CS375 / Psych 249:
Large-Scale Neural Network Models for Neuroscience

Lecture 8: Other Sensory Domains (Audition, Somatosensation, etc)

2026.02.09

Daniel Yamins

Departments of Computer Science and of Psychology
Stanford Neuroscience and Artificial Intelligence Laboratory
Wu Tsai Neurosciences Institute
Stanford University

Date	Session
01/05	Introduction to NeuroAI
01/07	Visual Systems Neuroscience Background
01/12	DNN Models of the Visual System -- Part 1
01/14	Model-Brain Mapping Methods
01/19	[NO CLASS-MLK DAY]
01/21	DNN Models of the Visual System -- Part 2
01/26	Unsupervised Learning and the Brain
01/27	
01/28	**Cliona O'Doherty (Stanford):** Modeling Infant Development
02/02	Recurrent Models of Vision
02/04	**Andreas Tolias (Stanford):** The Enigma Project
02/09	Auditory and Somatosensory Models
02/11	**Andrew Saxe (UCL):** Decision Making
02/16	[NO CLASS-PRESIDENT'S DAY] (BBScore Evening Session)
02/18	Navigation, Memory, and the MEC
02/23	**Aran Nayebi (CMU):** Models of Agents
02/25	The Motor System
02/26	
03/02	**Scott Lindermann (Stanford):** Dynamical Systems in the Brain
03/04	**Greta Tuckute (Harvard):** Language, LLMs, and the Brain
03/09	**Tony Zador (Cold Spring Harbor):** Models of Brain Evolution
03/11	Functional Organization and Learning Rules in the Brain
03/16	Project Presentations
03/22	

Vision, as a core "worked example"

Date	Session
01/05	Introduction to NeuroAI
01/07	Visual Systems Neuroscience Background
01/12	DNN Models of the Visual System -- Part 1
01/14	Model-Brain Mapping Methods
01/19	[NO CLASS-MLK DAY]
01/21	DNN Models of the Visual System -- Part 2
01/26	Unsupervised Learning and the Brain
01/27	
01/28	**Cliona O'Doherty (Stanford):** Modeling Infant Development
02/02	Recurrent Models of Vision
02/04	**Andreas Tolias (Stanford):** The Enigma Project
02/09	Auditory and Somatosensory Models
02/11	**Andrew Saxe (UCL):** Decision Making
02/16	[NO CLASS-PRESIDENT'S DAY] (BBScore Evening Session)
02/18	Navigation, Memory, and the MEC
02/23	**Aran Nayebi (CMU):** Models of Agents
02/25	The Motor System
02/26	
03/02	**Scott Lindermann (Stanford):** Dynamical Systems in the Brain
03/04	**Greta Tuckute (Harvard):** Language, LLMs, and the Brain
03/09	**Tony Zador (Cold Spring Harbor):** Models of Brain Evolution
03/11	Functional Organization and Learning Rules in the Brain
03/16	Project Presentations
03/22	

Vision, as a core "worked example"

Going beyond vision

Audition

When objects in the world vibrate, they transmit acoustic energy through surrounding medium in the form of a wave.

Audition

When objects in the world vibrate, they transmit acoustic energy through surrounding medium in the form of a wave.

The ears measure this sound energy and transmit it to the brain.

Audition

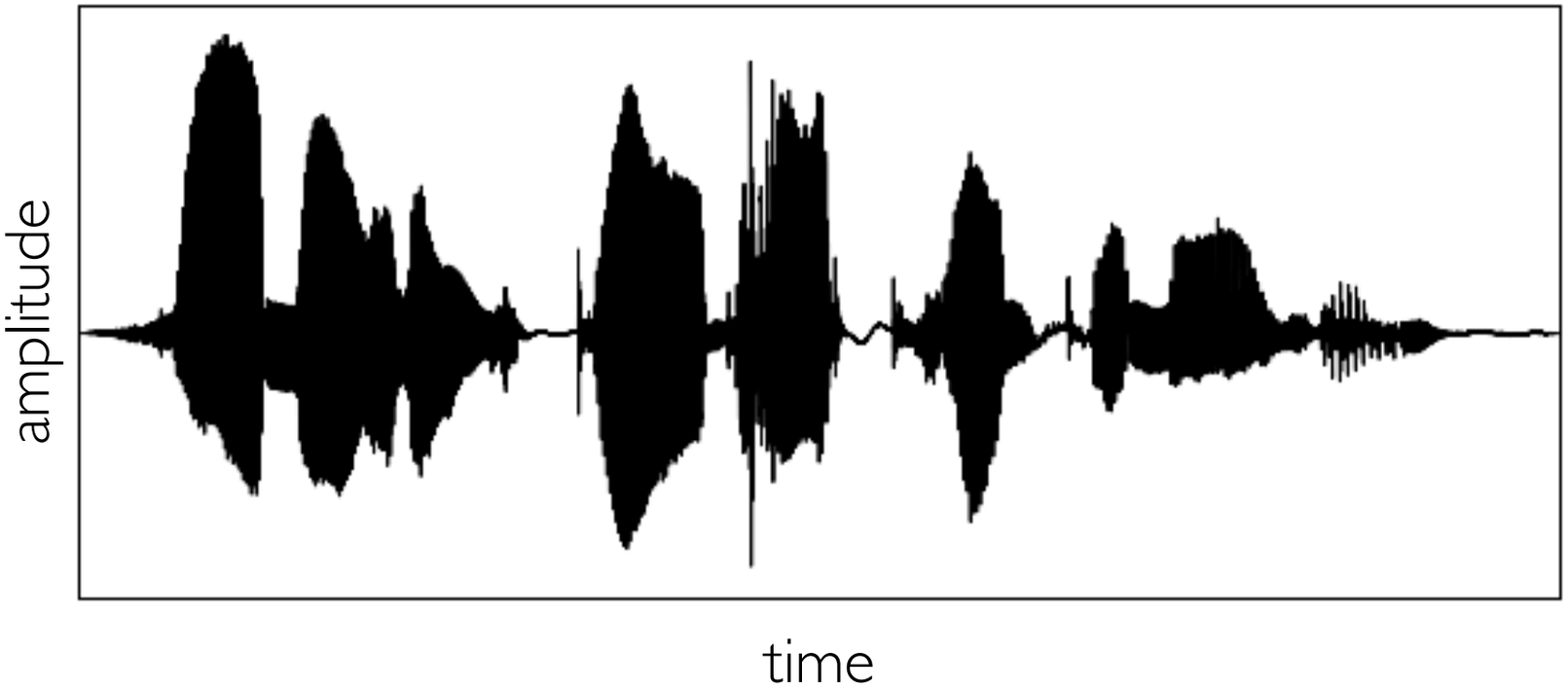
When objects in the world vibrate, they transmit acoustic energy through surrounding medium in the form of a wave.

The ears measure this sound energy and transmit it to the brain.

The task of the brain is to interpret this signal, and use it to figure out what is out there in the world.

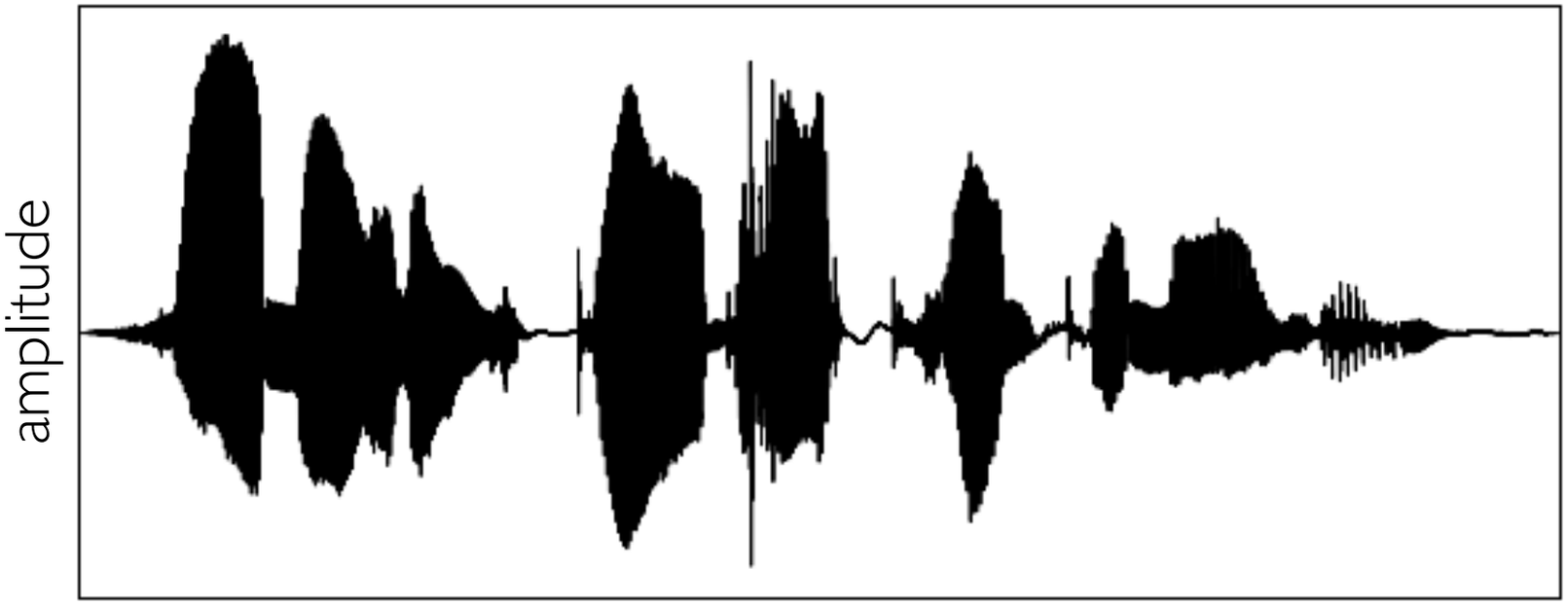
Problem: Entity Extraction

Understanding complex, noisy data streams is a critical part of cognition.



Problem: Entity Extraction

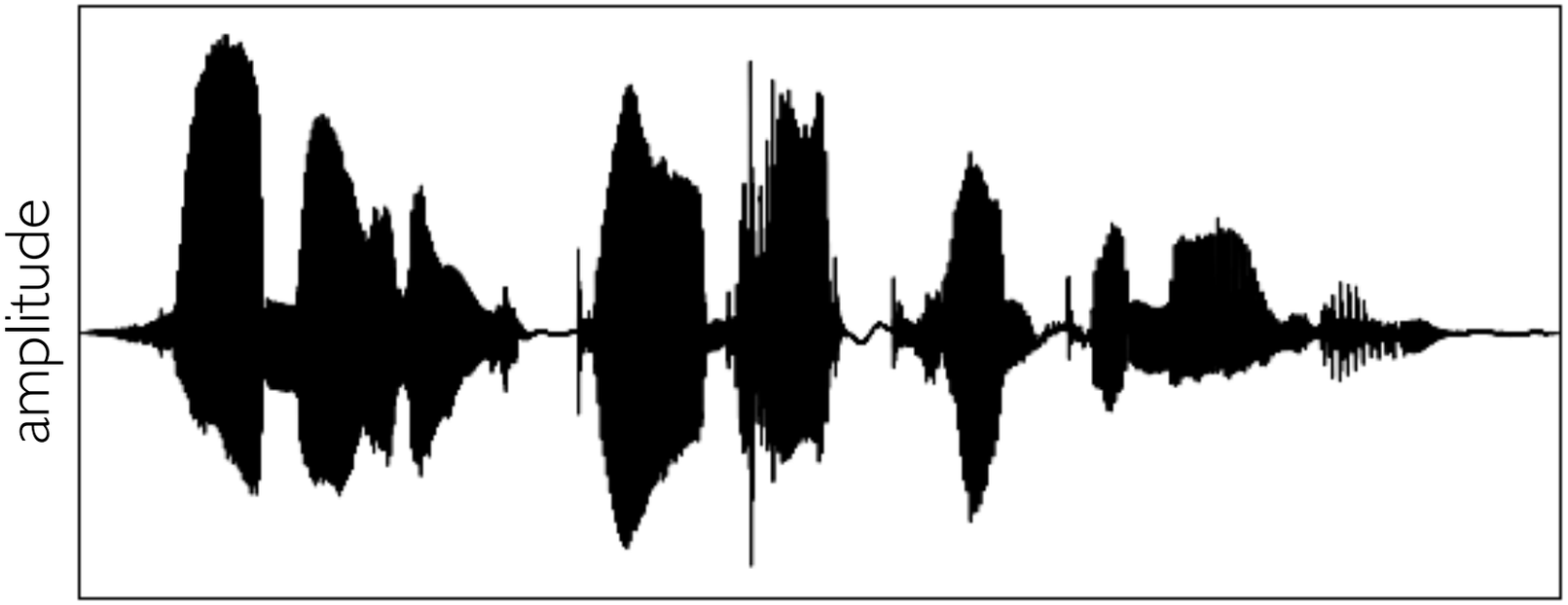
Understanding complex, noisy data streams is critical part of cognition.



"Hannah is good at compromising."

Problem: Entity Extraction

Understanding complex, noisy data streams is critical part of cognition.



"Hannah is good at compromising."

variation sources: speaker identity

background noise

reverberation

…

Audition

0 dB Threshold of hearing
10 dB Normal breathing

30 dB Soft whisper

50 dB Quiet conversation

70 dB Busy traffic

90 dB Shouting

110 dB <--- prolonged exposure can cause hearing loss
120 dB Propeller plane at takeoff

140 dB Jet at takeoff, threshold of pain

160 dB Instant perforation of eardrum, 10^{16} times something at 0 dB.

Audition

Common sounds …

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
Ringtone
Microwave
Dog barking

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
Pouring water
Pop song
Water boiling

Guitar
Coughing
Crumpling paper
Siren
Splashing water
Computer speech
Alarm clock
Walking with heels
Vacuum
Wind
Boy speaking
Chair rolling
Rock song
Door knocking

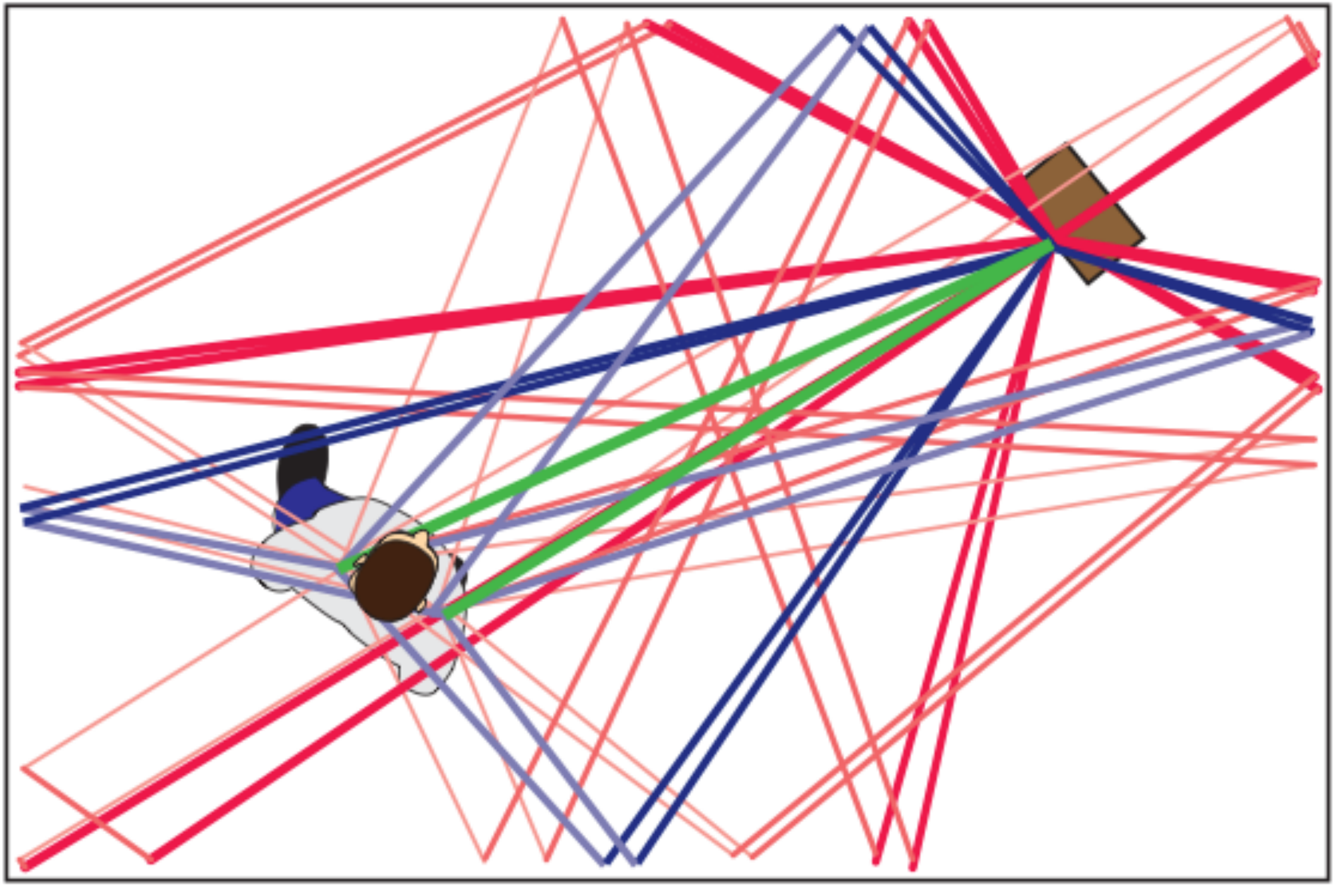
-
-
-

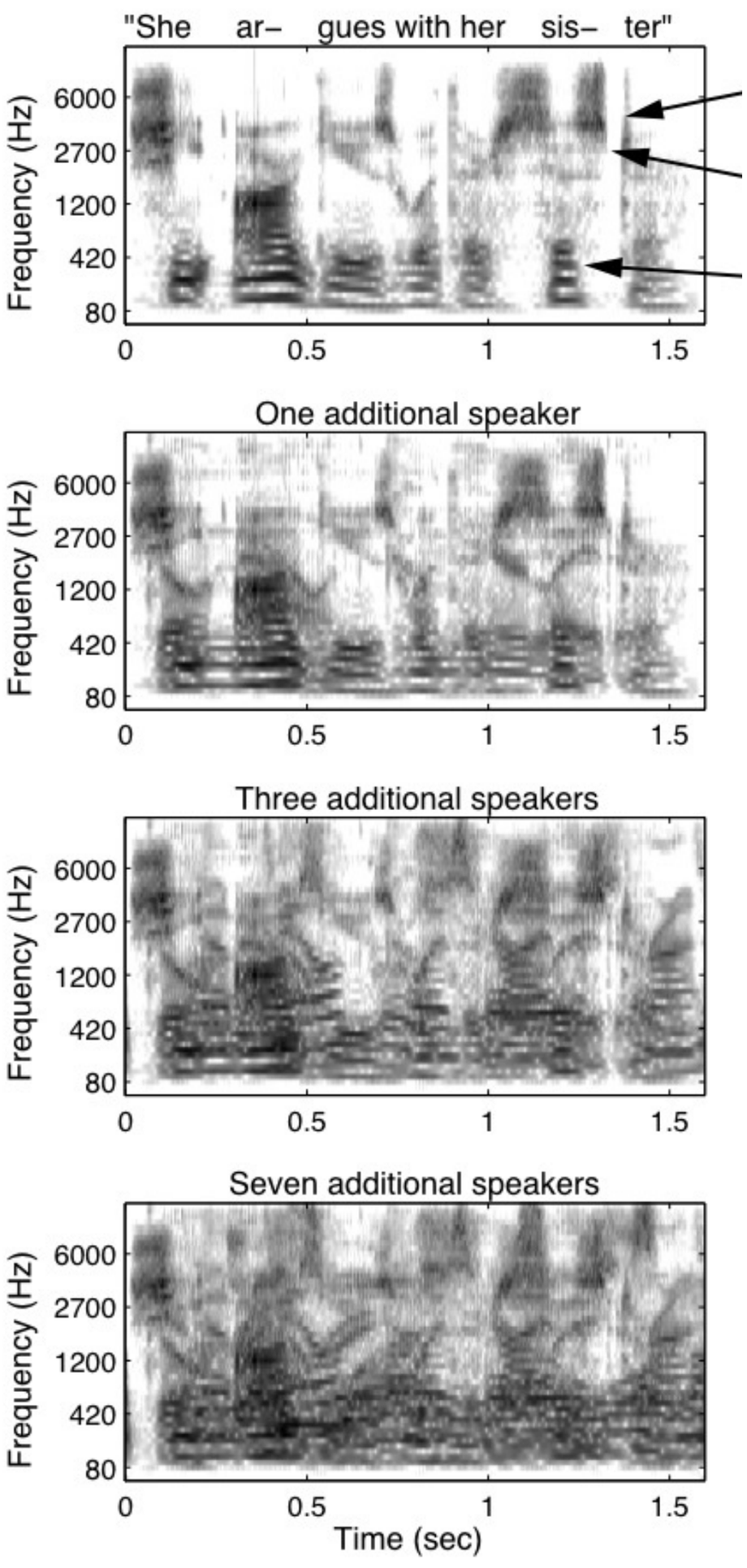
*Sam Norman-Haignere, Nancy Kanwisher, and Josh McDermott

The Cocktail Party Problem

Real-world settings often involve concurrent sounds.

What happens to sound in a room:

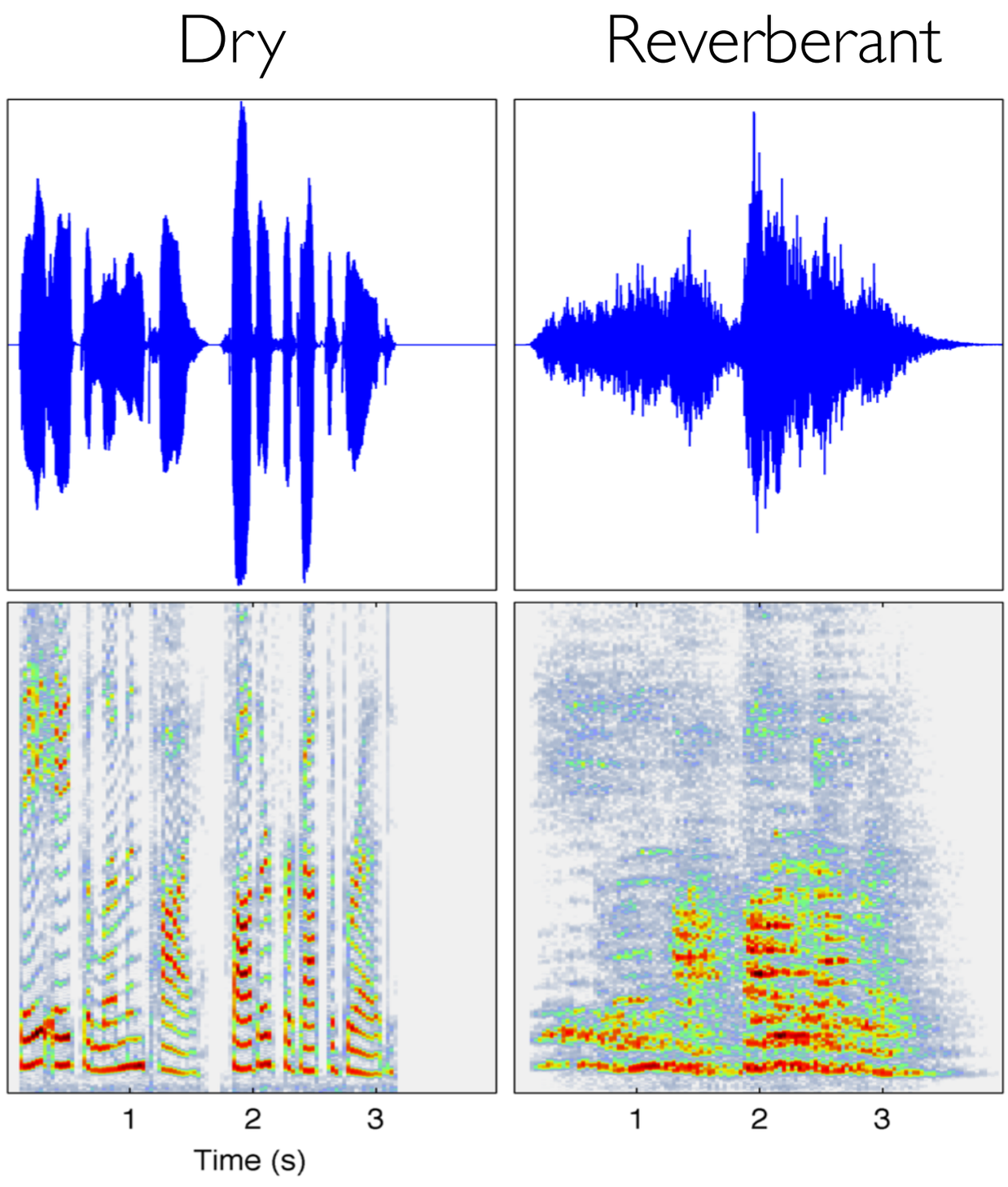




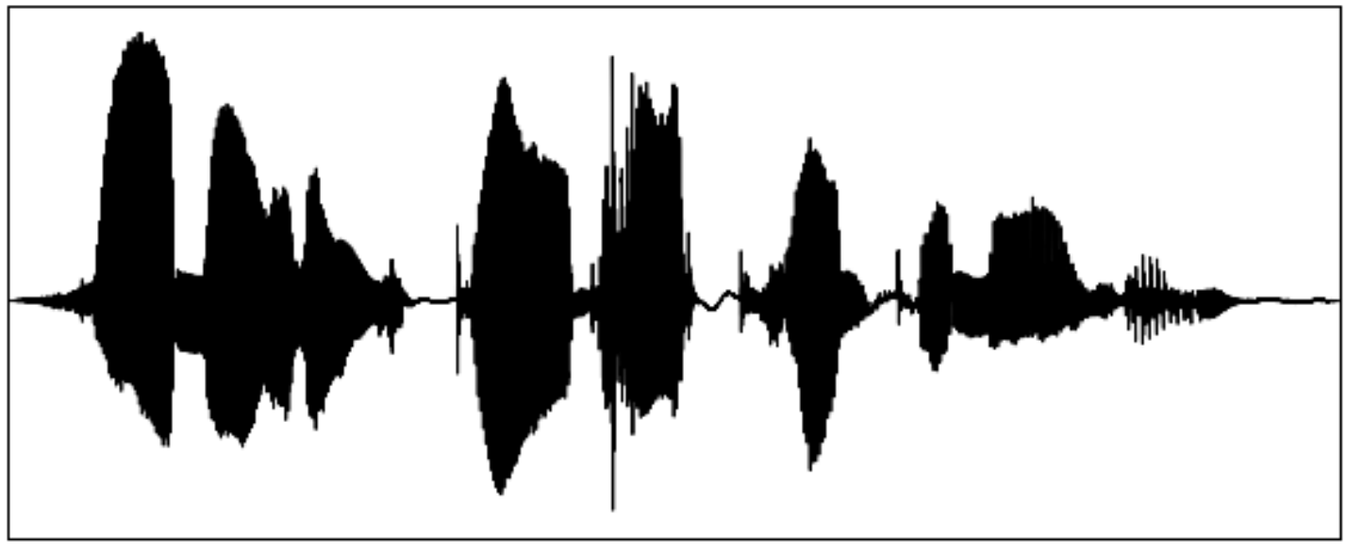
- Presence of other speakers obscures much structure of target utterance, but speech remains intelligible.

- Speech recognition algorithms circa ~2013 fell apart in such circumstances.

Human speech recognition is remarkably invariant:



Problem: Entity Extraction



visual cortex

auditory cortex

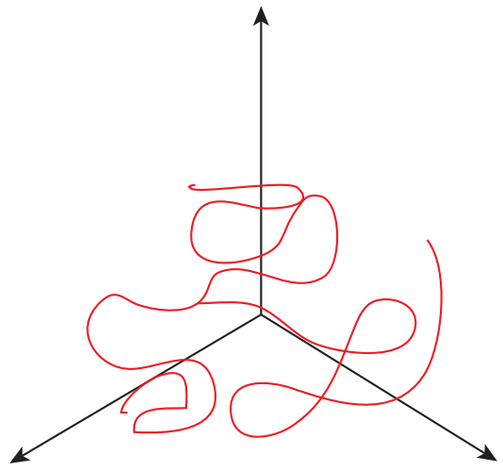
"Mercedes behind Lamborghini, on a field in front of mountains."

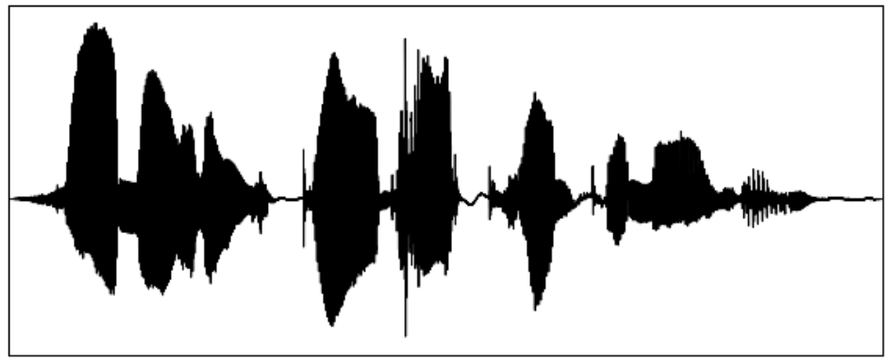
"Hannah is good at compromising"

Problem: Entity Extraction

visual cortex

"Mercedes behind Lamborghini, on a field in front of mountains."

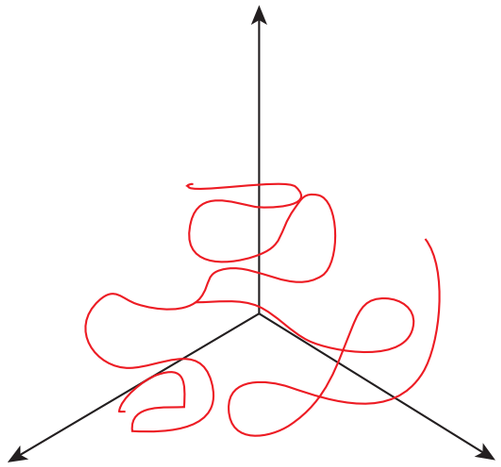


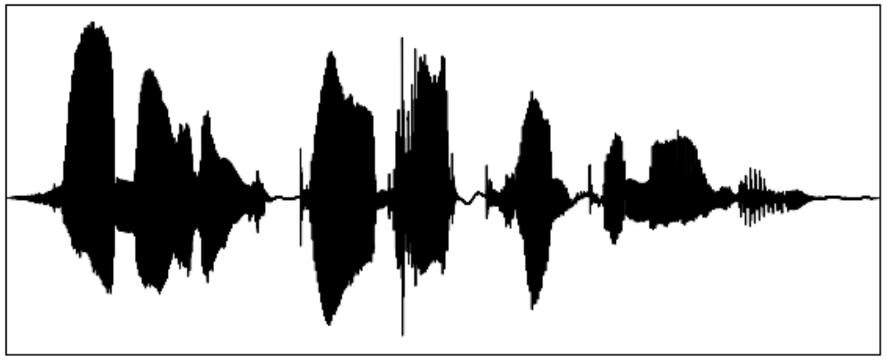


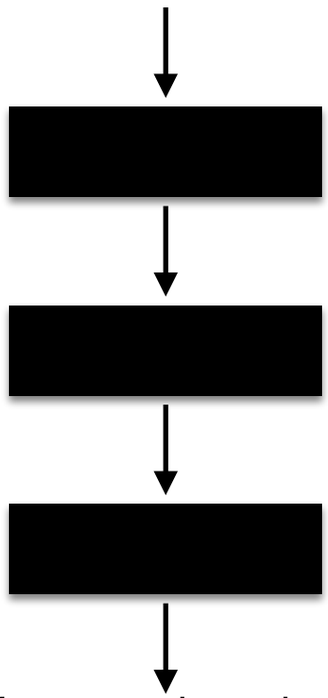
auditory cortex

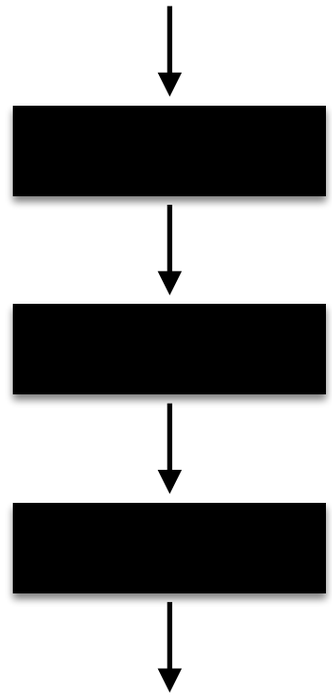
"Hannah is good at compromising"

Problem: Entity Extraction

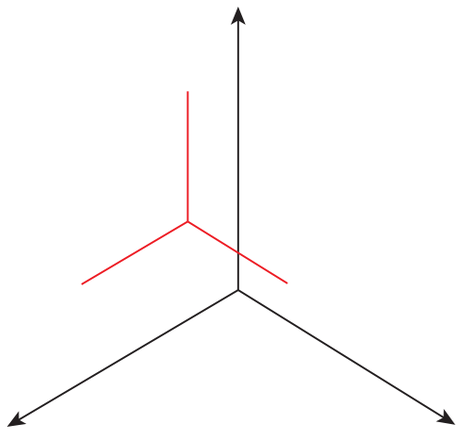






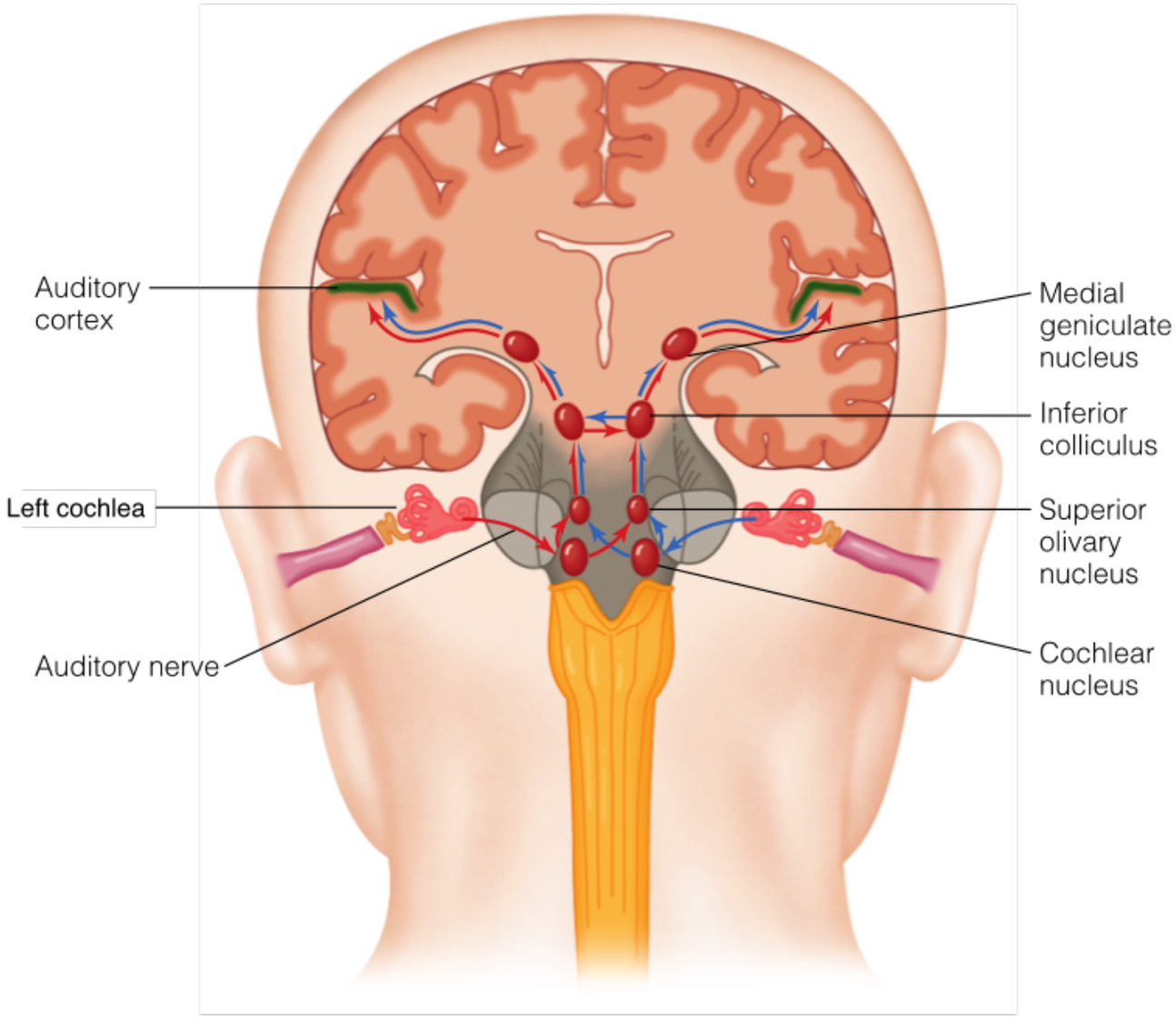


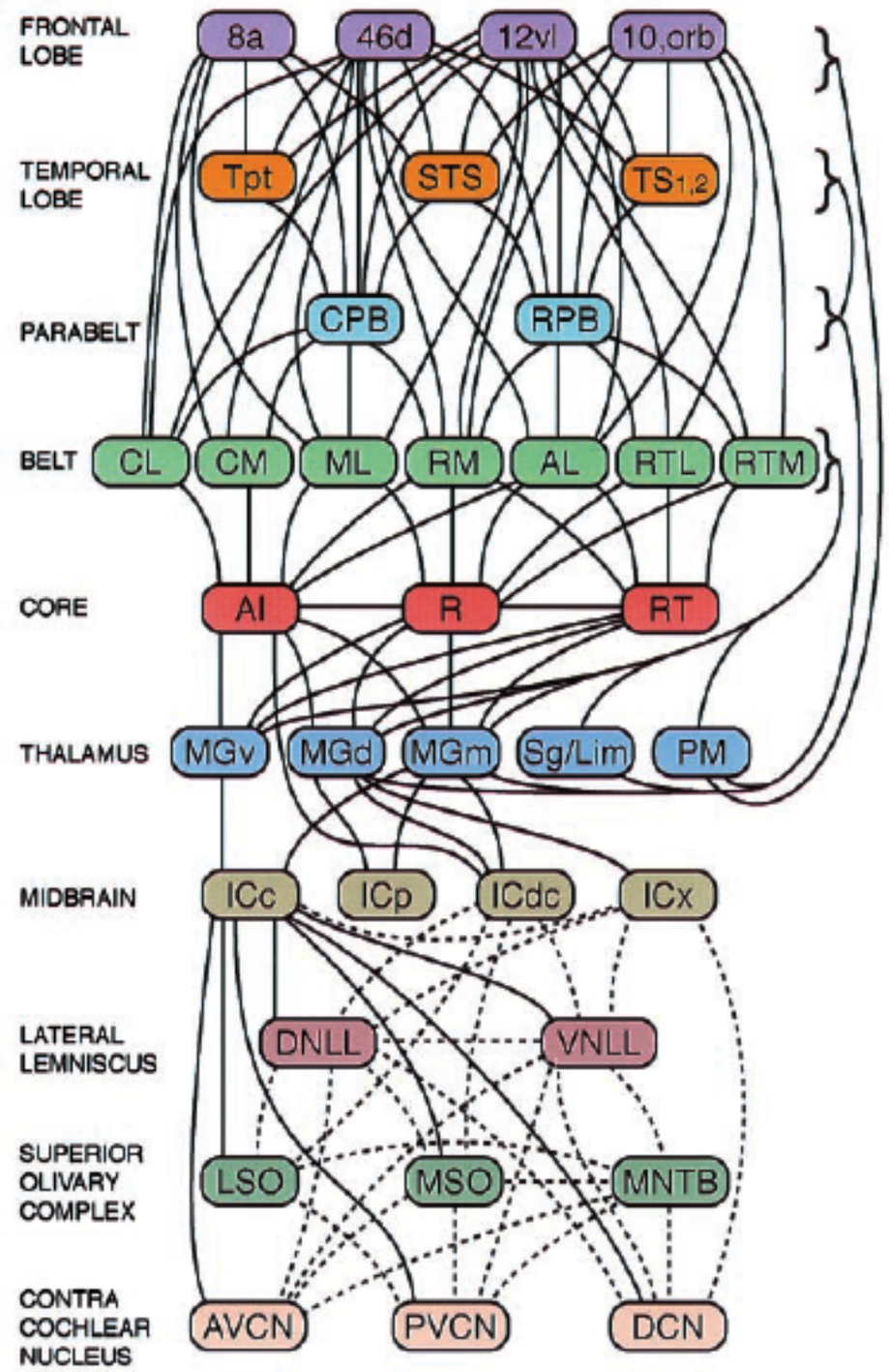
"Mercedes behind Lamborghini, on a field in front of mountains."



