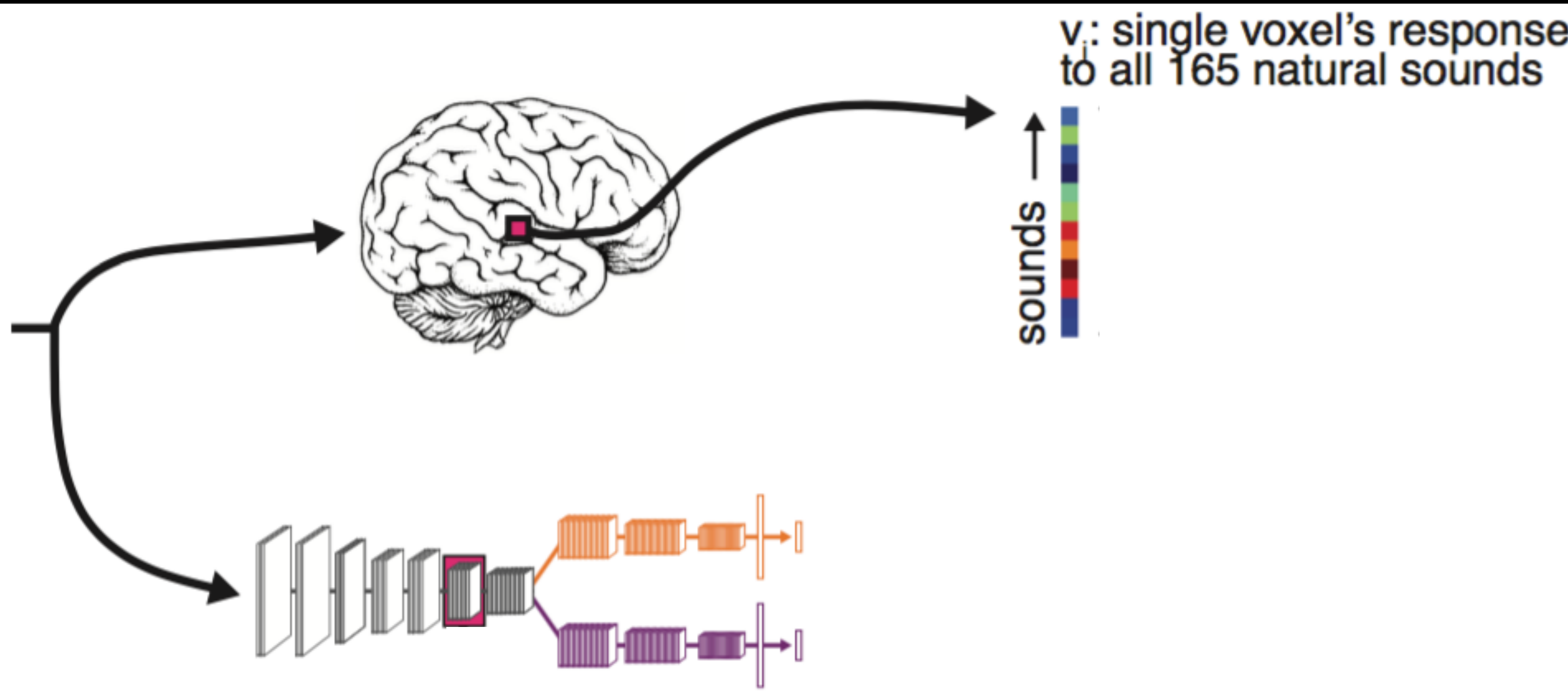
*Norman-Haignere, Kanwisher, McDermott Neuron 2015 (Thanks!)



Each voxel:
Mean response to each of 165 sounds

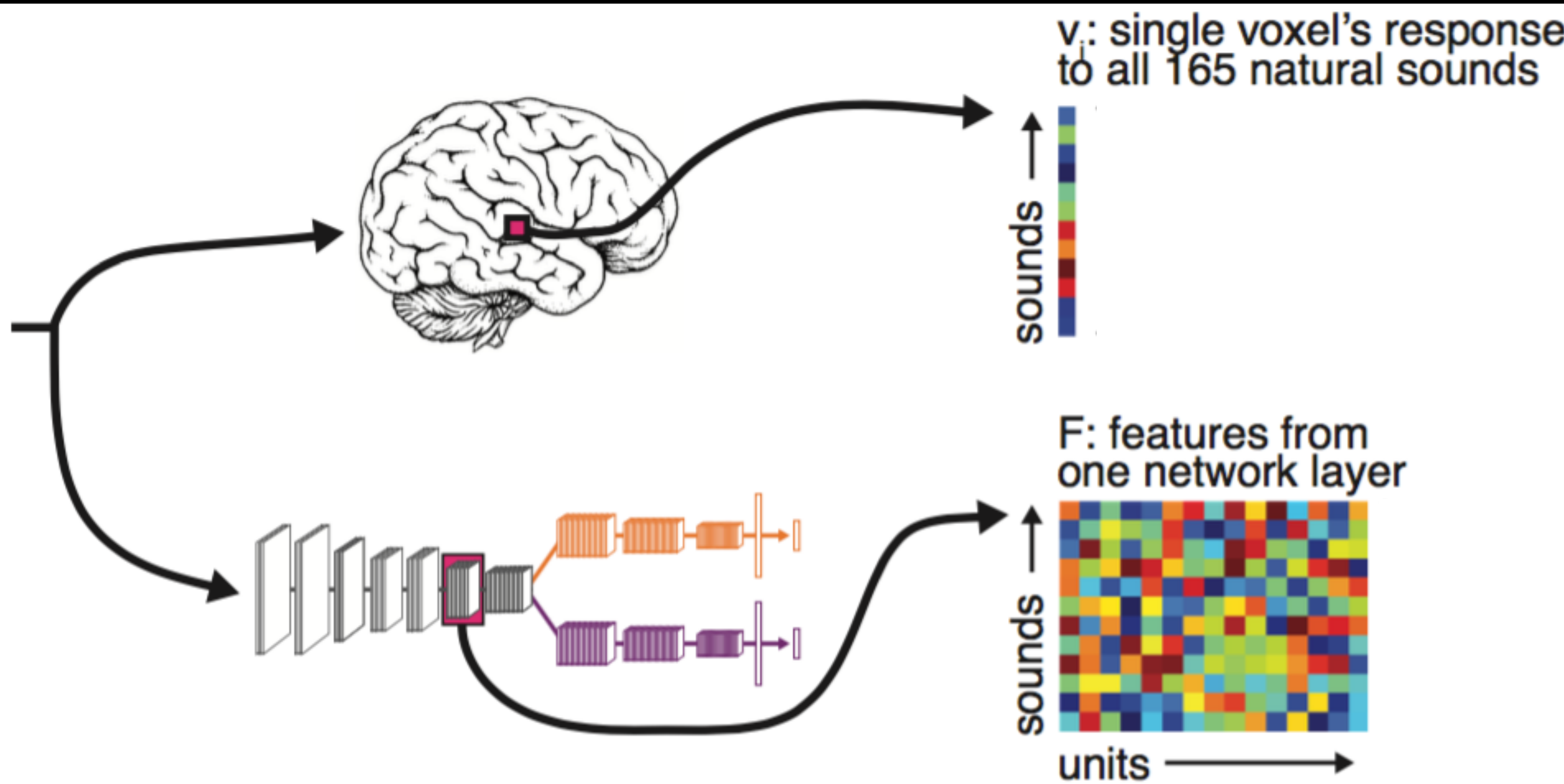
165 everyday sounds:

person screaming
velcro
whistling
frying pan sizzling
alarm clock
cat purring
guitar riff
... etc. ...



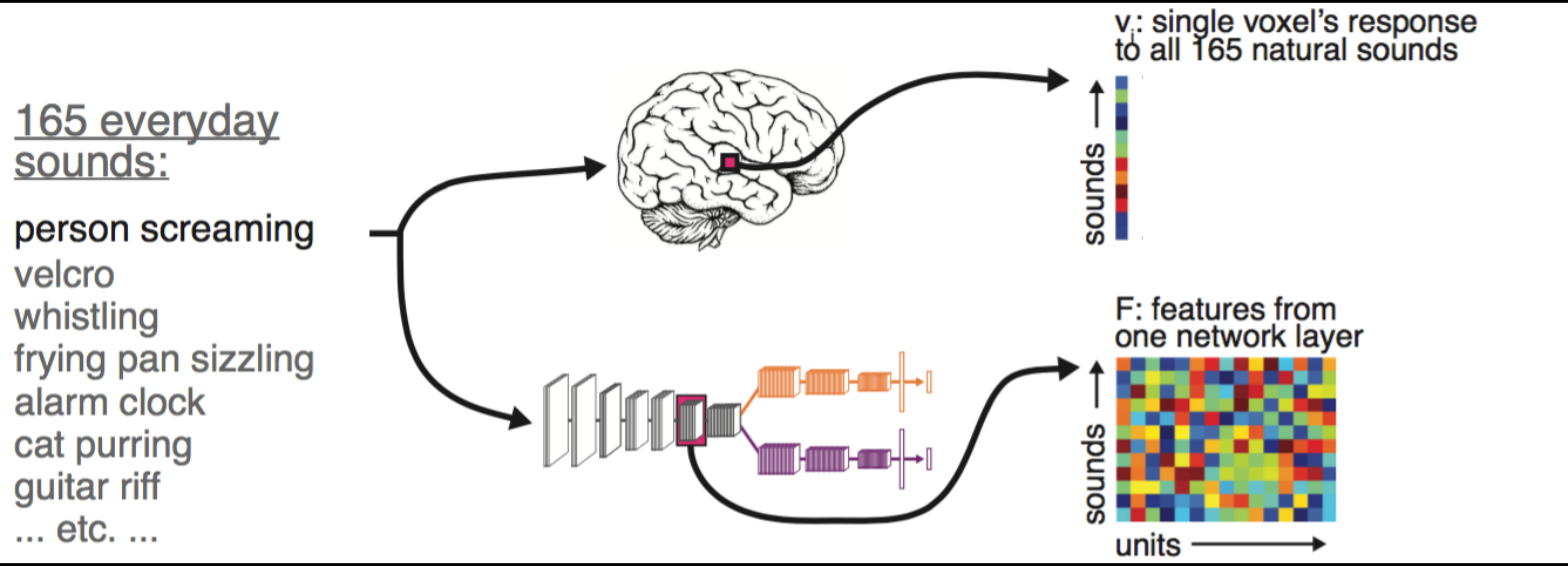
165 everyday sounds:

person screaming
velcro
whistling
frying pan sizzling
alarm clock
cat purring
guitar riff
... etc. ...



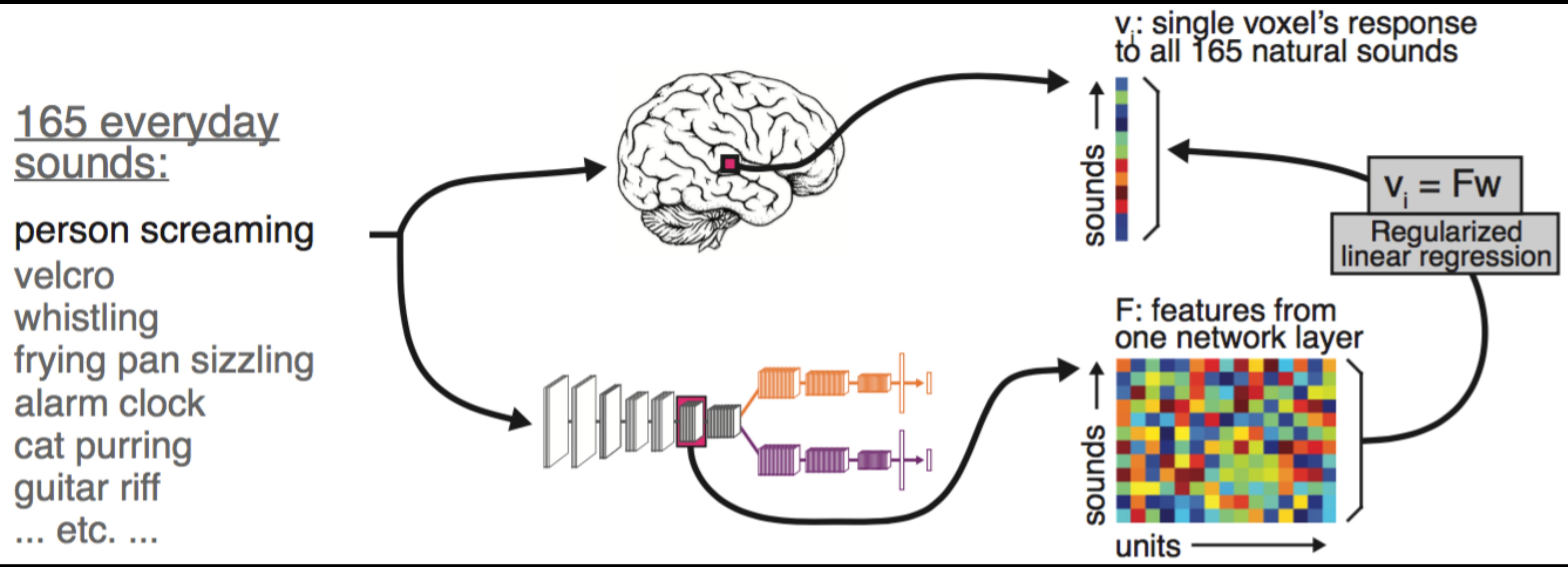
CNN as encoding model

Each voxel = weighted sum of units in a given layer



CNN as encoding model

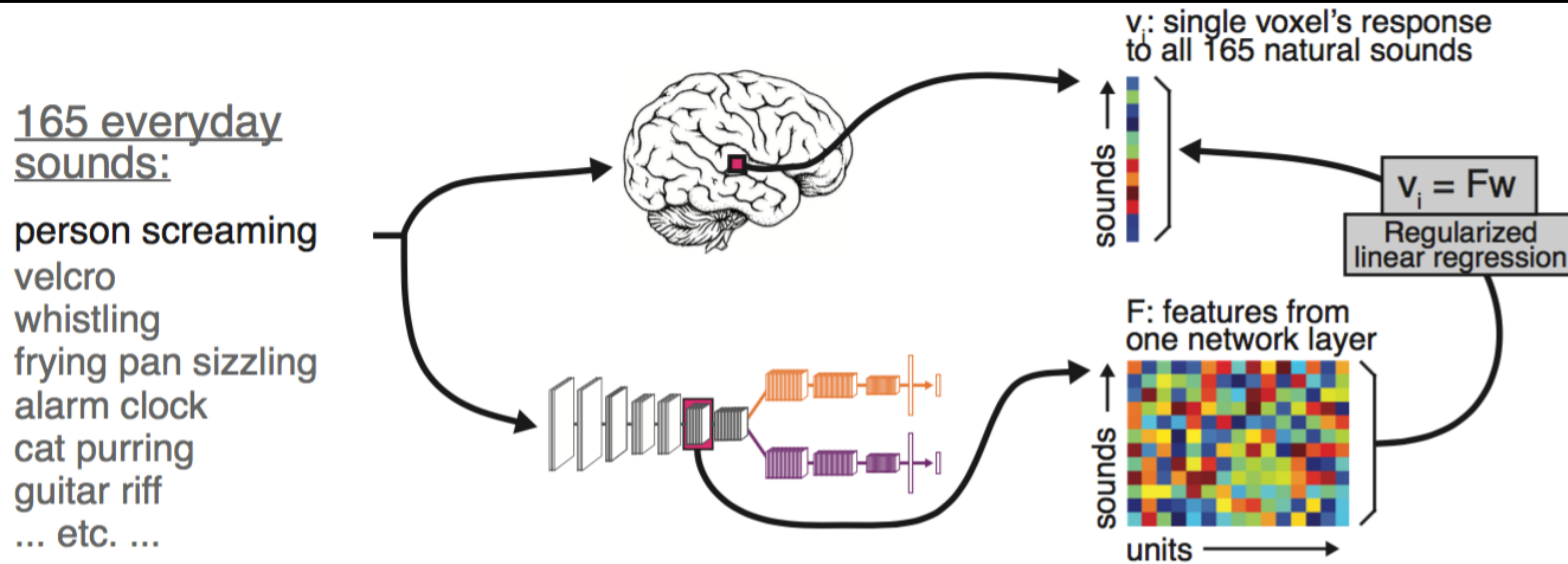
Each voxel = weighted sum of units in a given layer