"Hannah is good at compromising"

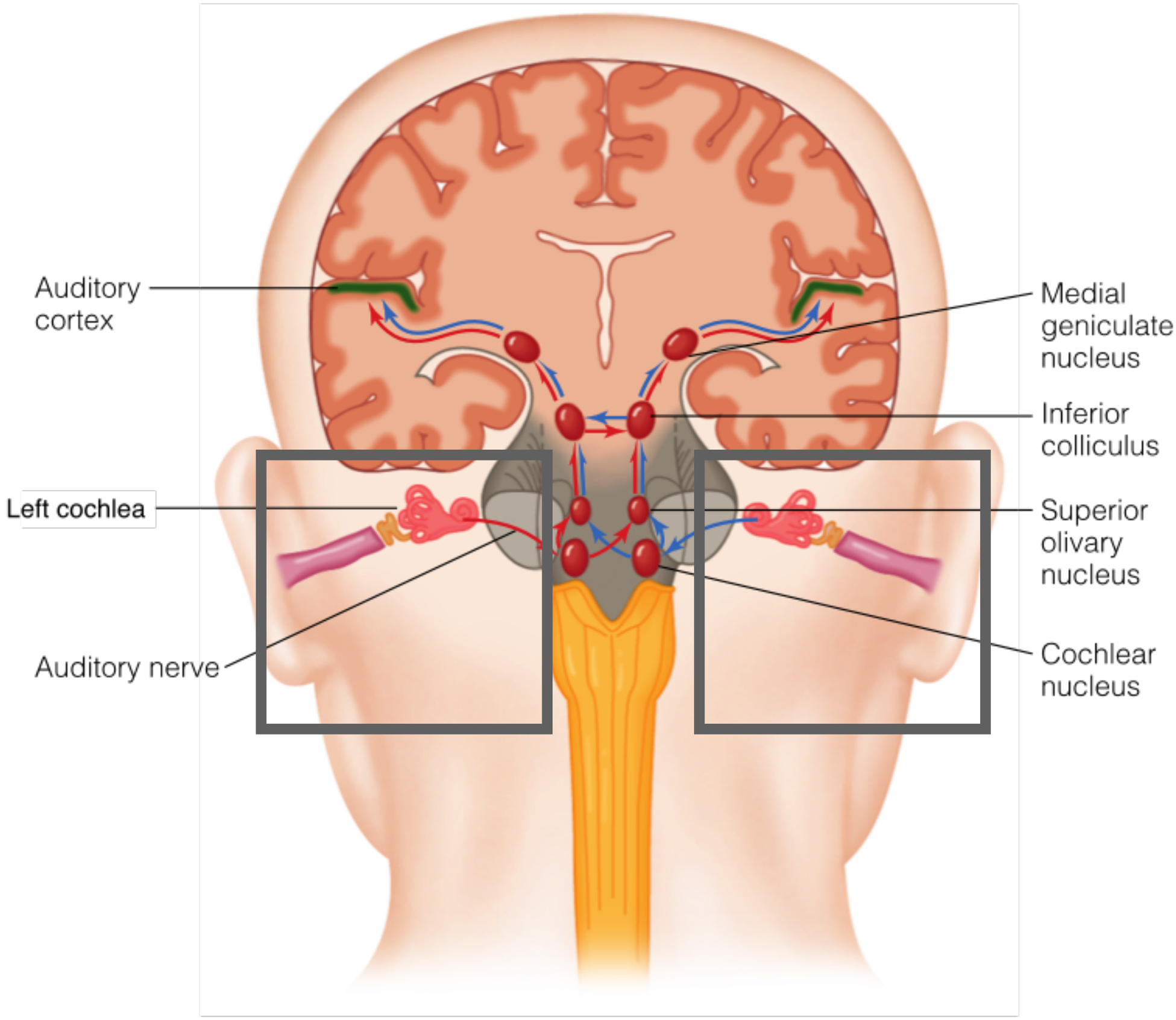
The Auditory System



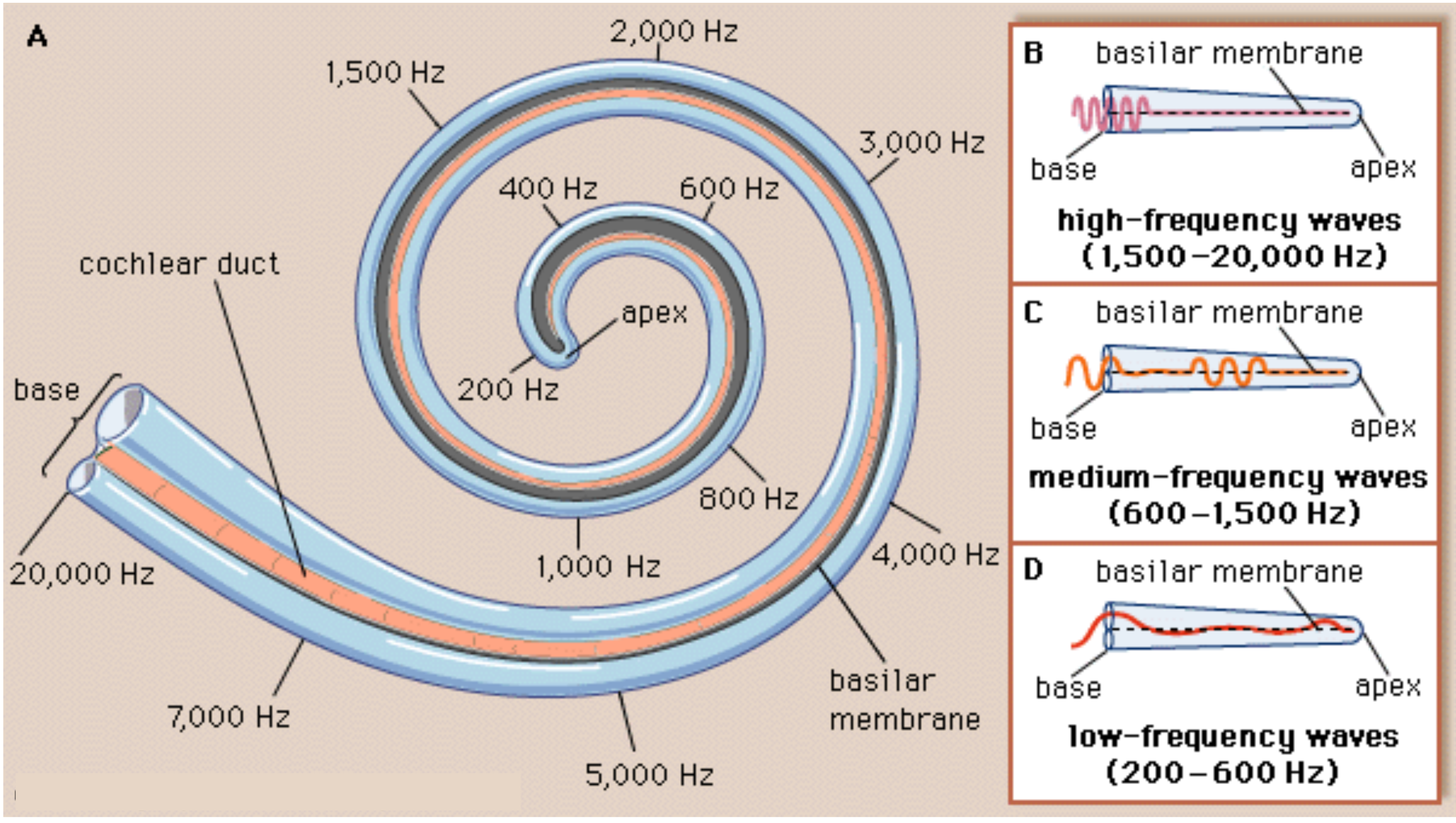


Kaas & Hackett 2000

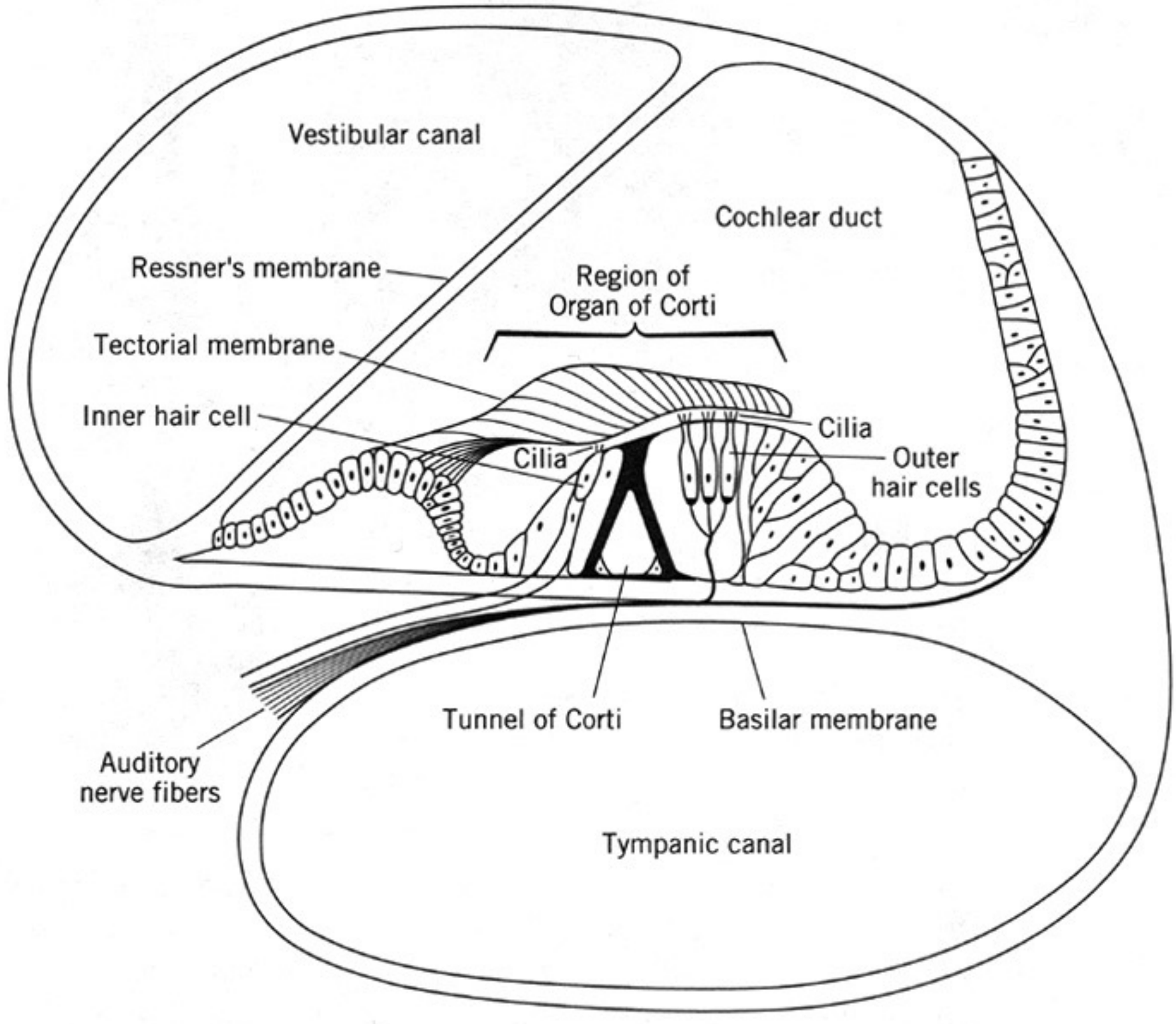
The Auditory System



The Cochlea

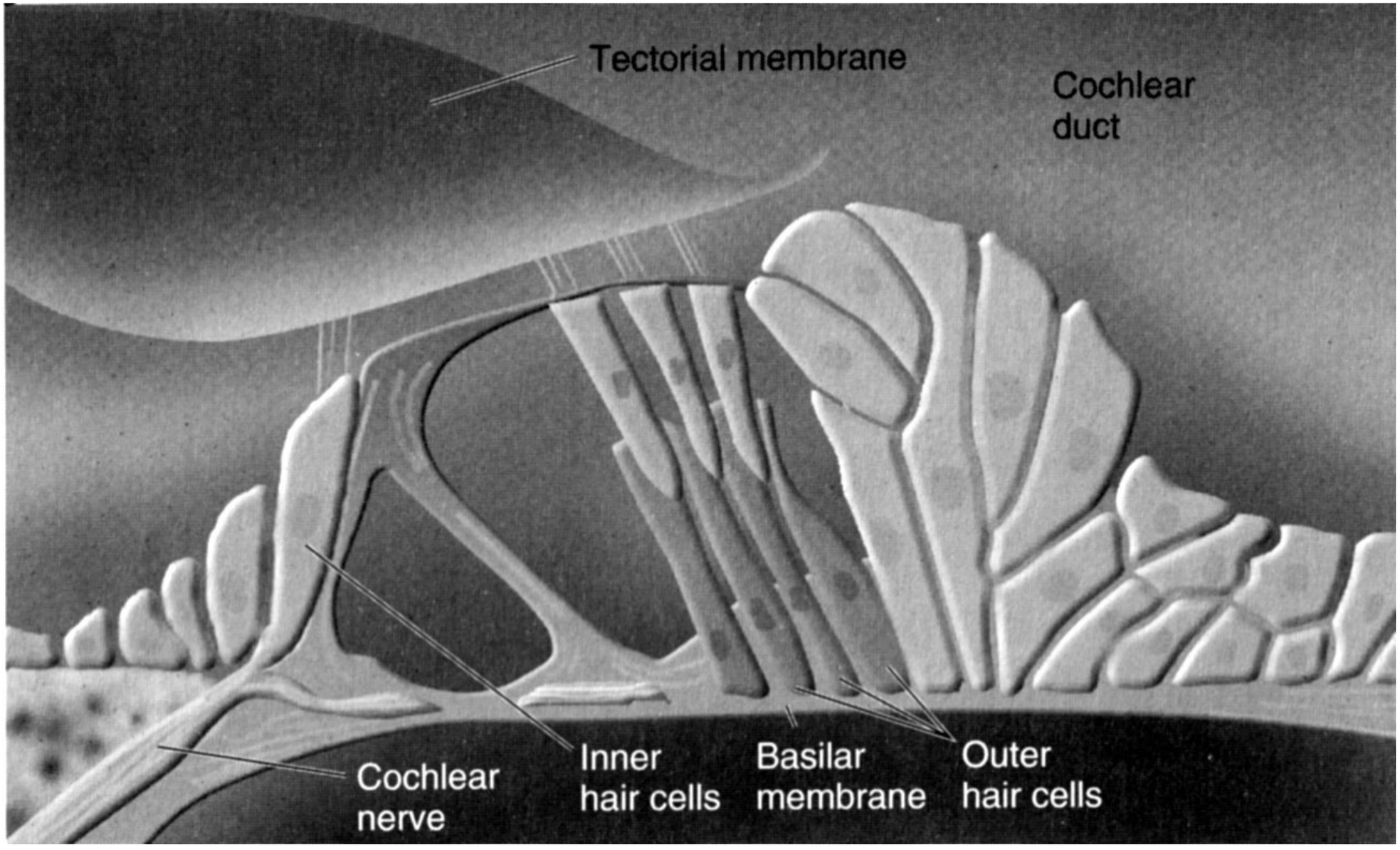

A
2,000 Hz
1,500 Hz
3,000 Hz
400 Hz
600 Hz
cochlear duct
apex
base
200 Hz
800 Hz
1,000 Hz
20,000 Hz
4,000 Hz
7,000 Hz
basilar membrane
5,000 Hz
B
basilar membrane
base
apex
high-frequency waves (1,500–20,000 Hz)
C
basilar membrane
base
apex
medium-frequency waves (600–1,500 Hz)
D
basilar membrane
base
apex
low-frequency waves (200–600 Hz)

The Cochlea


Vestibular canal
Cochlear duct
Ressner's membrane
Region of
Organ of Corti
Tectorial membrane
Inner hair cell
Cilia
Cilia
Outer
hair cells
Tunnel of Corti
Basilar membrane
Auditory
nerve fibers
Tympanic canal

The Cochlea

Movement of the basilar membrane causes the hair cells to move against the tectorial membrane, which causes the cilia to bend.

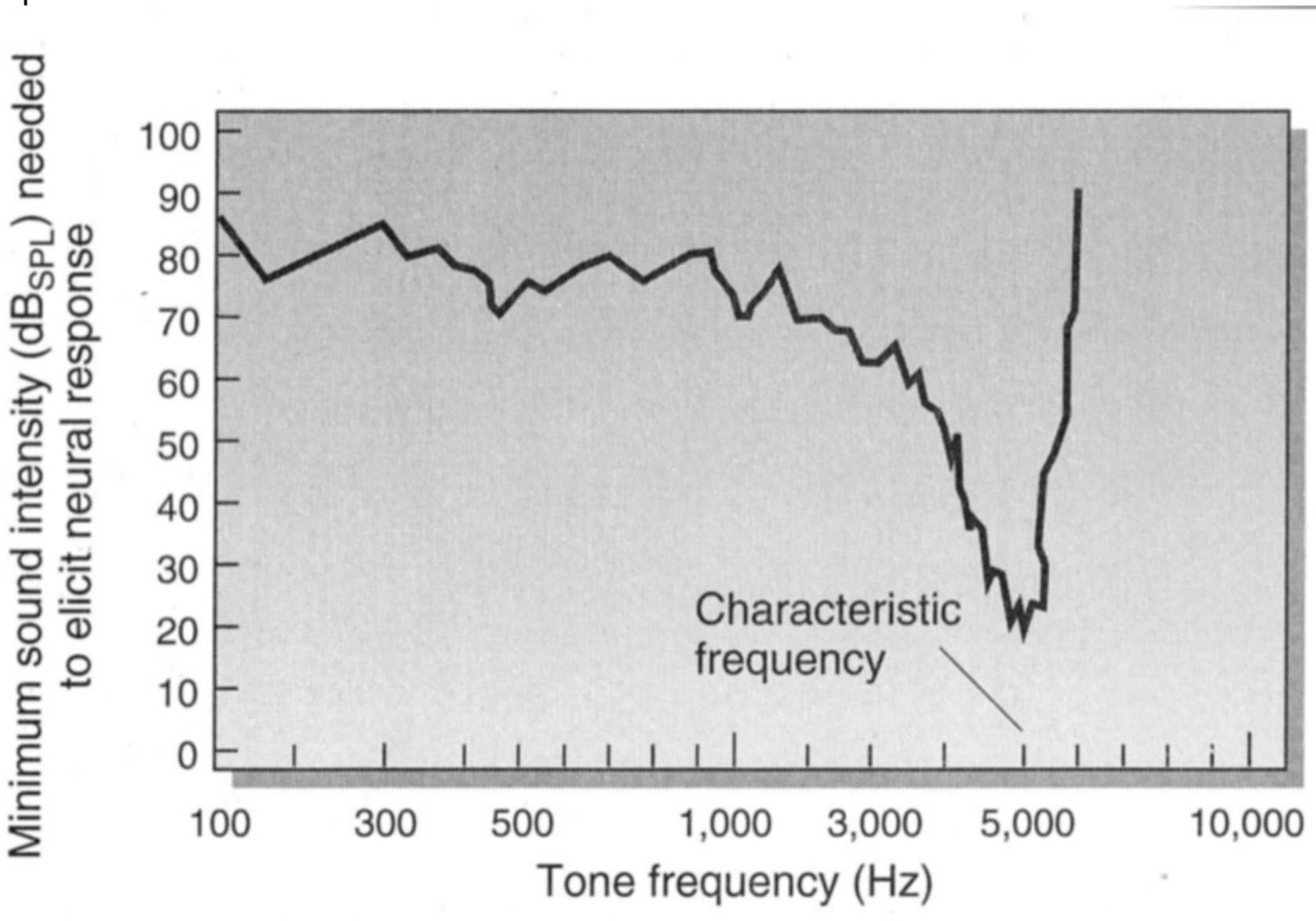


When the cilia bend, the hair cells release neurotransmitter onto synapses with auditory nerve fibers that send signals to the brain.

The Cochlea

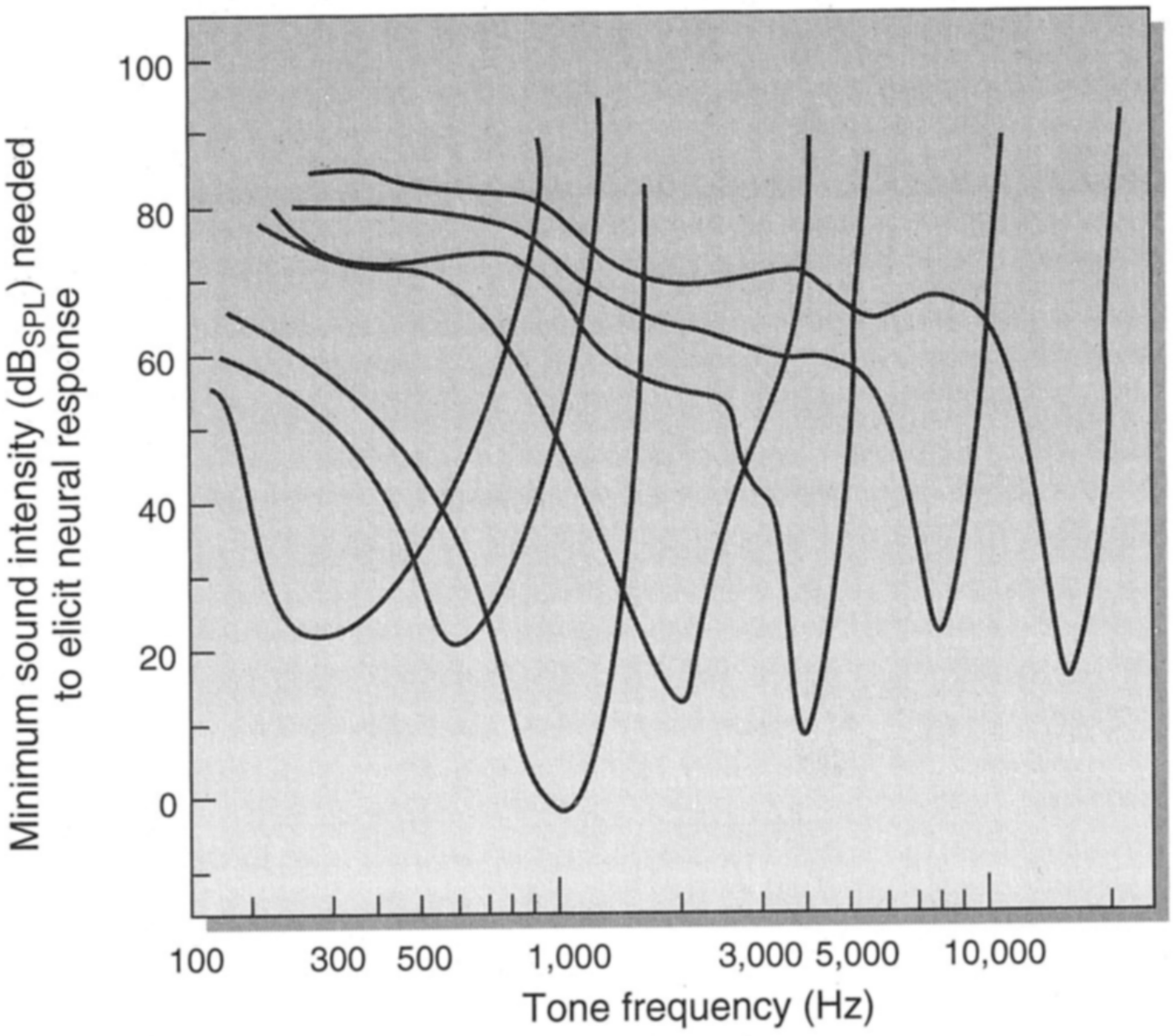
But because only part of the basilar membrane moves for a given frequency of sound, each hair cell and auditory nerve fiber signal only particular frequencies of sound.

One example:



The Cochlea

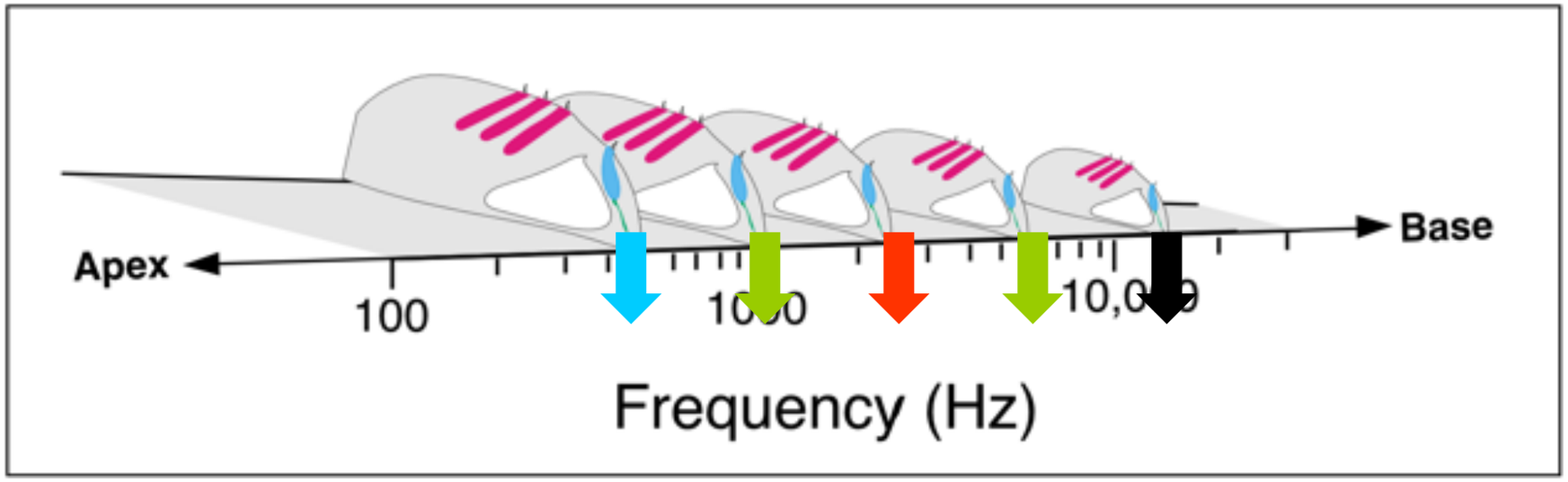
Different auditory nerve fibers encode different frequencies:



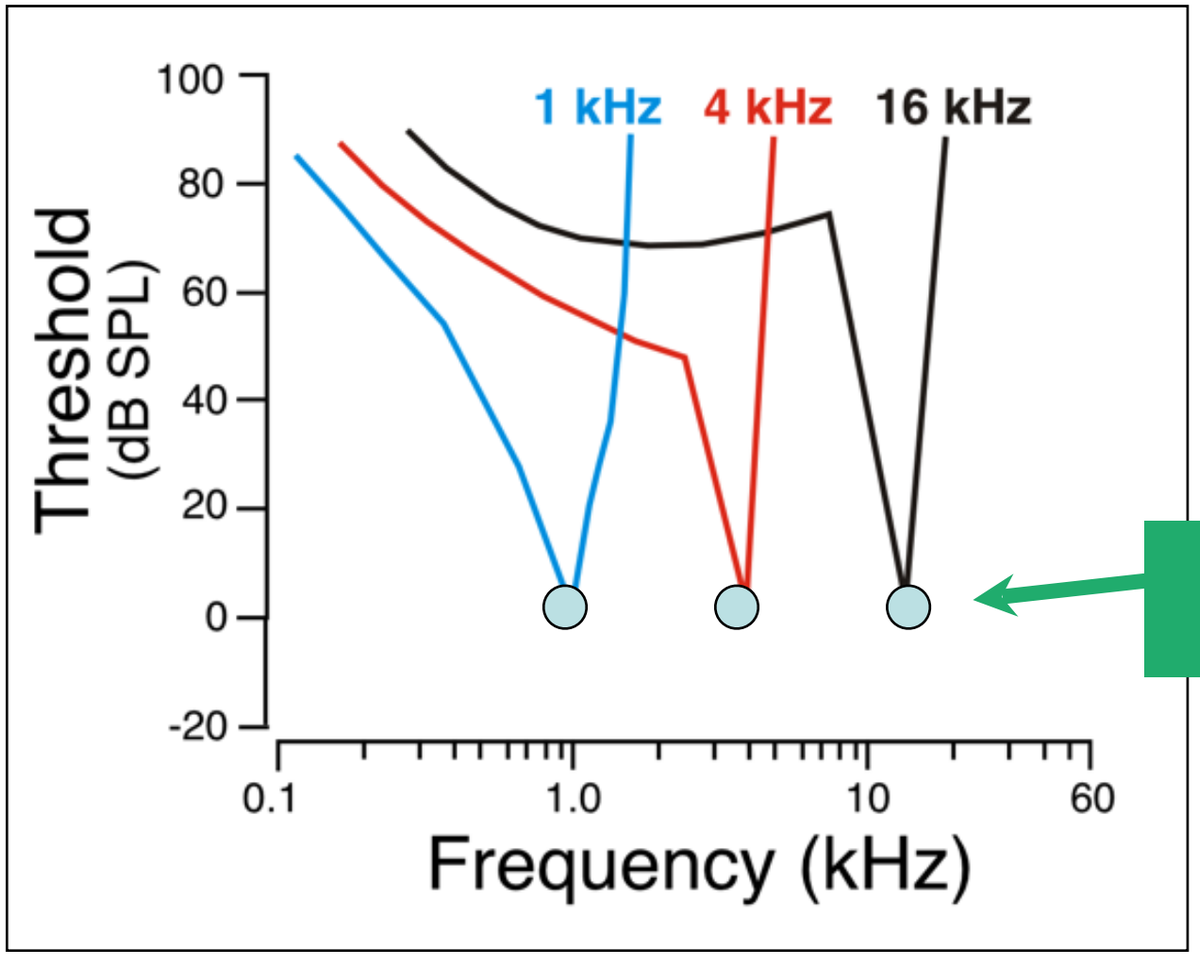
The cochlea is doing a frequency analysis of the sound signal!

The Cochlea

Auditory nerve: Frequency map (tonotopy)



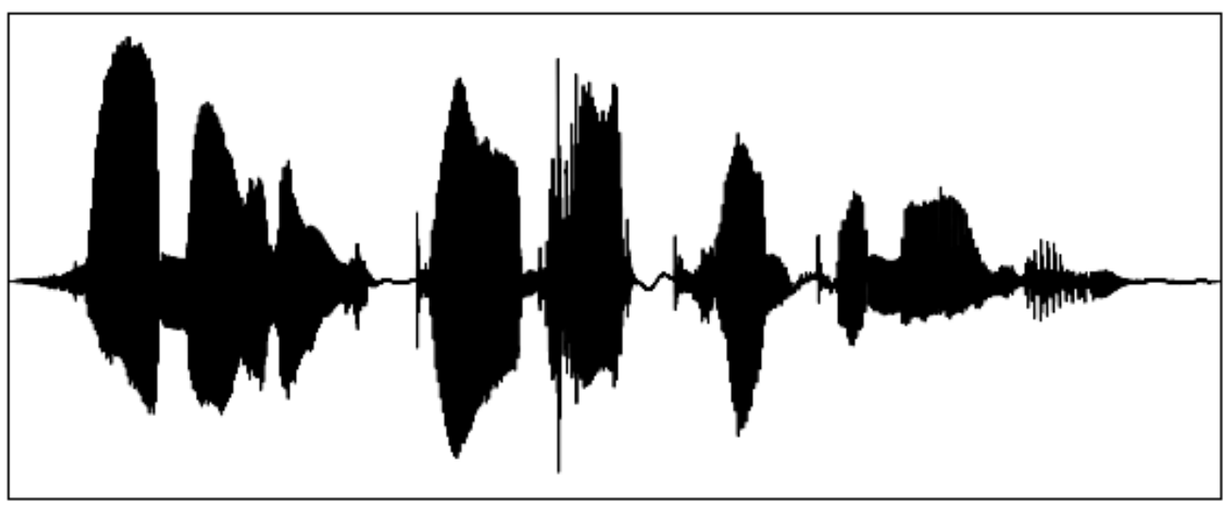
Frequency Selectivity



Characteristic Frequency (CF)

The Cochlea

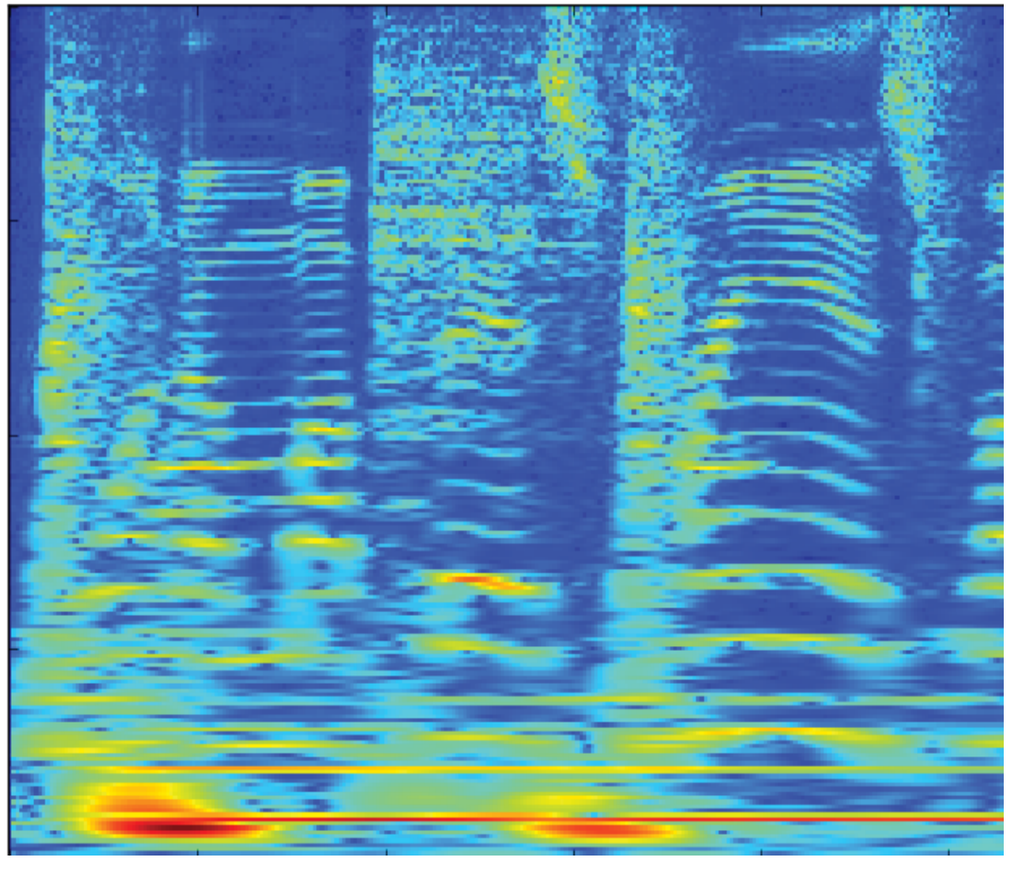
Waveform representation



Time →

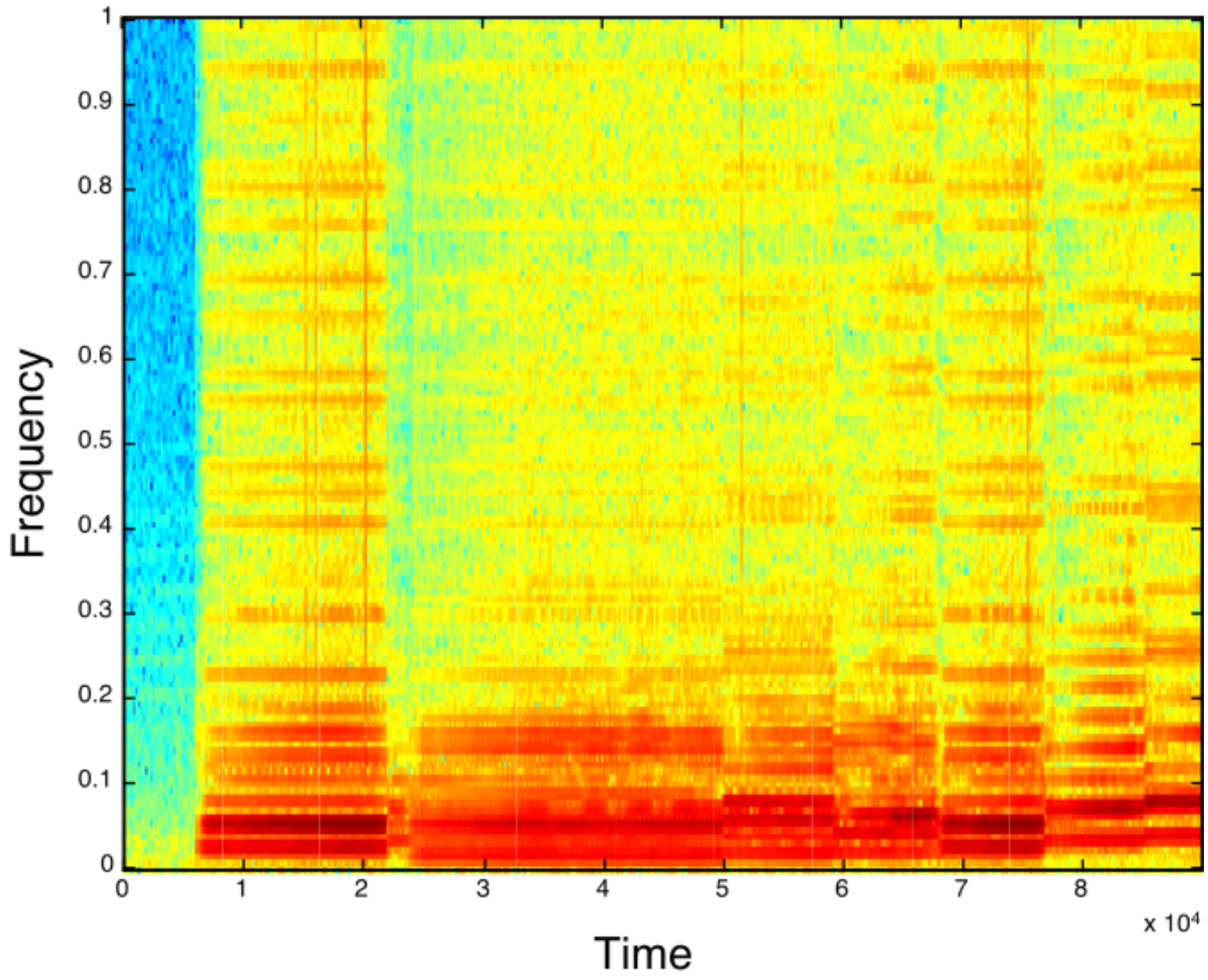
Coarse model of the cochlea

Cochleagram representation



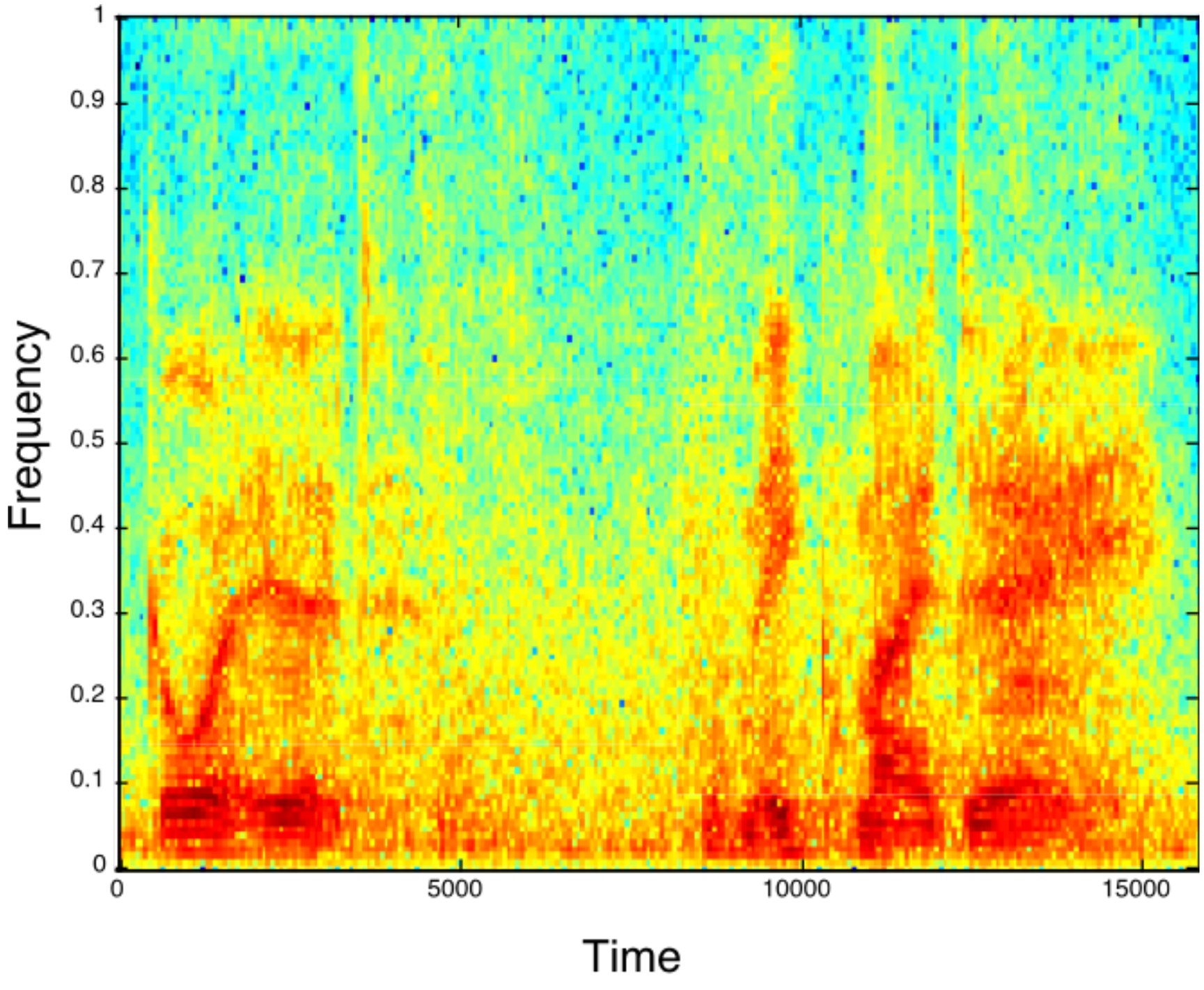
Amplitude spectrum as a function of time is called a spectrogram.

Oboe melody



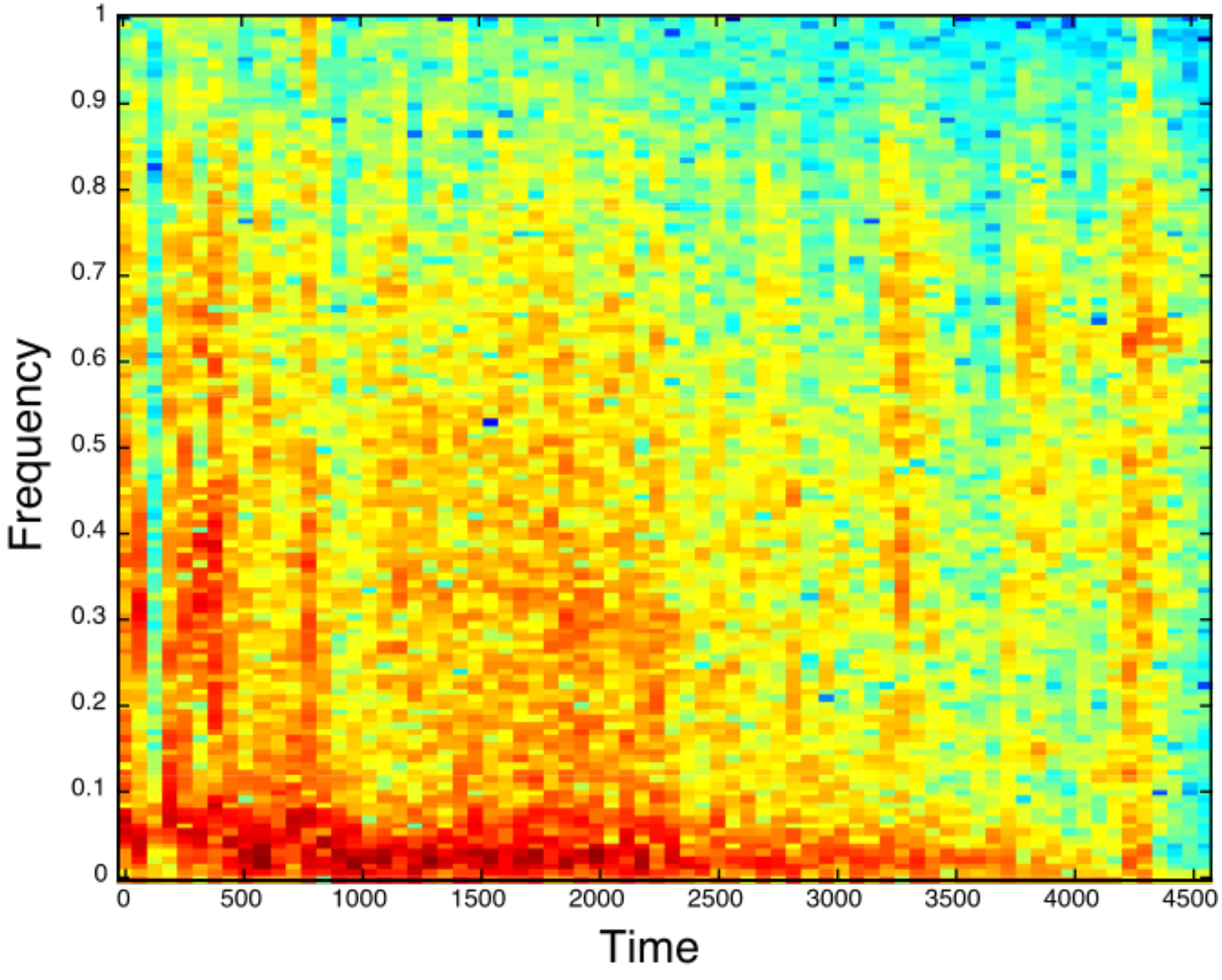
The Cochlea

"Go ahead … make my day."