Cross-validated regularized linear regression
to predict voxel's response

CNN as encoding model

Each voxel = weighted sum of units in a given layer

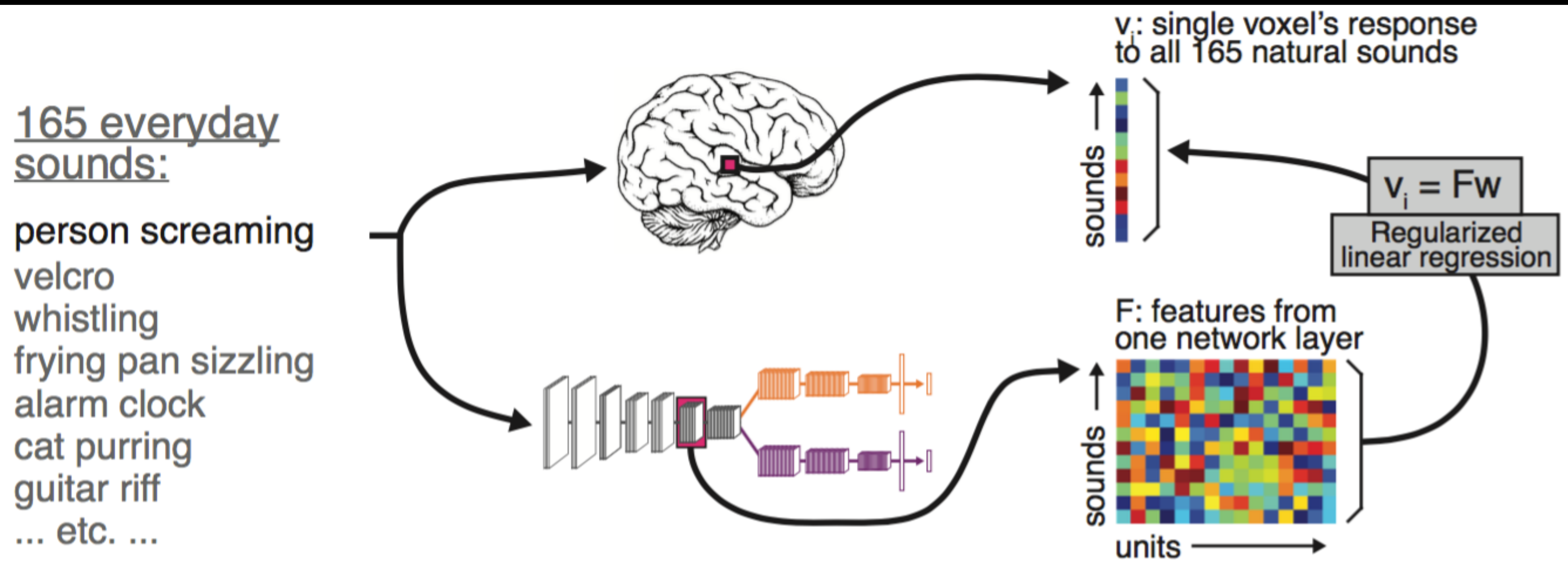


Cross-validated regularized linear regression to predict voxel's response

Dependent measure: Variance explained

CNN as encoding model

Each voxel = weighted sum of units in a given layer

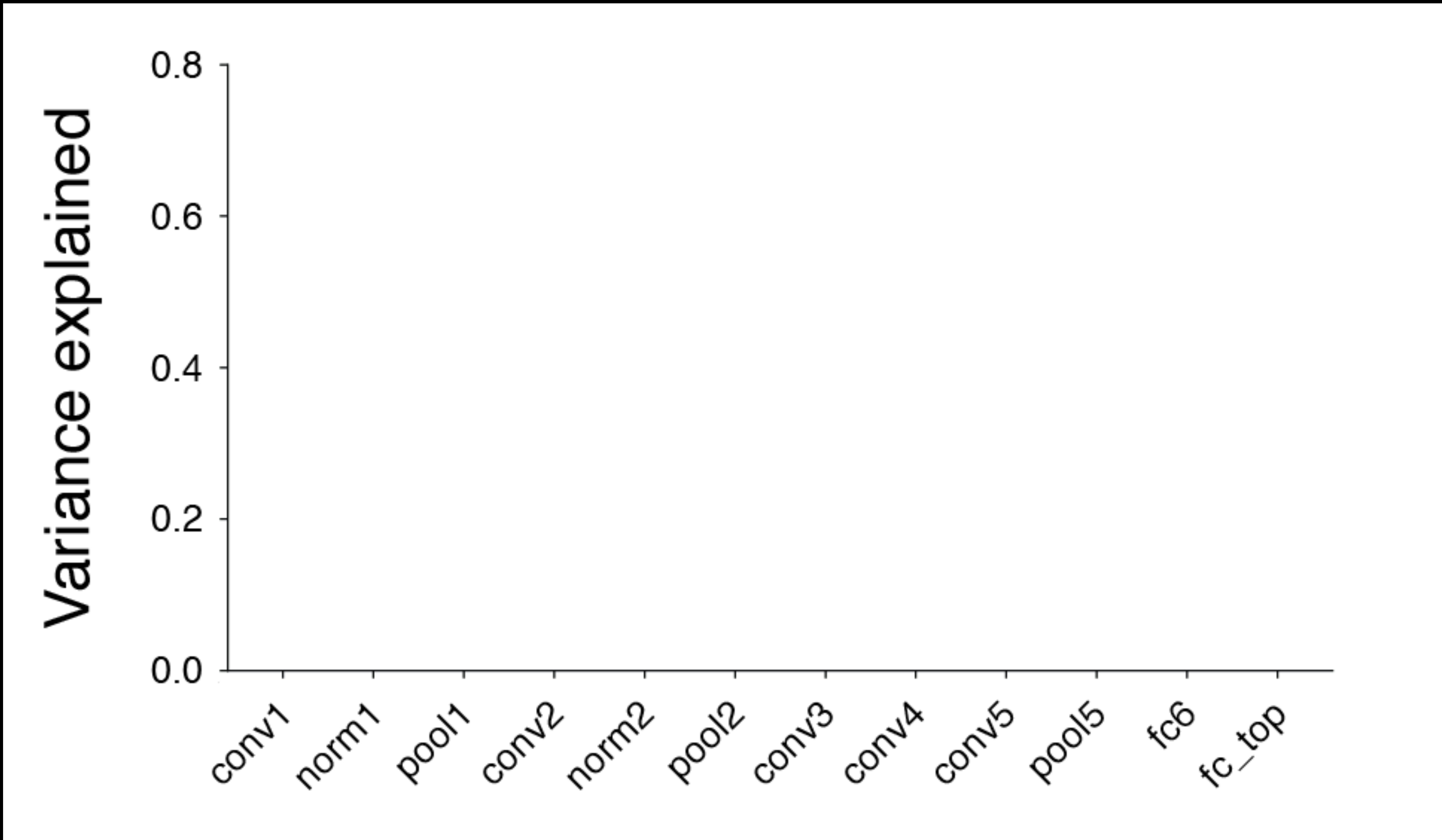


Cross-validated regularized linear regression to predict voxel's response

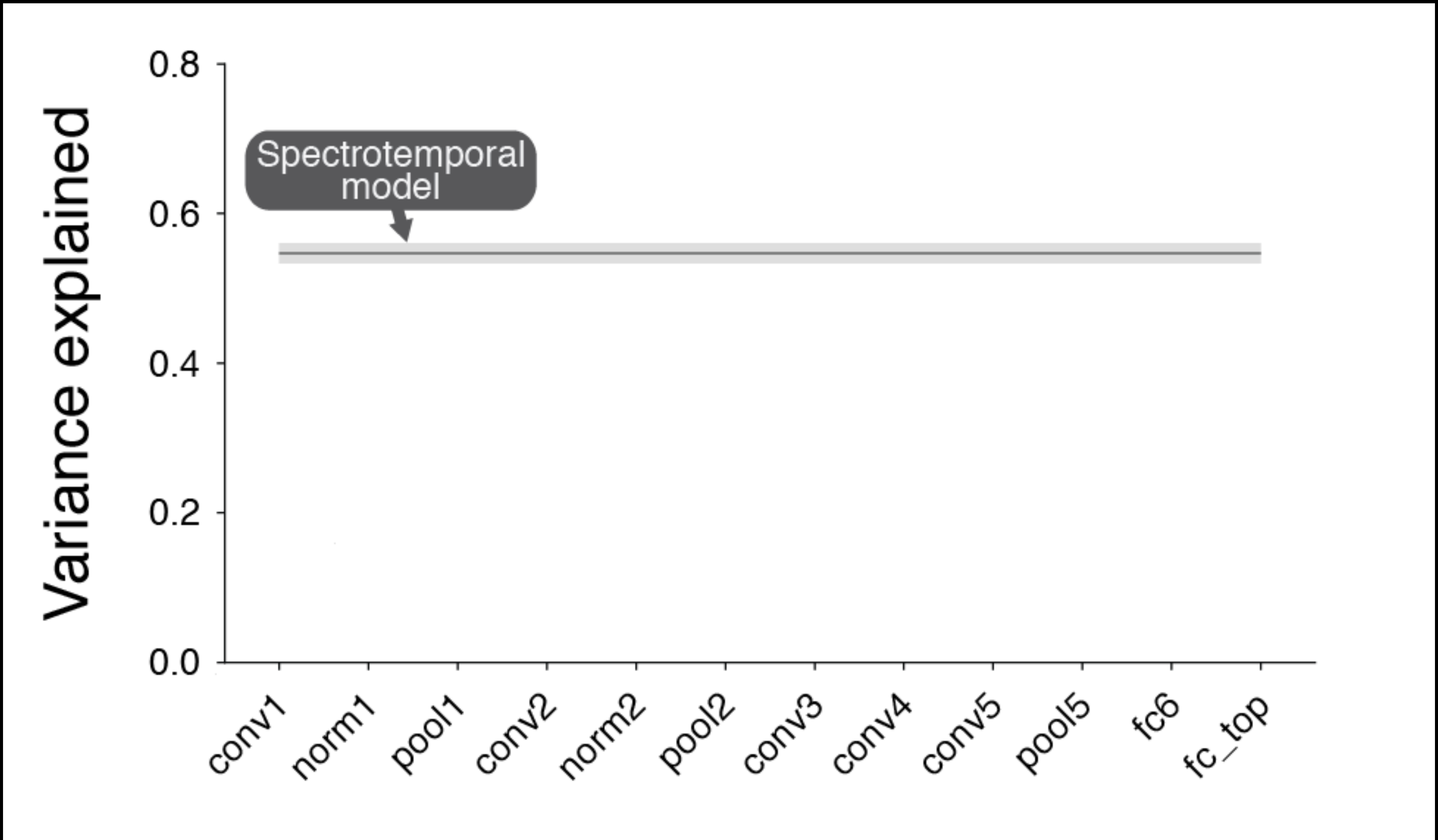
Dependent measure: Variance explained

Baseline:
Identical procedure with a spectrotemporal filter model

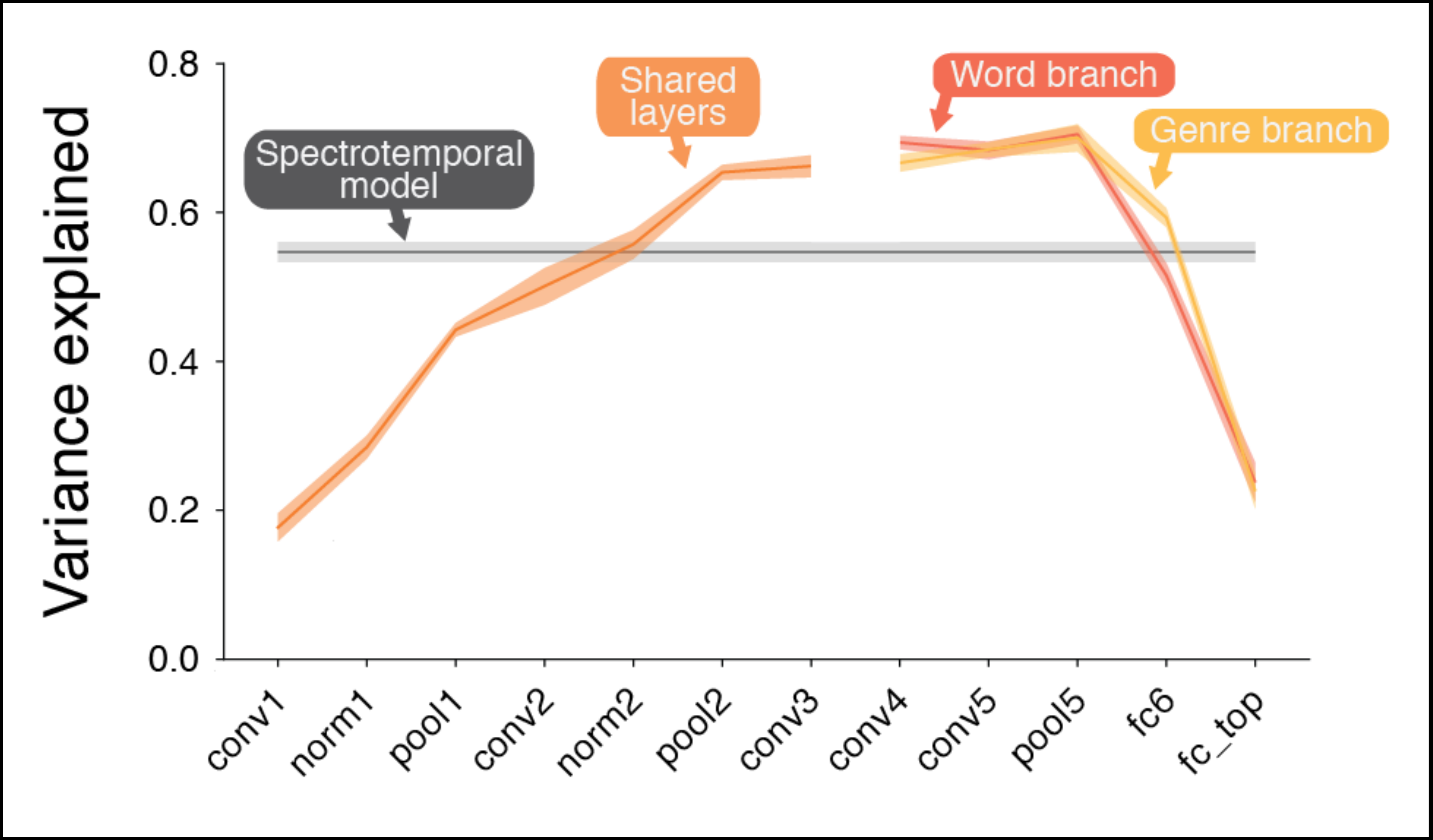
Variance explained across all of auditory cortex