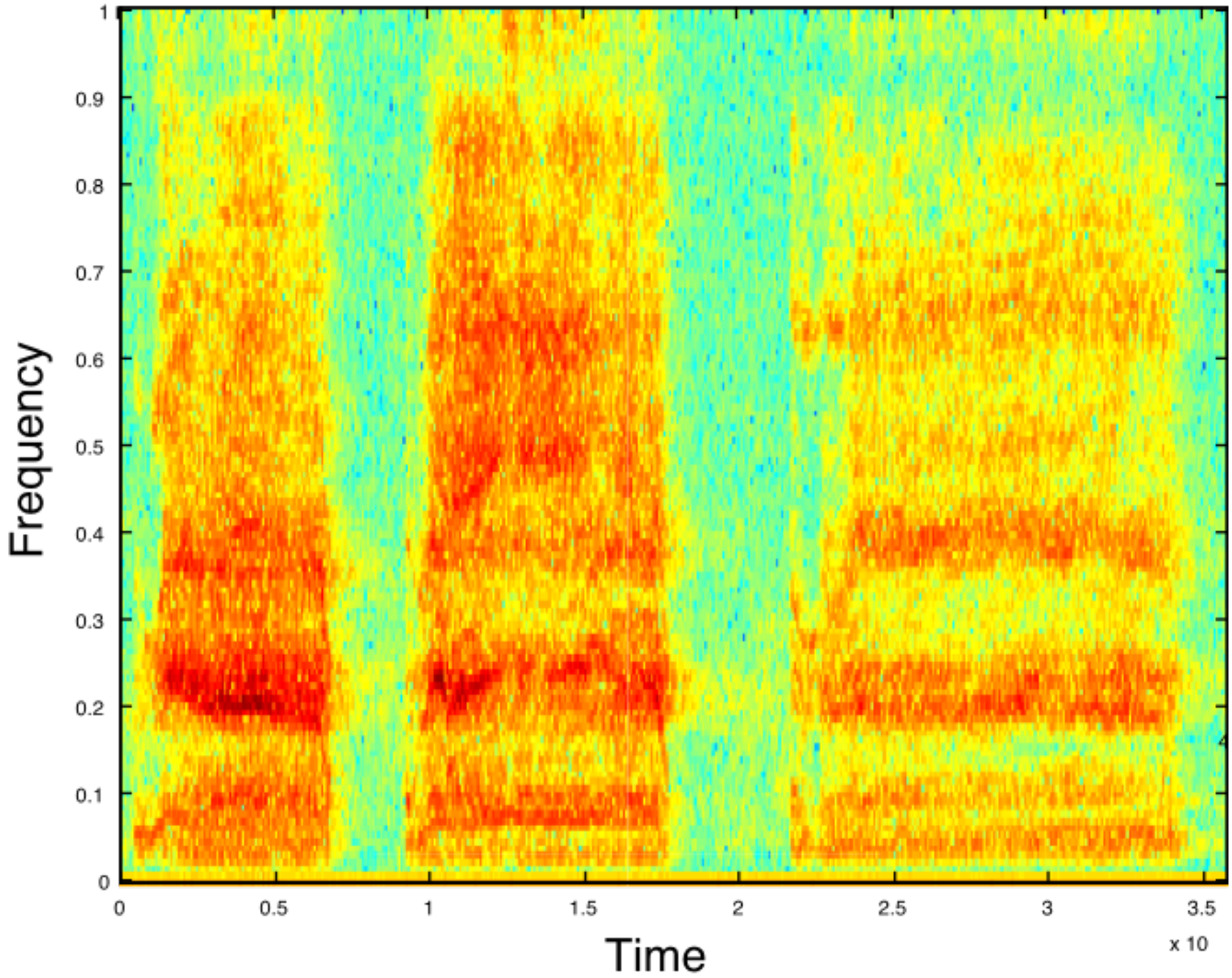
The Cochlea

A shotgun blast

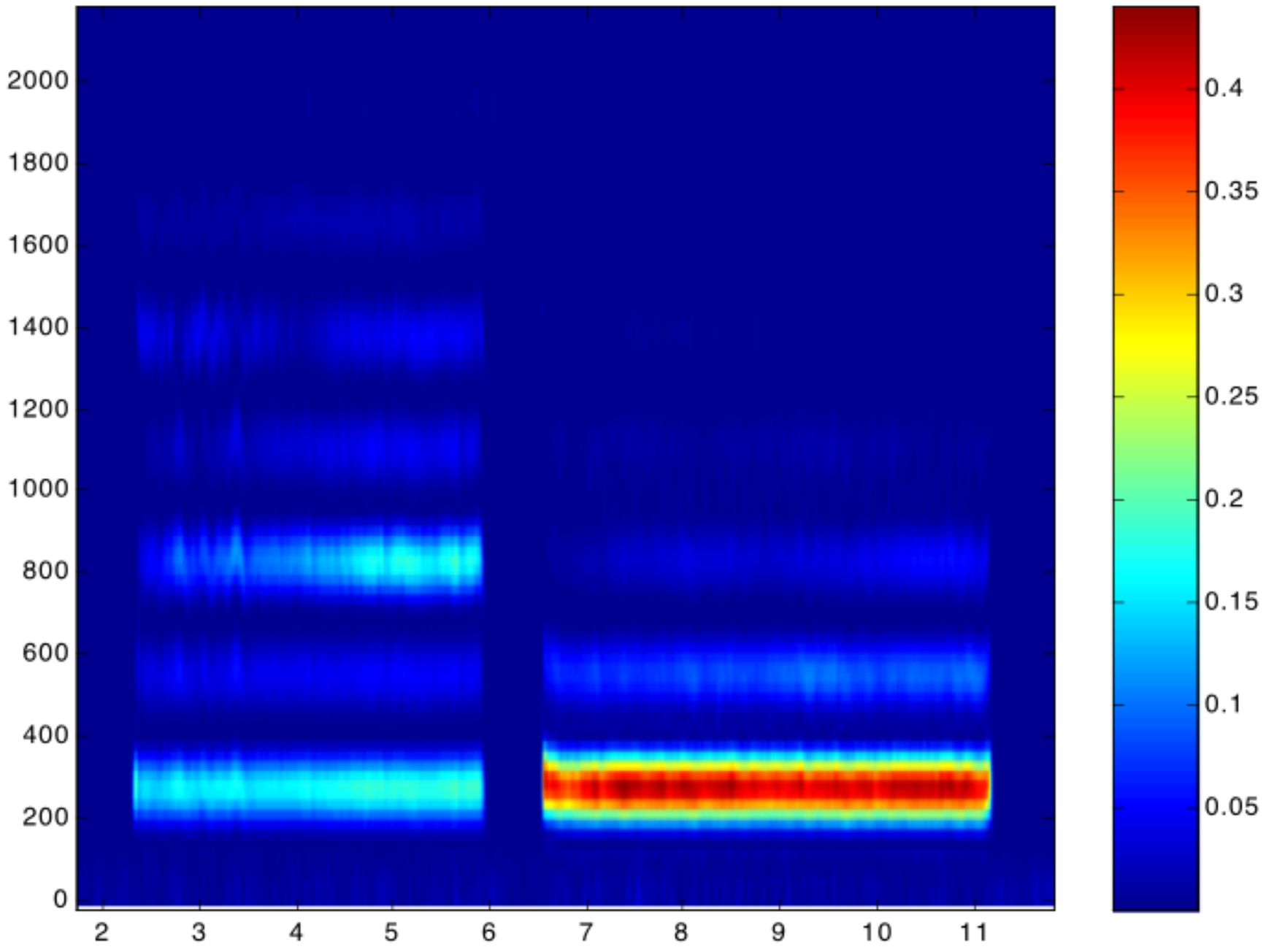


The Cochlea

A pig squealing



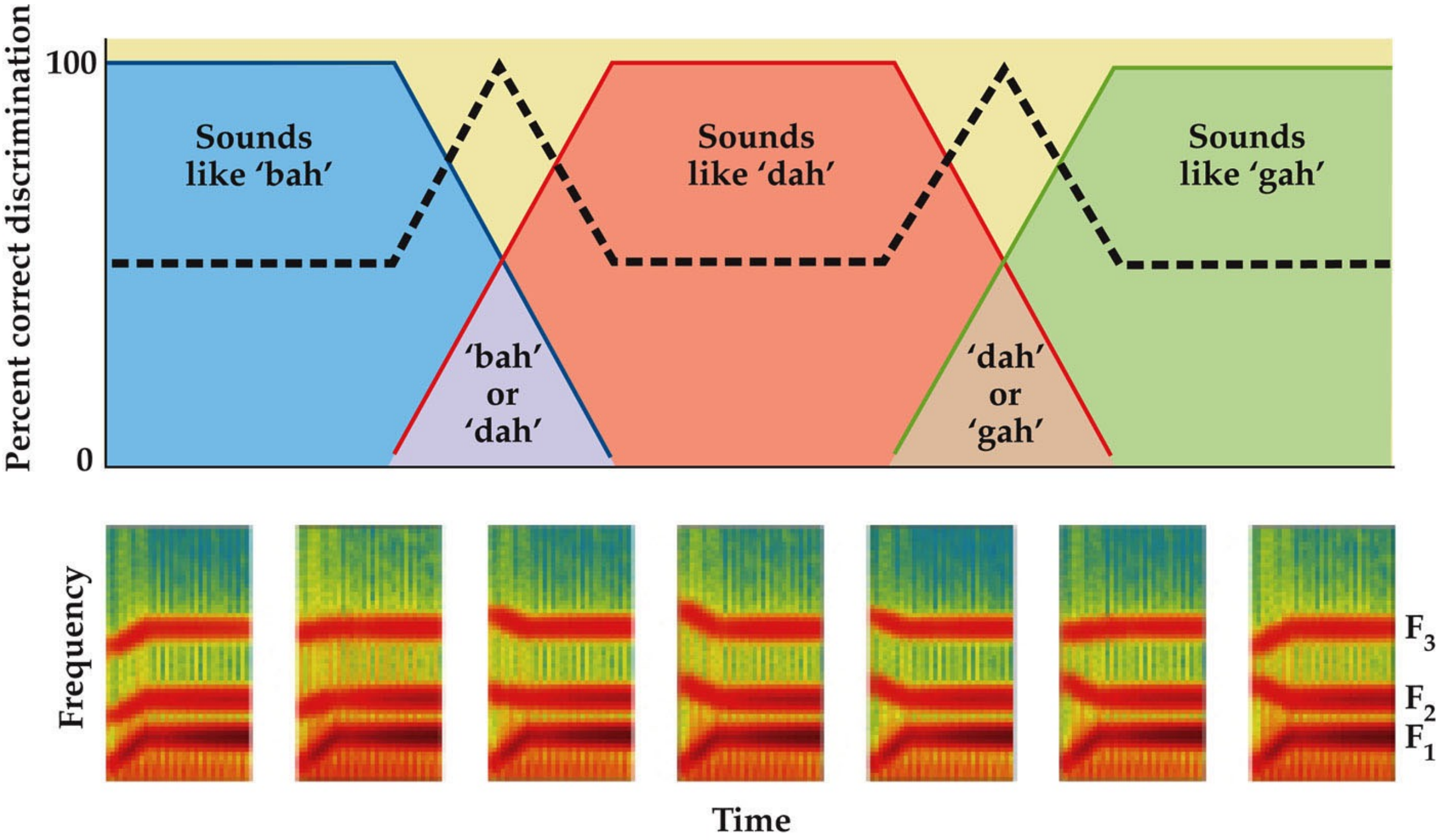
Cochlear Representation



Laaaaaaaaaaa Looooooooooooooh

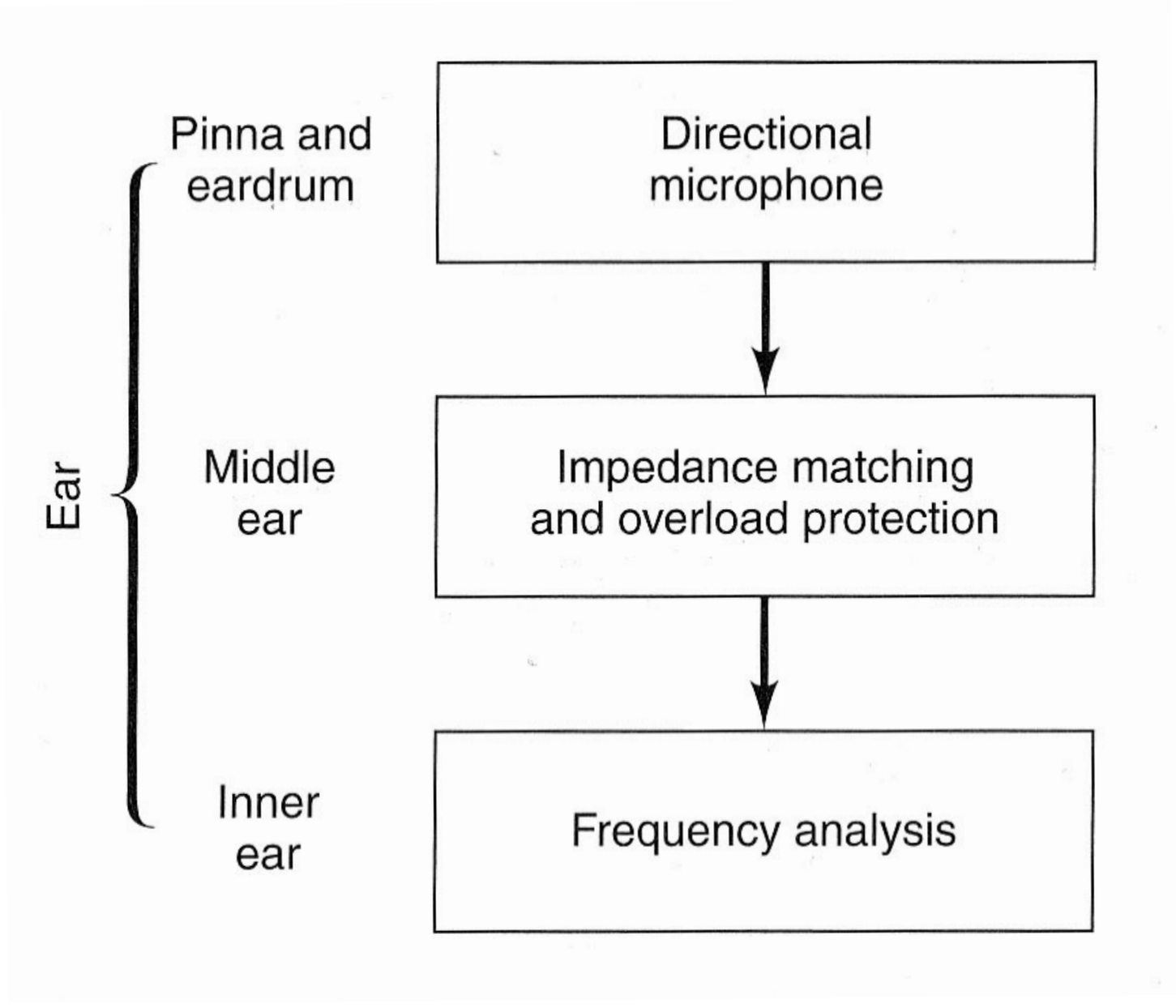
Cochlear Representation

A little bit of behavior can be explained in this representation.

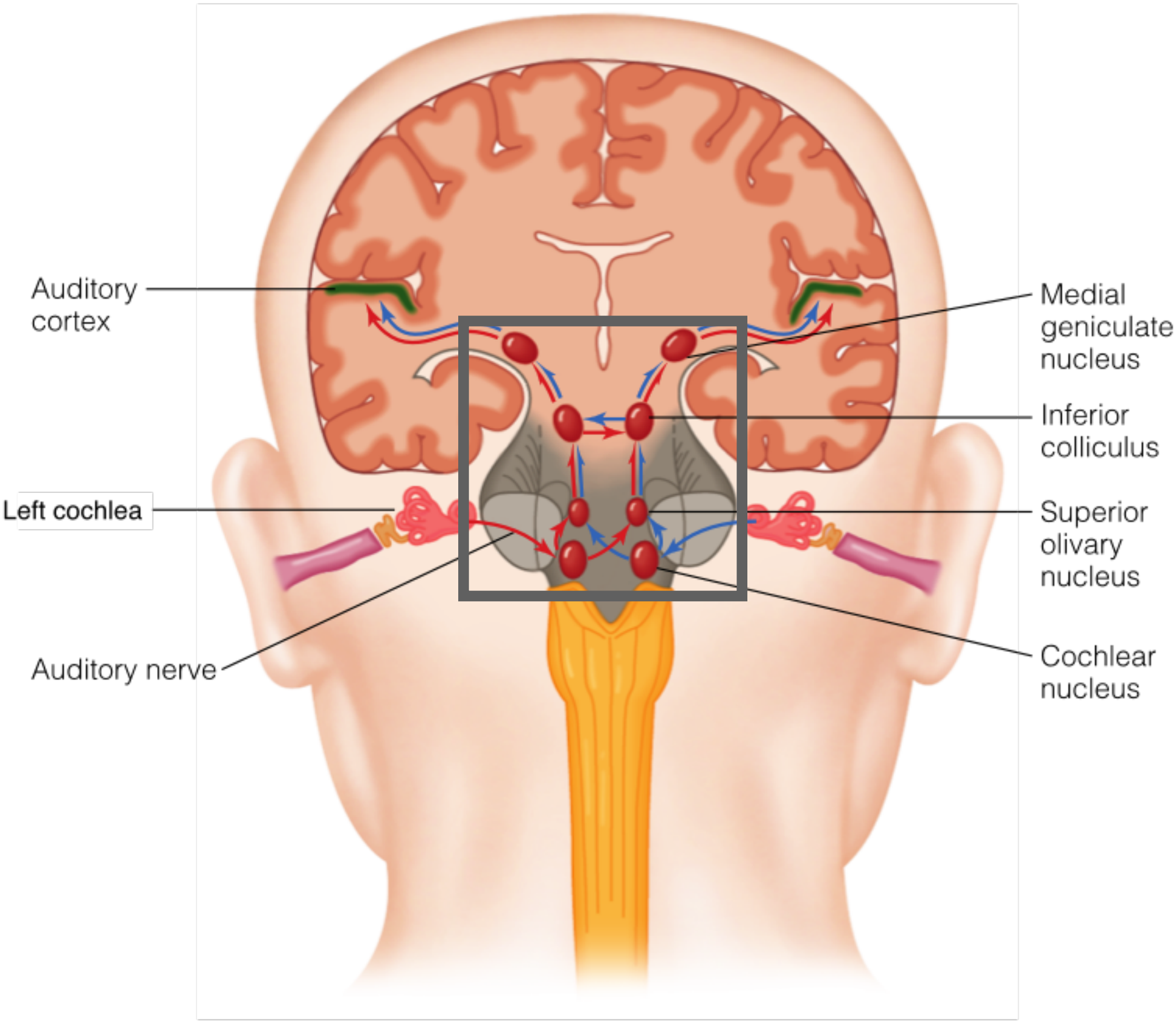


***SENSATION & PERCEPTION 3e,* Figure 11.13**

Functional schematic of the ear:

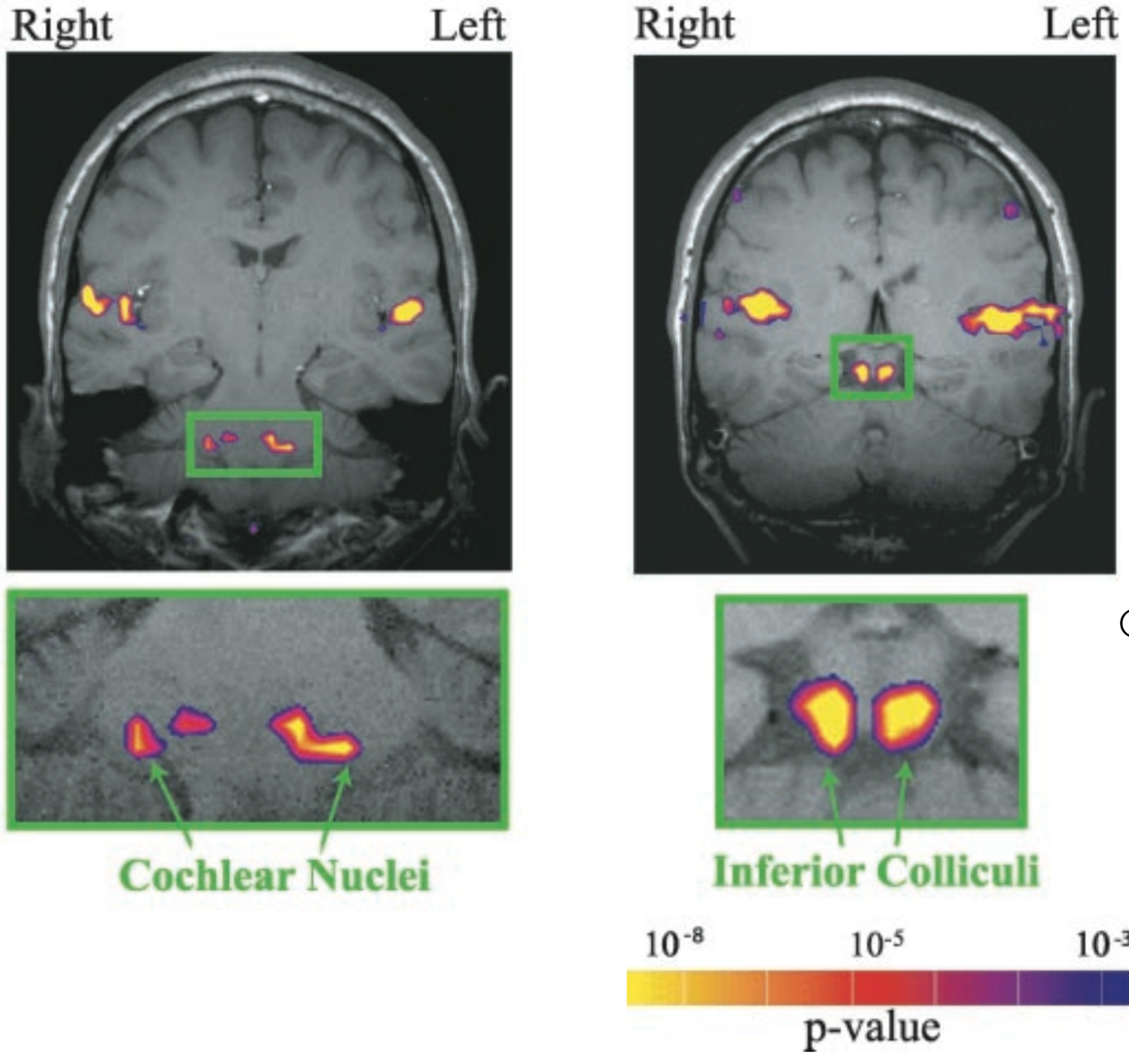


The Auditory Midbrain



The Auditory Midbrain

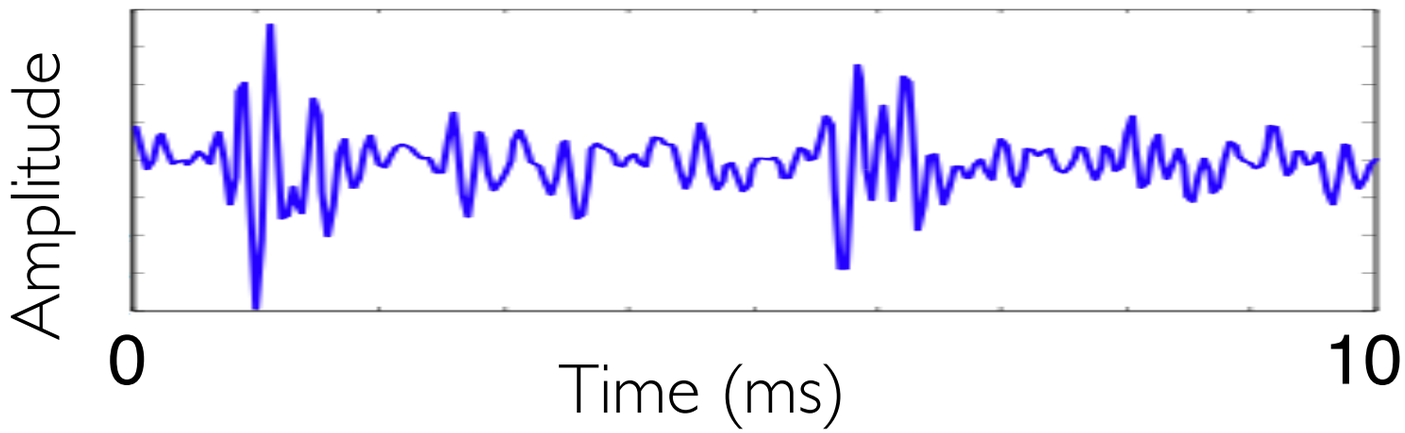
Contrast of Sound vs Silence

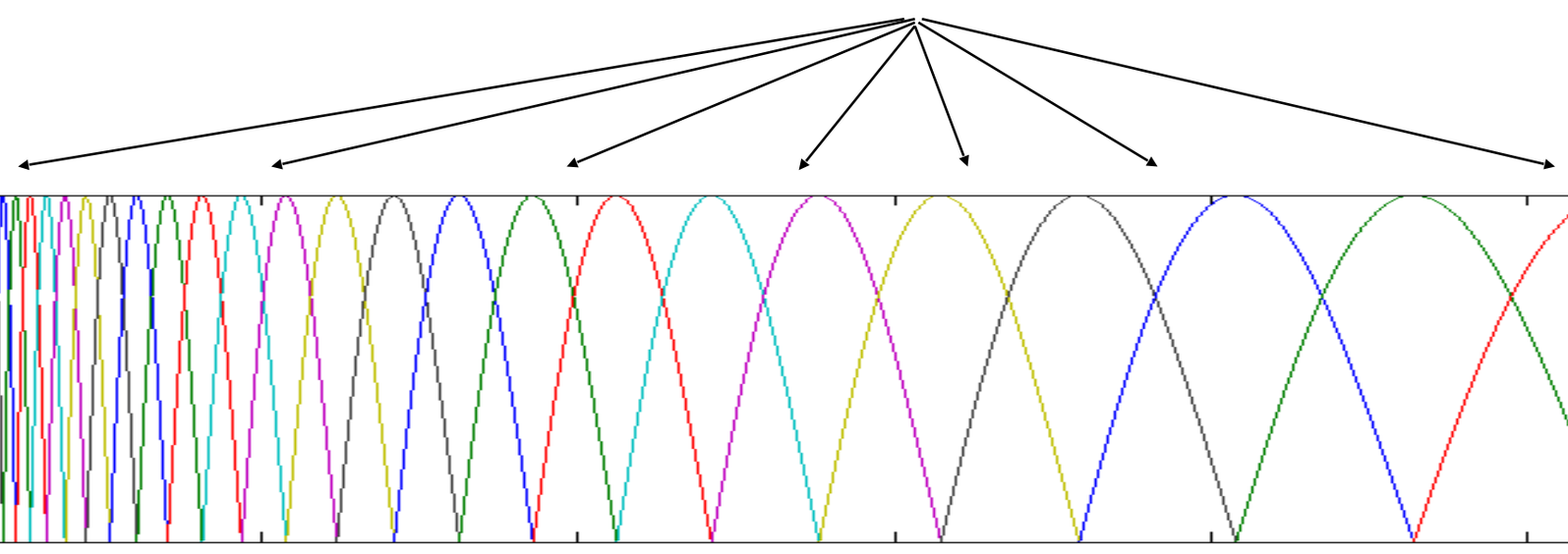


Penagos et al. 2004

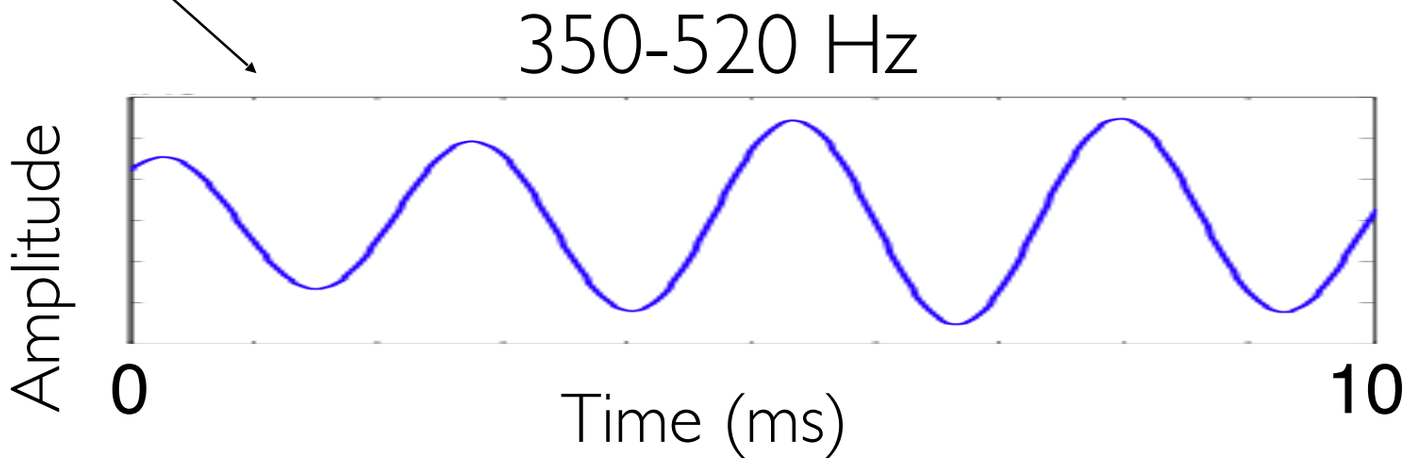
Subband Representation

Original Sound Signal



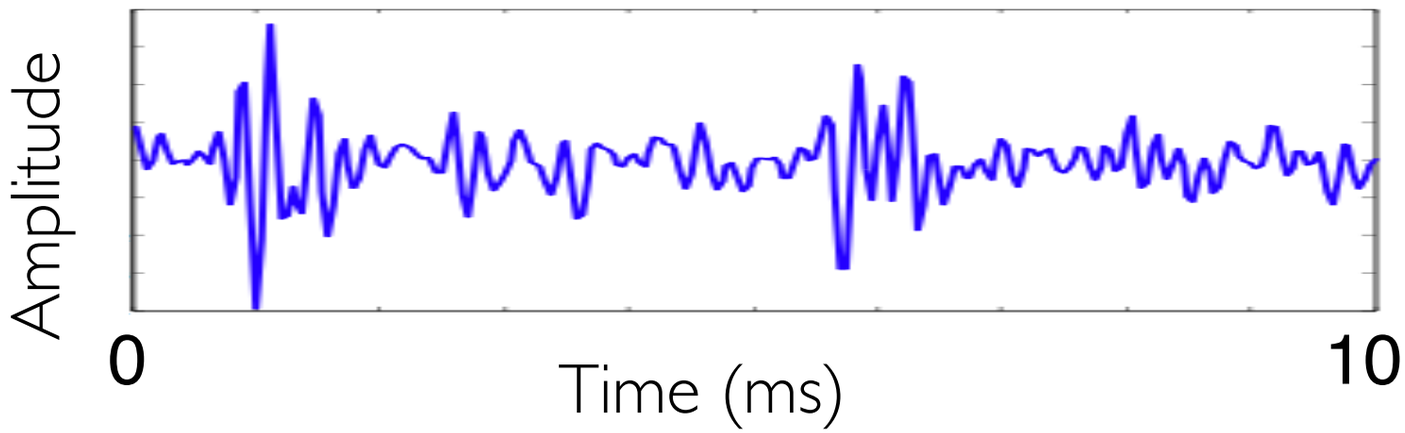


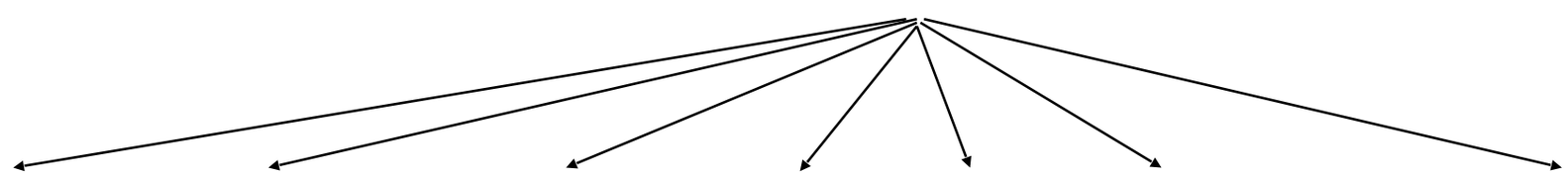
Subband



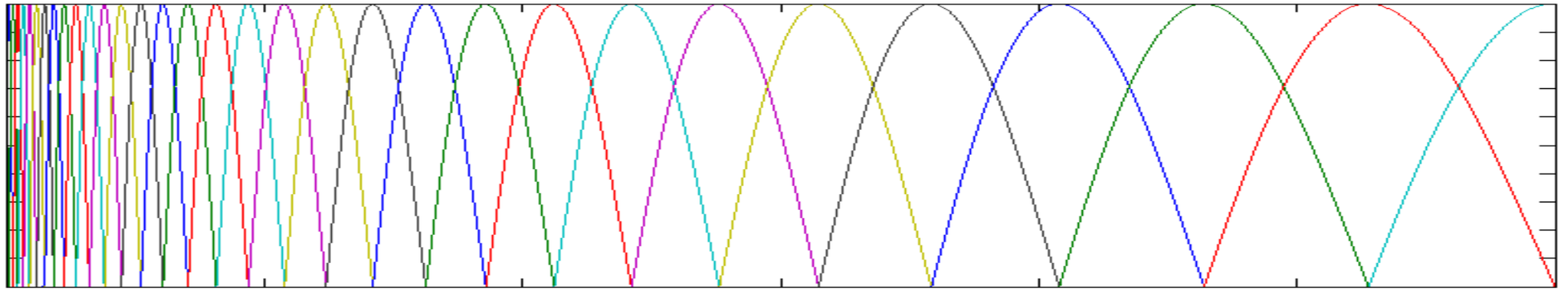
Subband Representation

Original Sound Signal

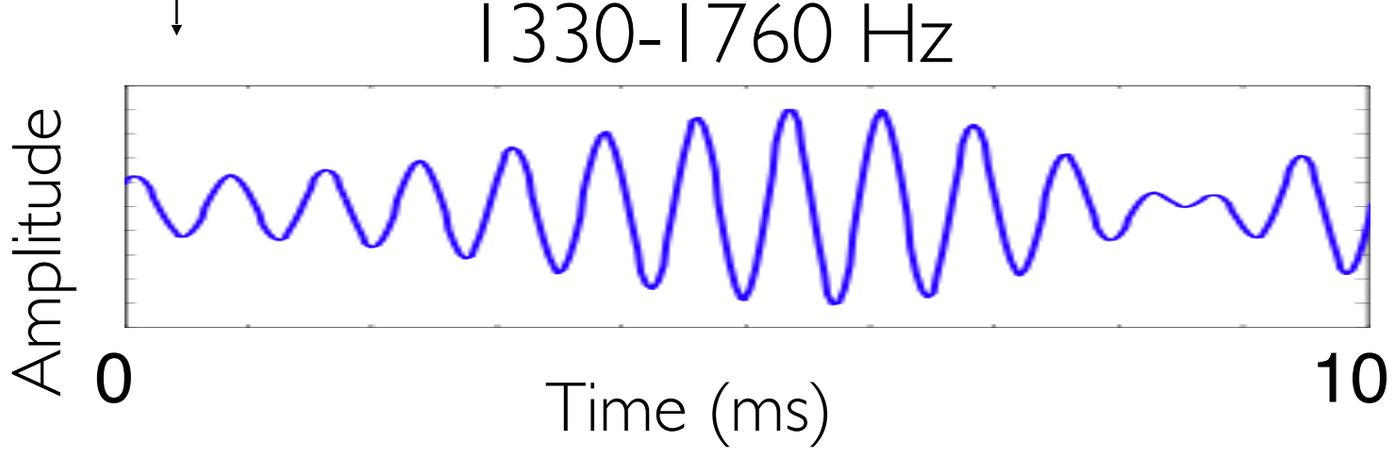




Cochlear Filter Bank

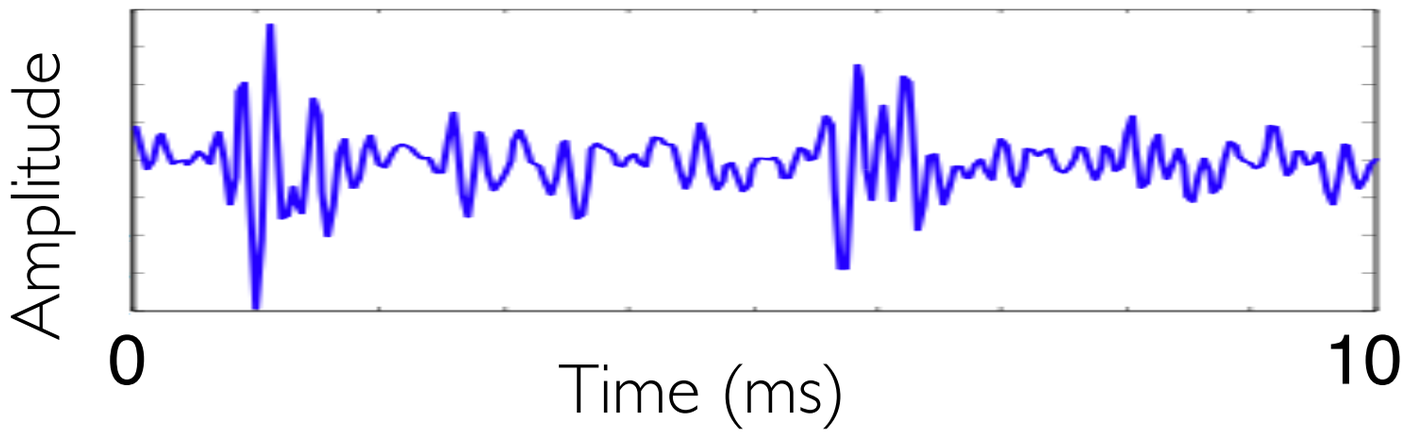


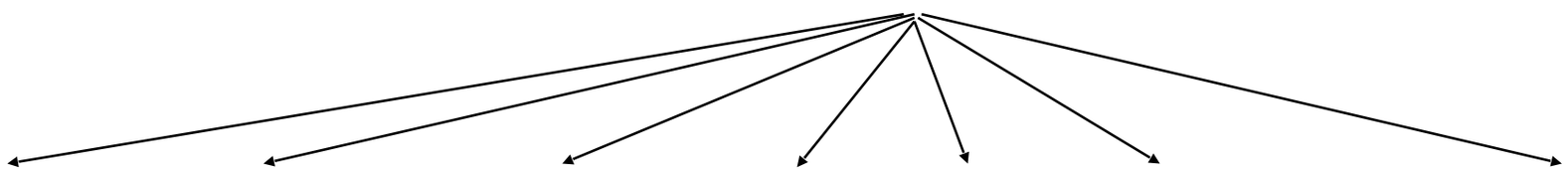
Subband



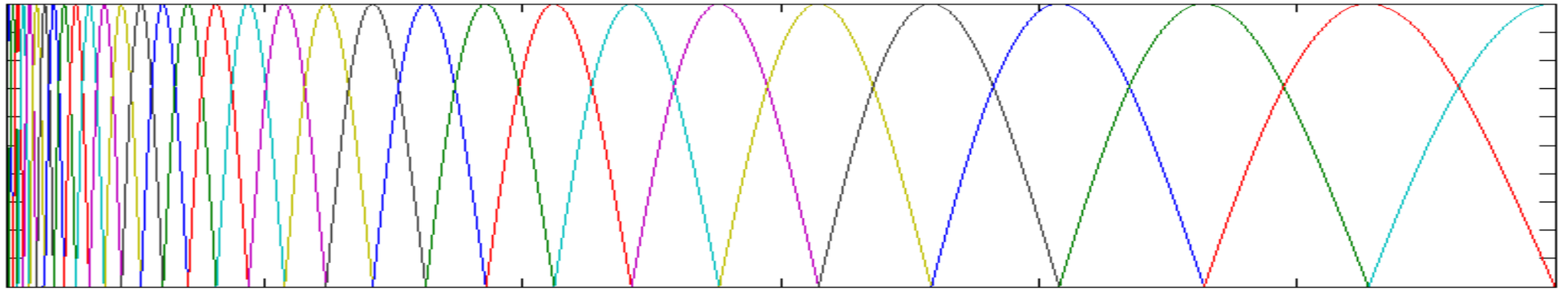
Subband Representation

Original Sound Signal

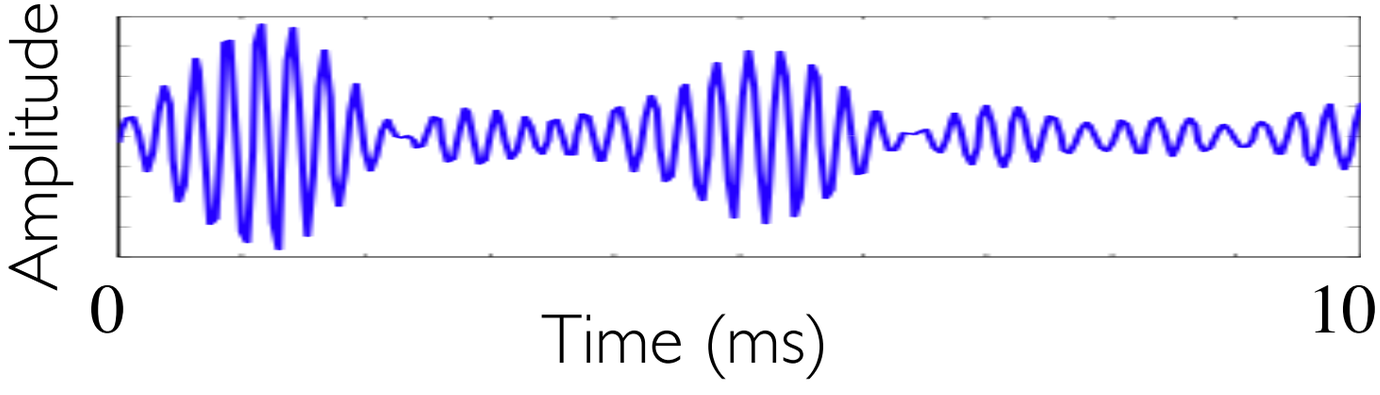




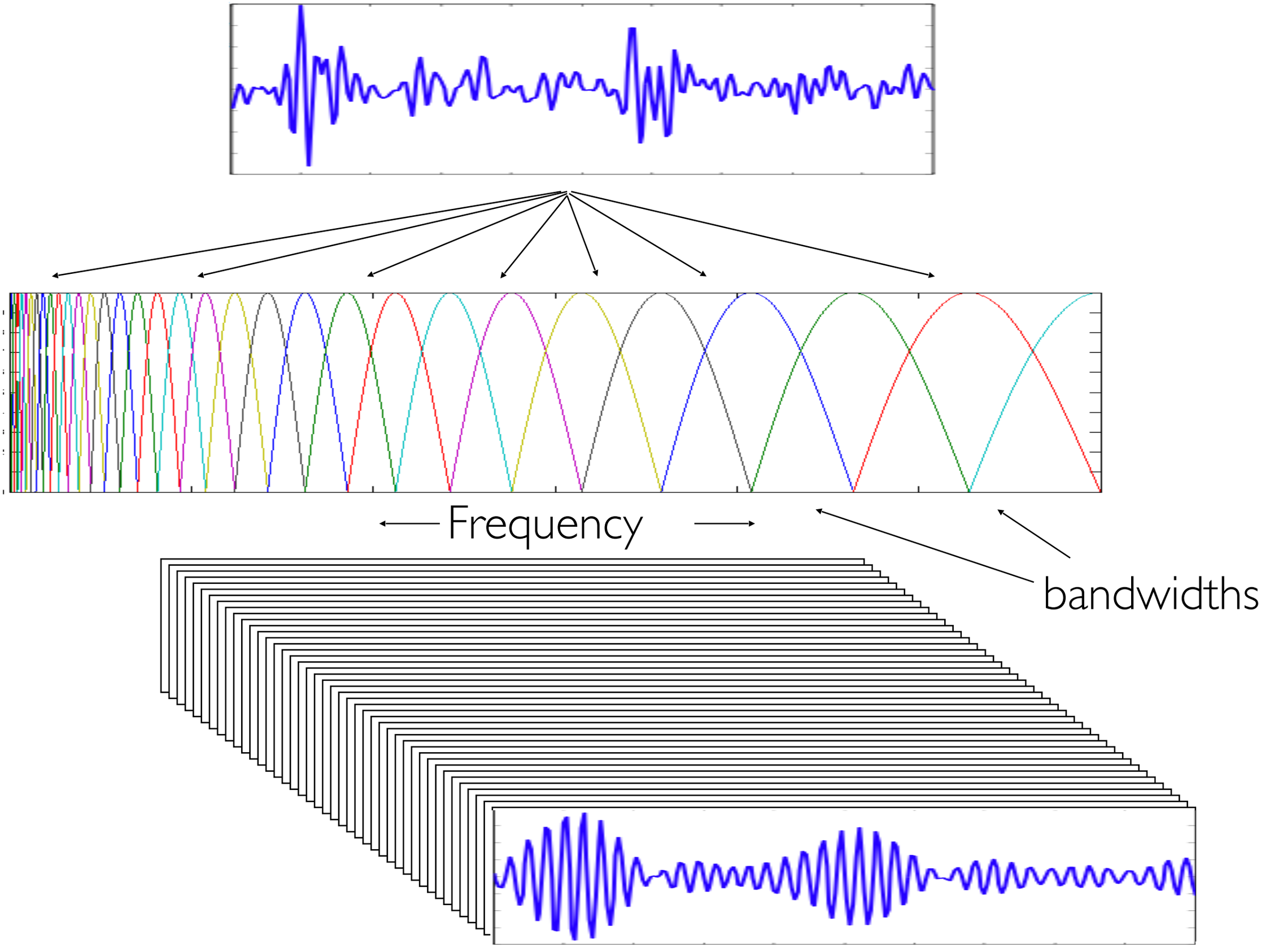
Cochlear Filter Bank



Subband

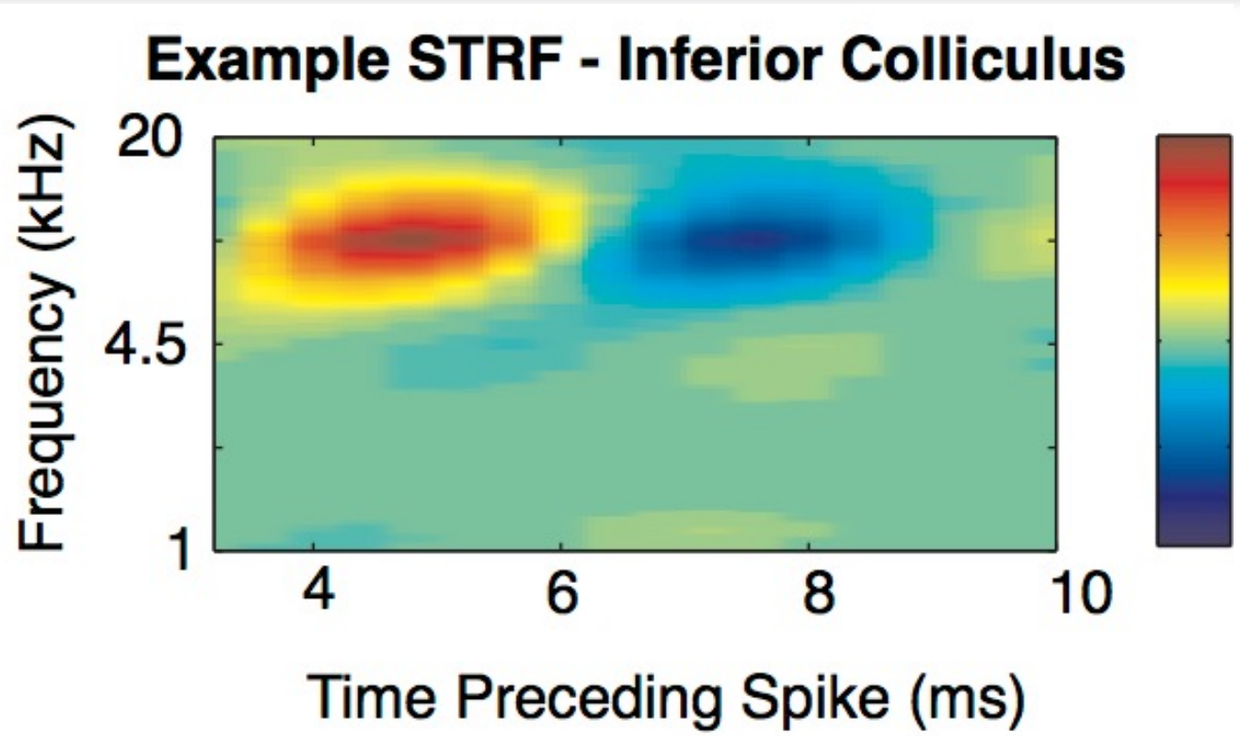


Subband Representation



Modulation in Midbrain STRF

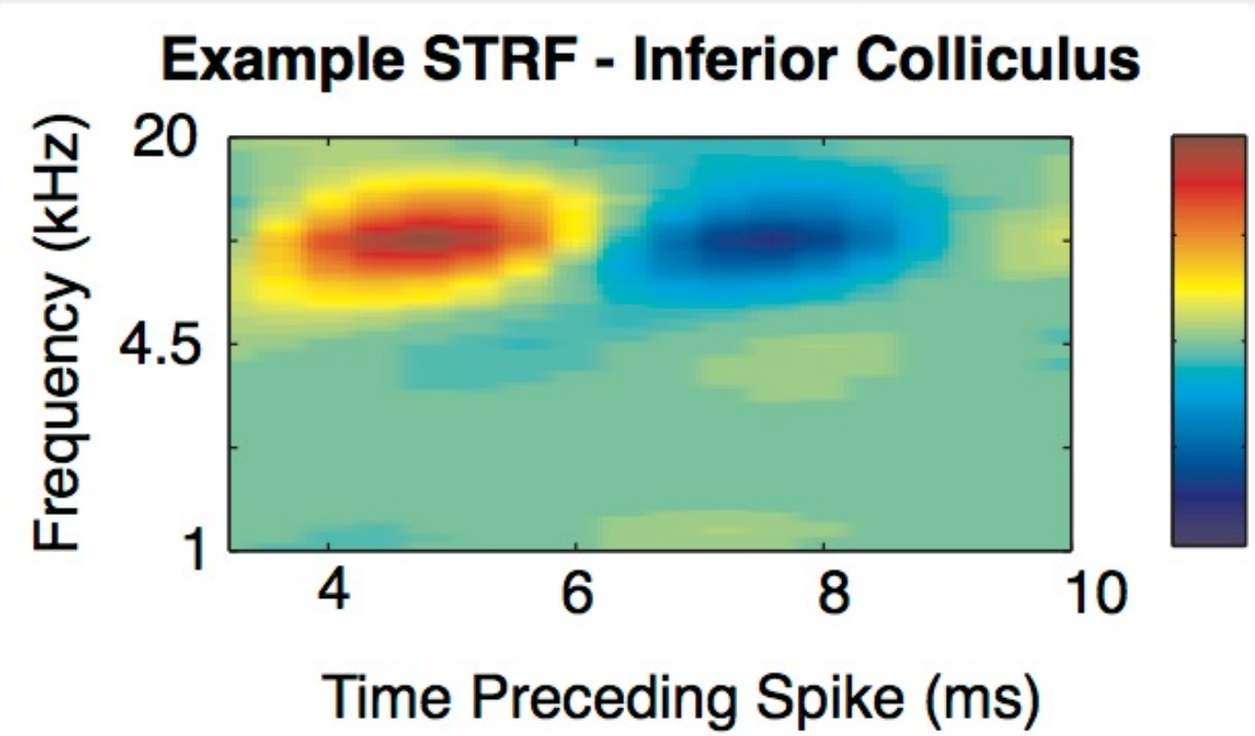
STRF = spectro-temporal receptive field.



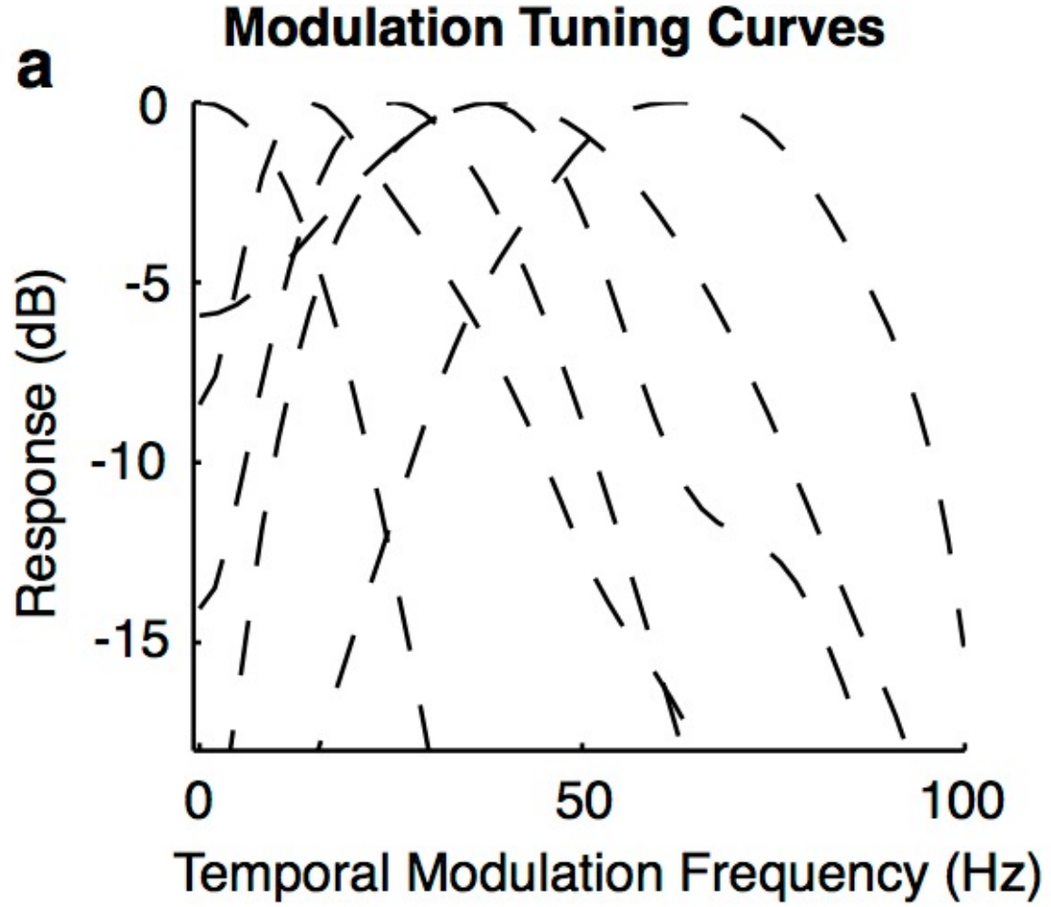
Neuron is responding to changes in amplitude in particular frequency range.

Modulation in Midbrain STRF

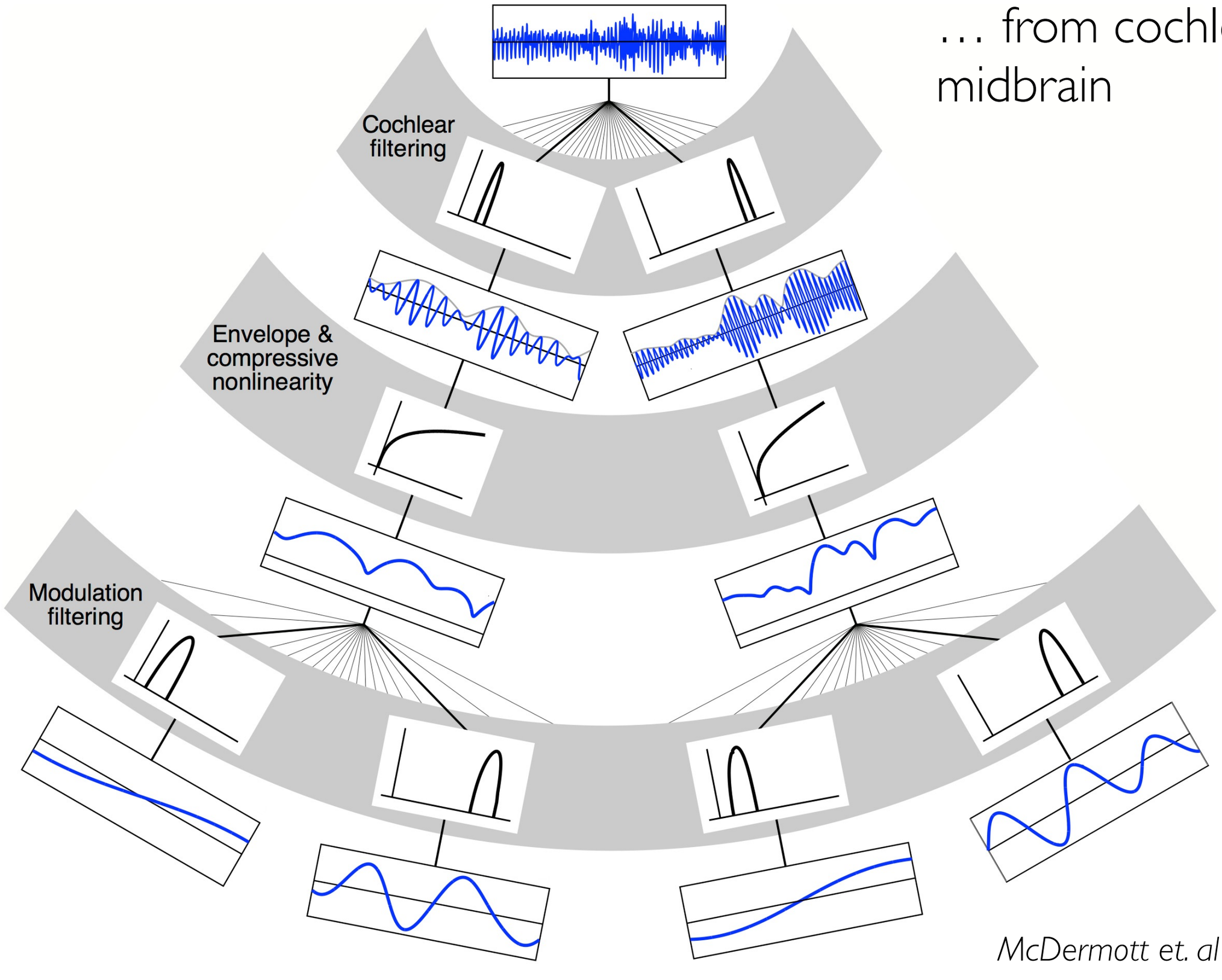
STRF = spectro-temporal receptive field.



As early as the midbrain, auditory neurons are tuned to particular modulation rates.

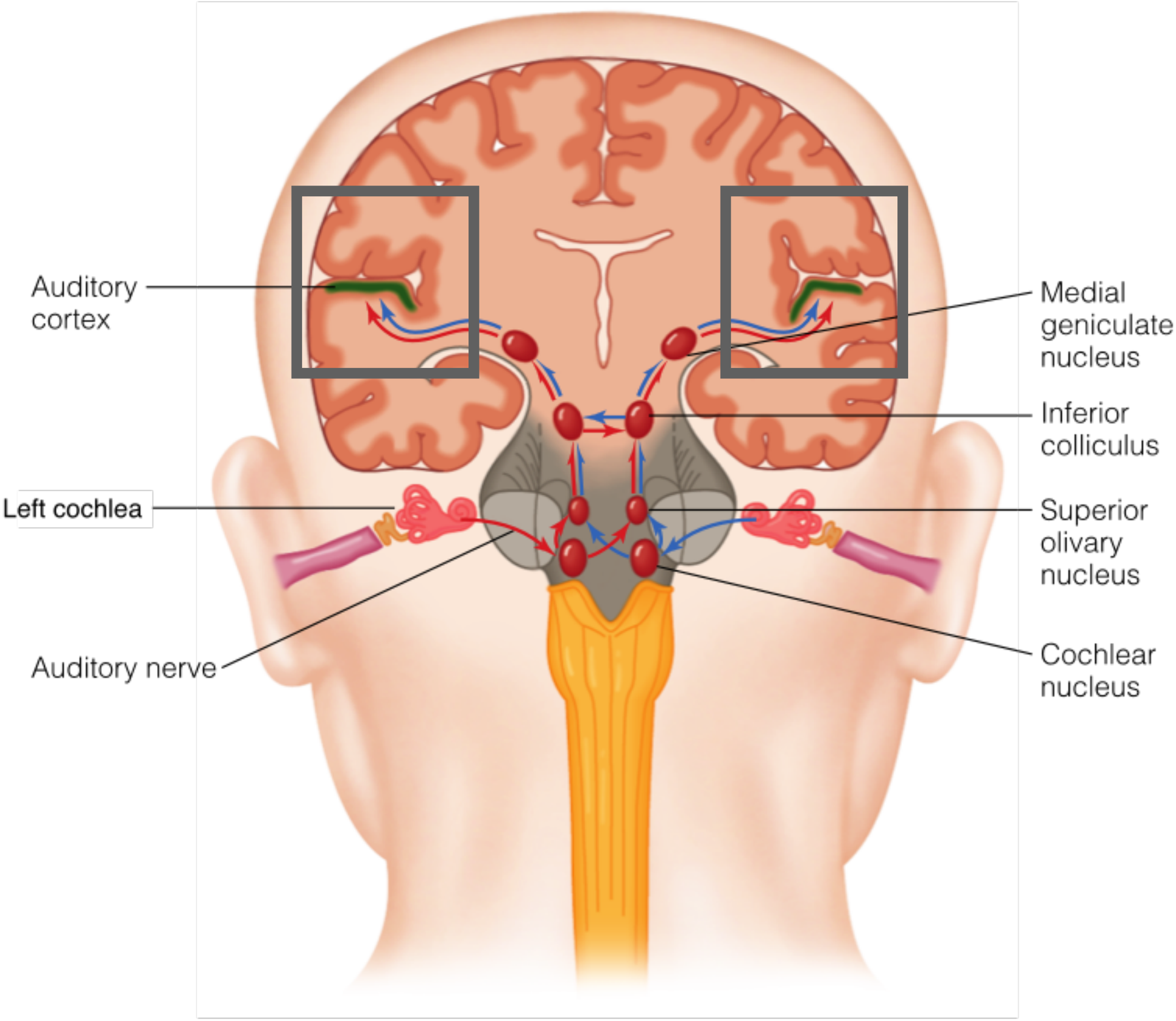


Hierarchical Processing Model

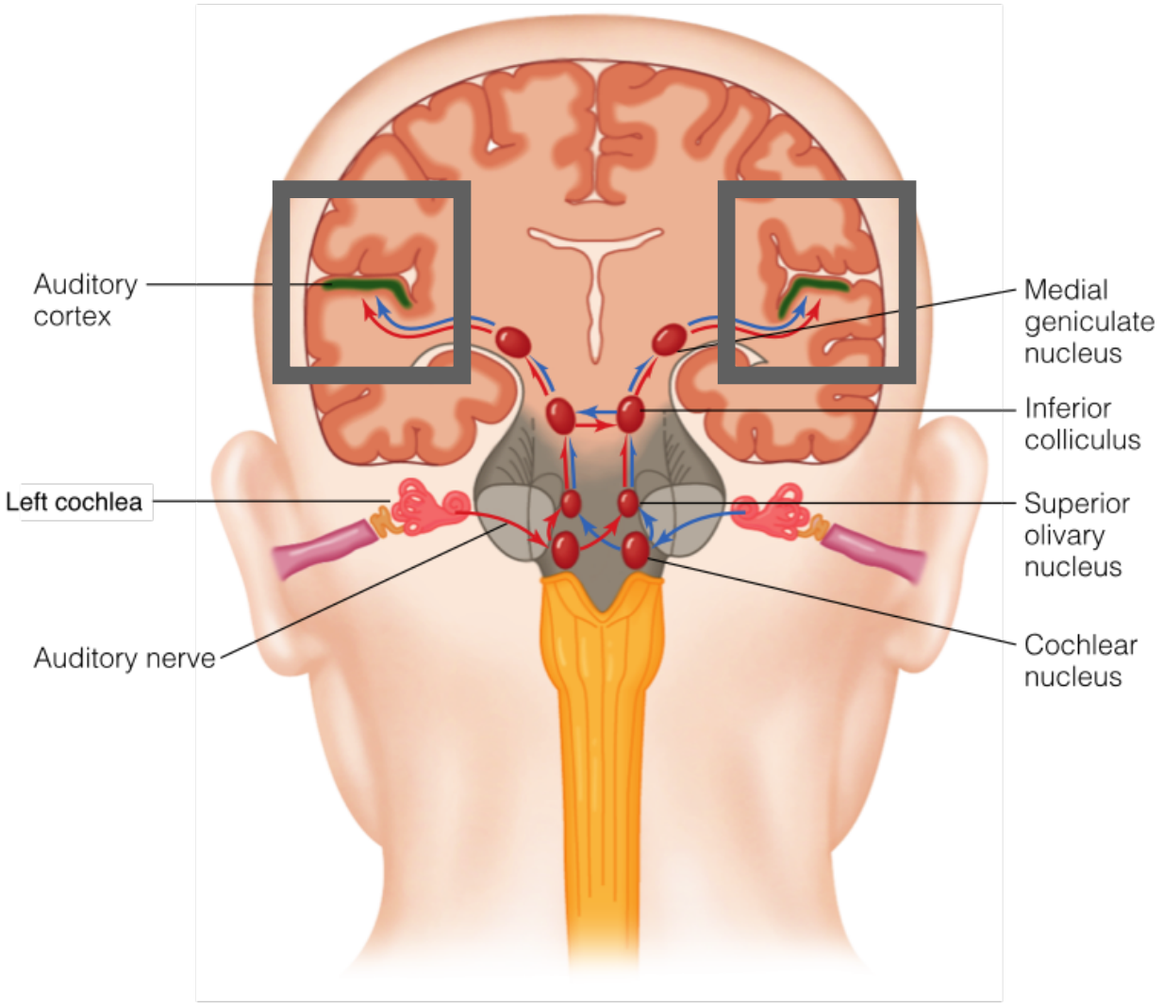


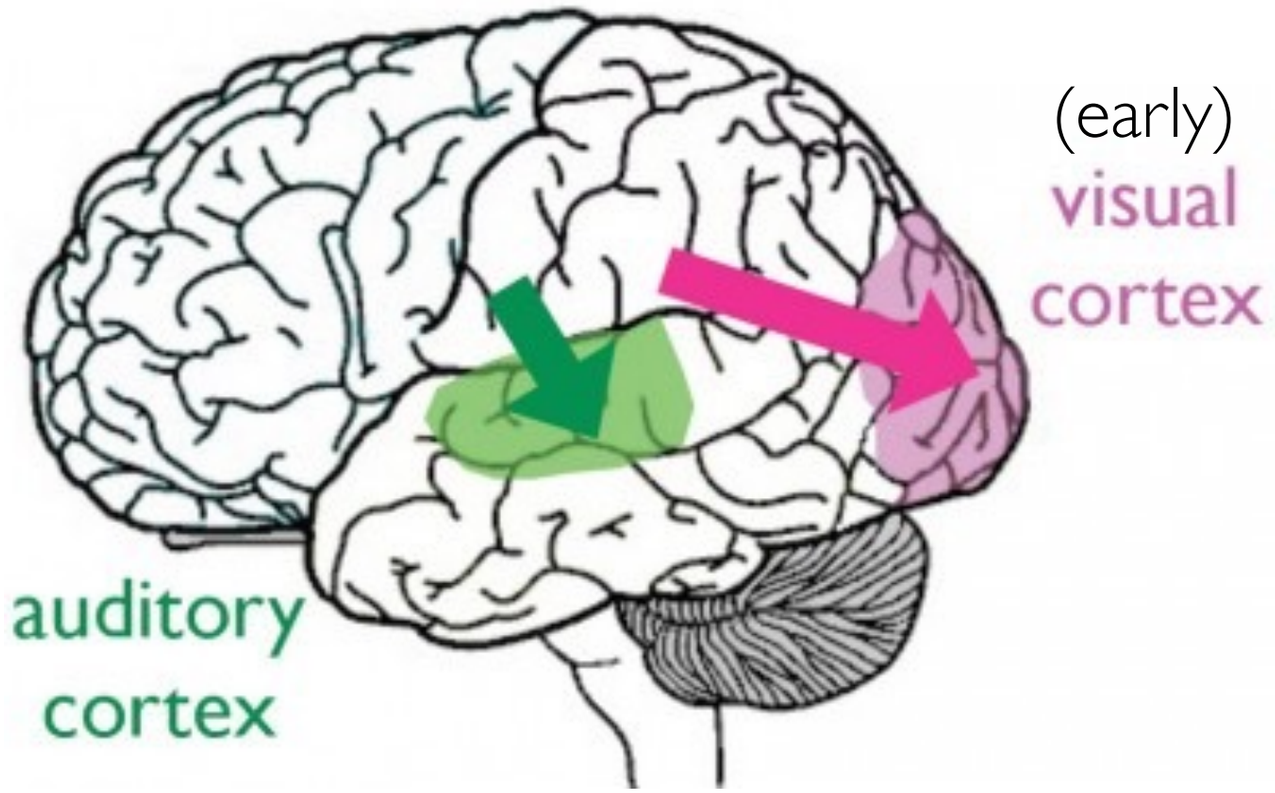
… from cochlea to midbrain

McDermott et. al 2011


Auditory cortex
Medial geniculate nucleus
Inferior colliculus
Left cochlea
Superior olivary nucleus
Auditory nerve
Cochlear nucleus

Auditory Cortex

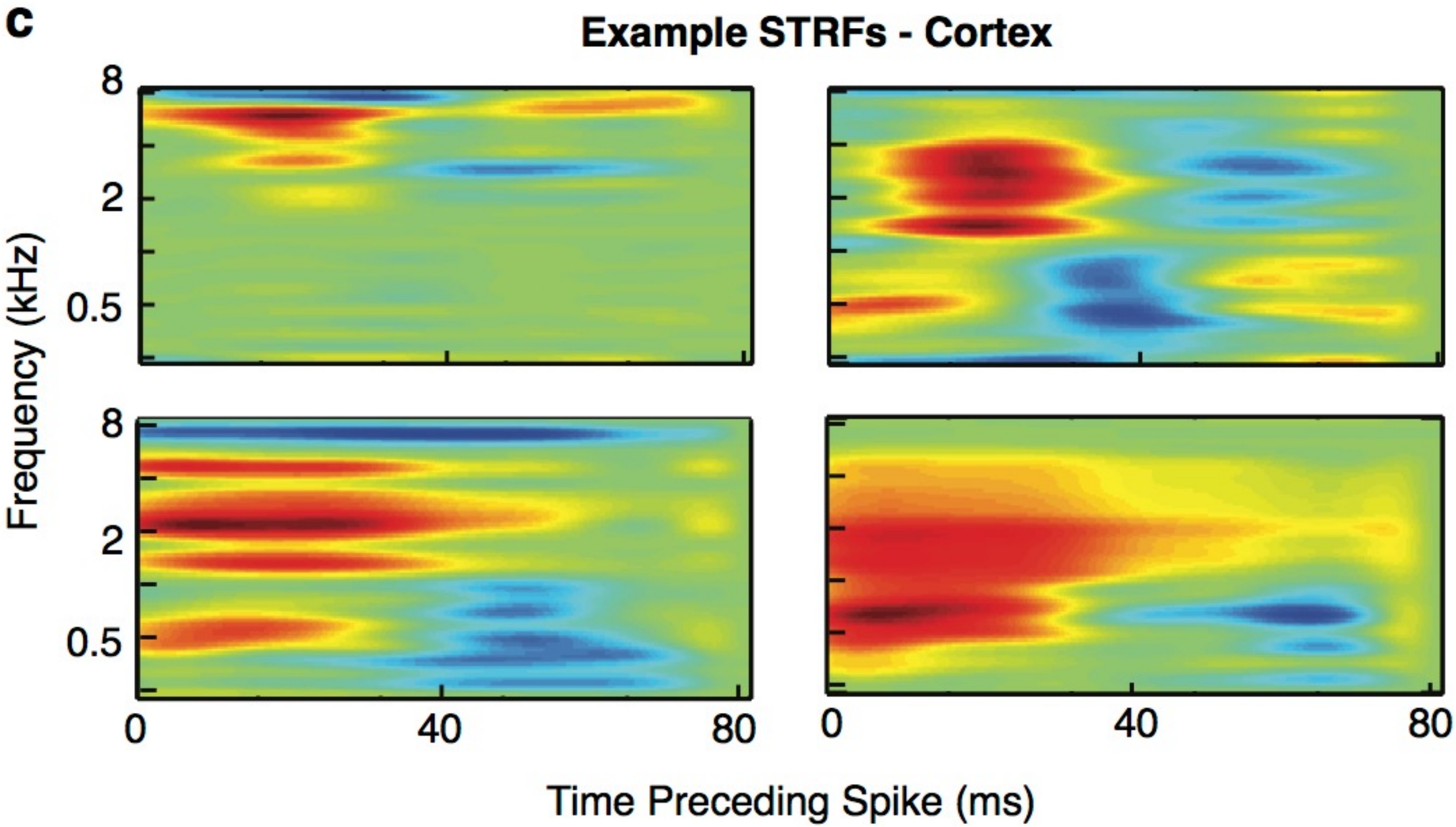




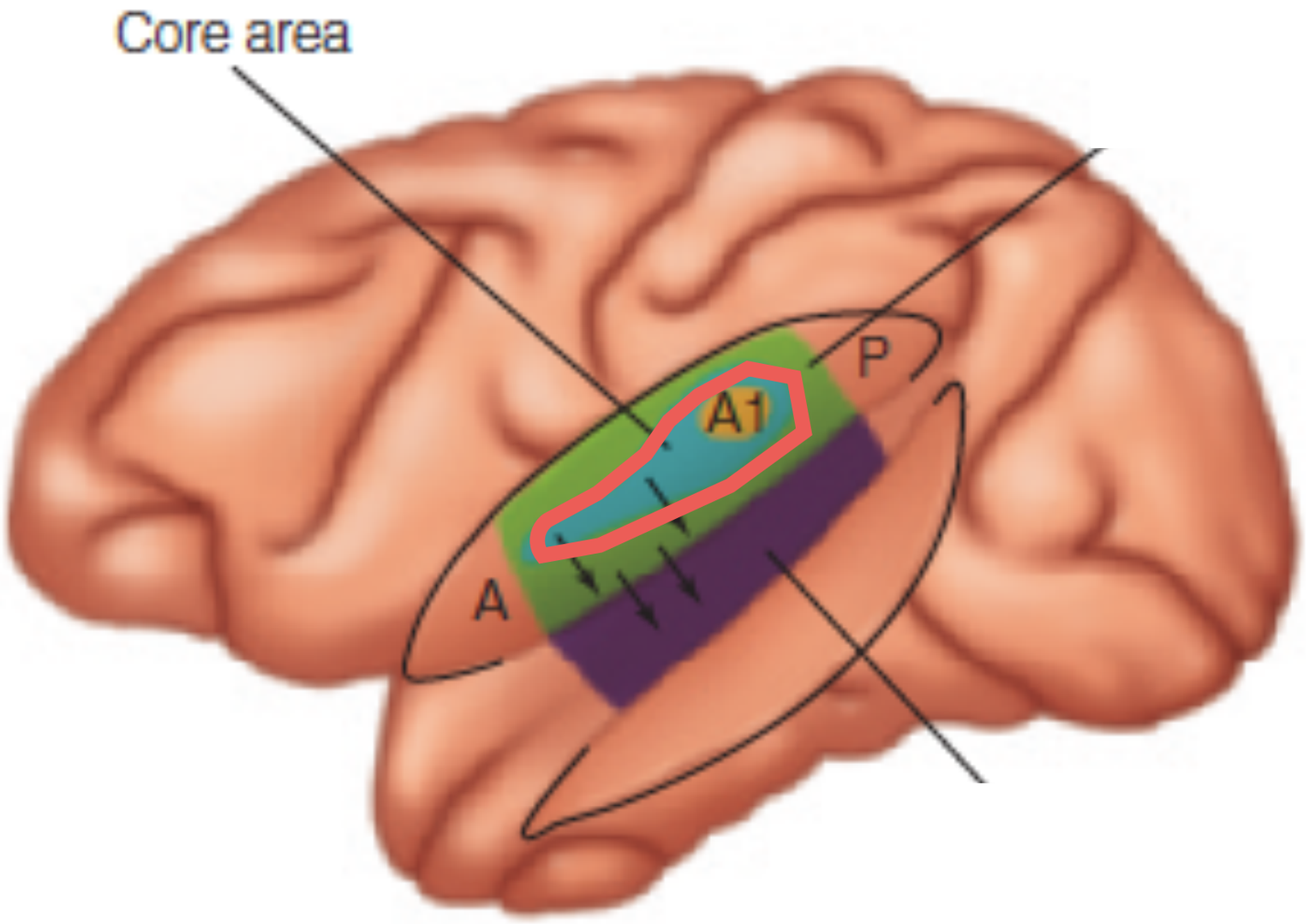
How are circuits making sense of complex sound patterns?

Auditory Cortex

Cortical STRFs are often more complex than those in the midbrain and thalamus:



Auditory Cortex



*monkey
*

Tramo et. al, Curr. Opin. Neuro. (1999)

Auditory Cortex

Macaque

Human

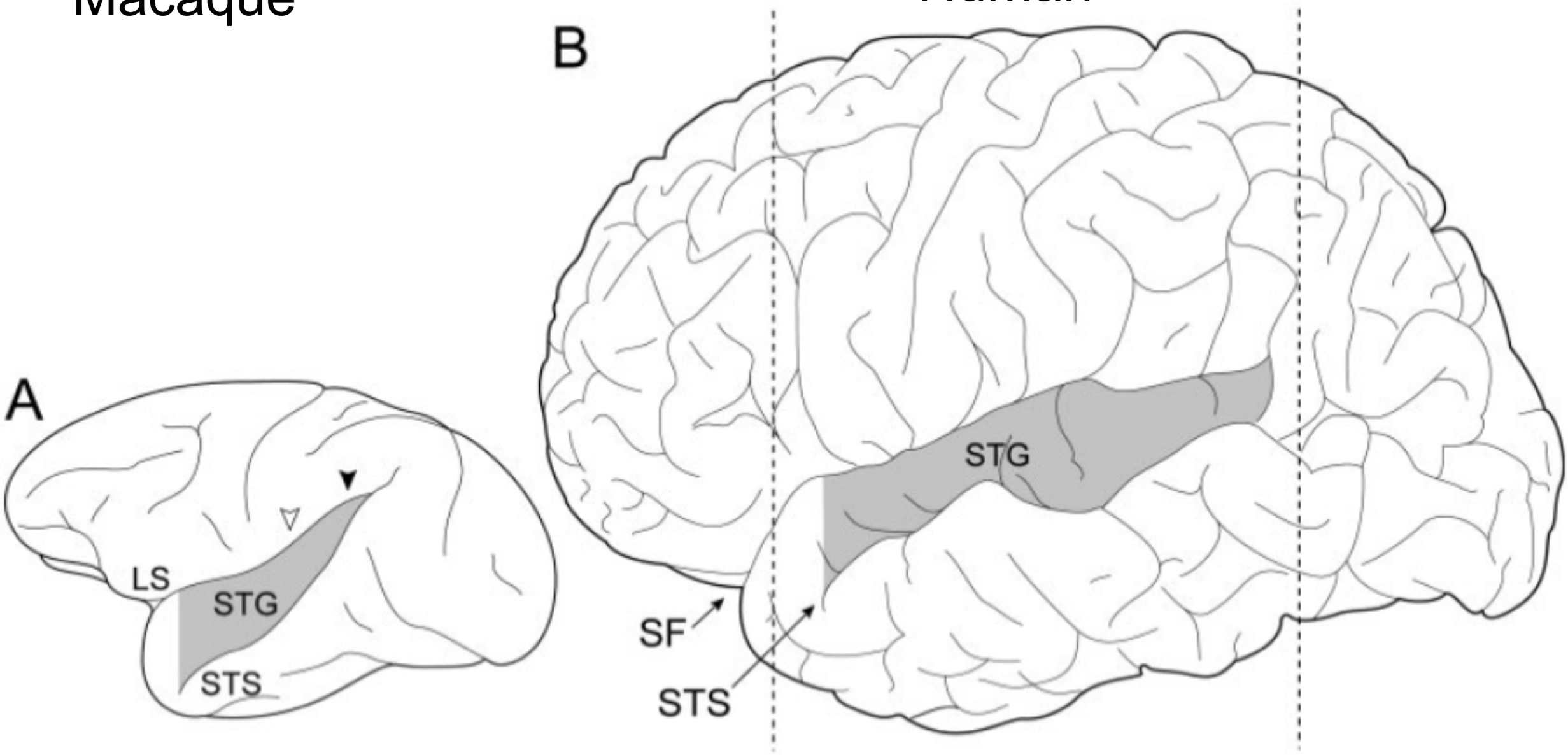
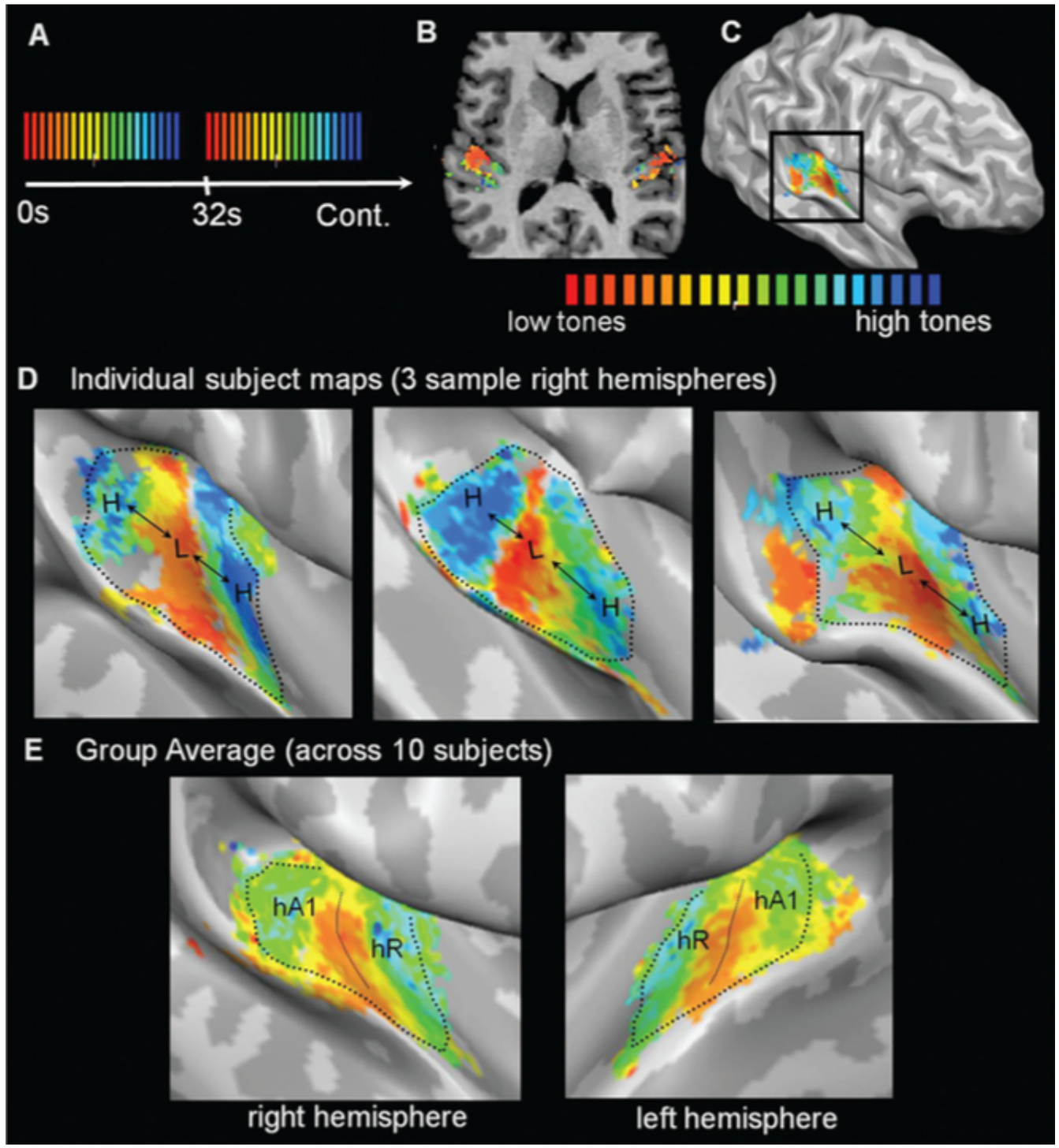


Fig. 1. Lateral view of the left hemisphere in macaque (**A**) and human (**B**). The STG is shaded in both species. In A, open arrowhead indicates approximate level of sections shown in Figure 3 and closed arrowhead indicates approximate level of sections shown in Figure 4. In B, dashed lines indicate approximate rostral and caudal boundaries of existing coronal blocks containing the entire superior temporal gyrus (STG) in human subjects. LS, lateral sulcus; STS, superior temporal sulcus; SF, sylvian fissure.

Sweet et al. 2005

Auditory Cortex

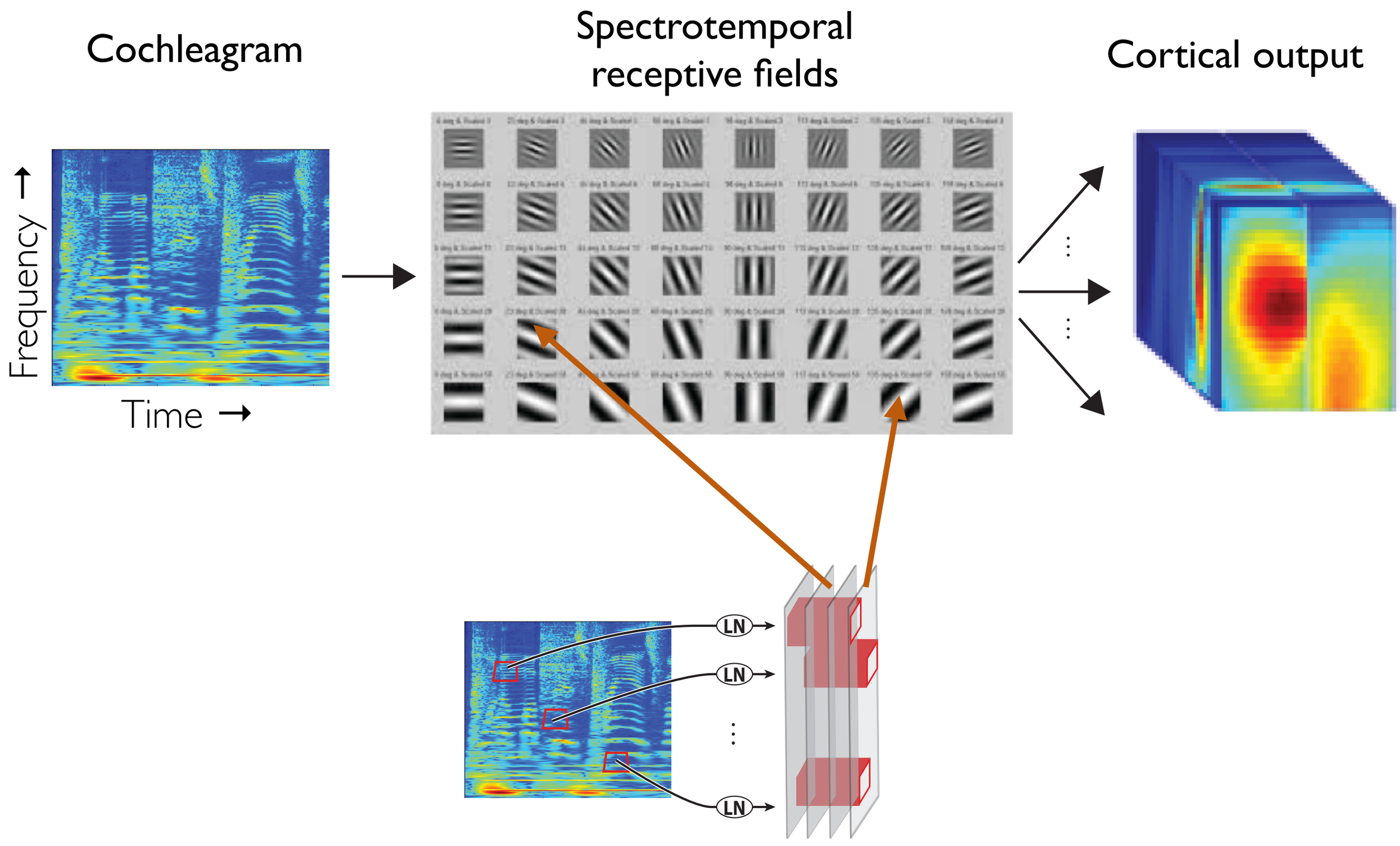
Primary core area A1 is tonotopic.



Da Costa et al. 2011

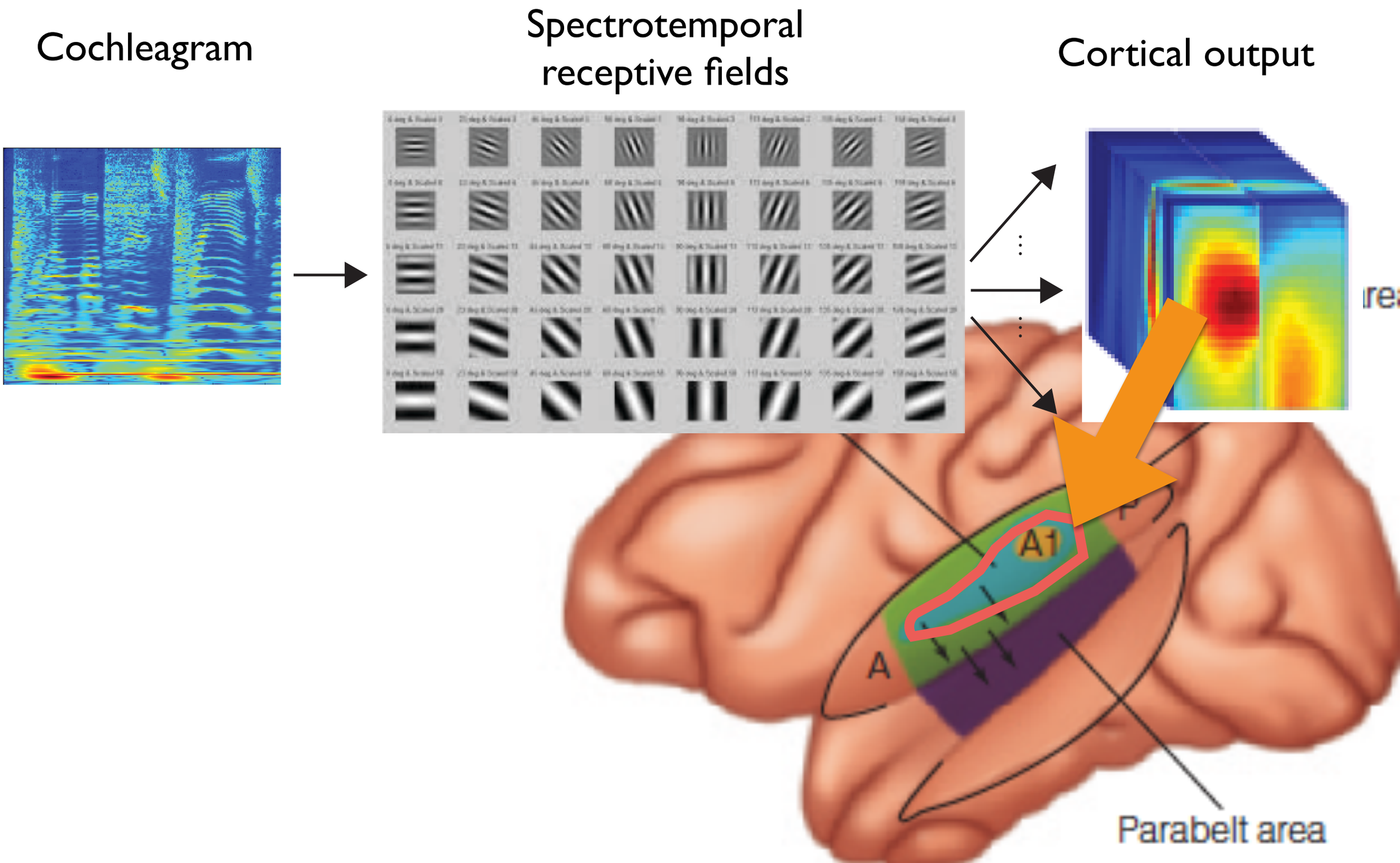
Spectrotemporal model

Spectrotemporal model (Shamma, 2005) of early auditory cortex is of this form:



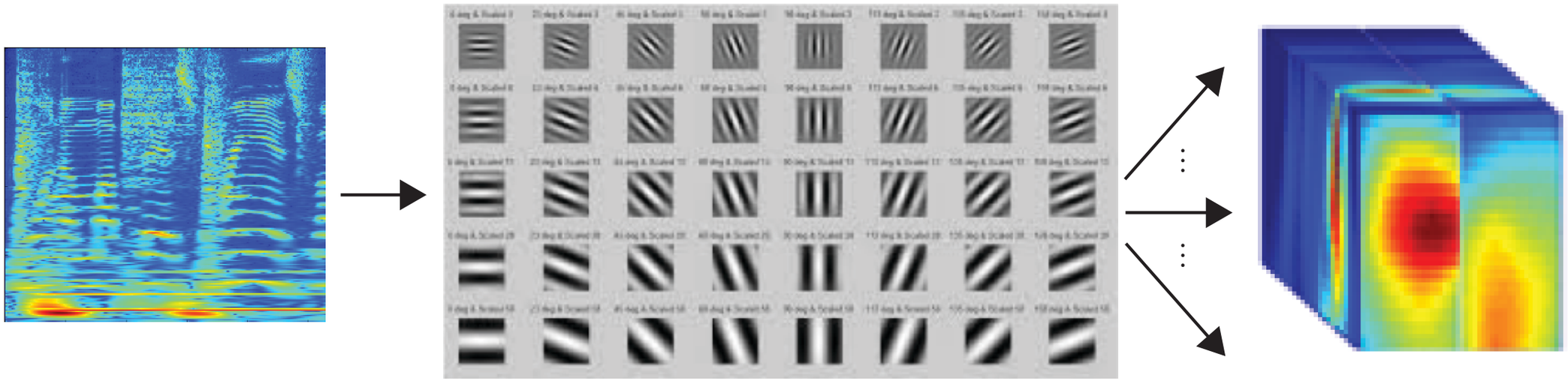
Spectrotemporal model

Spectrotemporal model (Shamma, 2005) of early auditory cortex is of this form:

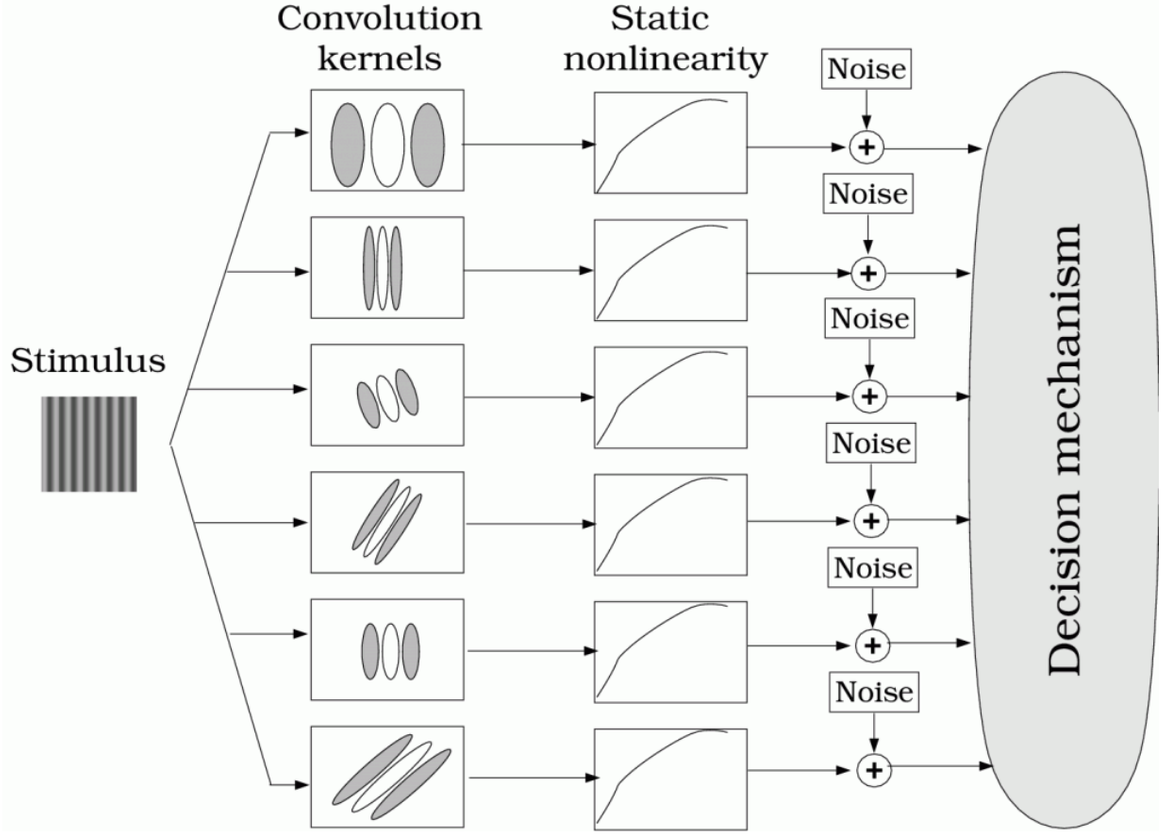


Hierarchical Processing Model

Primary auditory cortex: Shamma 2005

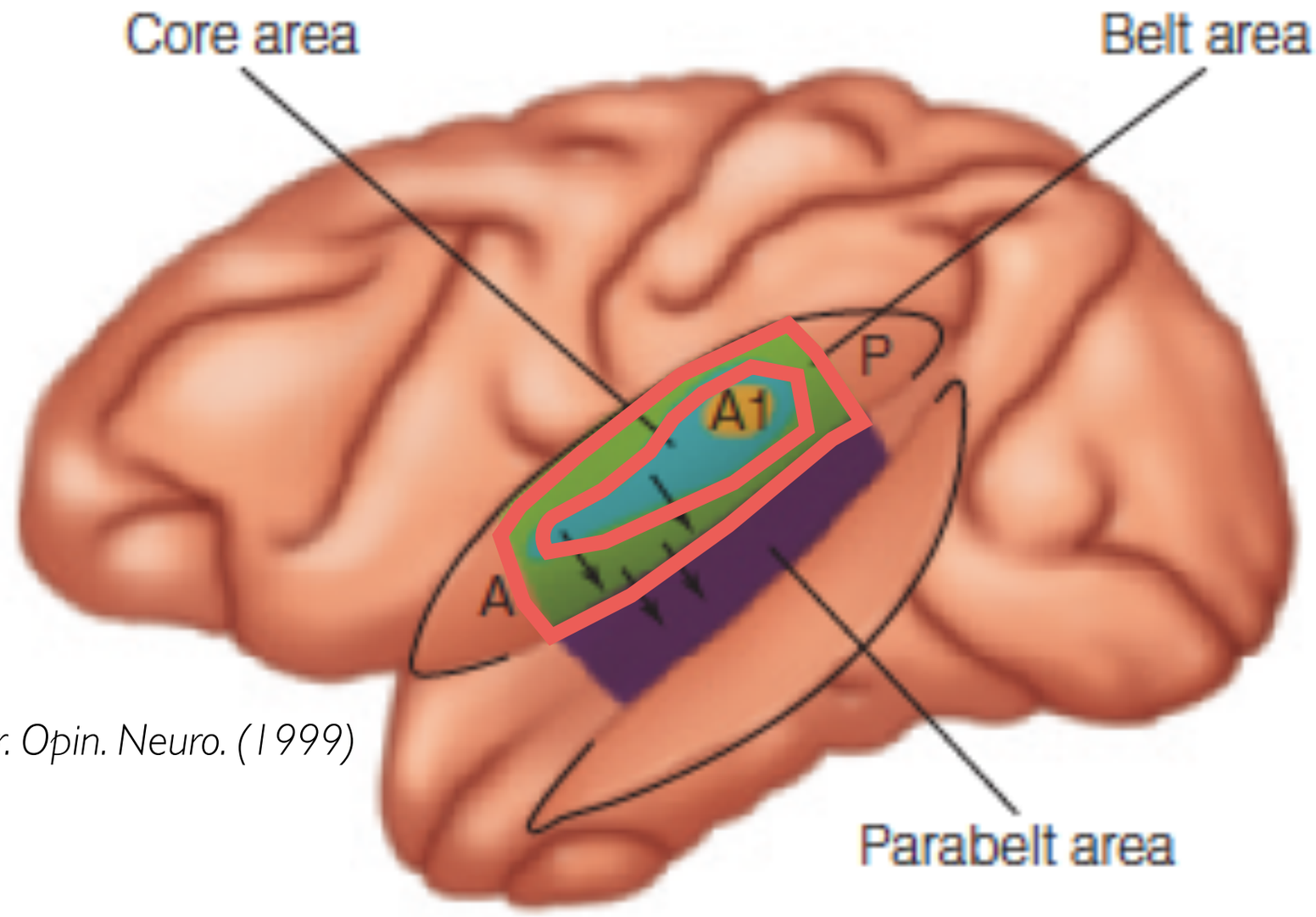


… compare to



Primary visual cortex: Wandell 1996

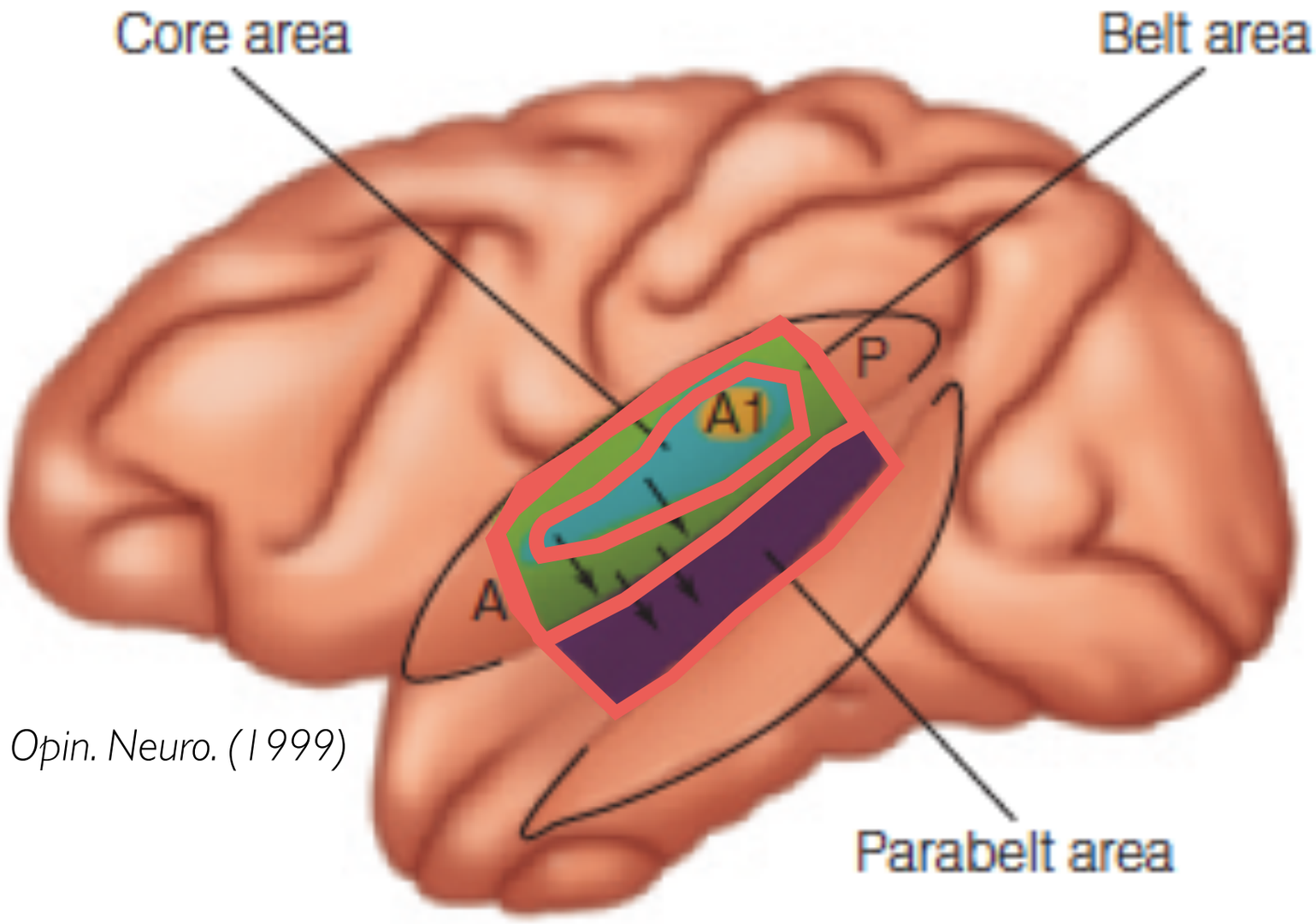
Auditory Cortex and Audition



Tramo et. al, Curr. Opin. Neuro. (1999)

*__monkey__
*

Auditory Cortex and Audition

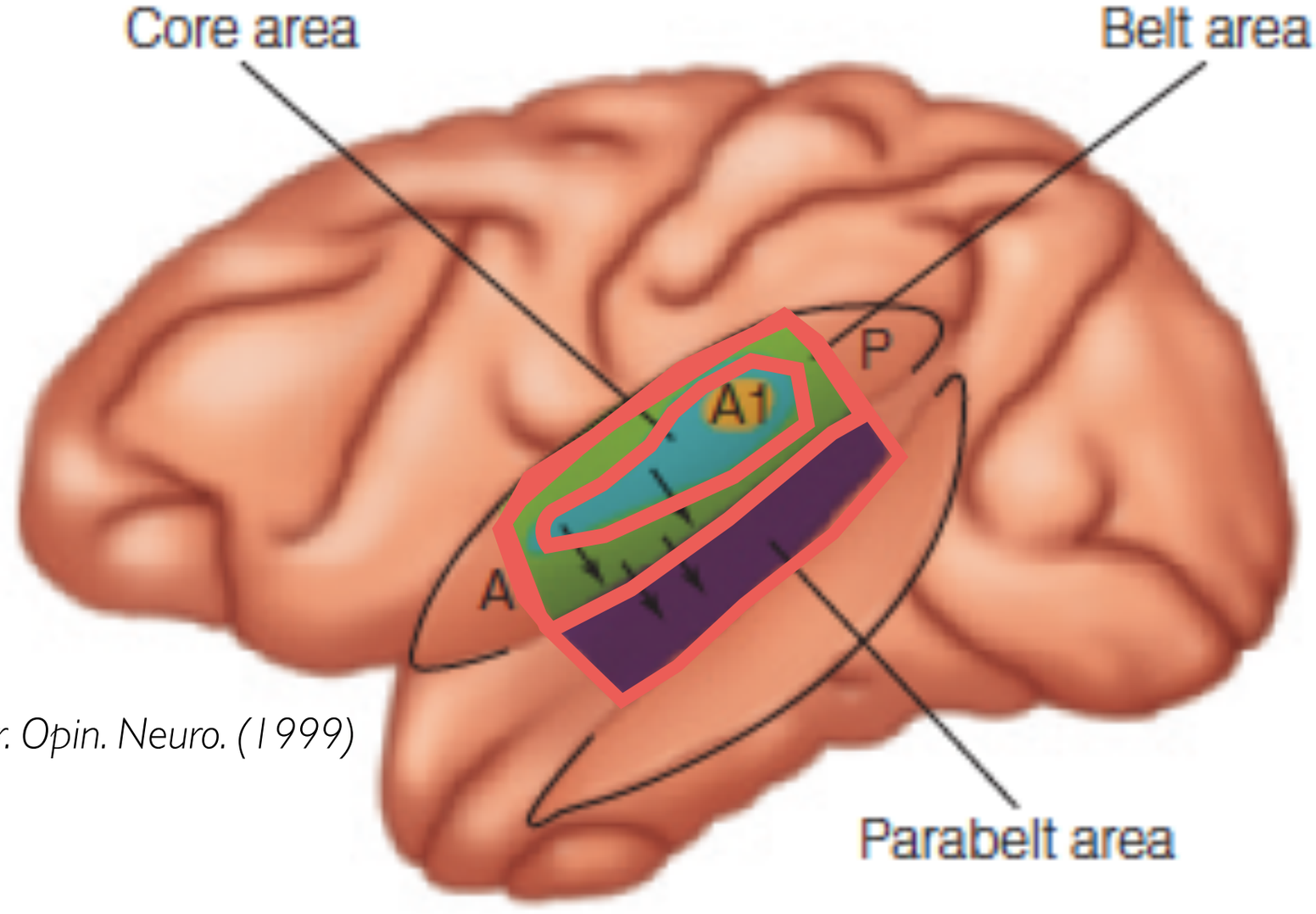


Tramo et. al, Curr. Opin. Neuro. (1999)

*__monkey__
*

Auditory Cortex and Audition

Spectrotemporal filtering? *Shamma, 2005*

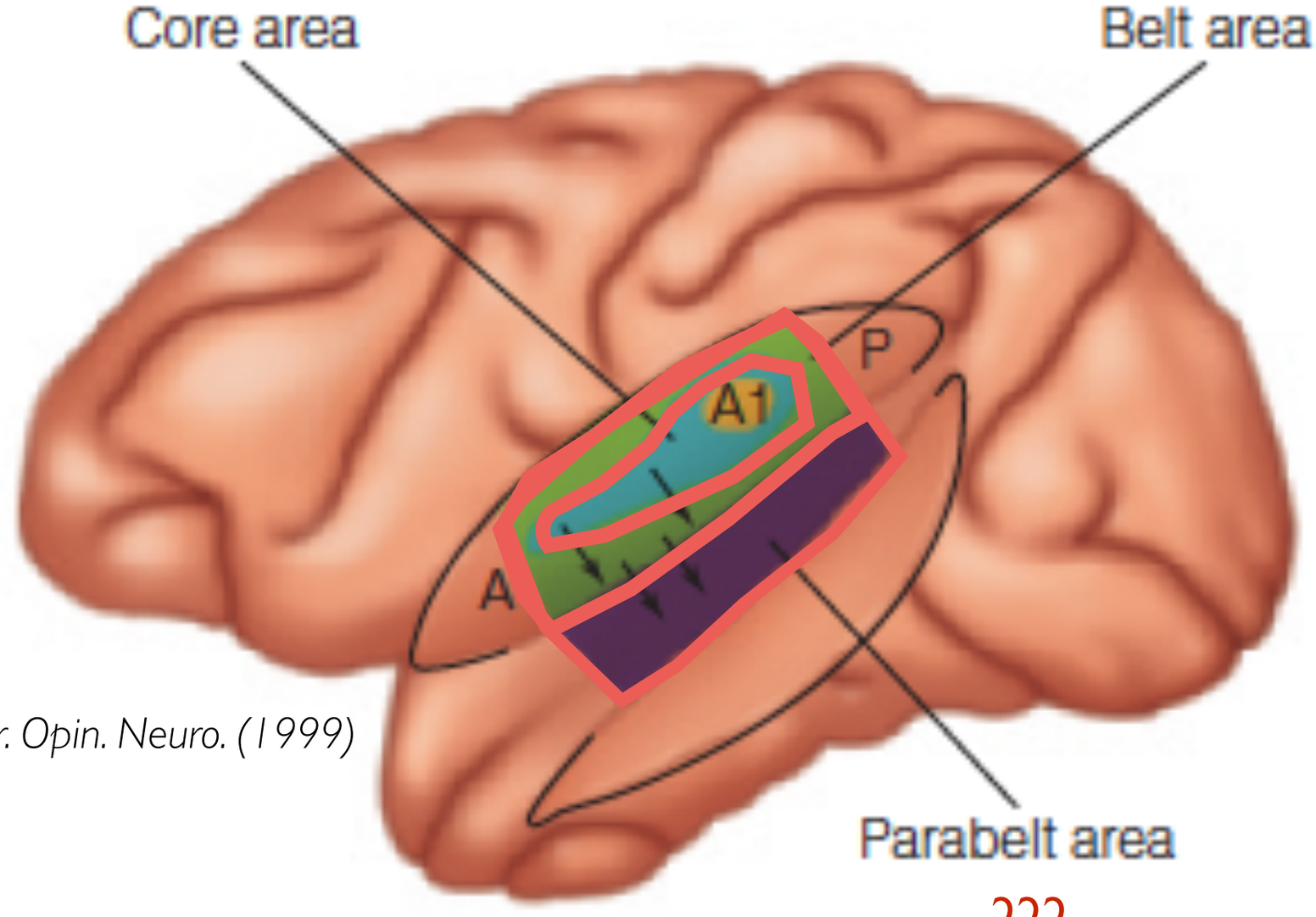


Tramo et. al, Curr. Opin. Neuro. (1999)

*__monkey__
*

Auditory Cortex and Audition

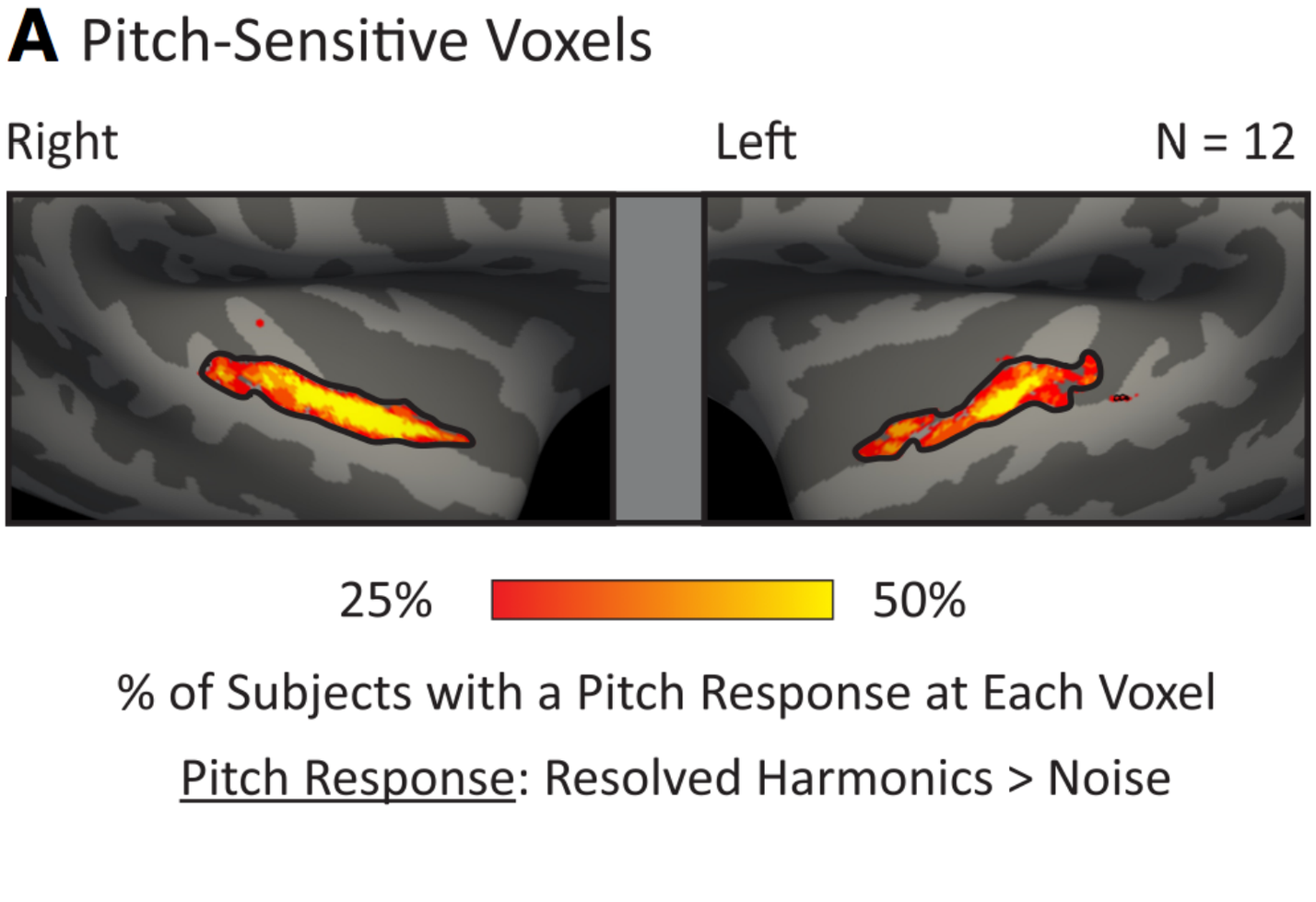
Spectrotemporal filtering? *Shamma, 2005*



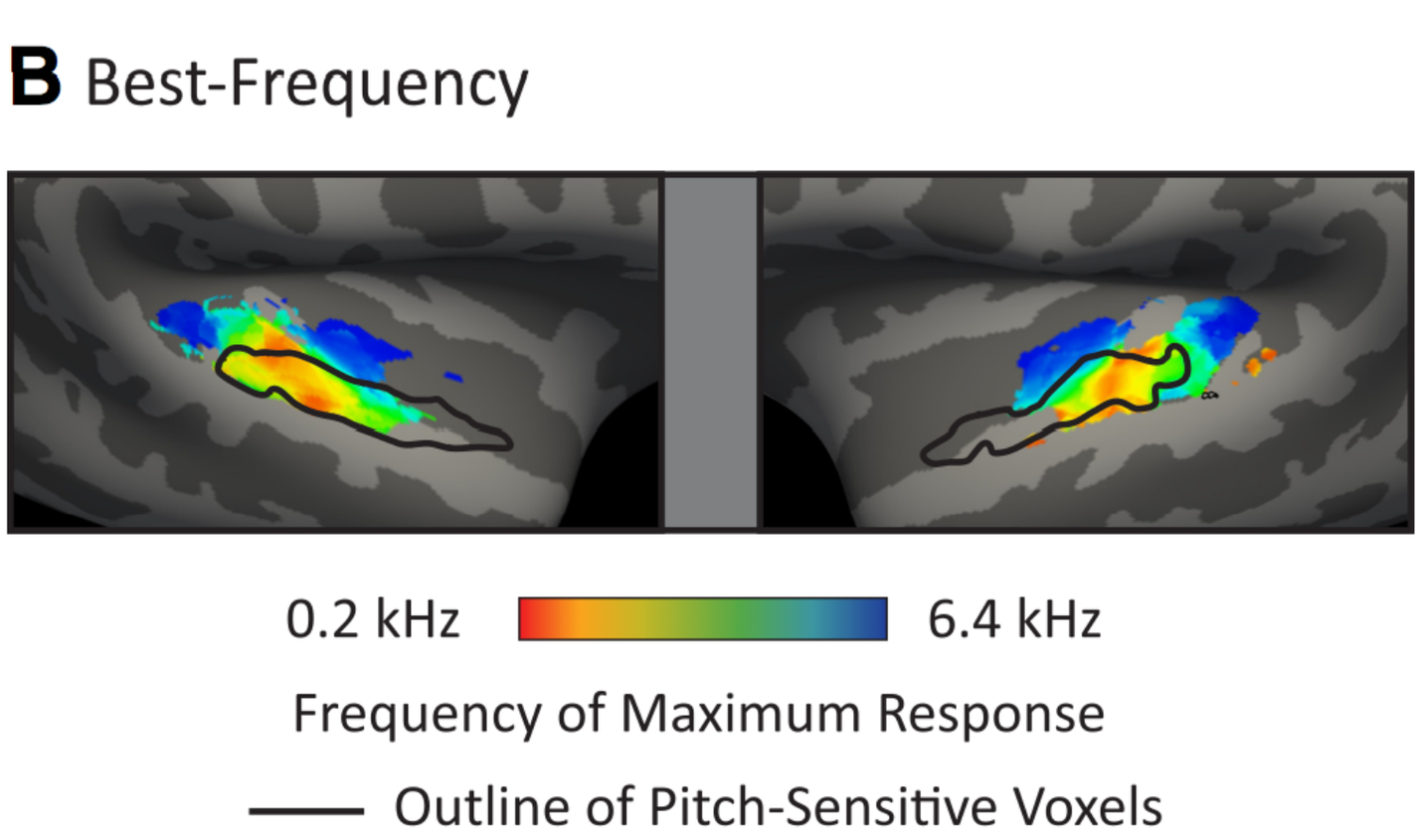
Tramo et. al, Curr. Opin. Neuro. (1999)

*__monkey__
*

Human auditory cortex contains a region that responds more to tones than noise:



Extends out of tonotopic cortex:



Brain (2000), **123,** 2400–2406

Identification of a pathway for intelligible speech in the left temporal lobe

Sophie K. Scott,[1] C. Catrin Blank,[3] Stuart Rosen[2] and Richard J. S. Wise[3]

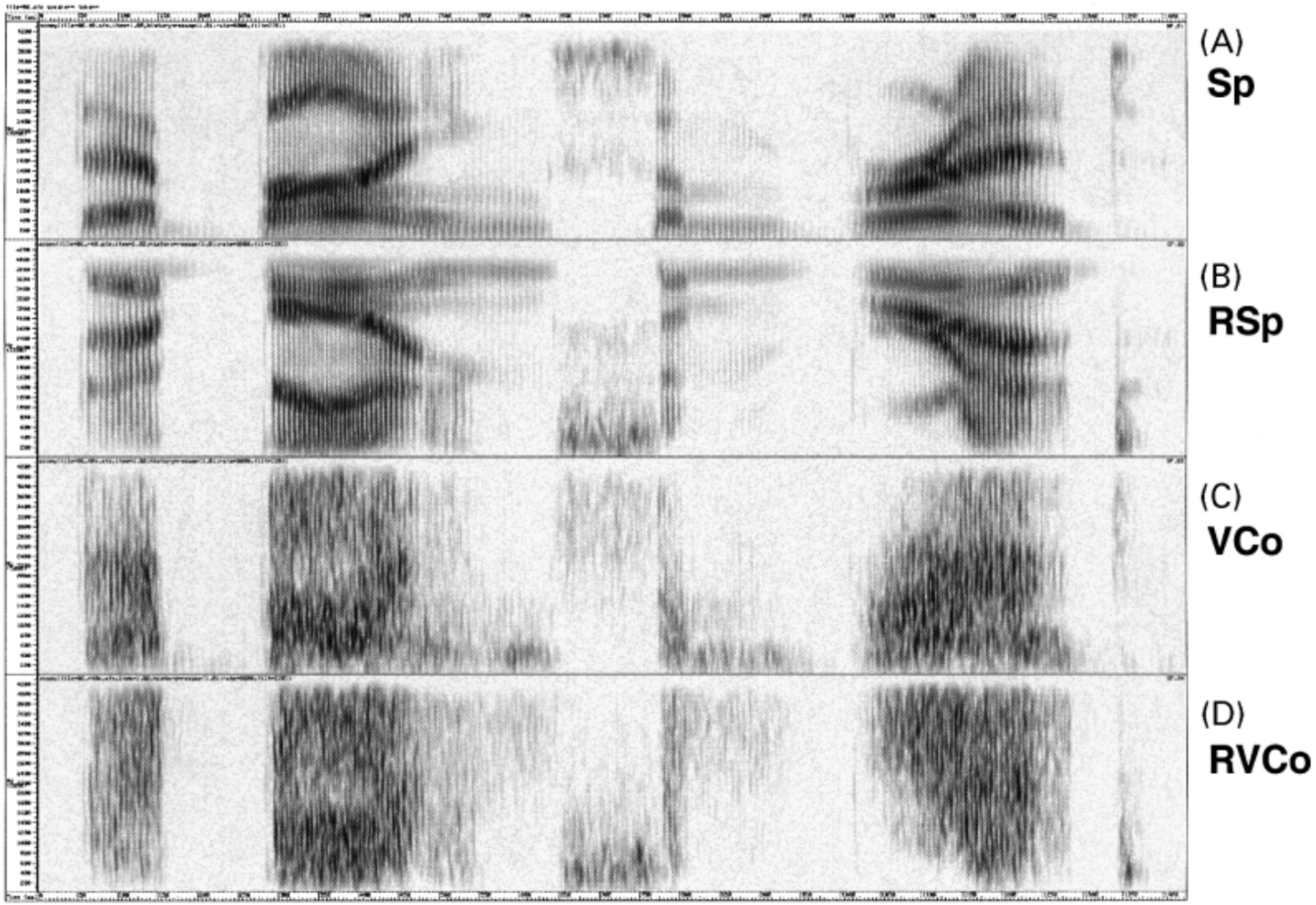
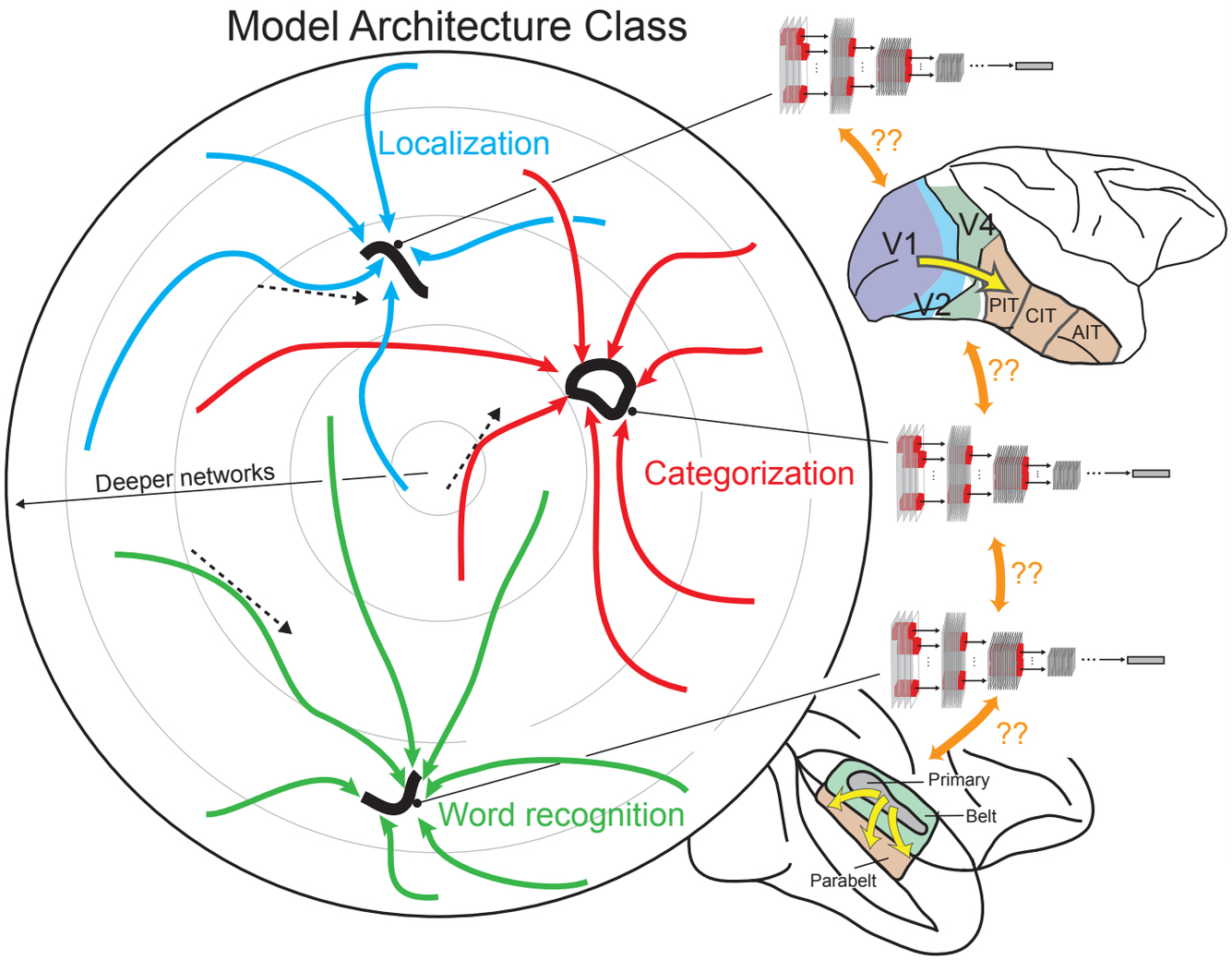


Fig. 1 Spectrograms of 'They're buying some bread'. Time is represented on the abscissa (0.0–1.43 s) and frequency on the ordinate (0.0–4.4 kHz). The darkness of the trace in each time/frequency region is controlled by the amount of energy in the signal at that particular frequency and time. (**A**) Normal speech (Sp) is intelligible with clear intonation. (**B**) Spectrally rotated speech (RSp) is not intelligible without extensive training, though some phonetic features and some of the original intonation are preserved. (**C**) Noise-vocoded speech (VCo) is intelligible, has very weak intonation and a rough sound quality. (**D**) Spectrally rotated noise-vocoded speech (RVCo) is completely unintelligible and does not sound like a voice.



3.

$$\underset{a \in \mathcal{A}}{\operatorname{argmin}}[L(p_a^*)]$$

where p is result of*

$$\frac{dp_a}{dt} = -\lambda(t) \cdot \langle \nabla_{p_a} L(x) \rangle_{x \in \mathcal{D}}$$

"learning rule"

1.

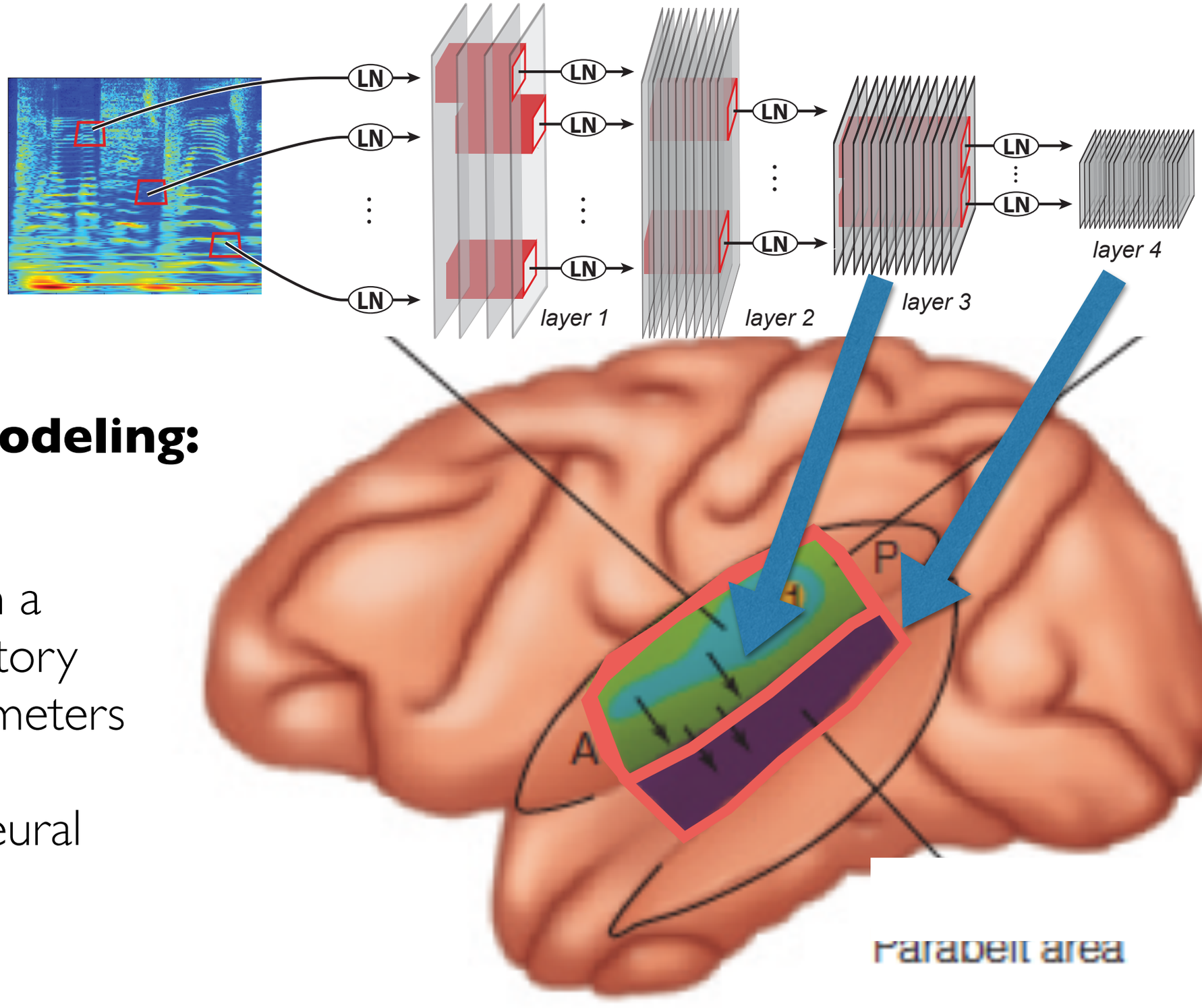
A = *architecture class*

2.

L = *loss function* ***D*** = *dataset*

"task"

Core Task-Driven Modeling Idea



Task-Driven Modeling:

1. Optimize for performance on a challenging auditory task, fixing parameters

2. Compare to neural data.

Apply to auditory tasks, where the regions themselves are less well known.

Core Task-Driven Modeling Idea

Alex Kell

Josh McDermott

Optimize for Performance: The Task

600-way word-recognition task assembled by:

- Recordings from standard speech recognition databases (TIMIT, WSJ) with words spoken at least 20 times

- Combined with significant background noise

  - auditory scenes
  - speech babble
  - music clips

*"She **had** your*	*'had'*
*dark **suit** in*	*'suit'*
*greasy **wash** water*	*'wash'*
*all **year** … "*	*'year'*

Optimize for Performance: The Task

600-way word-recognition task assembled by:

- Recordings from standard speech recognition databases (TIMIT, WSJ) with words spoken at least 20 times

- Combined with significant background noise

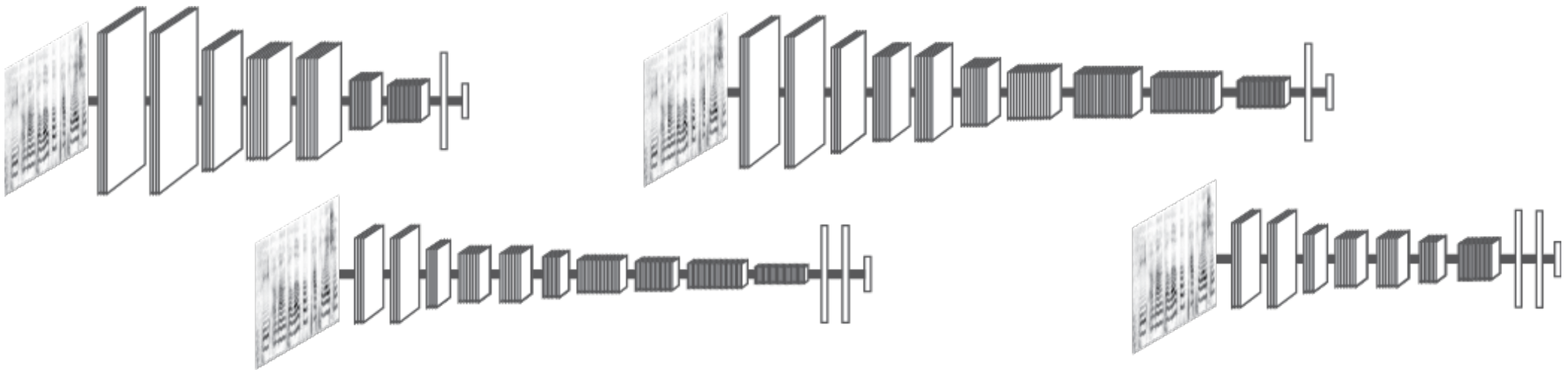
  - auditory scenes
  - speech babble
  - music clips

Backgrounds → humans not close to ceiling.

Optimize for Performance: The Task

Task: 600-way word-recognition task.

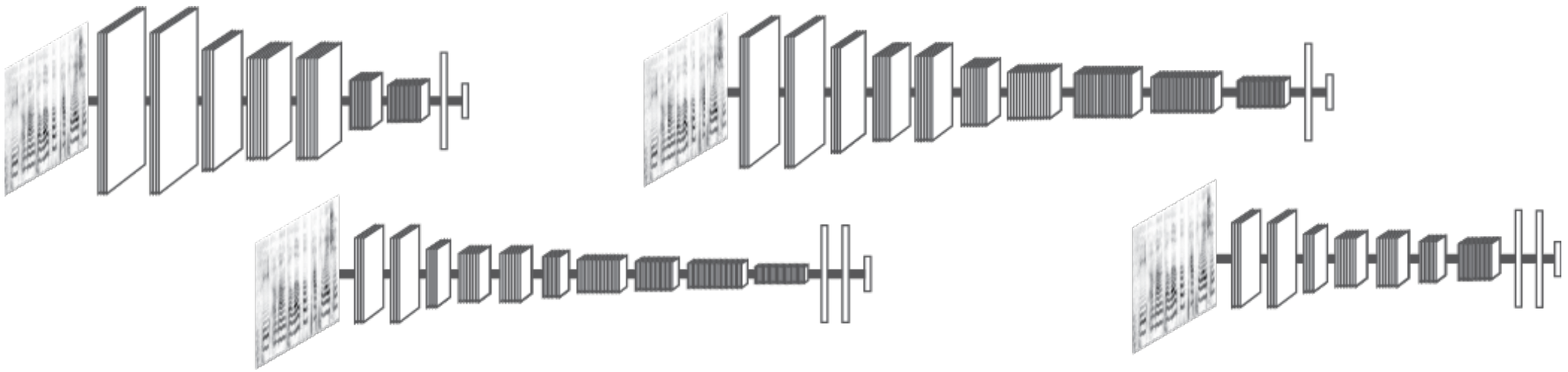
Architecture: Hyperparameter search over 1-D and 2-D convolutional structures, with different numbers of layers, kernel sizes, operations, &c.



Optimize for Performance: The Task

Task: 600-way word-recognition task.

Architecture: Hyperparameter search over 1-D and 2-D convolutional structures, with different numbers of layers, kernel sizes, operations, &c.



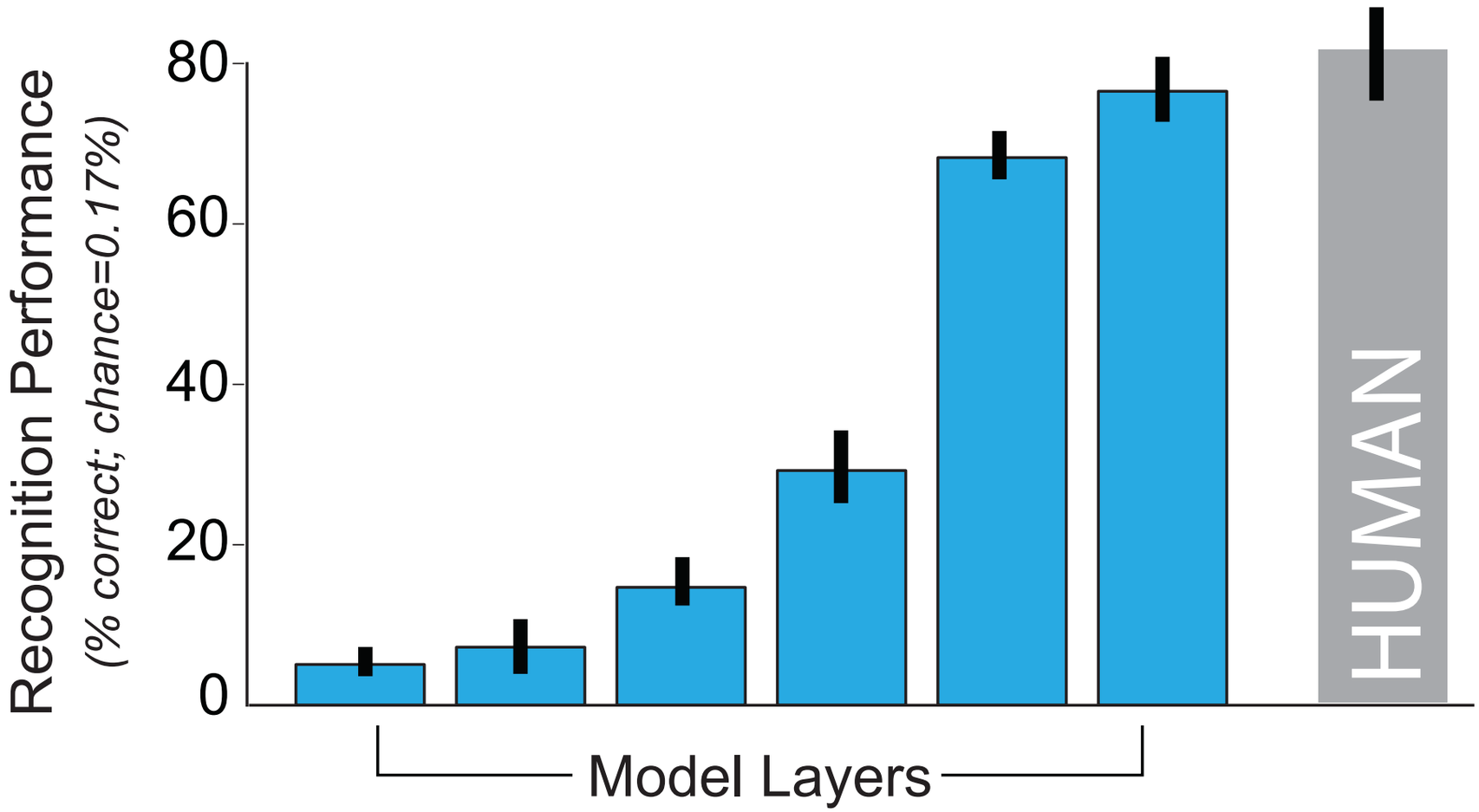
Crucial fact: use of some order 2 pooling rather than max:

$$y = \left(\frac{1}{|N_r|} \sum_{i \in N_r} x_i^p \right)^{1/p}$$

p = 2

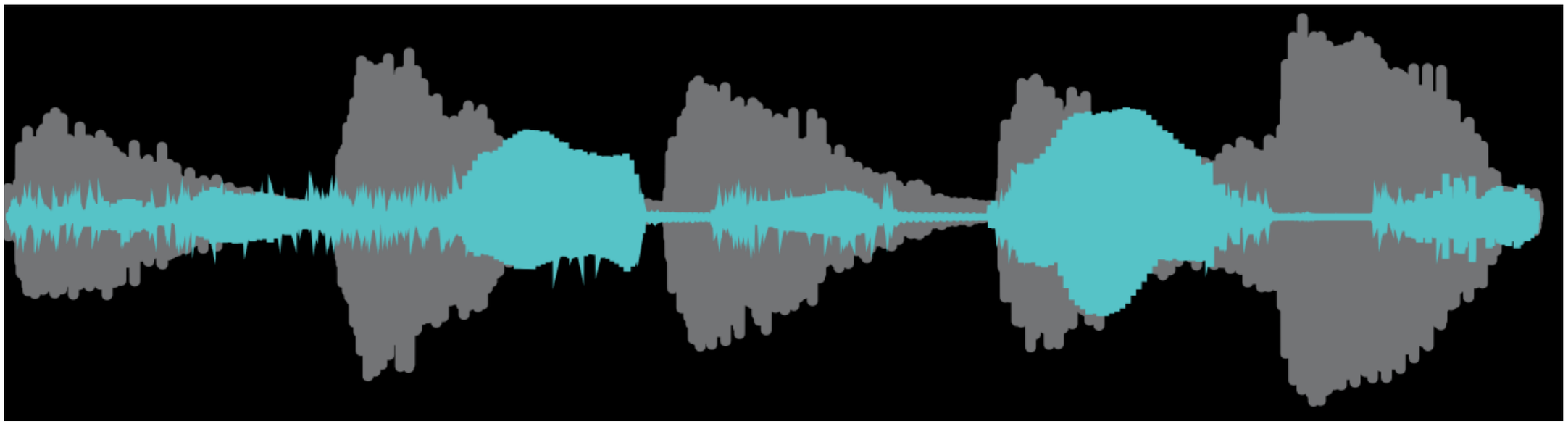
Performance Results

Performance on 600-way word-recognition task



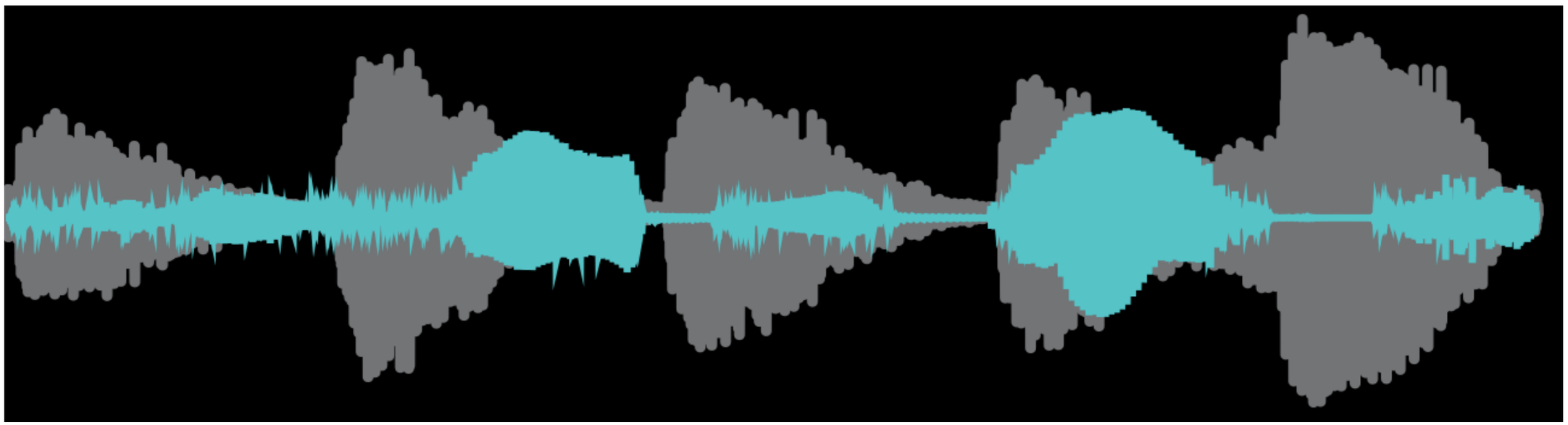
… for model, measured on held-out data with novel speakers and auditory background noise.