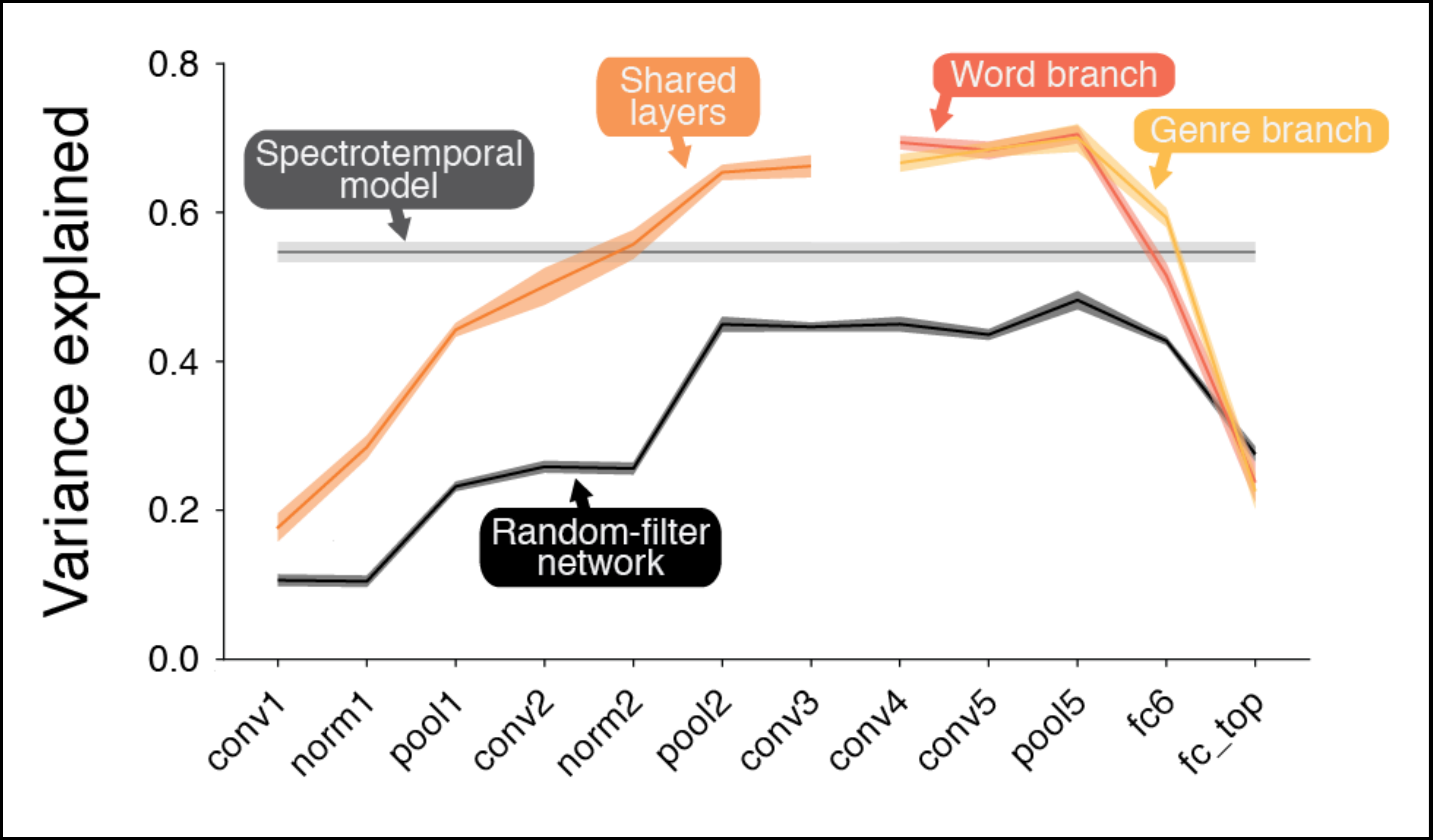
Variance explained across all of auditory cortex



Variance explained across all of auditory cortex



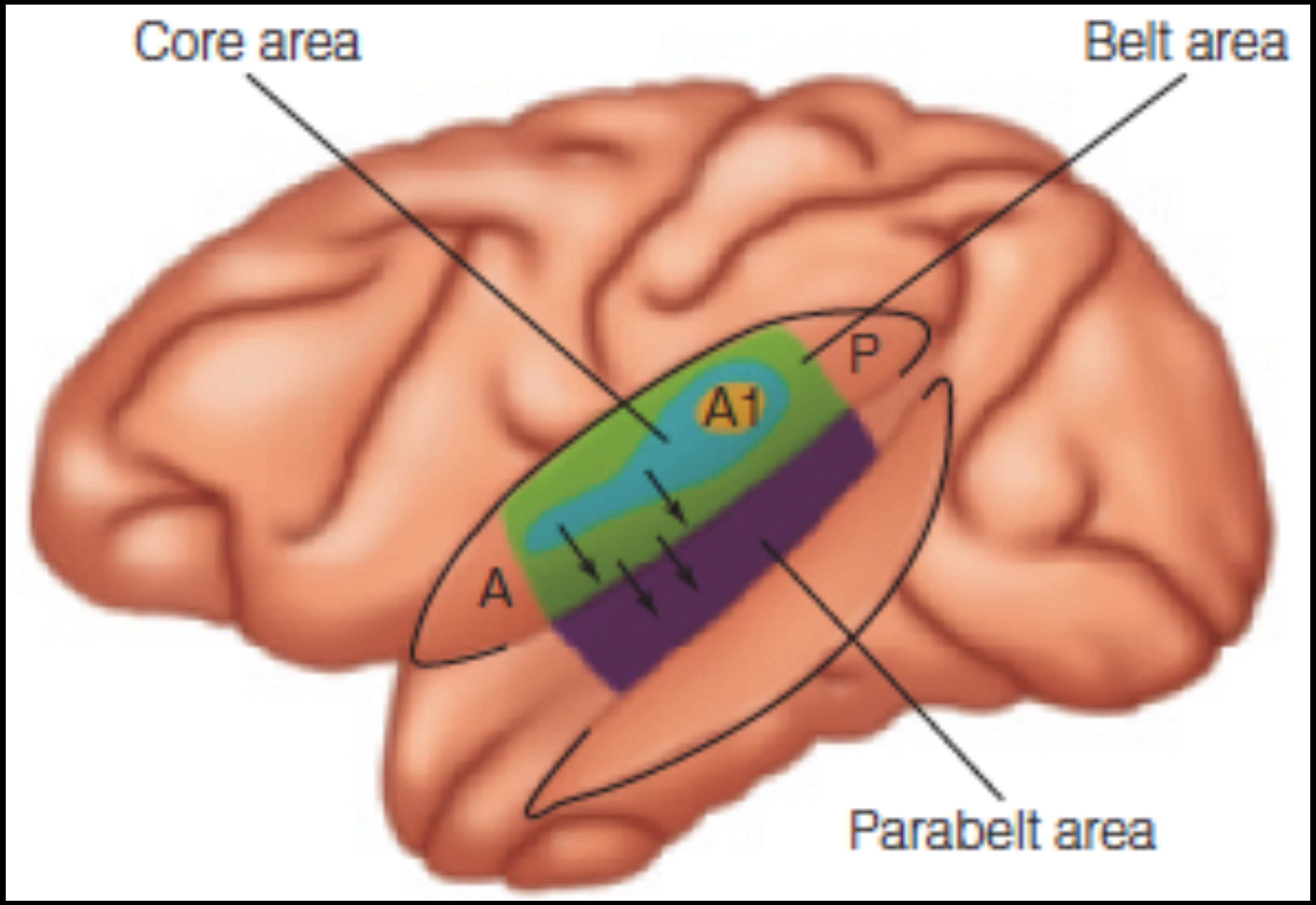
Variance explained across all of auditory cortex



Organization of human auditory cortex outside of primary areas?