Behavioral comparison:
CNN & humans on same task



Word recognition in complex backgrounds

Behavioral comparison:
CNN & humans on same task



Word recognition in complex backgrounds

21 conditions:

dry

+

4 different background types at 5 SNR levels:

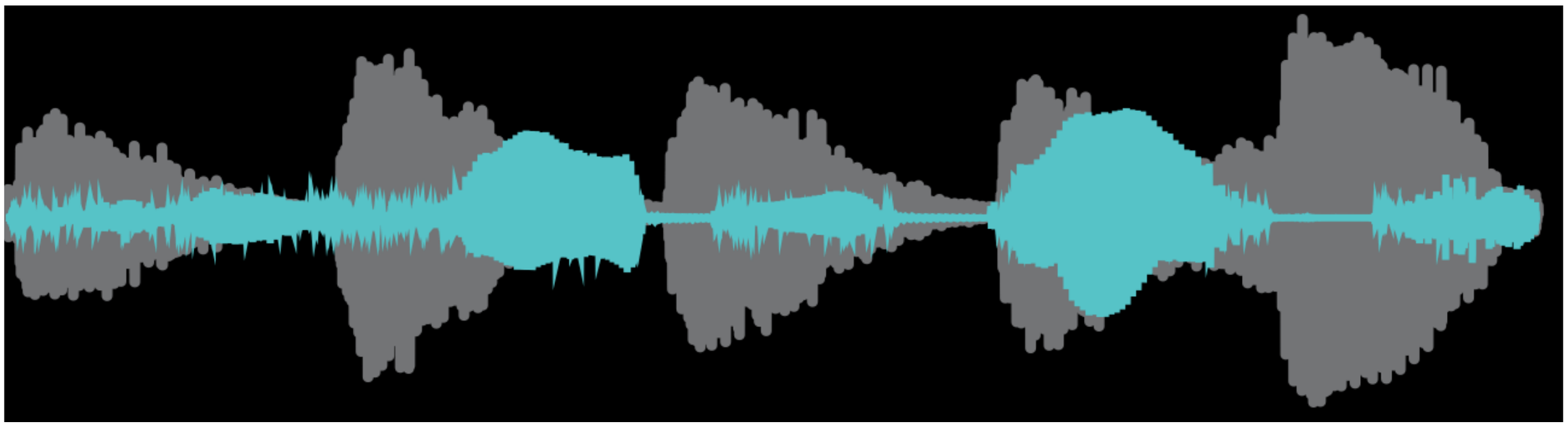
Auditory scenes

Speech babble

Music

Speech-shaped noise

Behavioral comparison:
CNN & humans on same task



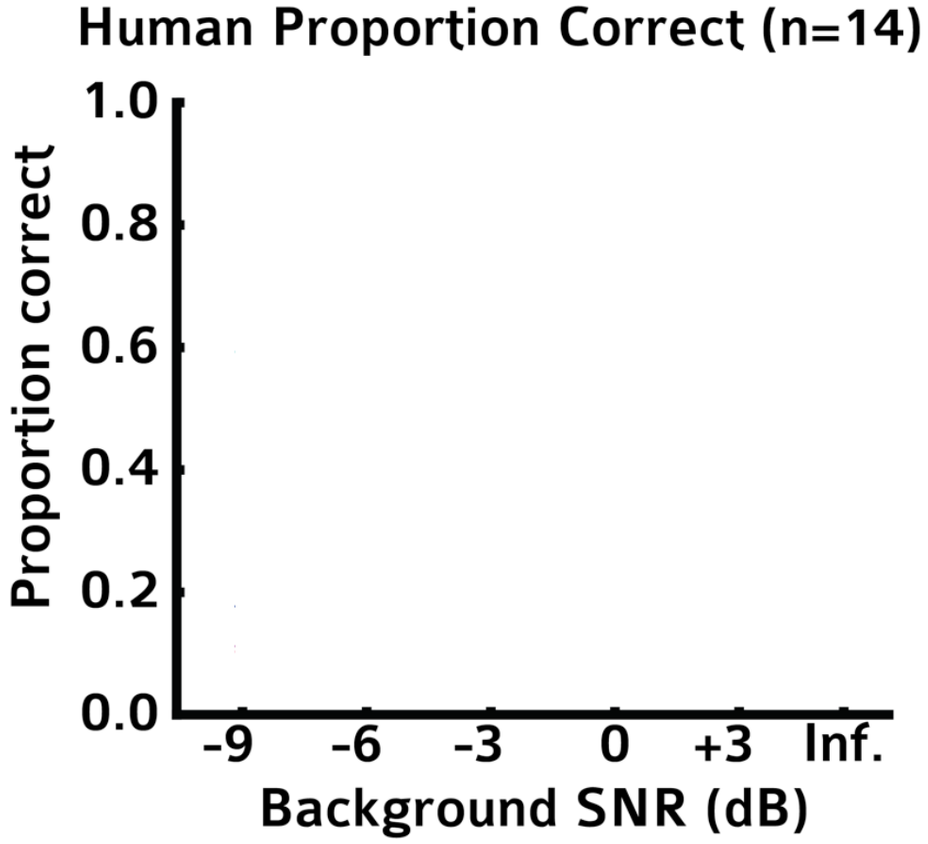
Word recognition in complex backgrounds

600
AFC

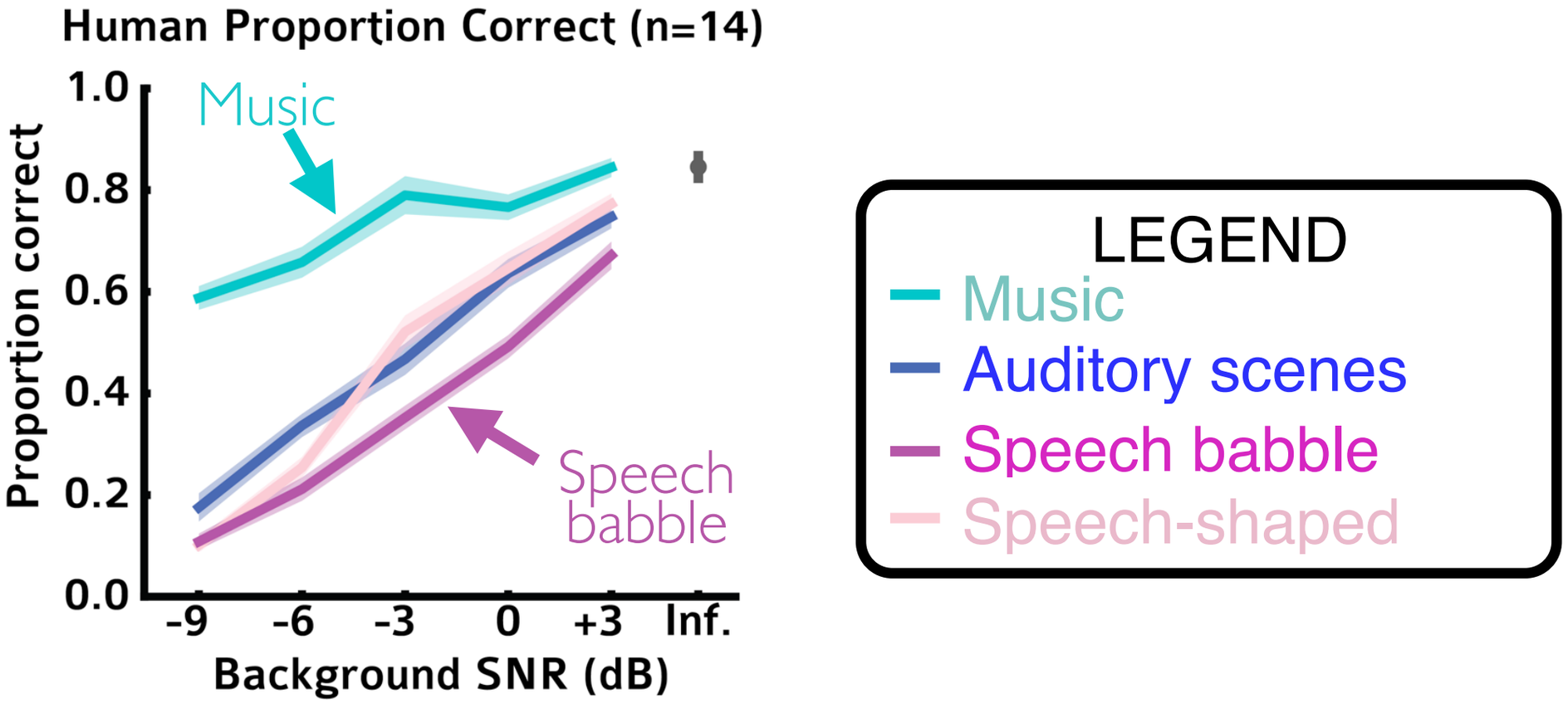
21 conditions:
dry
+
4 different background types at 5 SNR levels:

Auditory scenes
Speech babble
Music
Speech-shaped noise

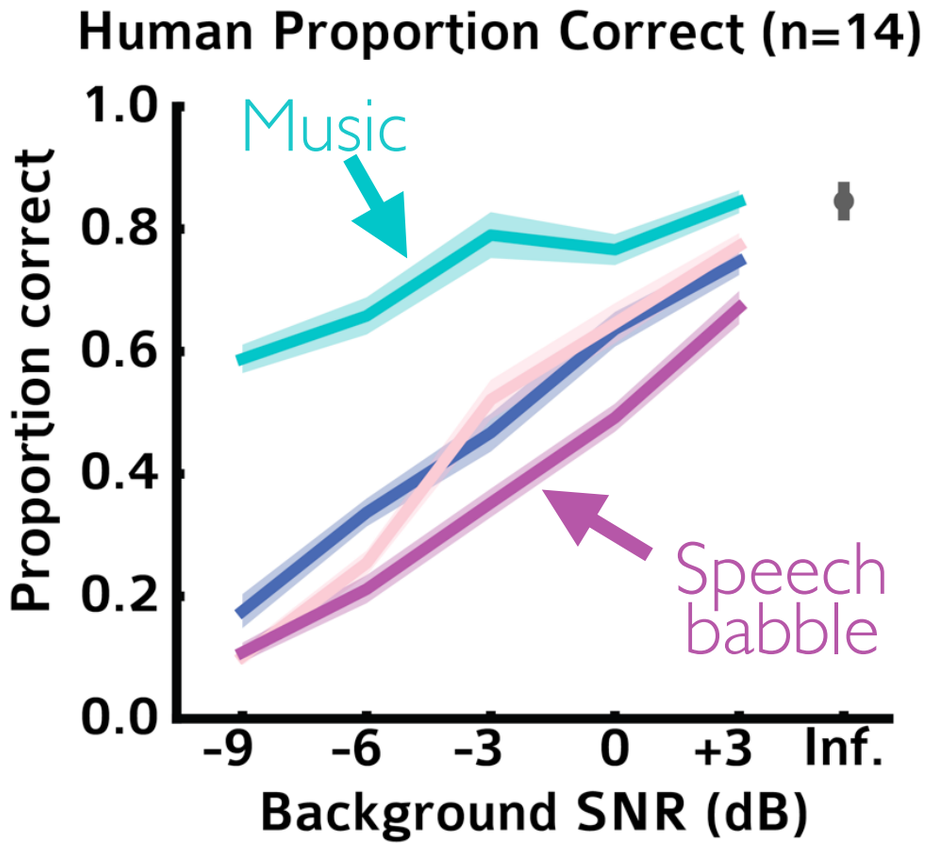
Behavioral comparison: CNN & humans on same task

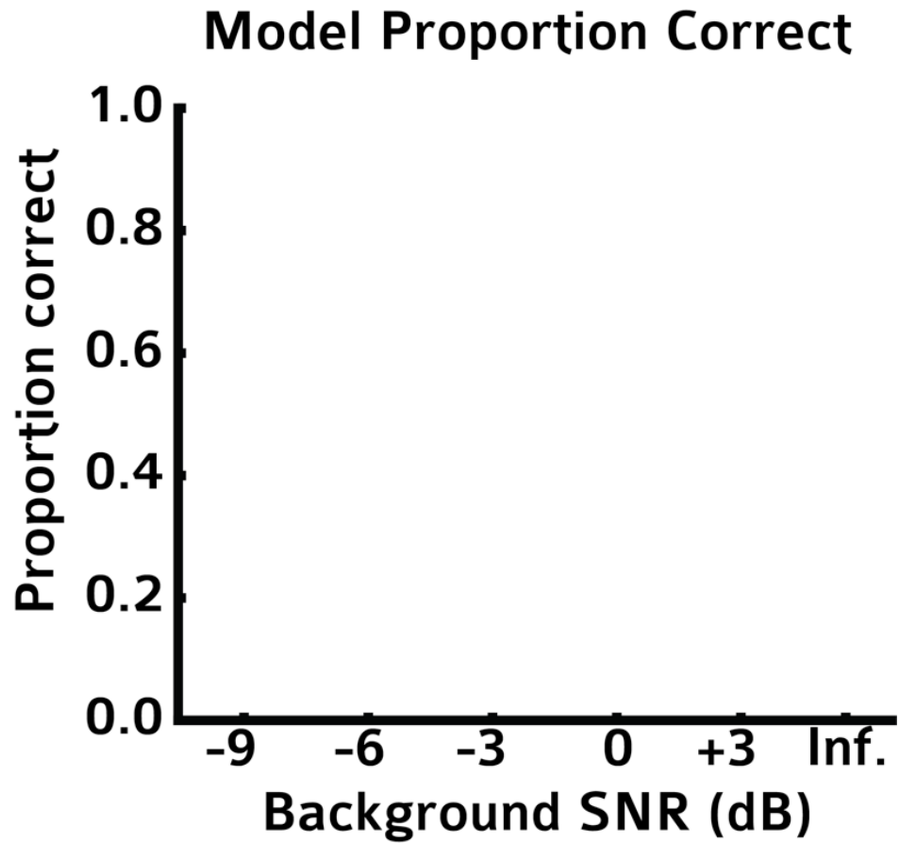


Behavioral comparison: CNN & humans on same task



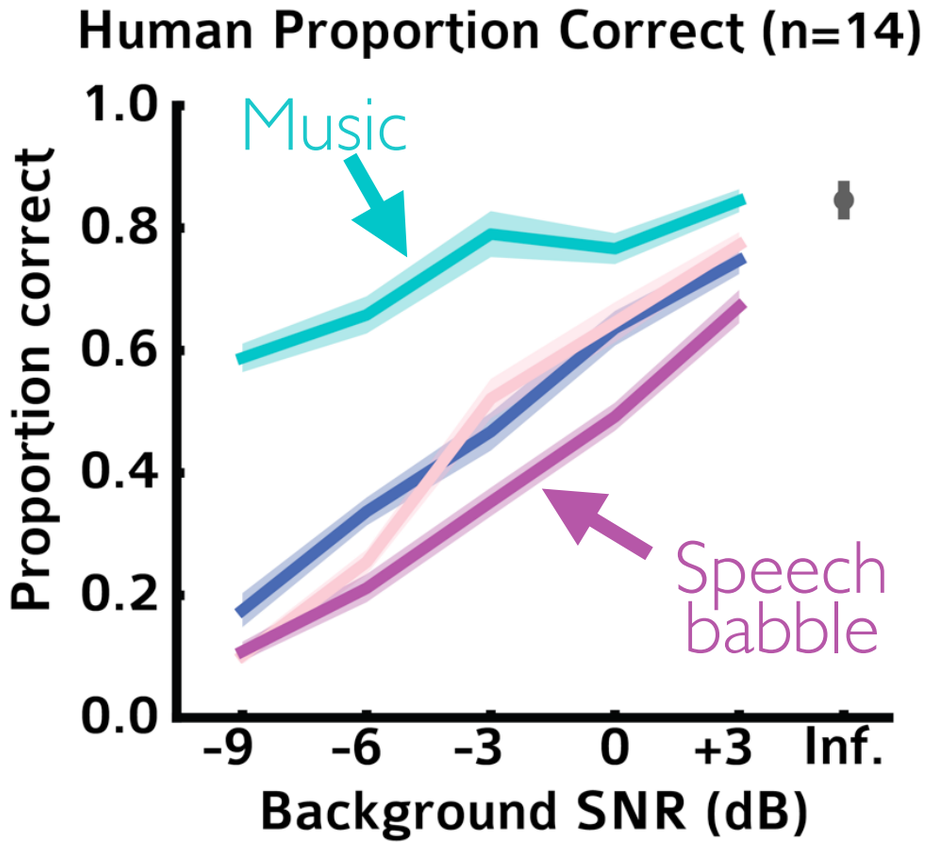
Behavioral comparison: CNN & humans on same task

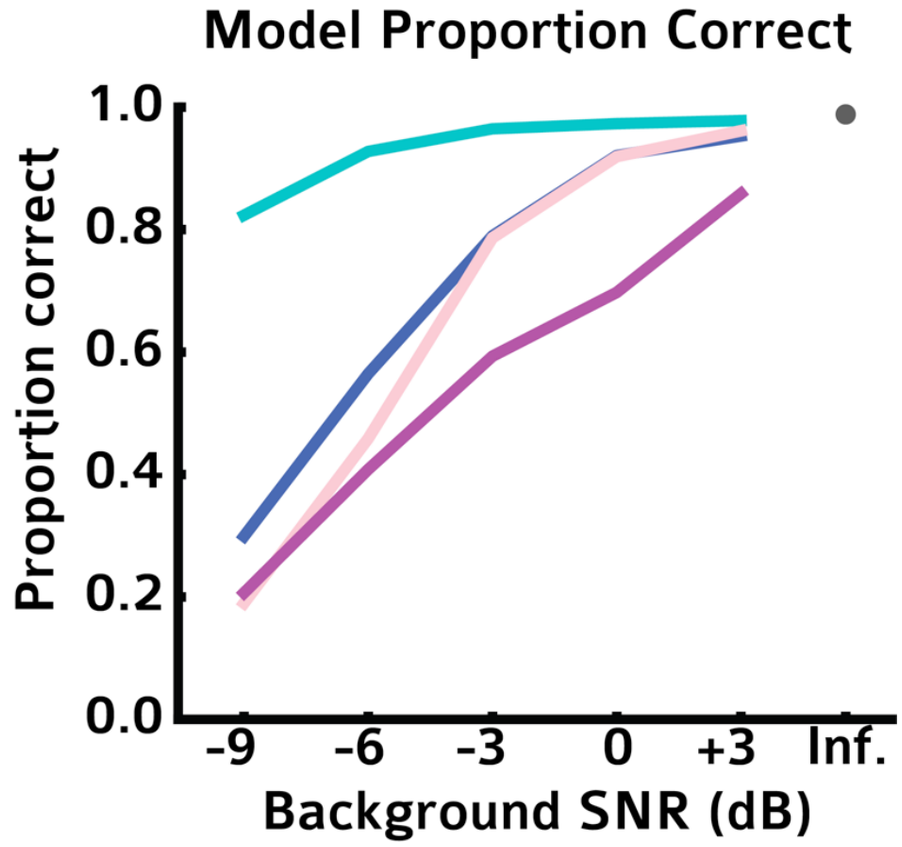




LEGEND
- Music
- Auditory scenes
- Speech babble
- Speech-shaped

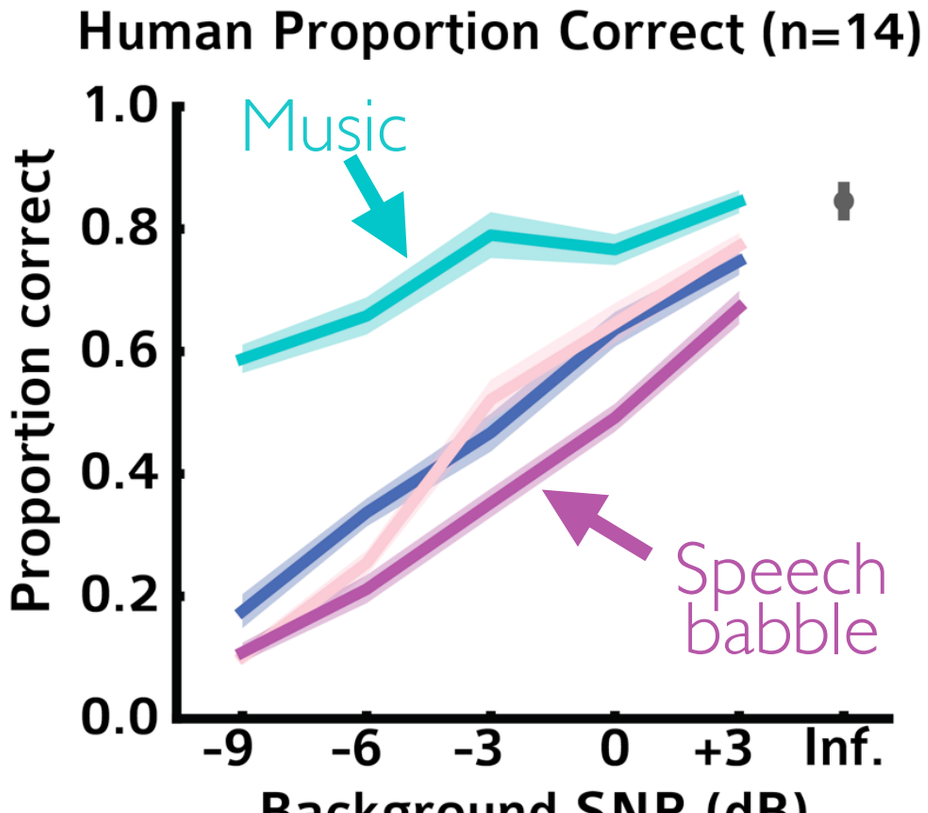
Behavioral comparison: CNN & humans on same task

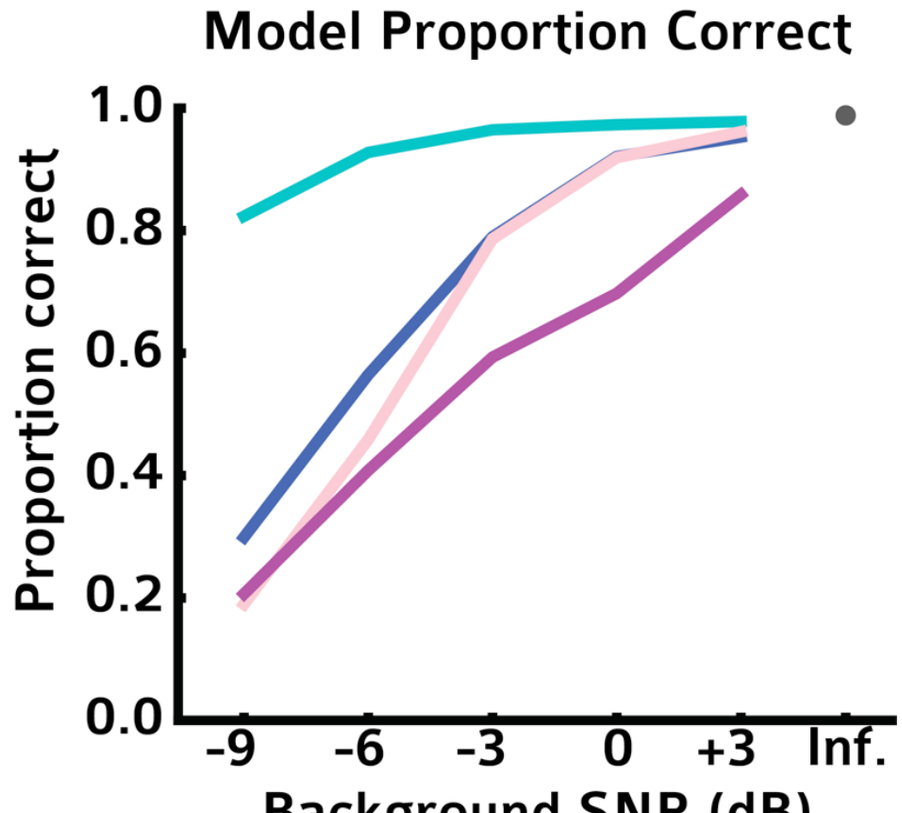




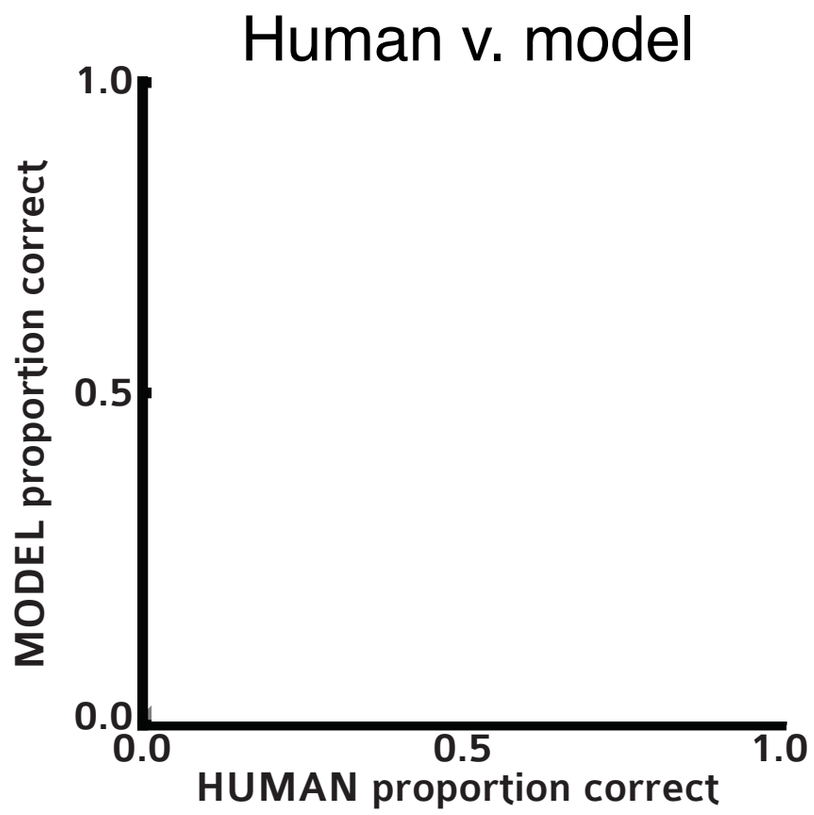
LEGEND
- Music
- Auditory scenes
- Speech babble
- Speech-shaped

Behavioral comparison: CNN & humans on same task

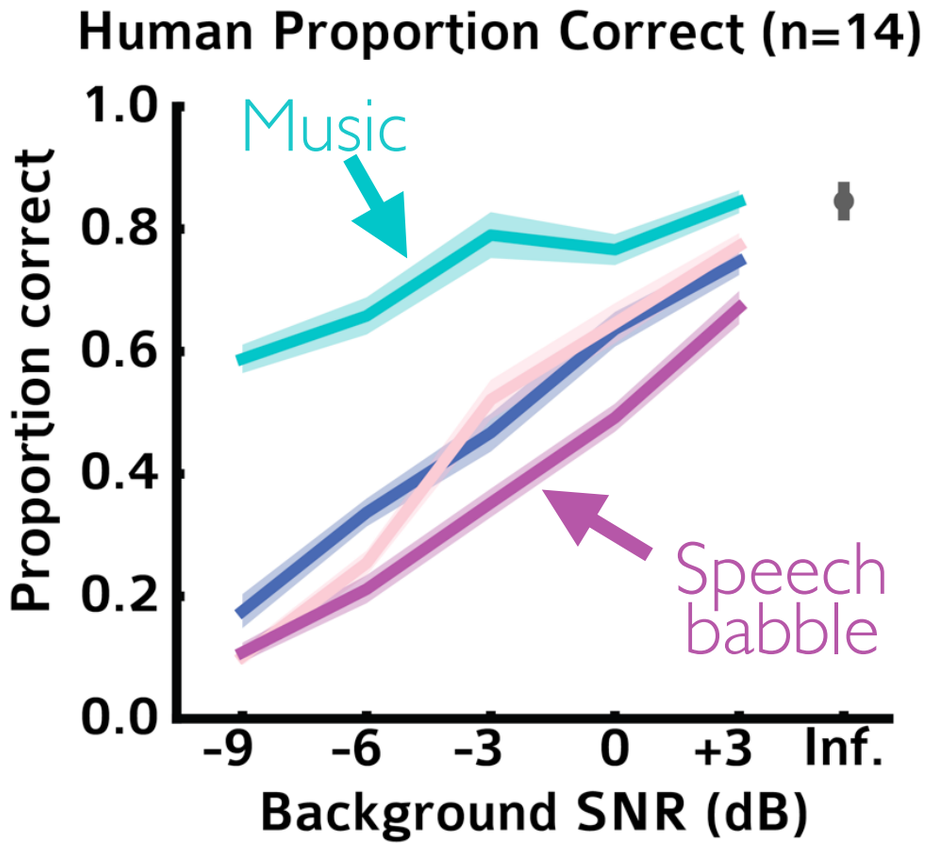


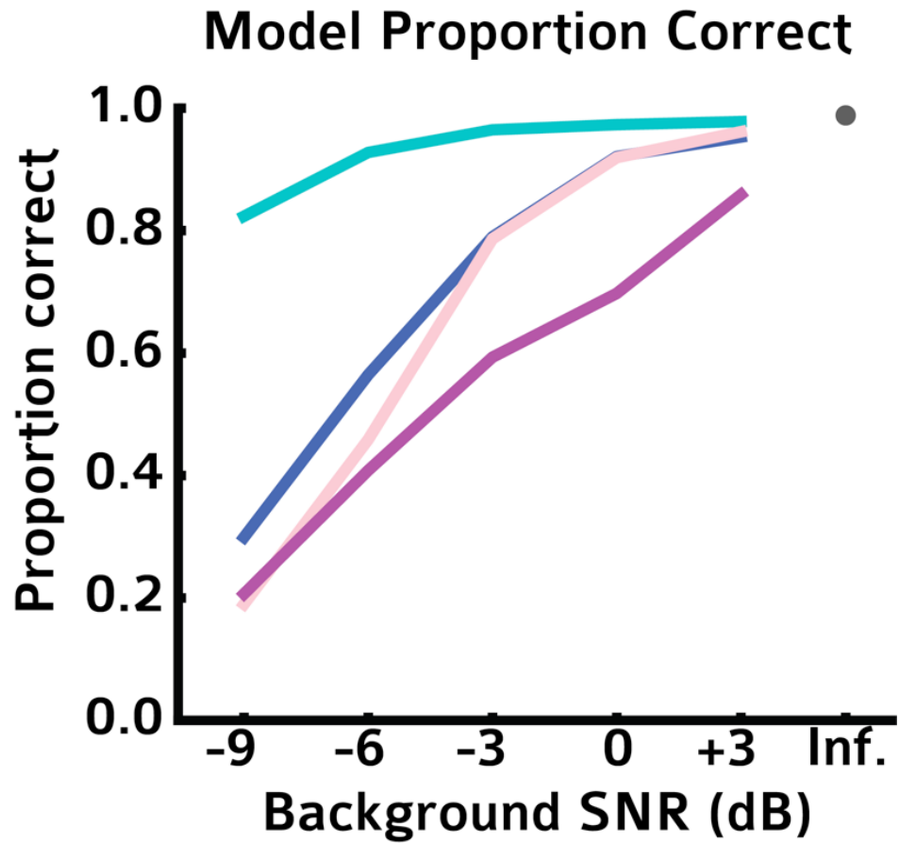


LEGEND
- Music
- Auditory scenes
- Speech babble
- Speech-shaped

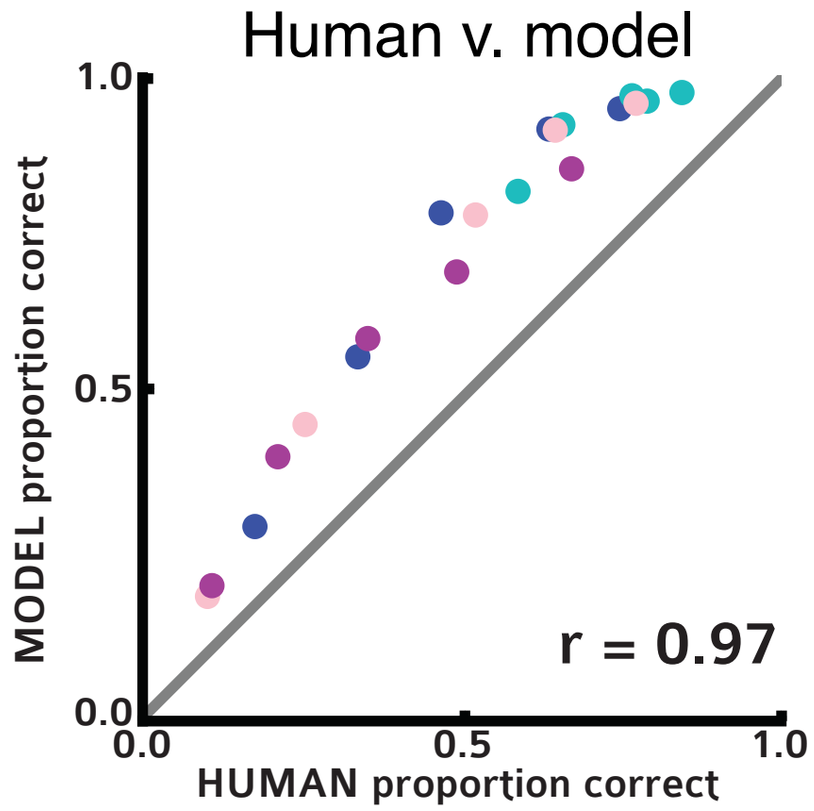


Behavioral comparison: CNN & humans on same task

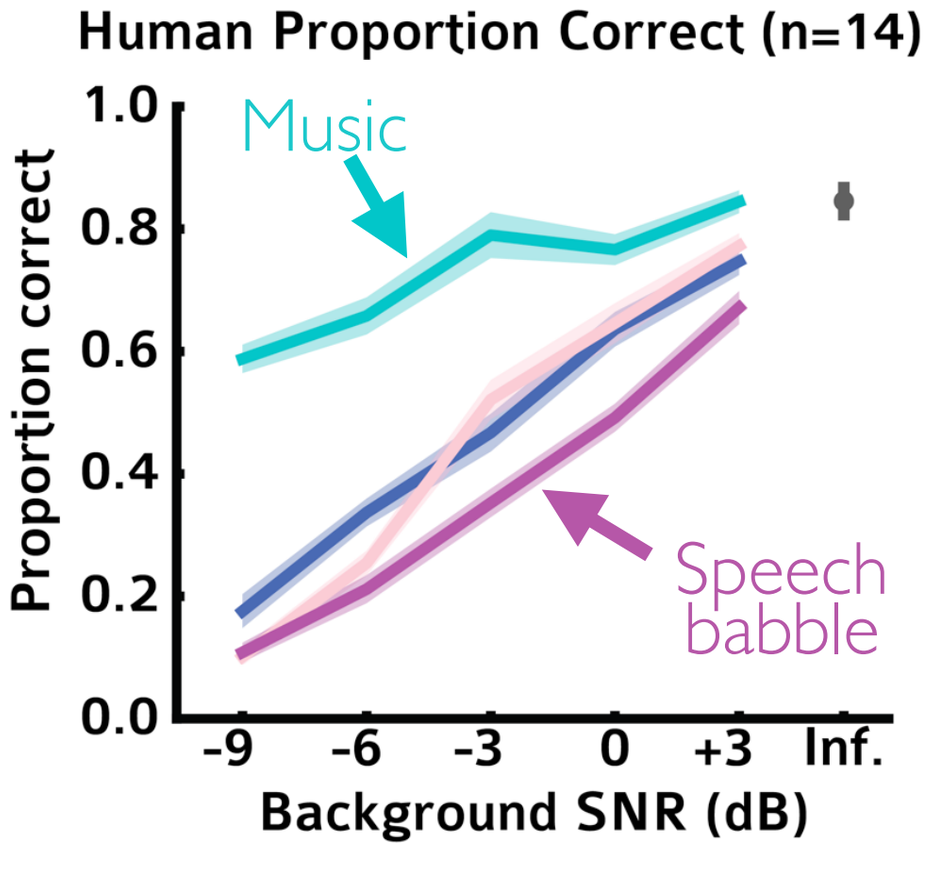


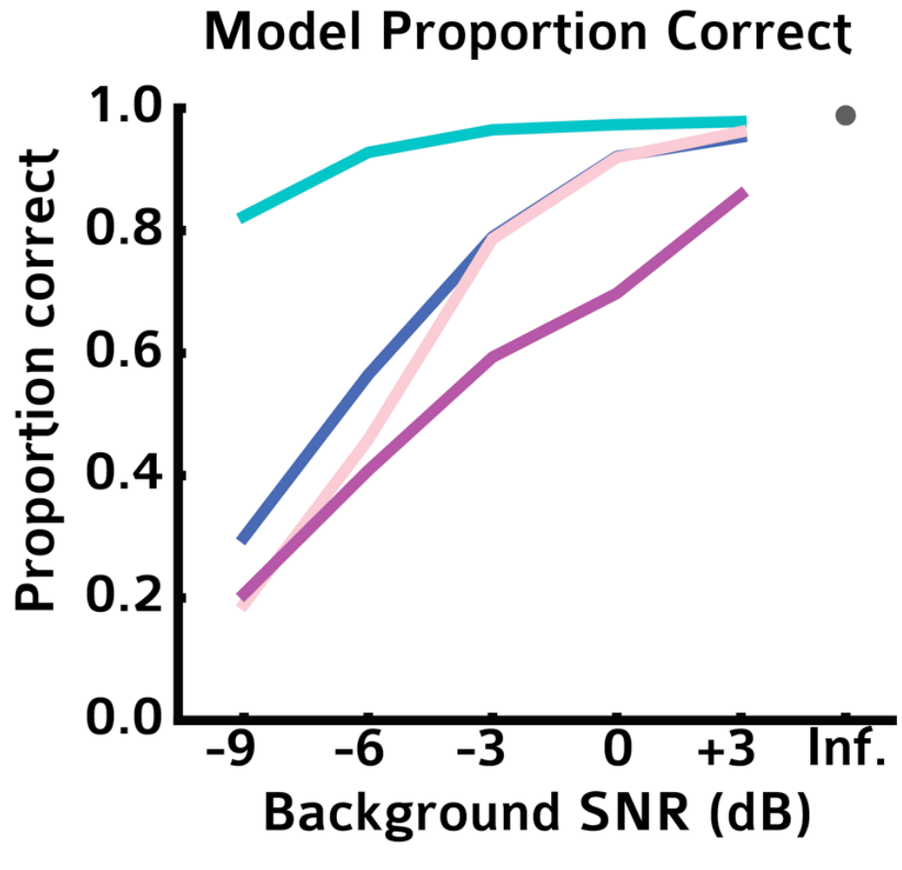


LEGEND
- Music
- Auditory scenes
- Speech babble
- Speech-shaped



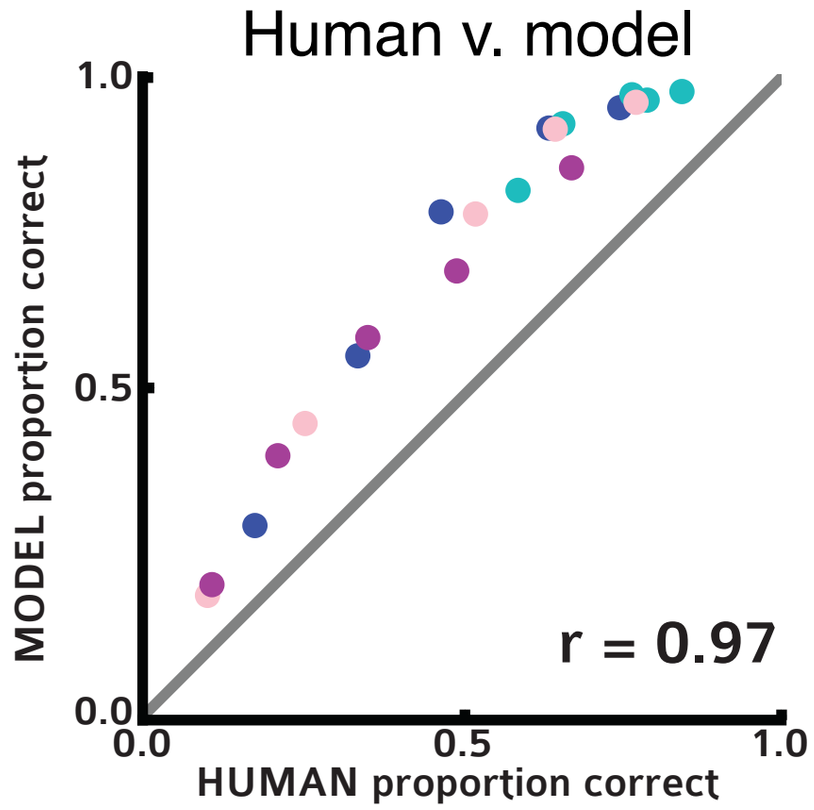
Behavioral comparison: CNN & humans on same task





LEGEND

- Music
- Auditory scenes
- Speech babble
- Speech-shaped



NB:

CNN optimized for task performance ***not*** for human behavior match

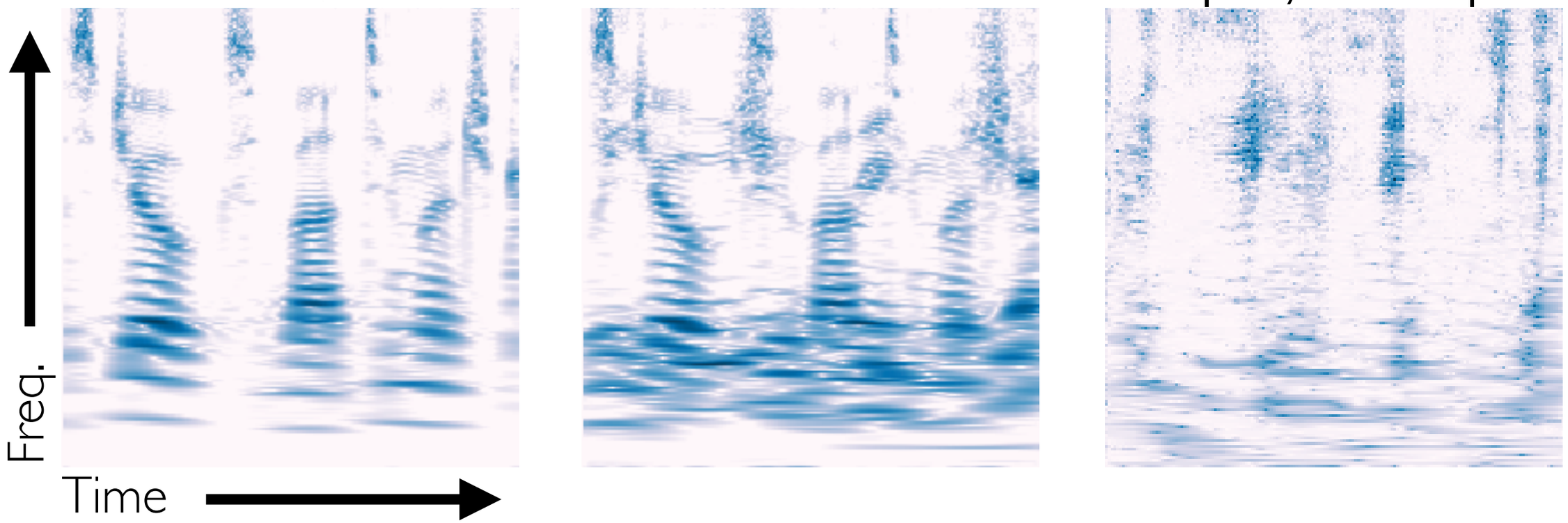
Does distortion in a periphery-like representation explain pattern of performance?

Measure physical distortion of background noise

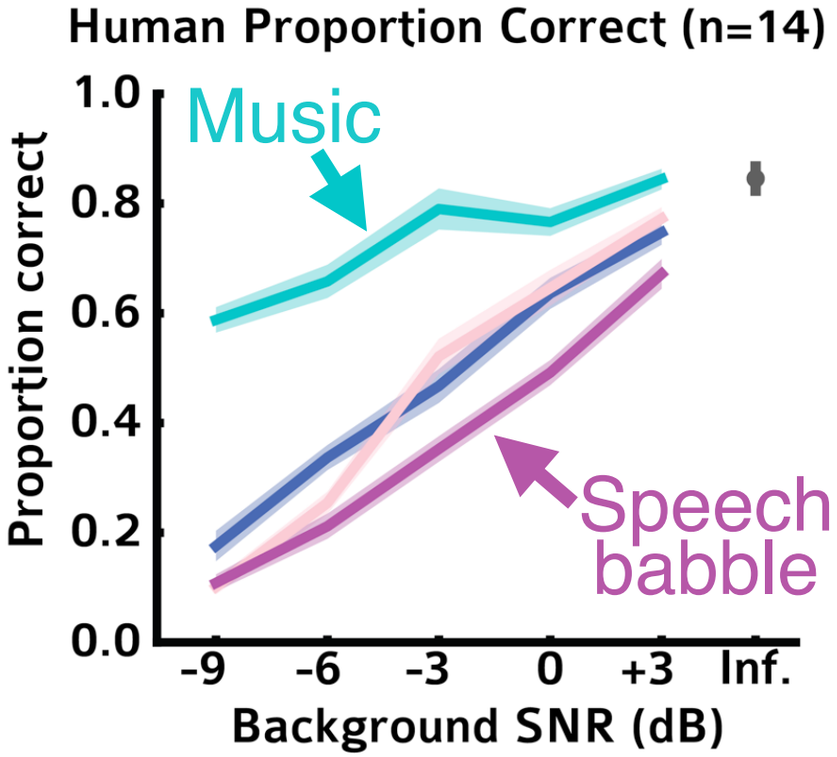
Dry

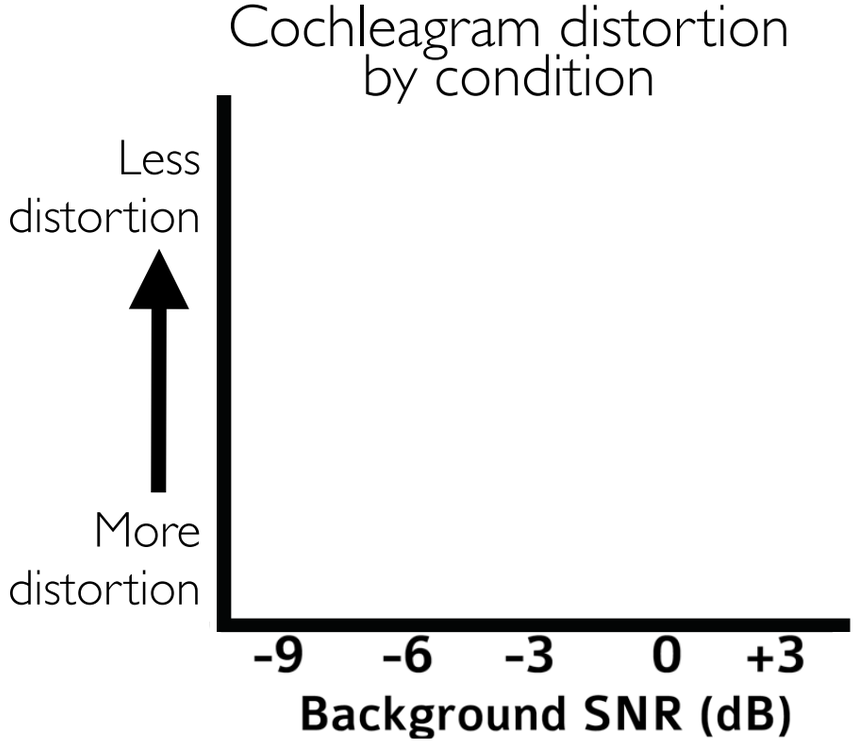
Wet

|Dry - Wet|



Distortion and pattern of performance.

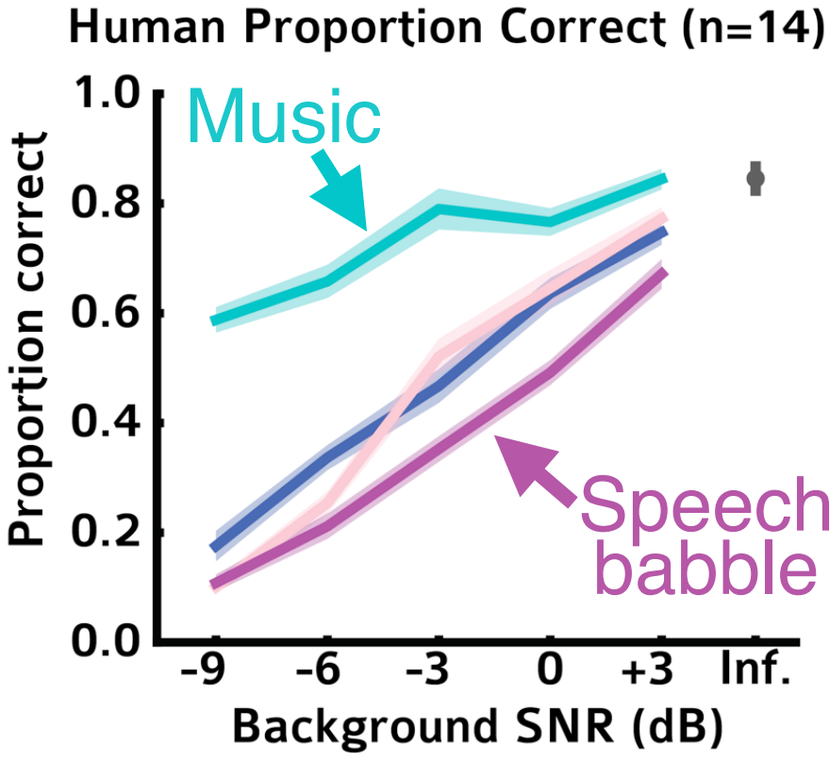


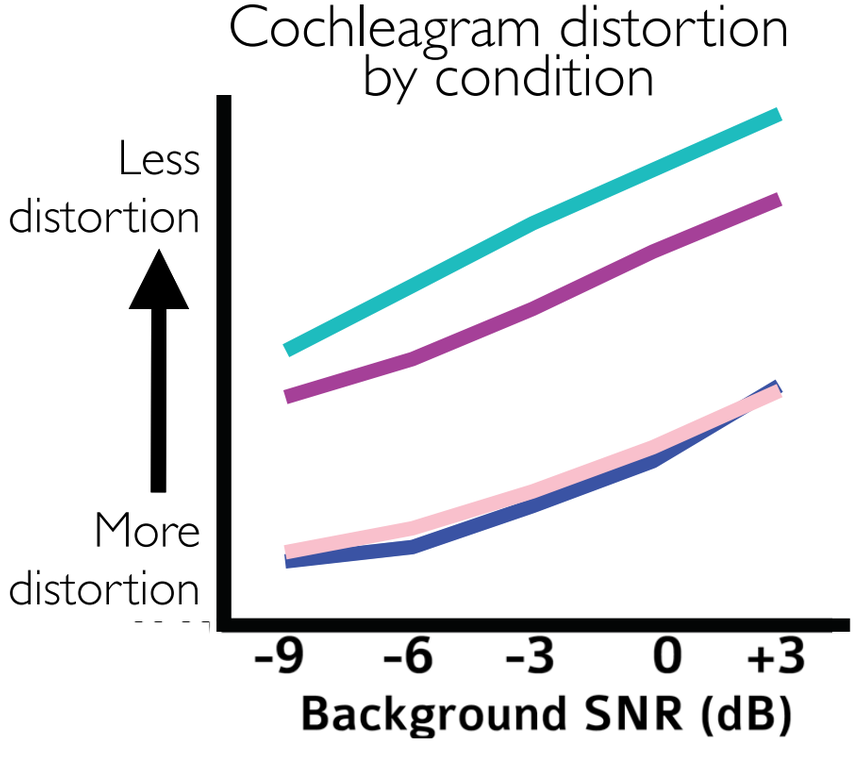


LEGEND

Music Auditory scenes Speech babble Speech-shaped noise

Distortion and pattern of performance.

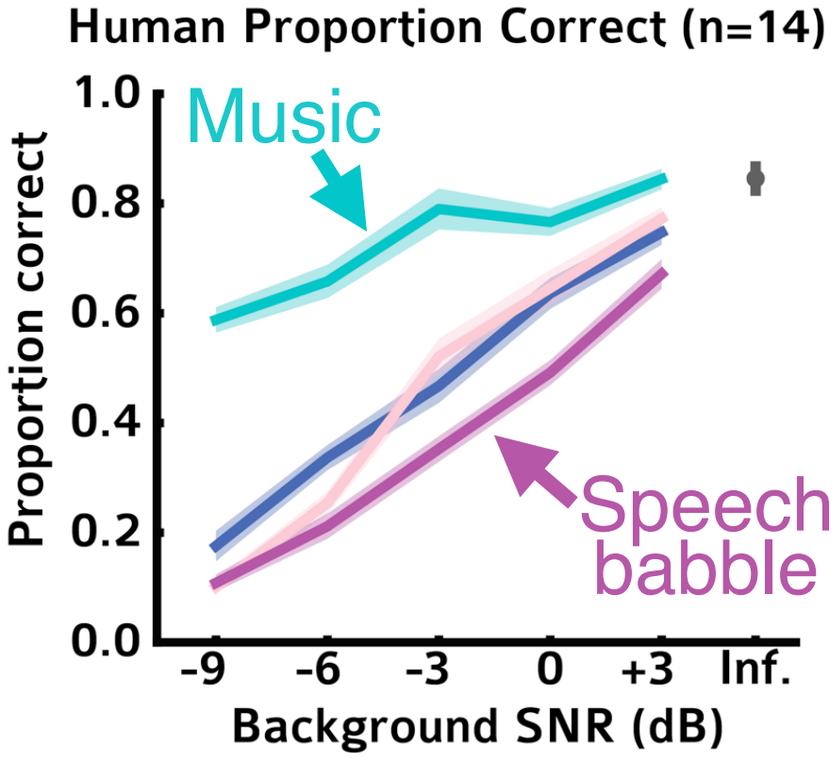


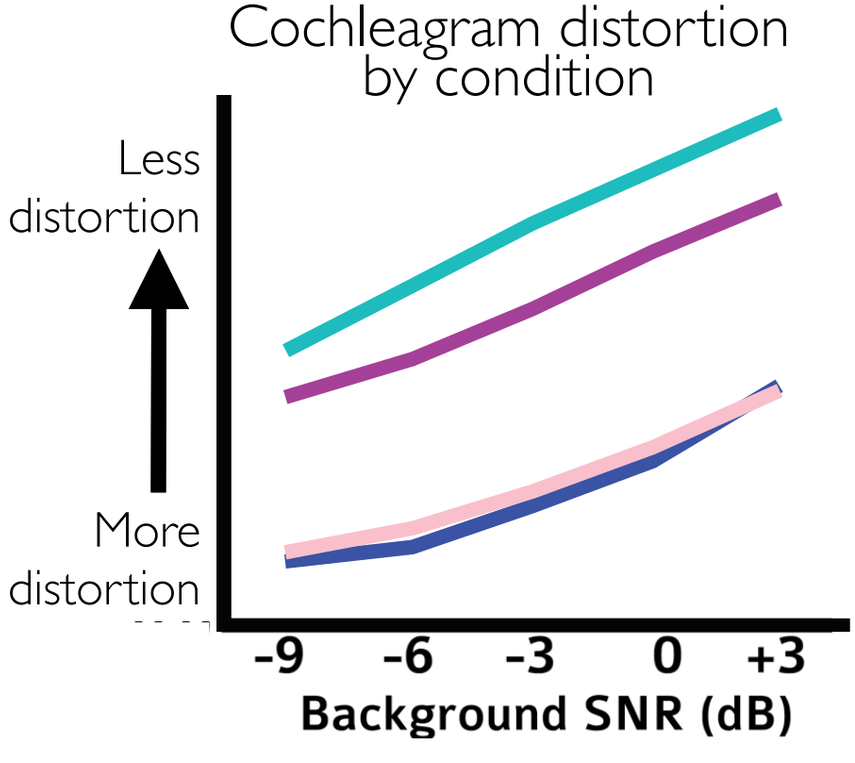


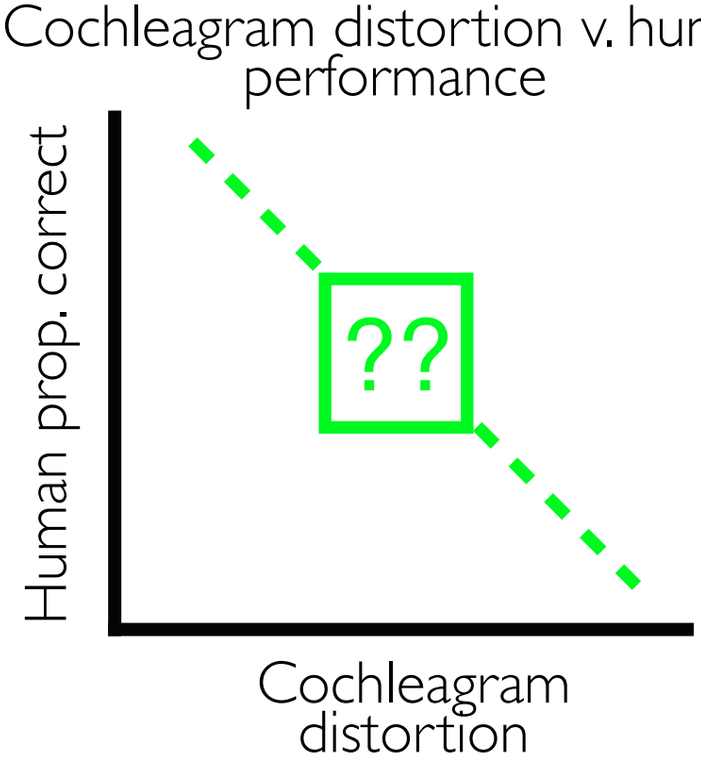
LEGEND

Music Auditory scenes Speech babble Speech-shaped noise

Distortion and pattern of performance.



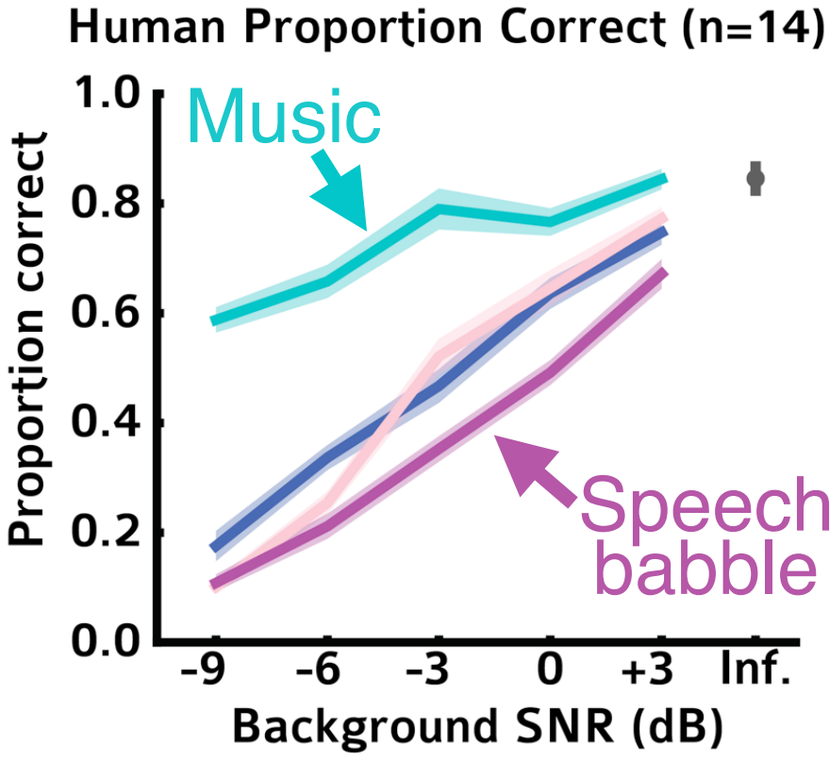


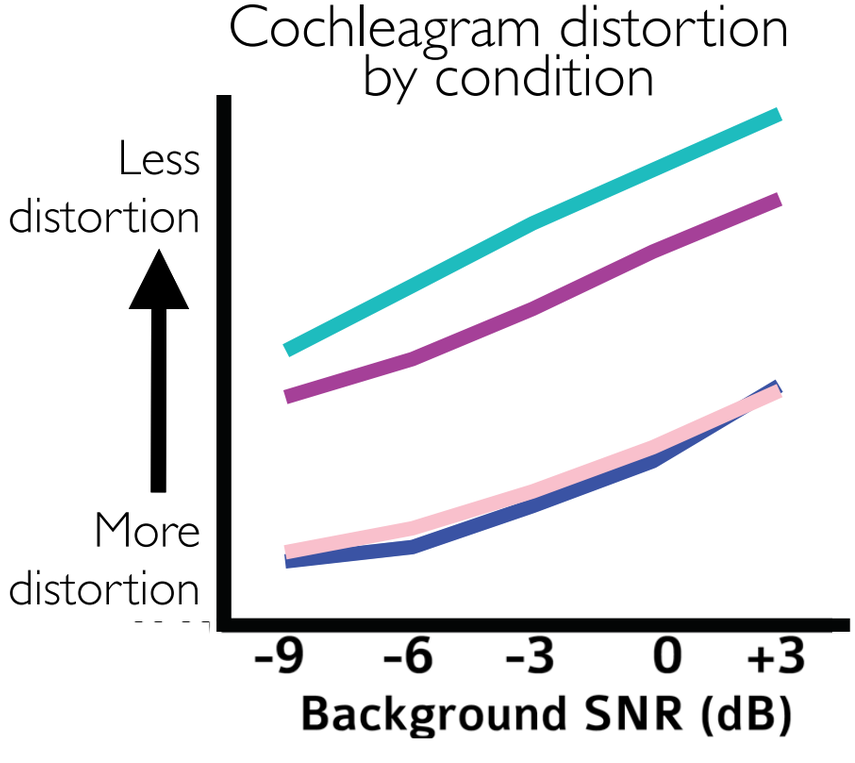


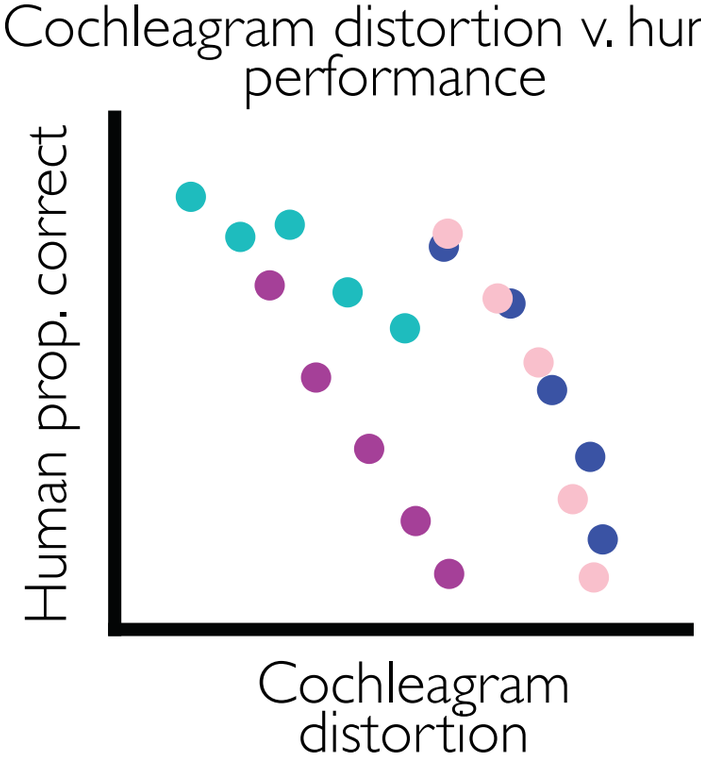
LEGEND

Music Auditory scenes Speech babble Speech-shaped noise

Distortion and pattern of performance.



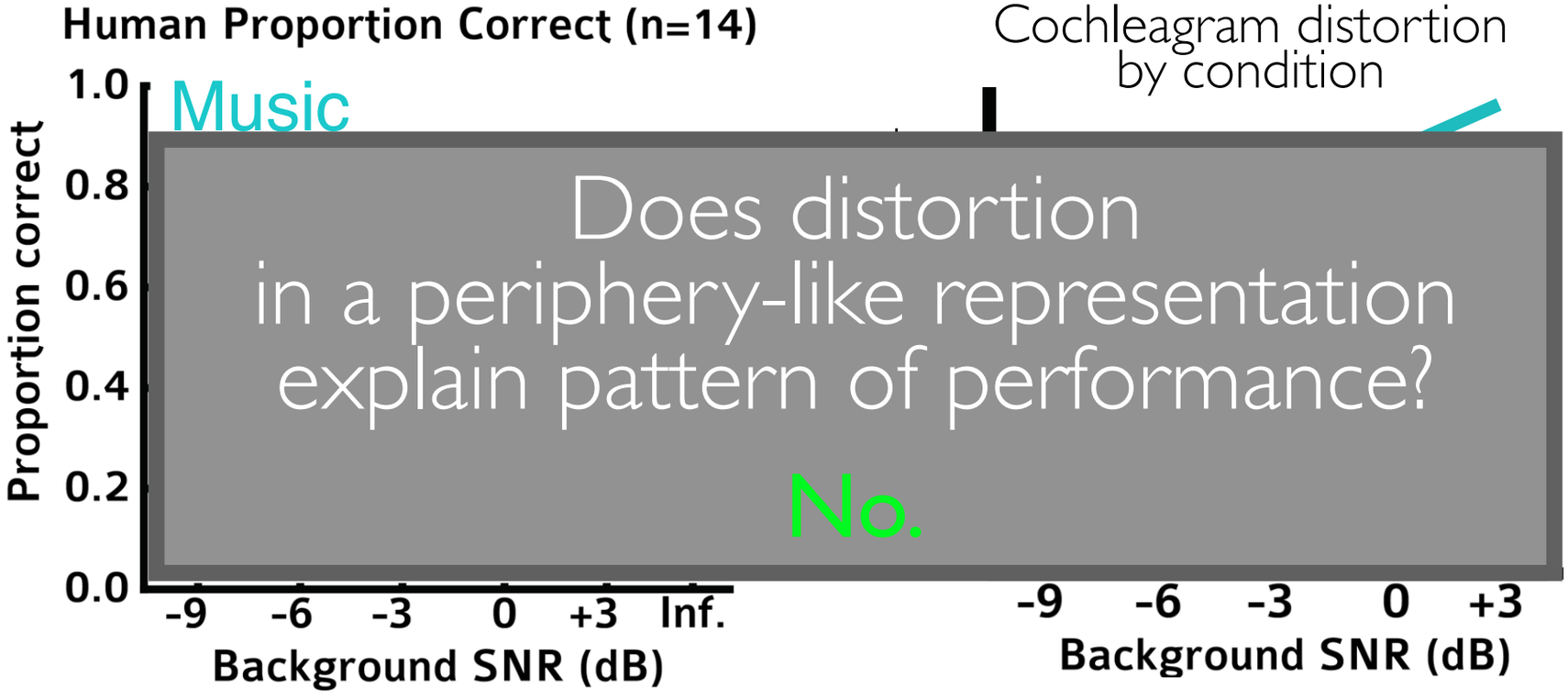




LEGEND

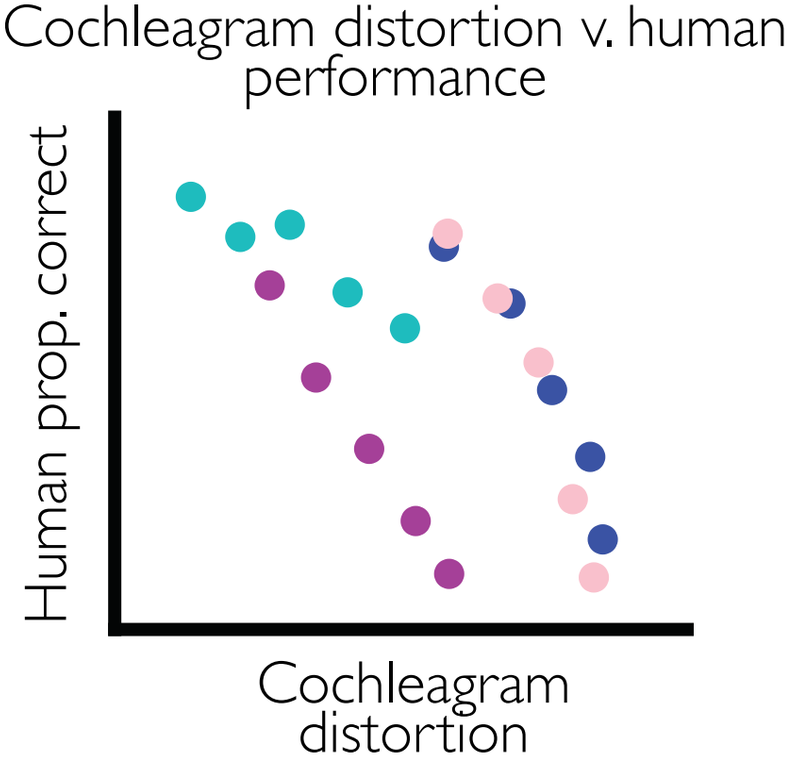
Music Auditory scenes Speech babble Speech-shaped noise

Distortion and pattern of performance.



Does distortion
in a periphery-like representation
explain pattern of performance?

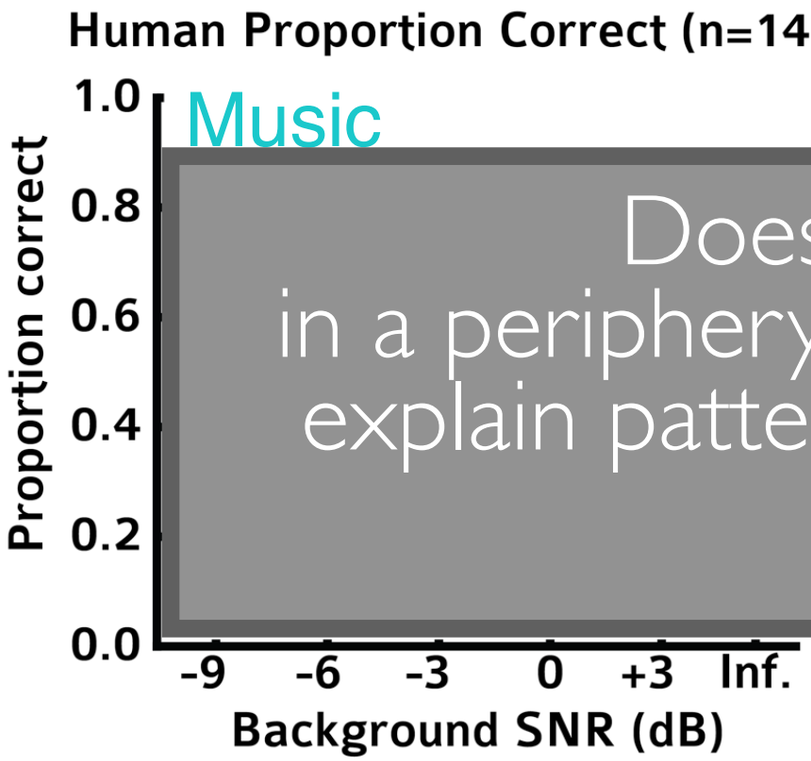
No.



LEGEND

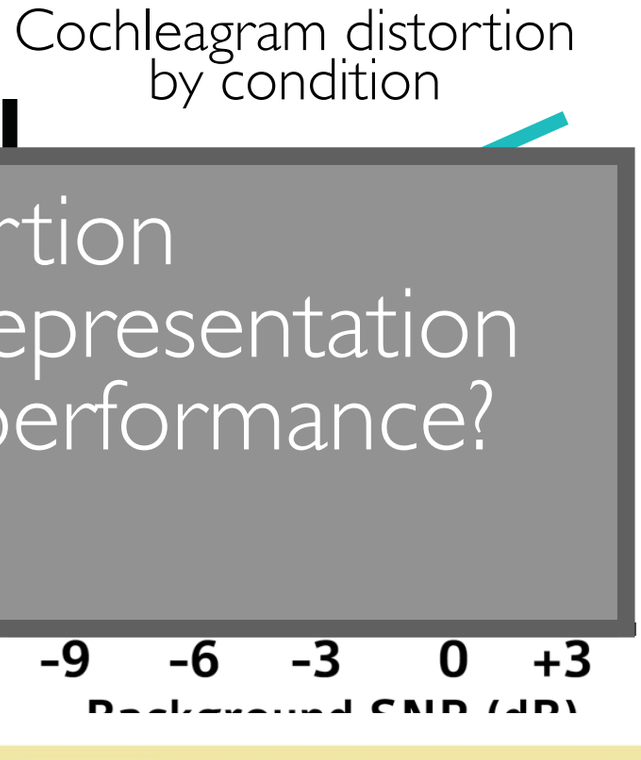
Music Auditory scenes Speech babble Speech-shaped noise

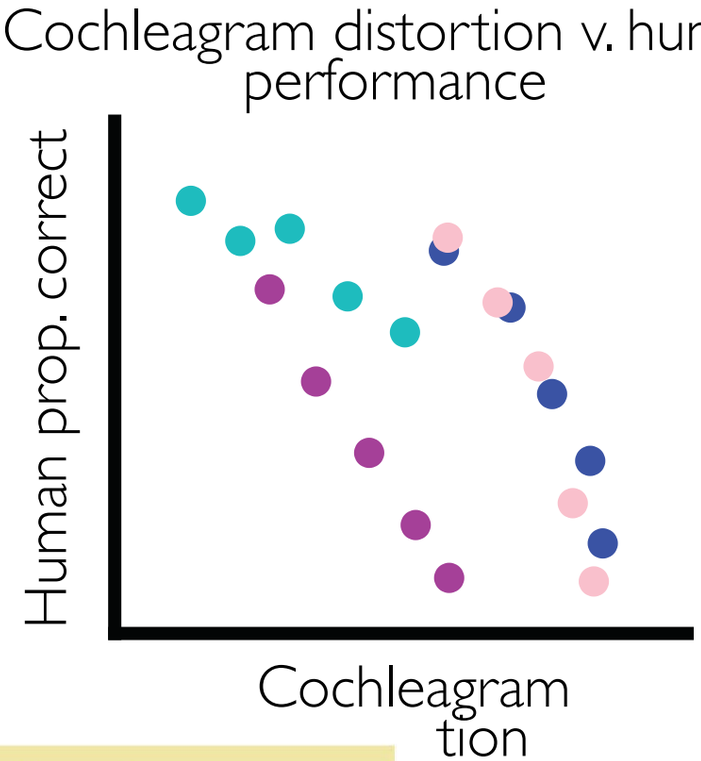
Distortion and pattern of performance.



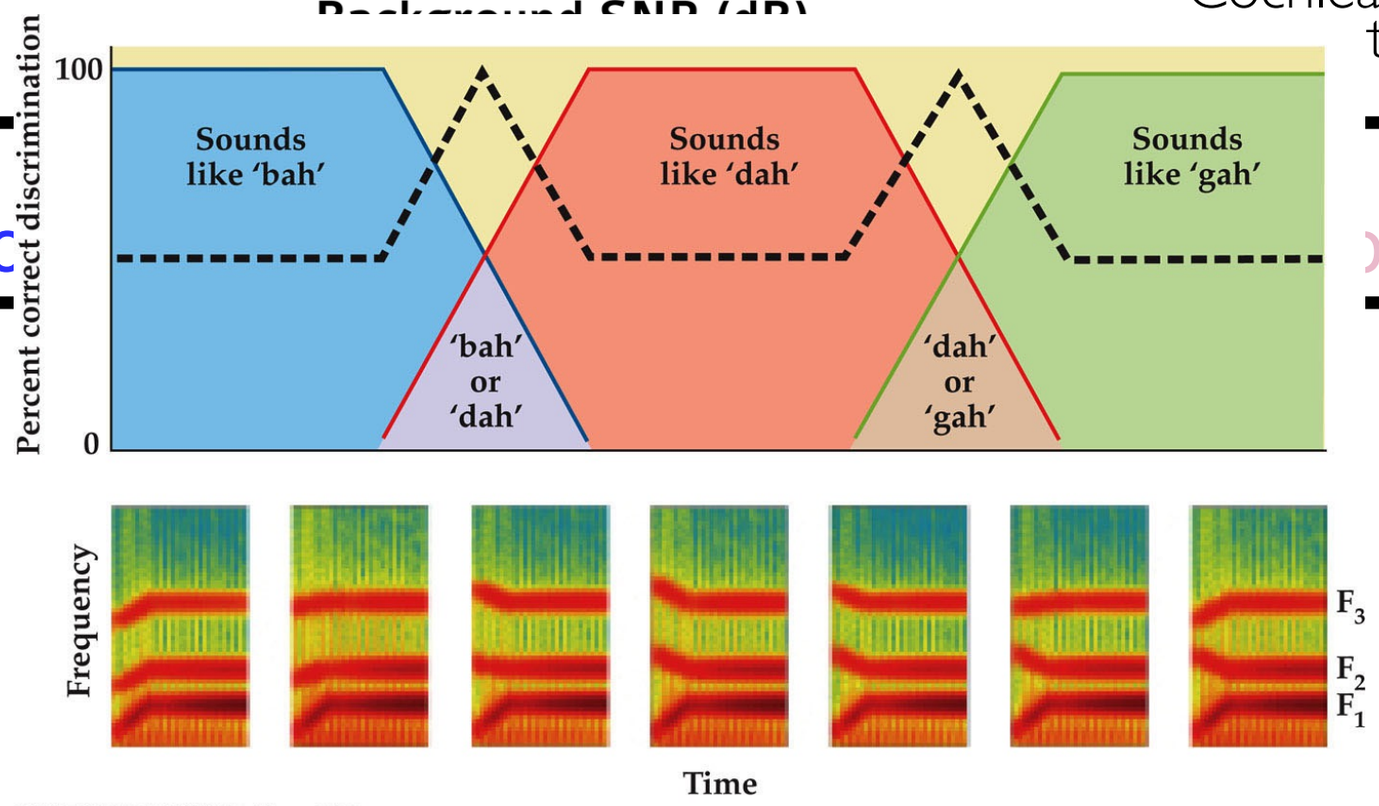
Does distortion
in a periphery-like representation
explain pattern of performance?

No.



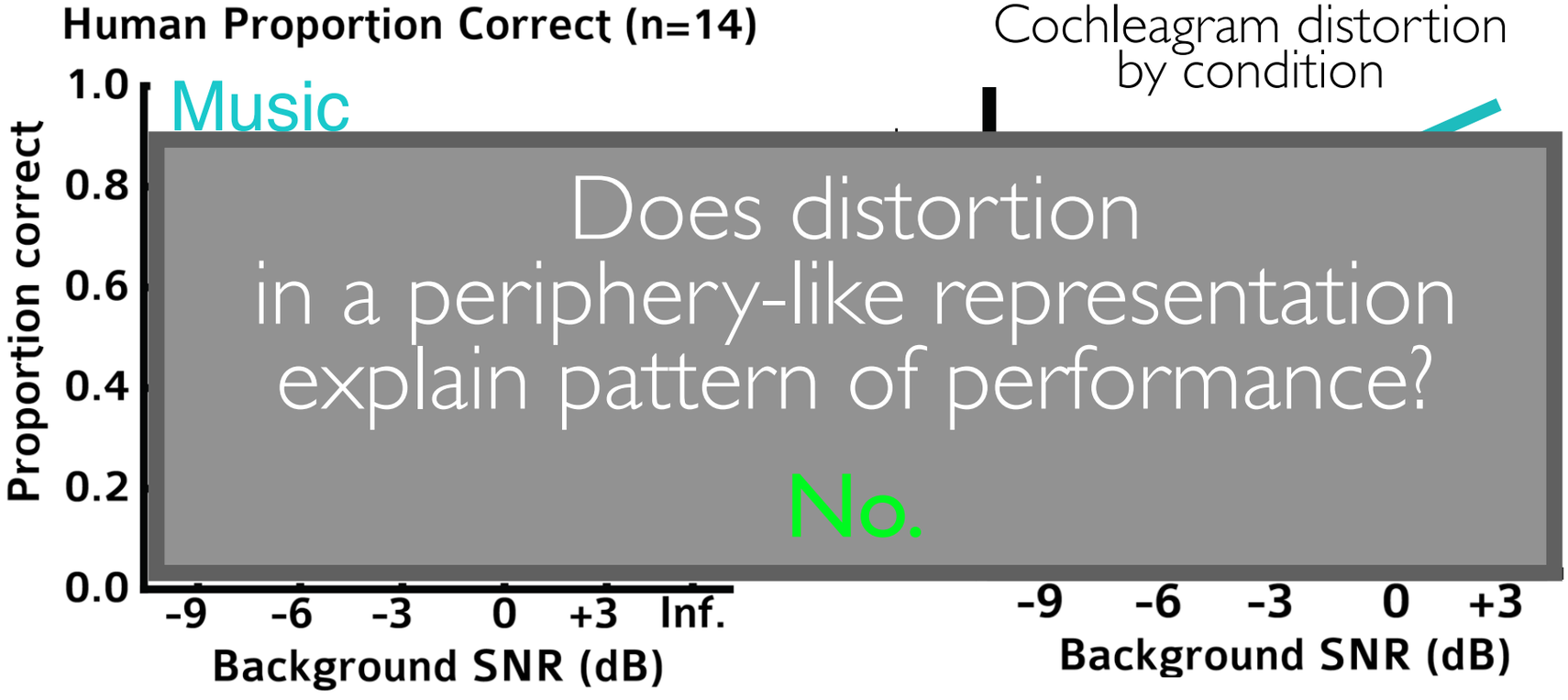


● Music ● Auditory sc ped noise



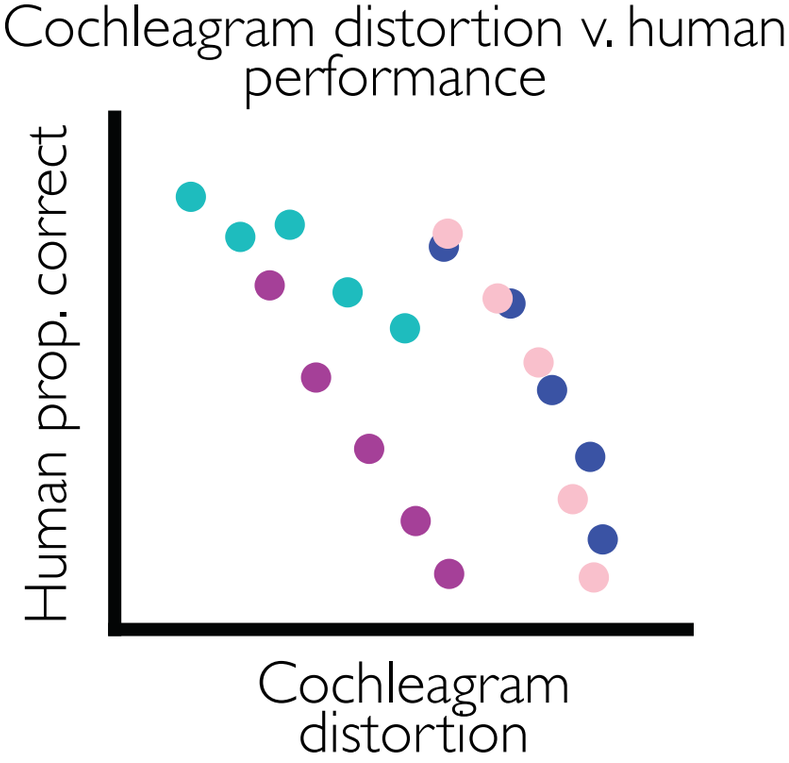
SENSATION & PERCEPTION 3e, Figure 11.13

Distortion and pattern of performance.



Does distortion
in a periphery-like representation
explain pattern of performance?

No.

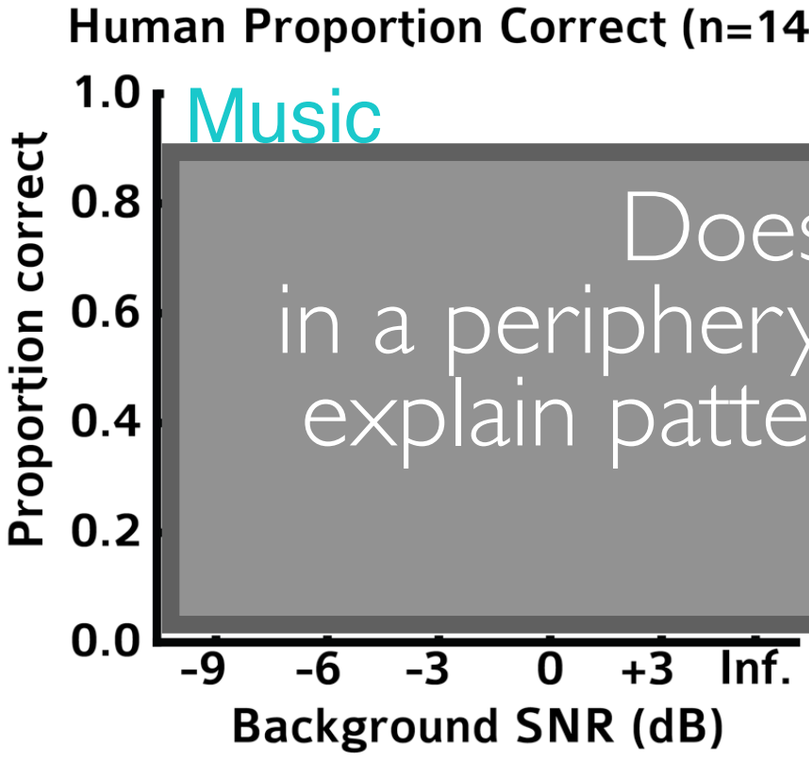


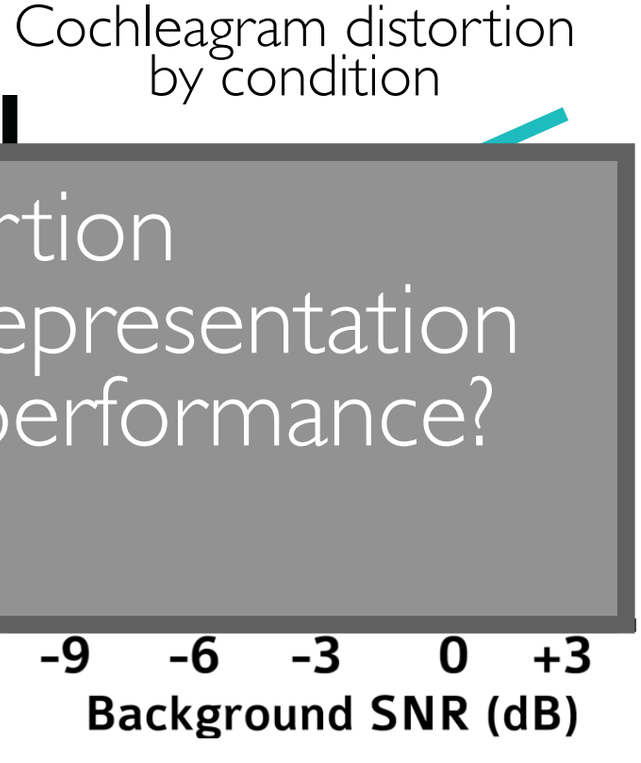
LEGEND

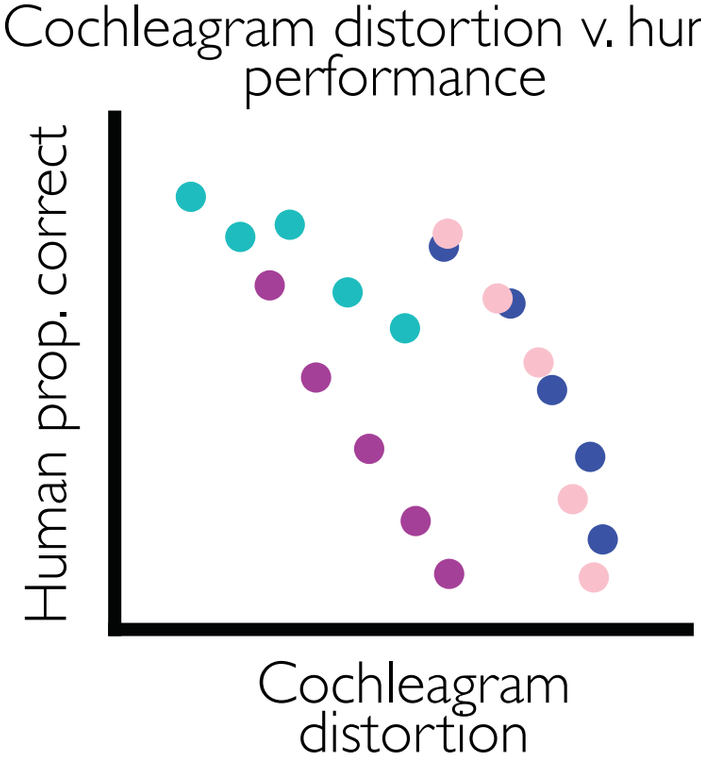
Music Auditory scenes Speech babble Speech-shaped noise

Does distortion of CNN representation explain pattern of performance?

Distortion and pattern of performance.







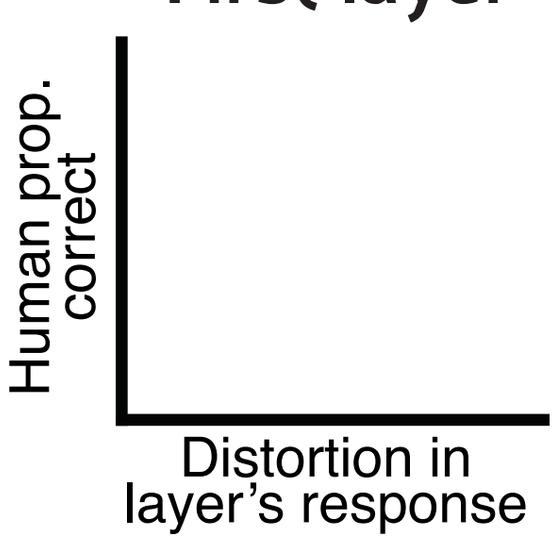
Does distortion
in a periphery-like representation
explain pattern of performance?

No.