Organization of human auditory cortex outside of primary areas?

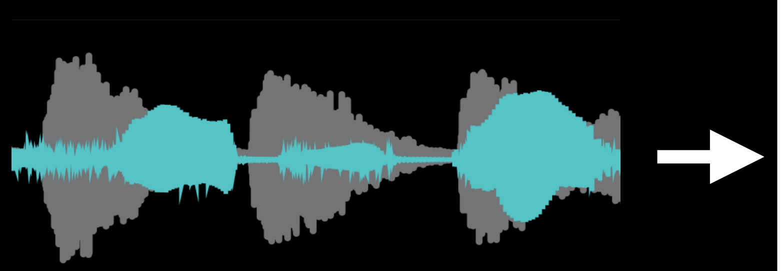
A proposal from macaque anatomy:
Tripartite hierarchical organization

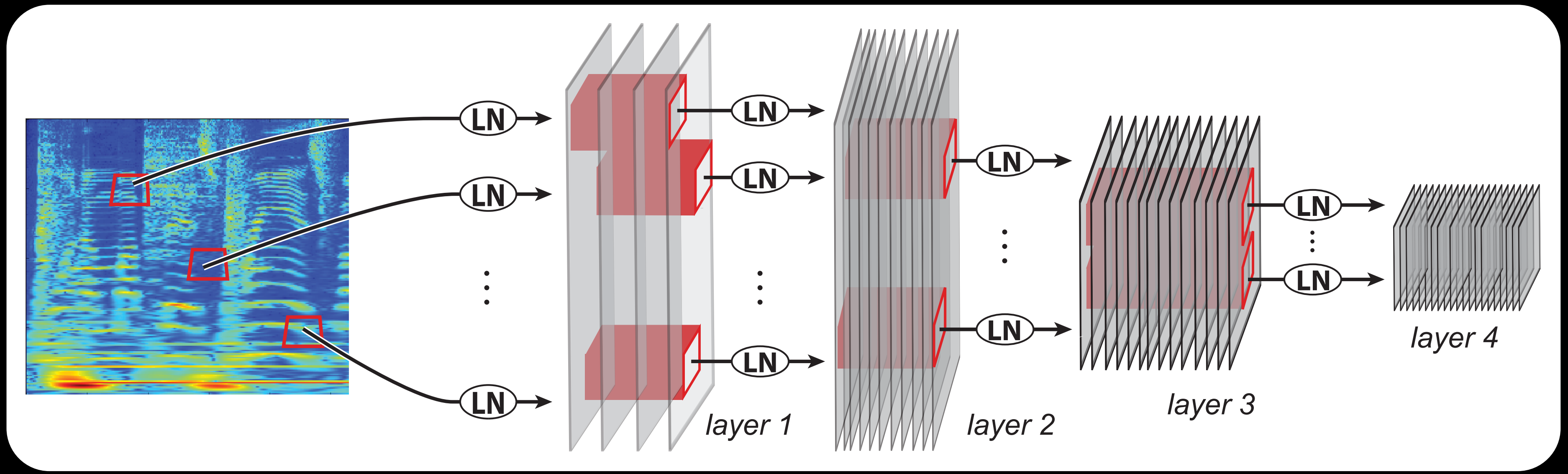


Tramo et al. (1999)

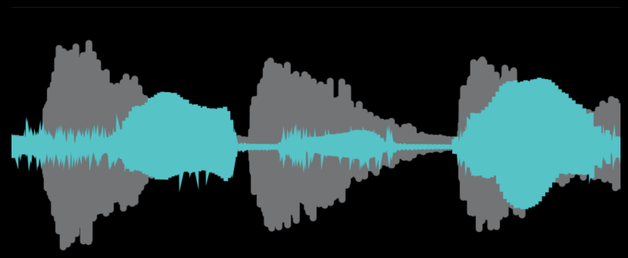
Evidence mostly anatomical

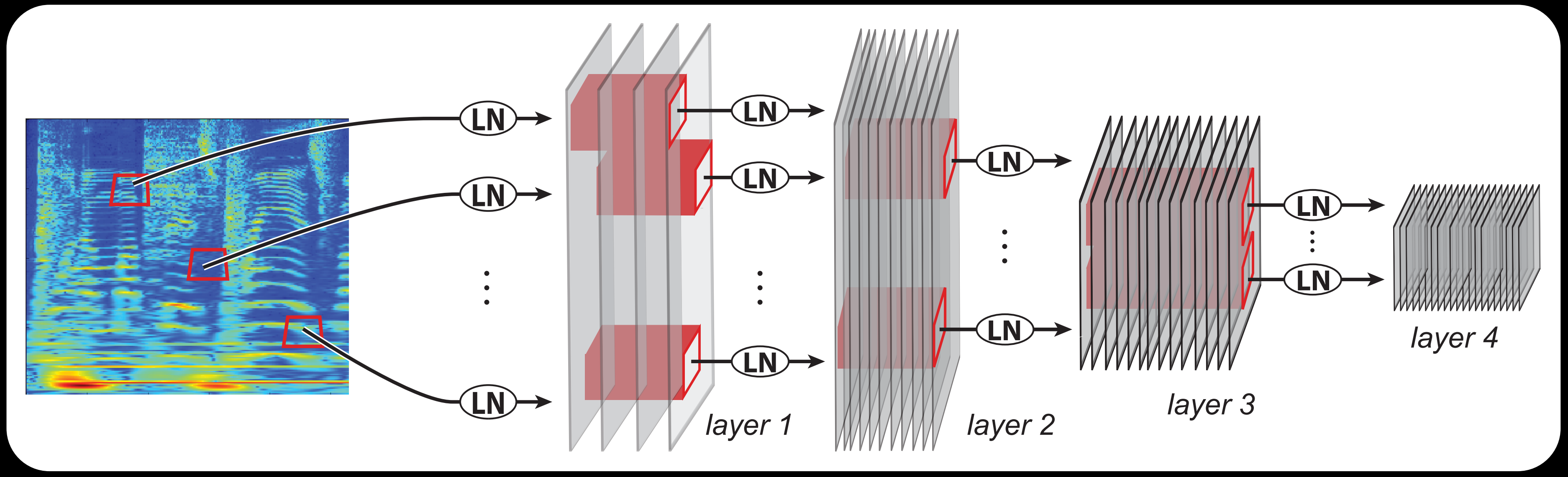
A measure of hierarchy?



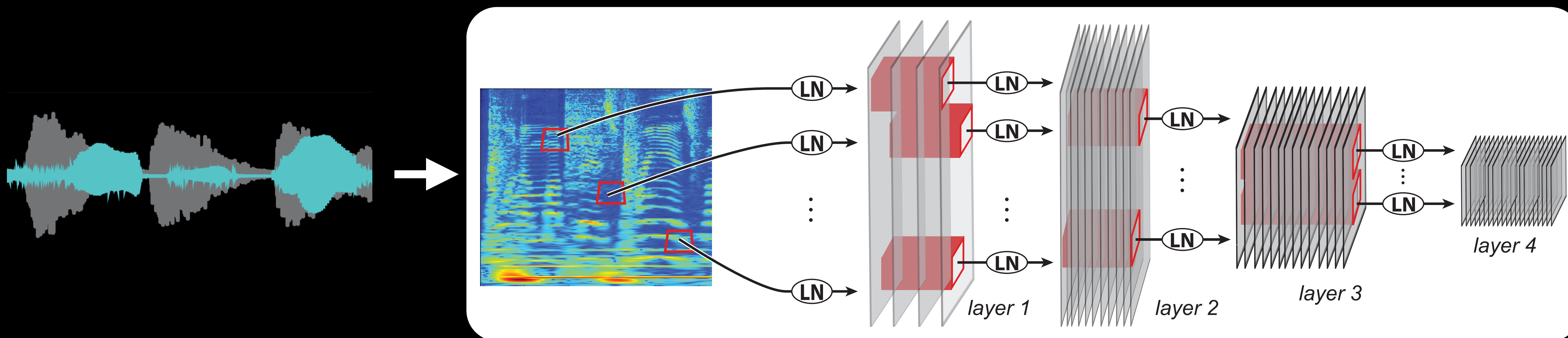

LN
LN
LN
LN
LN
LN
LN
LN
LN
LN
LN
layer 1
layer 2
layer 3
layer 4

CNN architecture:
Hierarchical and feedforward





CNN architecture:
Hierarchical and feedforward



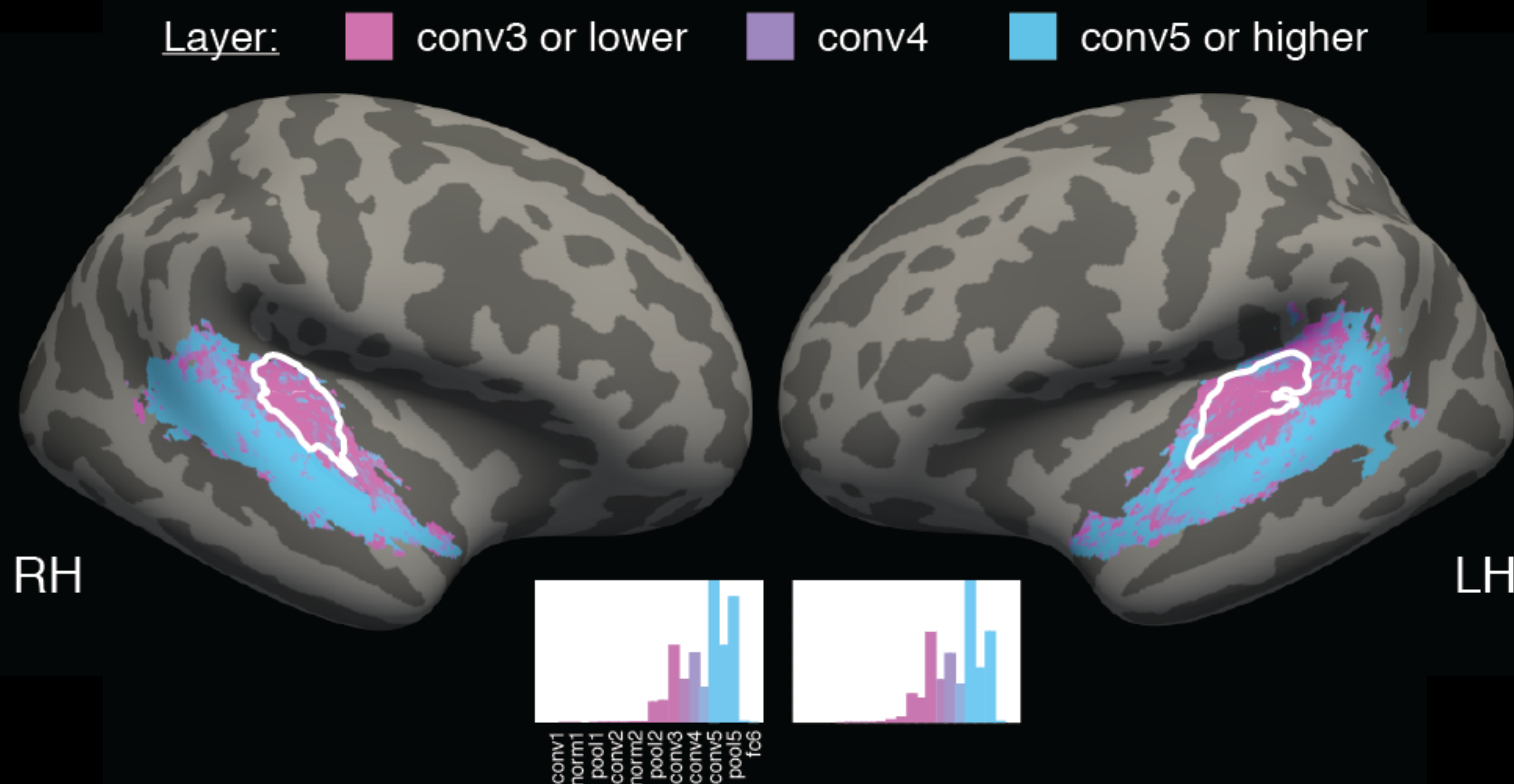
Which layer best predicts each voxel's response?
A measure of "complexity"

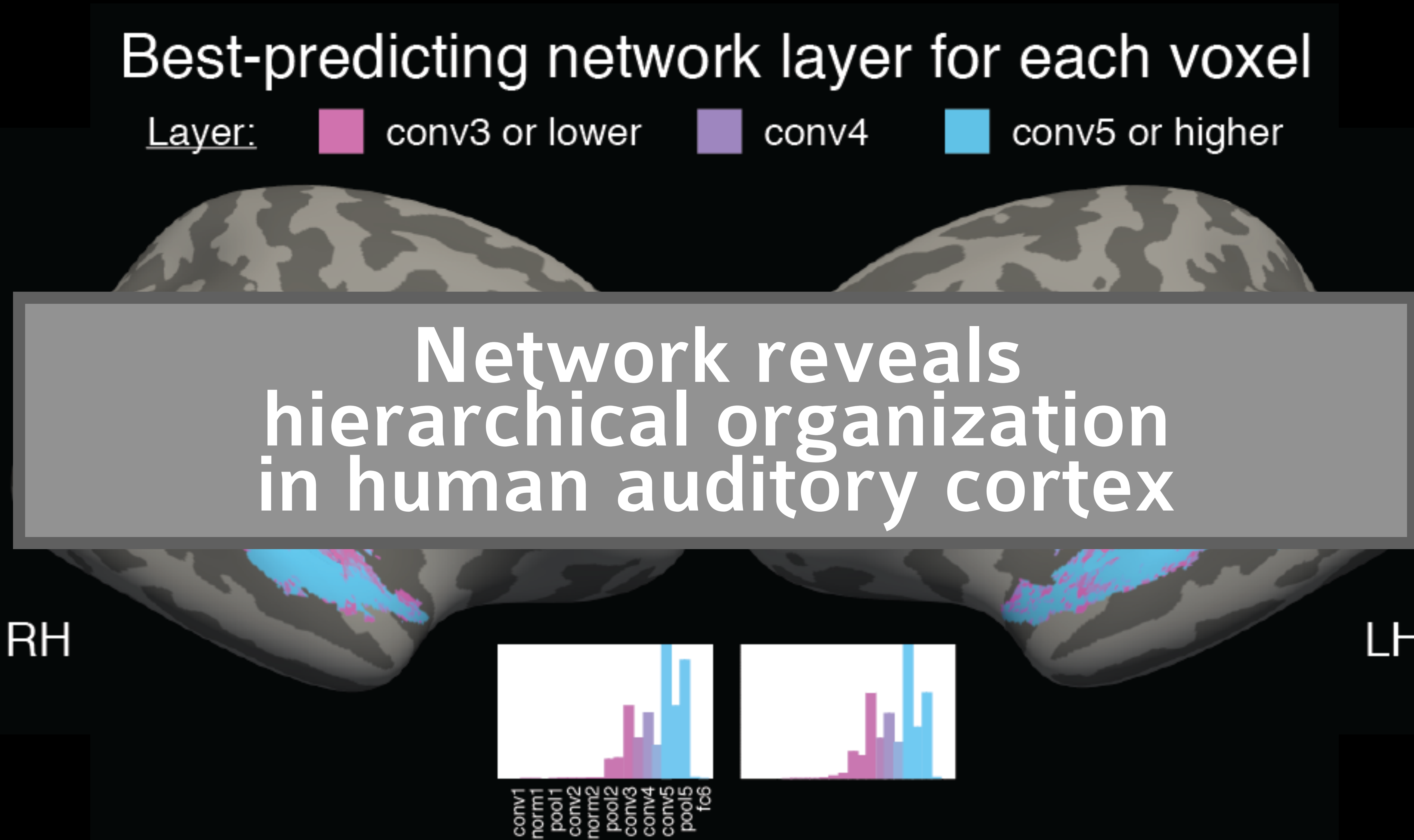
Best-predicting network layer for each voxel

Best-predicting network layer for each voxel

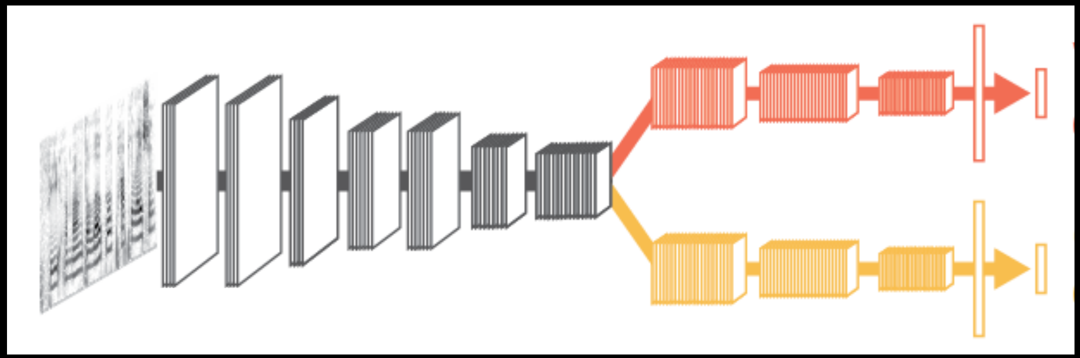
Layer: conv3 or lower · conv4 · conv5 or higher

Best-predicting network layer for each voxel

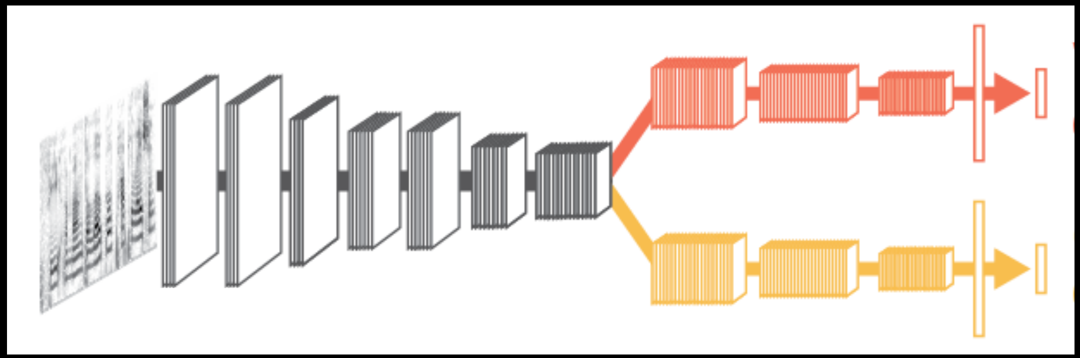




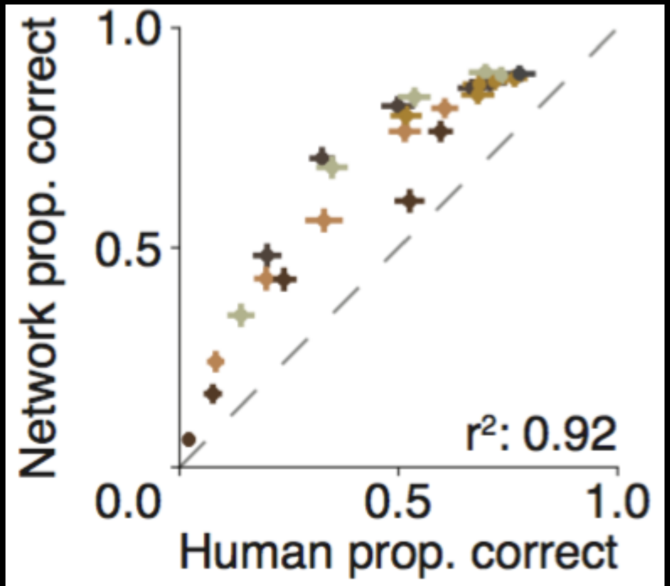
Network reveals hierarchical organization in human auditory cortex



Introduced multi-task networks as neural models

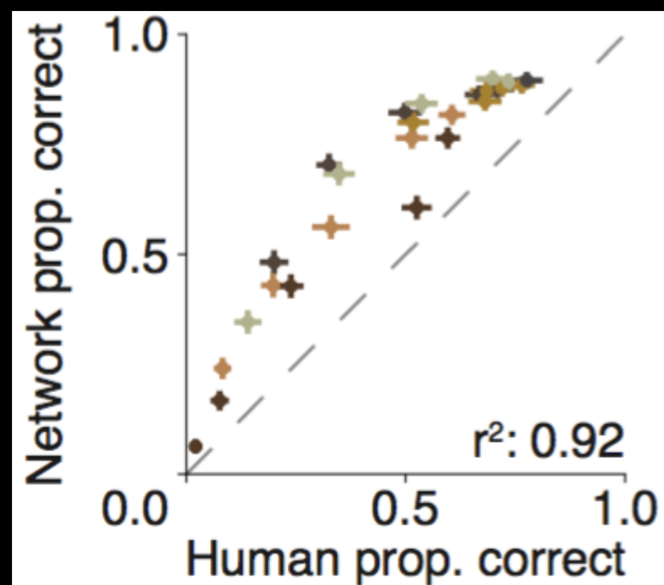


Introduced multi-task networks as neural models



Performs as well as humans, with similar pattern of errors

Introduced multi-task networks
as neural models



Performs as well as humans,
with similar pattern of errors

Reveals hierarchical organization
in human auditory cortex

Thanks.

Josh McDermott

Dan Yamins

Erica Shook

Department of Energy Computational Science Graduate Fellowship

DOE CSGF

McDermott lab for conversations + feedback.

Nancy Kanwisher + **Sam Norman-Haignere** for the fMRI data for CNN work.

Ariel Herbert-Voss for running behavioral subjects.

Atsushi Takahashi for help designing MR protocols.

Steve Shannon for MR support.

Satra Ghosh + **the Openmind team** for support with computational resources.