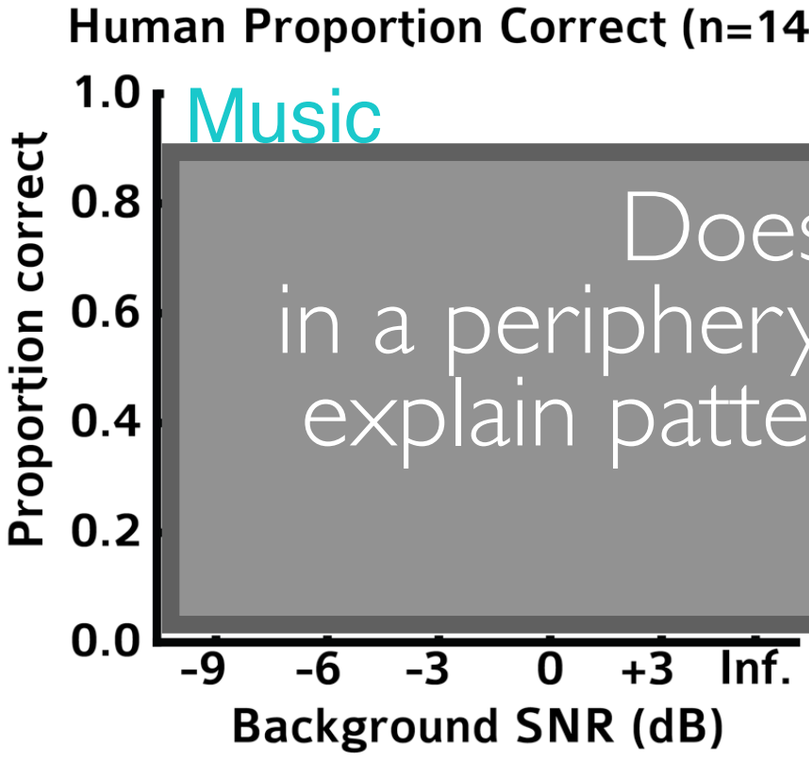
LEGEND

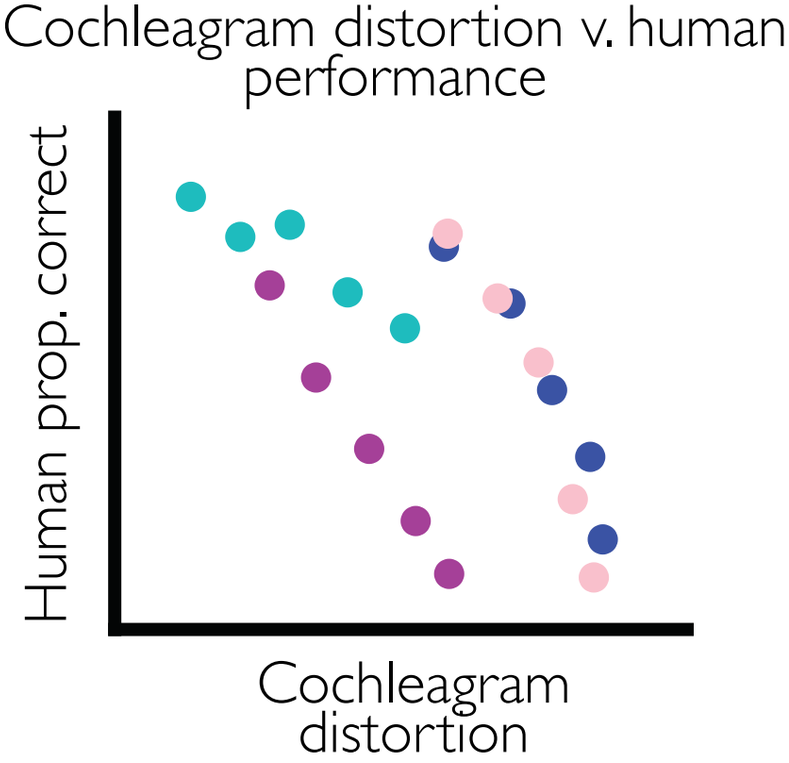
Music Auditory scenes Speech babble Speech-shaped noise

Does distortion of CNN representation explain pattern of performance?



Distortion and pattern of performance.





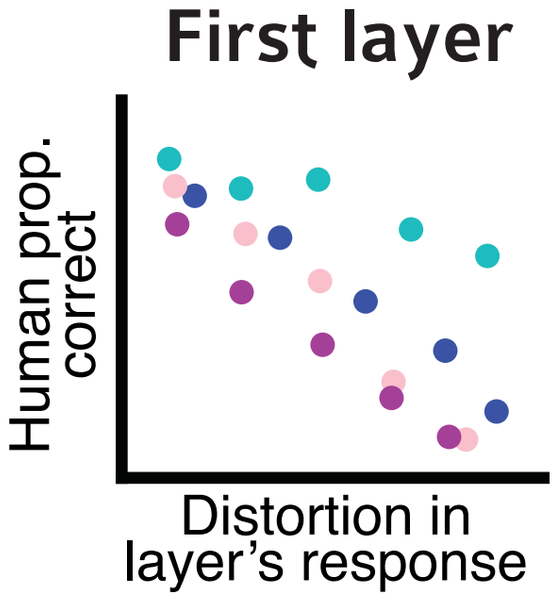
Does distortion
in a periphery-like representation
explain pattern of performance?

No.

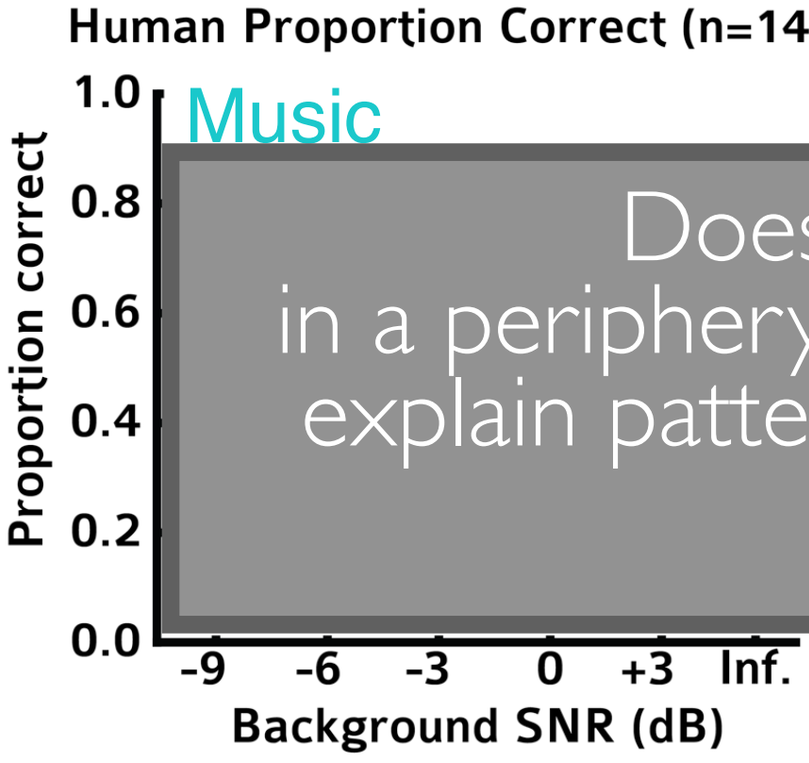
LEGEND

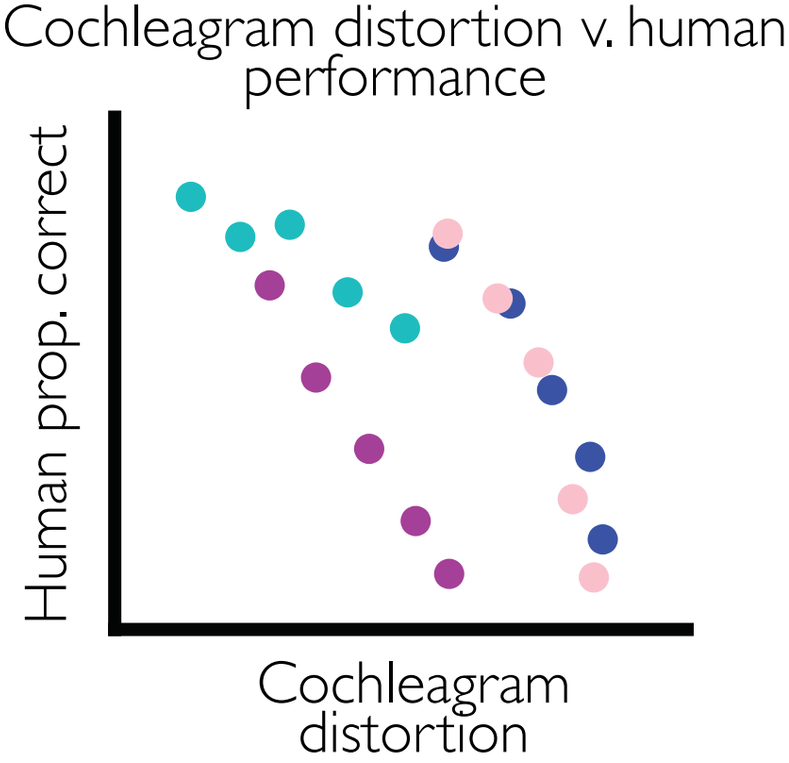
Music Auditory scenes Speech babble Speech-shaped noise

Does distortion of CNN representation explain pattern of performance?



Distortion and pattern of performance.





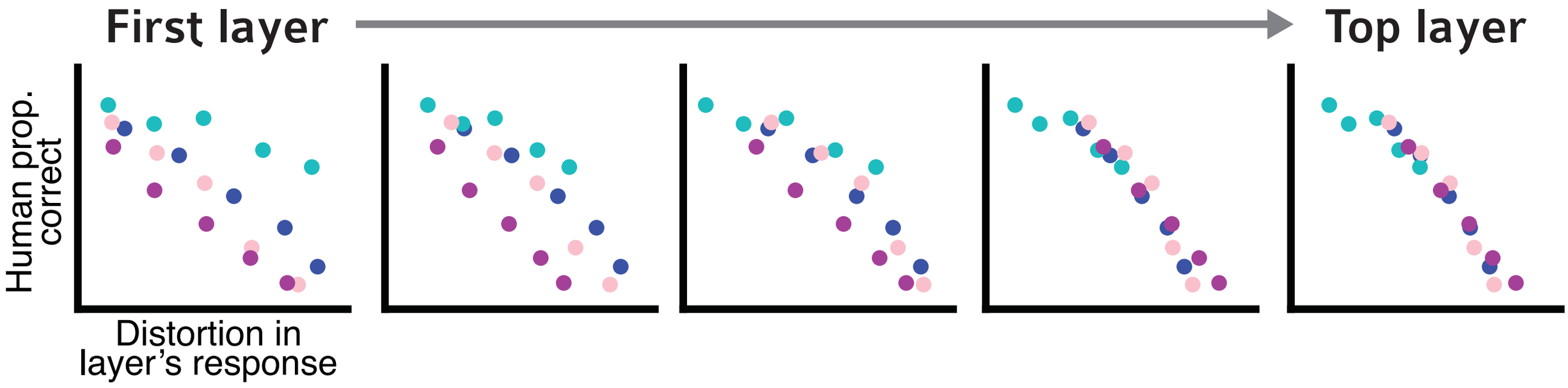
Does distortion
in a periphery-like representation
explain pattern of performance?

No.

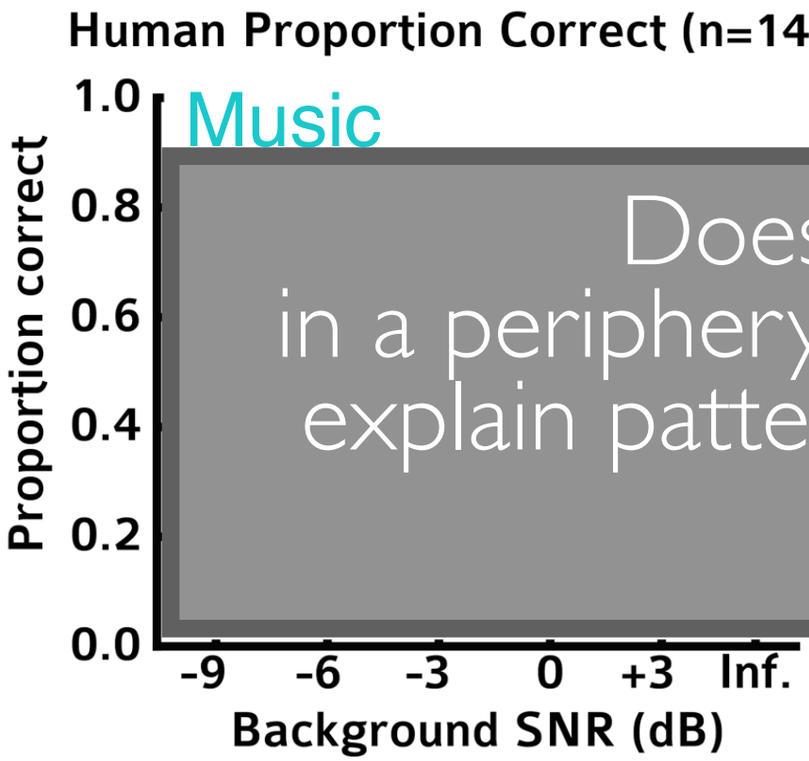
LEGEND

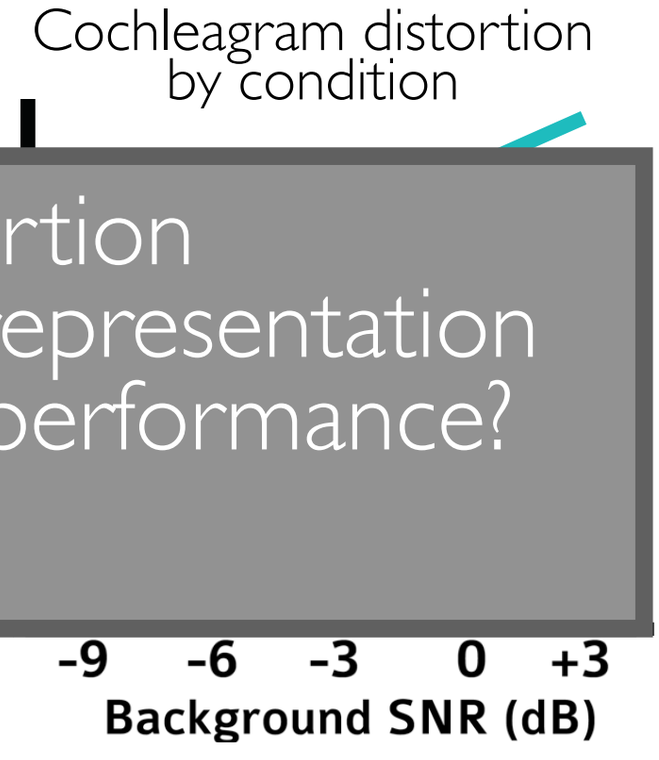
Music Auditory scenes Speech babble Speech-shaped noise

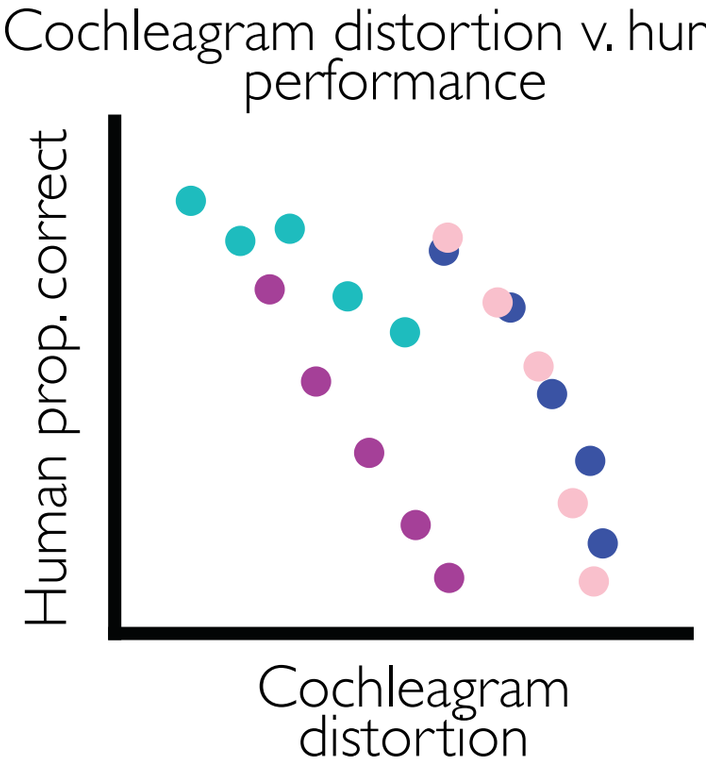
Does distortion of CNN representation explain pattern of performance?



Distortion and pattern of performance.





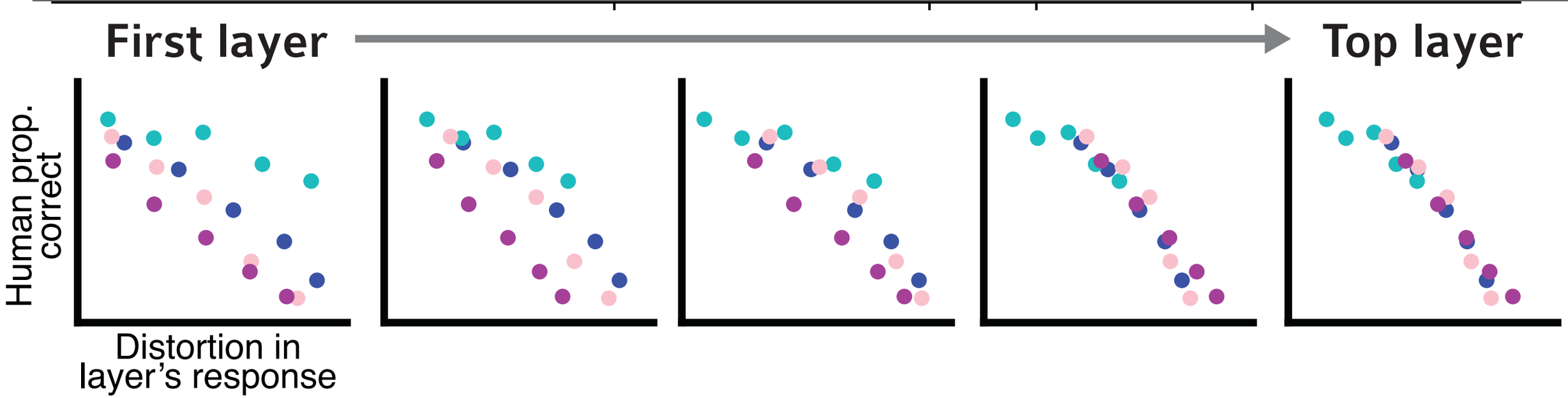


Does distortion
in a periphery-like representation
explain pattern of performance?

No.

Distortion in a highly nonlinear feature space
explains the pattern of performance.

Task-Optimized CNN has discovered proper space.



Imaging Experiment

fMRI response data collected* on 165 commonly heard natural sound stimuli.

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
Ringtone
Microwave
Dog barking

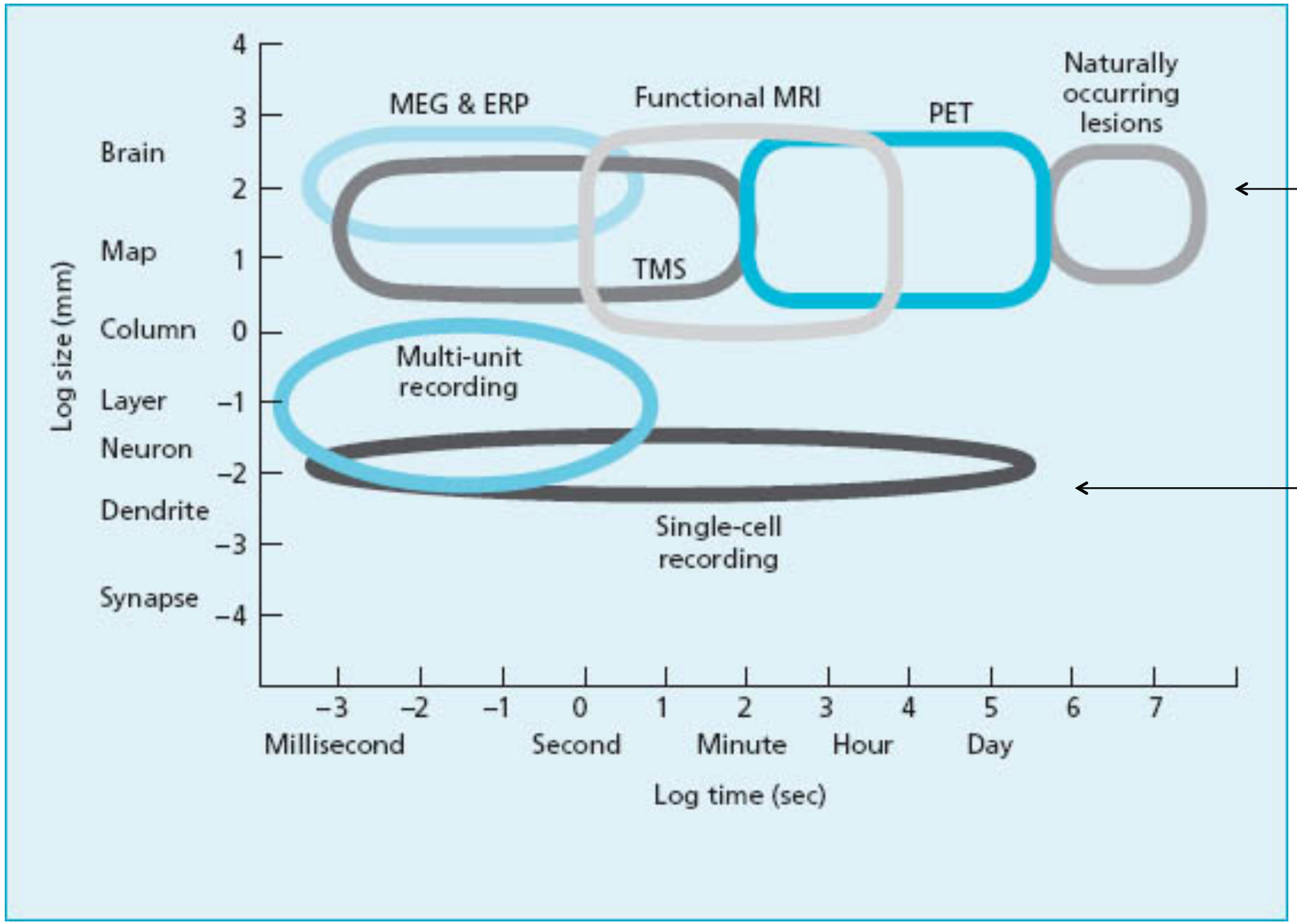
Road traffic
Zipper
Cellphone vibrating
Water dripping
Scratching
Car windows
Telephone ringing
Chopping food
Telephone dialing
Girl speaking
Car horn
Writing
Computer startup sound
Background speech
Songbird
Pouring water
Pop song
Water boiling

Guitar
Coughing
Crumpling paper
Siren
Splashing water
Computer speech
Alarm clock
Walking with heels
Vacuum
Wind
Boy speaking
Chair rolling
Rock song
Door knocking

•
•
•

*Sam Norman-Haignere, Nancy Kanwisher, and Josh McDermott

Neuroscience Methods

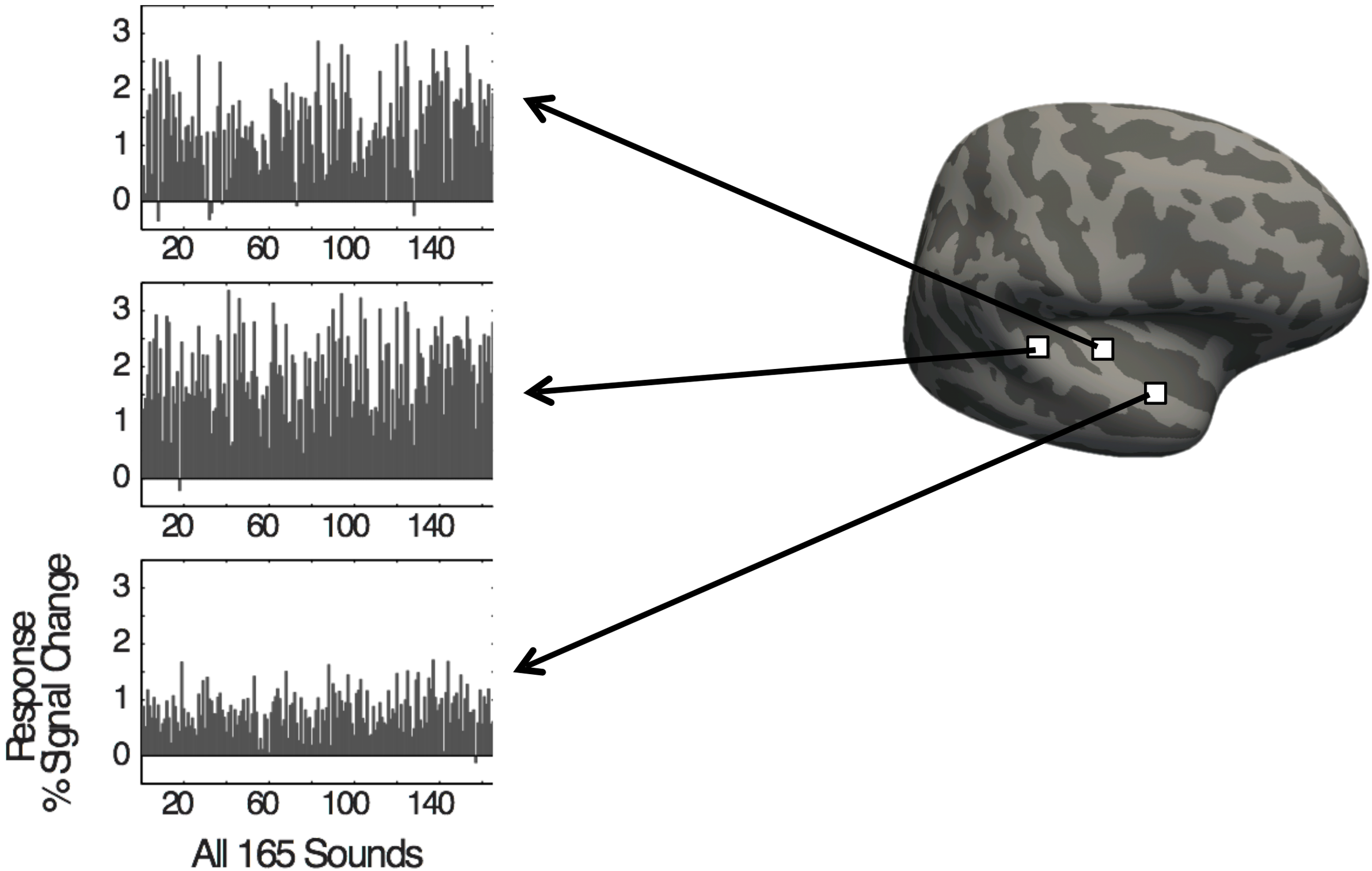


Methods available for studying awake behaving **humans**

can be used in awake behaving Macaques

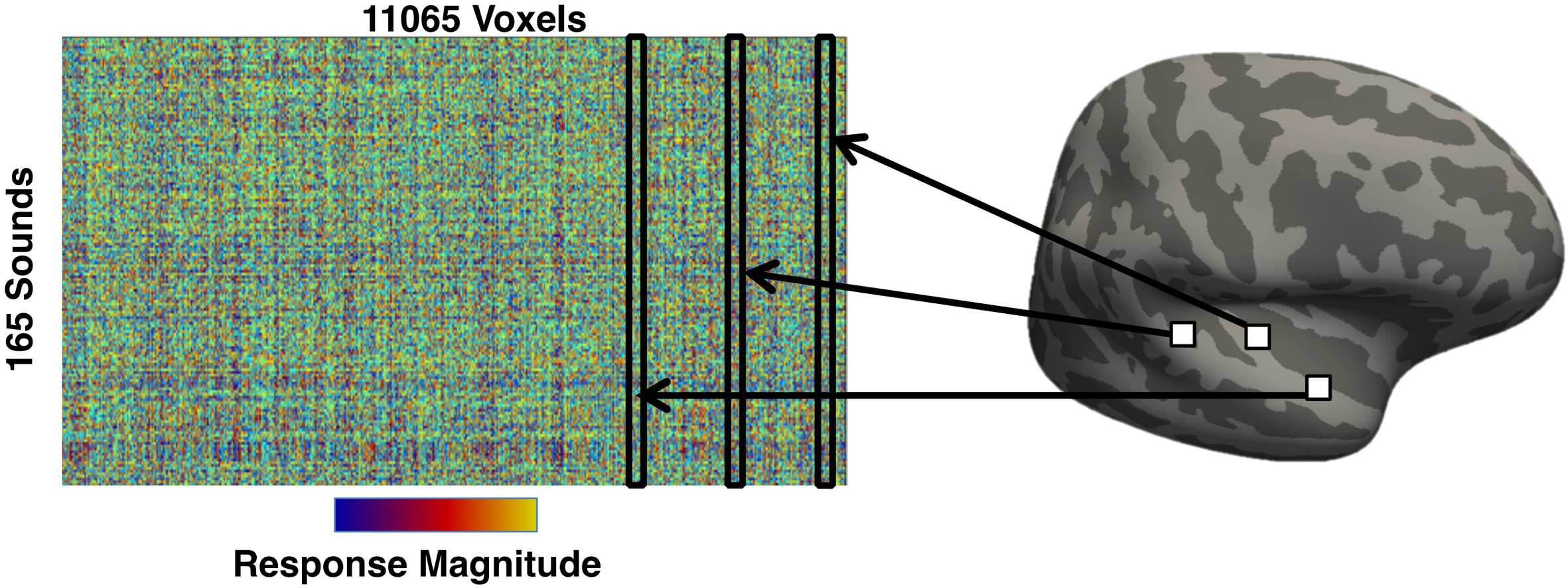
Imaging Experiment

For each voxel, measured average response to each sound:



Imaging Experiment

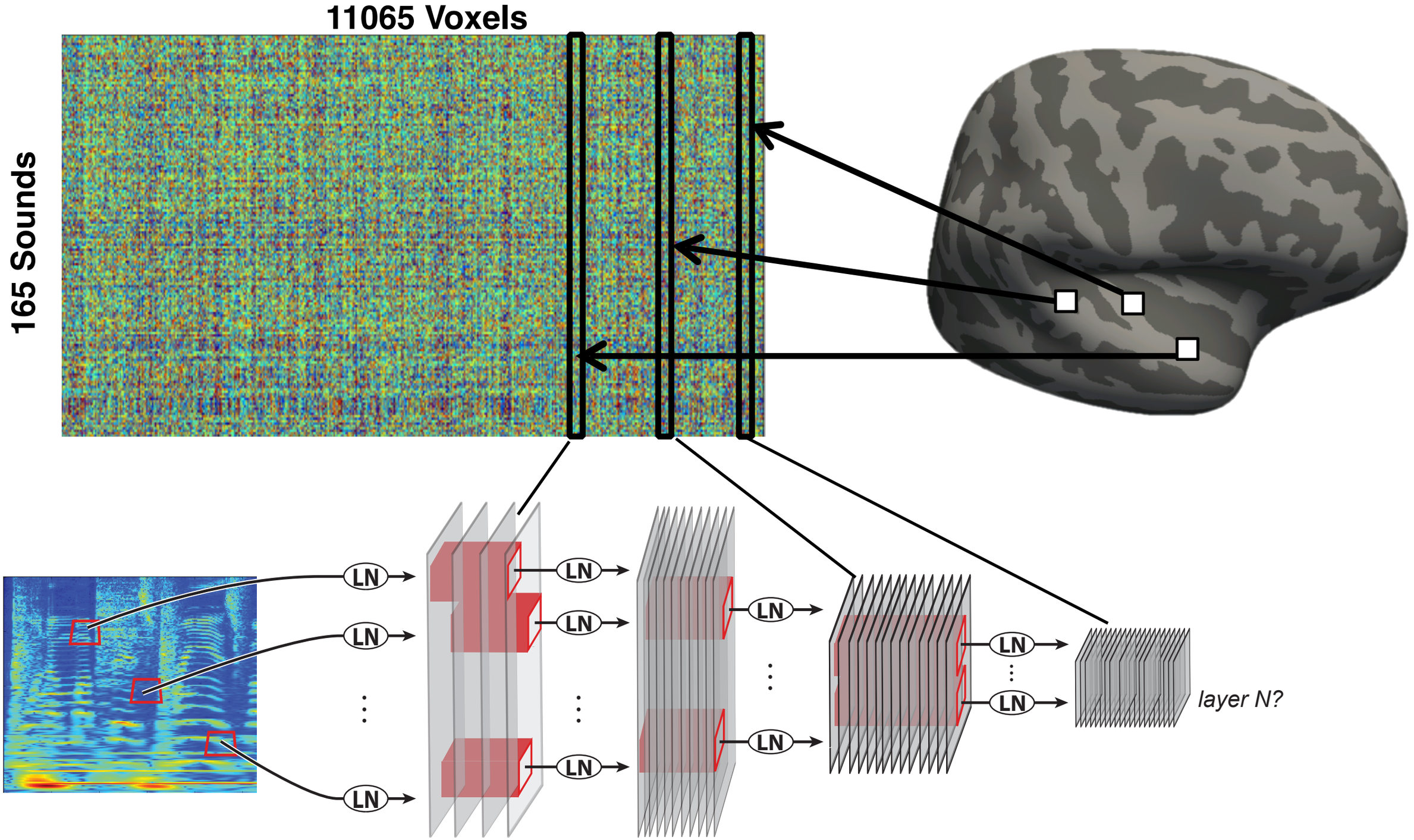
For each voxel, measured average response to each sound:



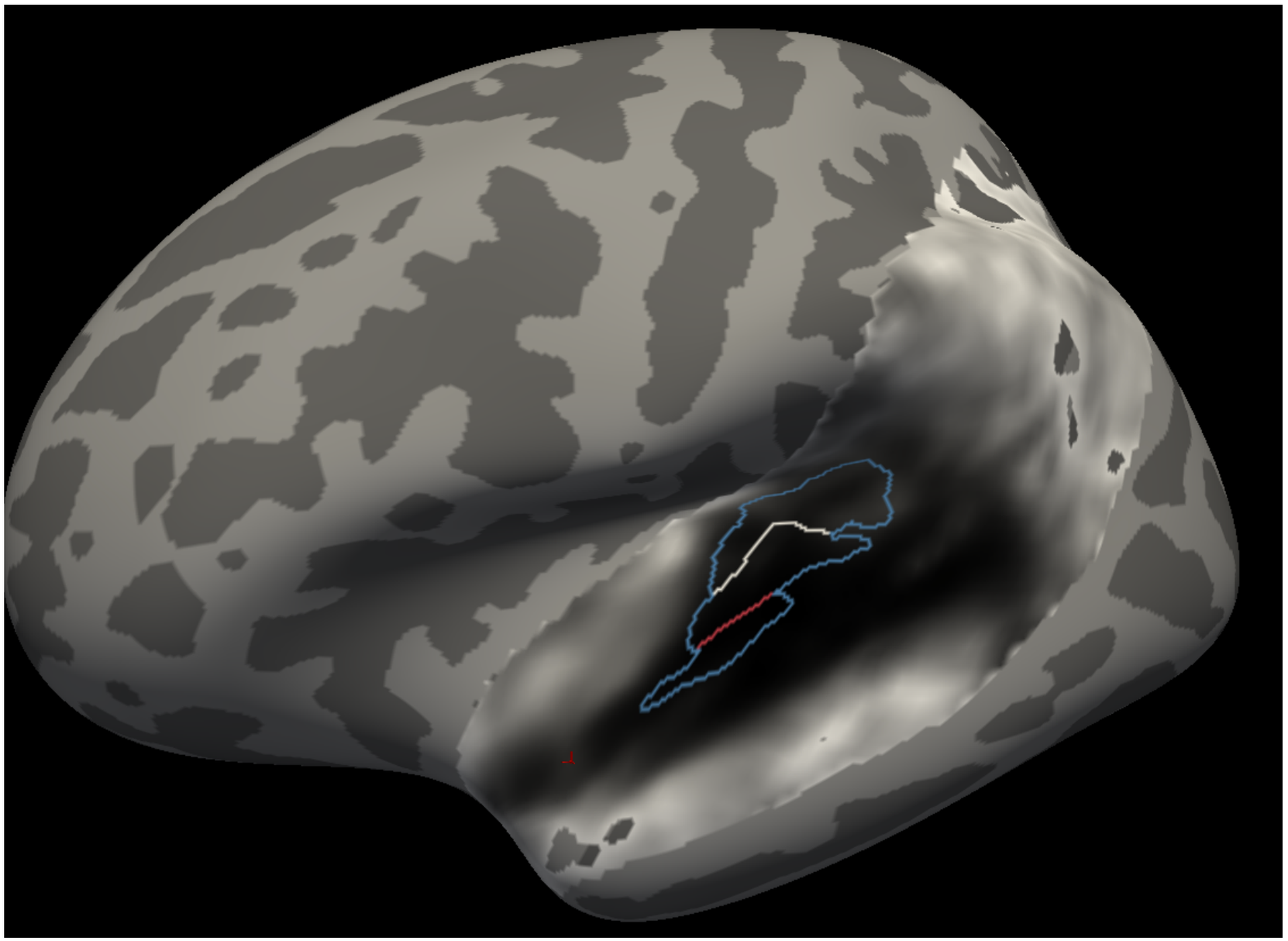
Data matrix: voxels X sounds.

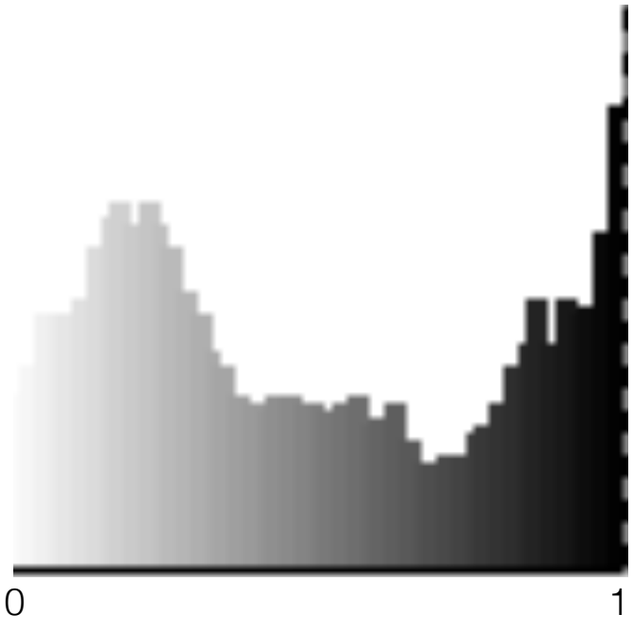
Imaging Experiment

Neural predictivity: the ability of model to predict each individual voxel's activity using linear regression.

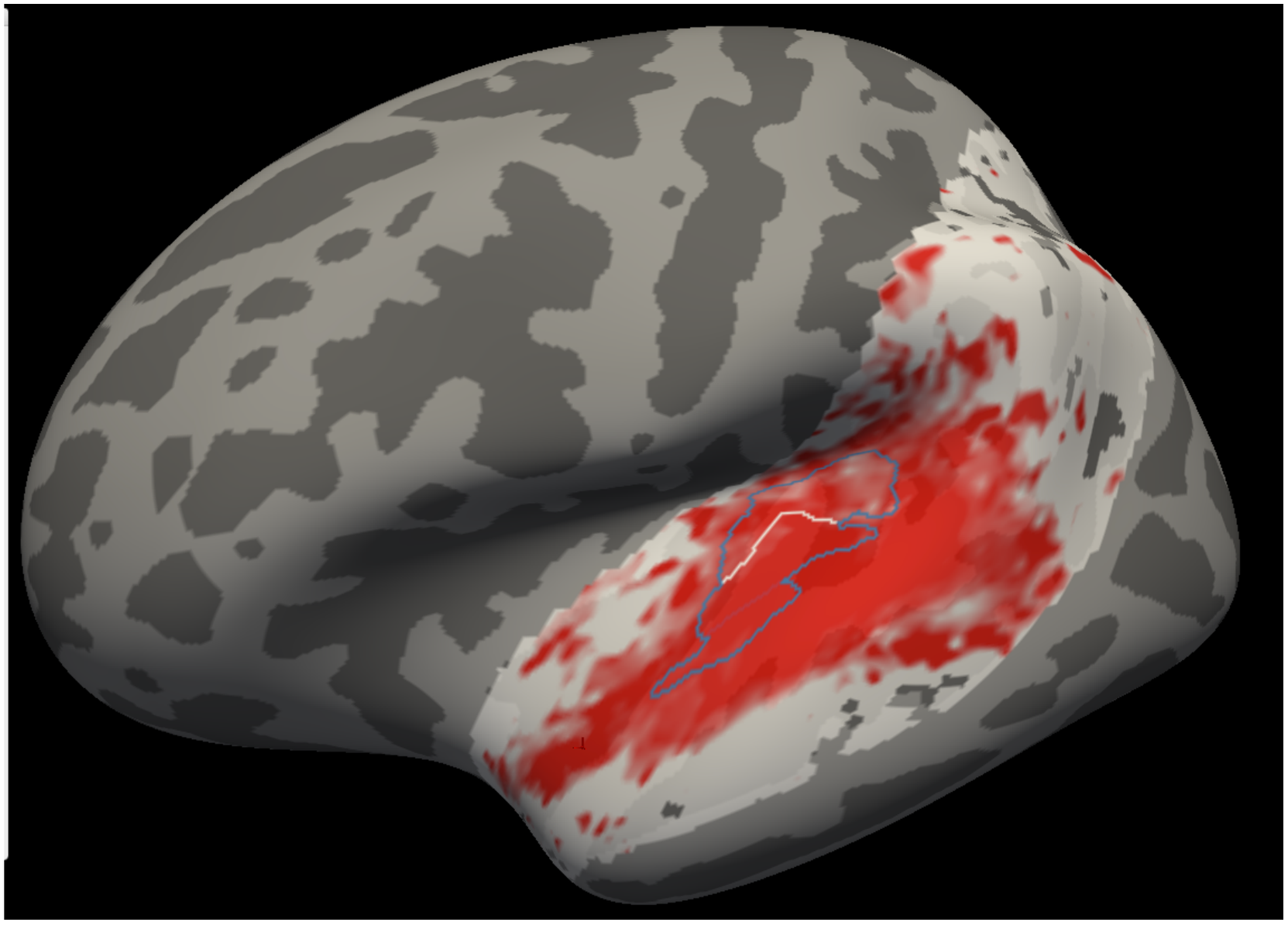


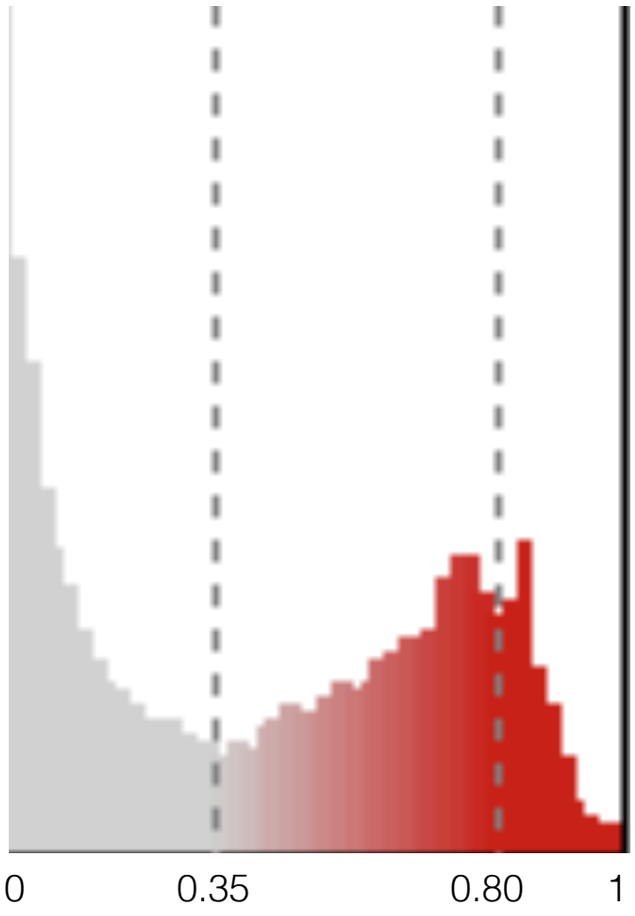
Response Reliability at Voxel Level



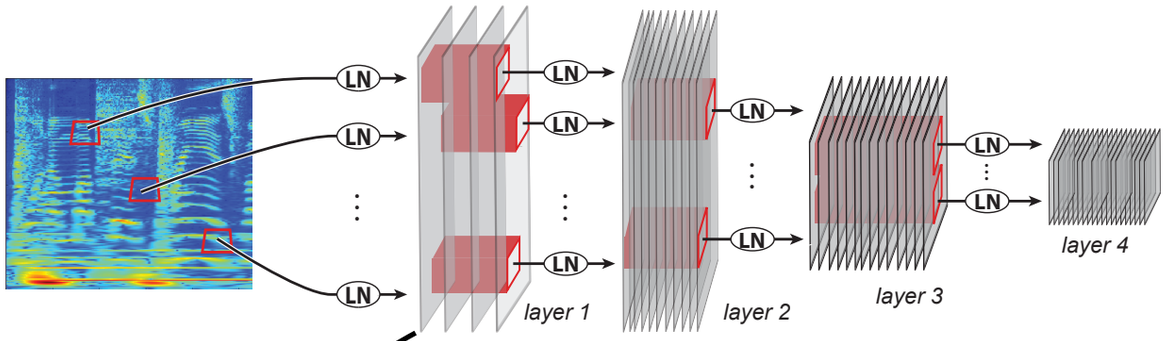


Model Productivity at Best Layer



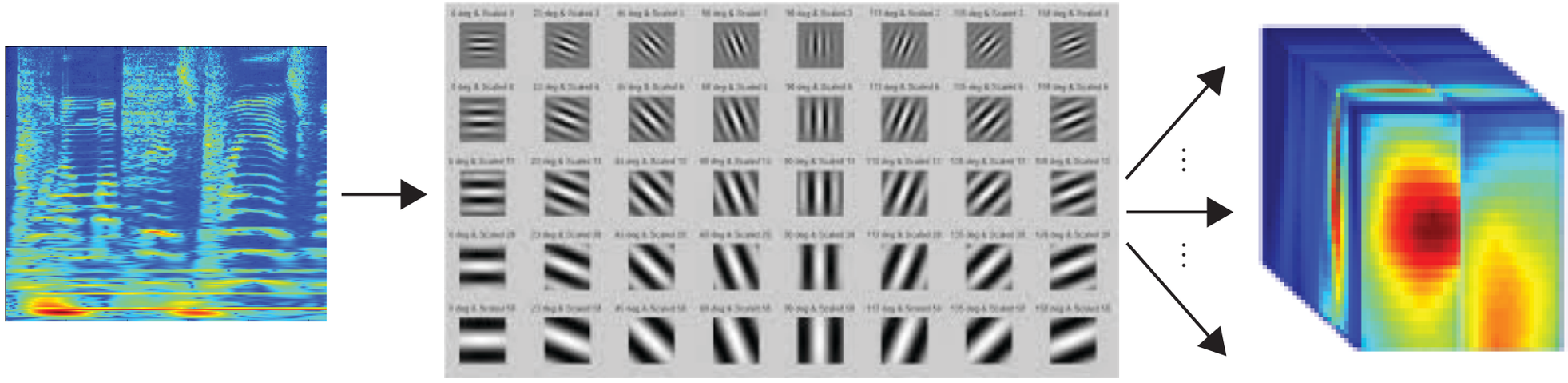


Model Productivity at Best Layer



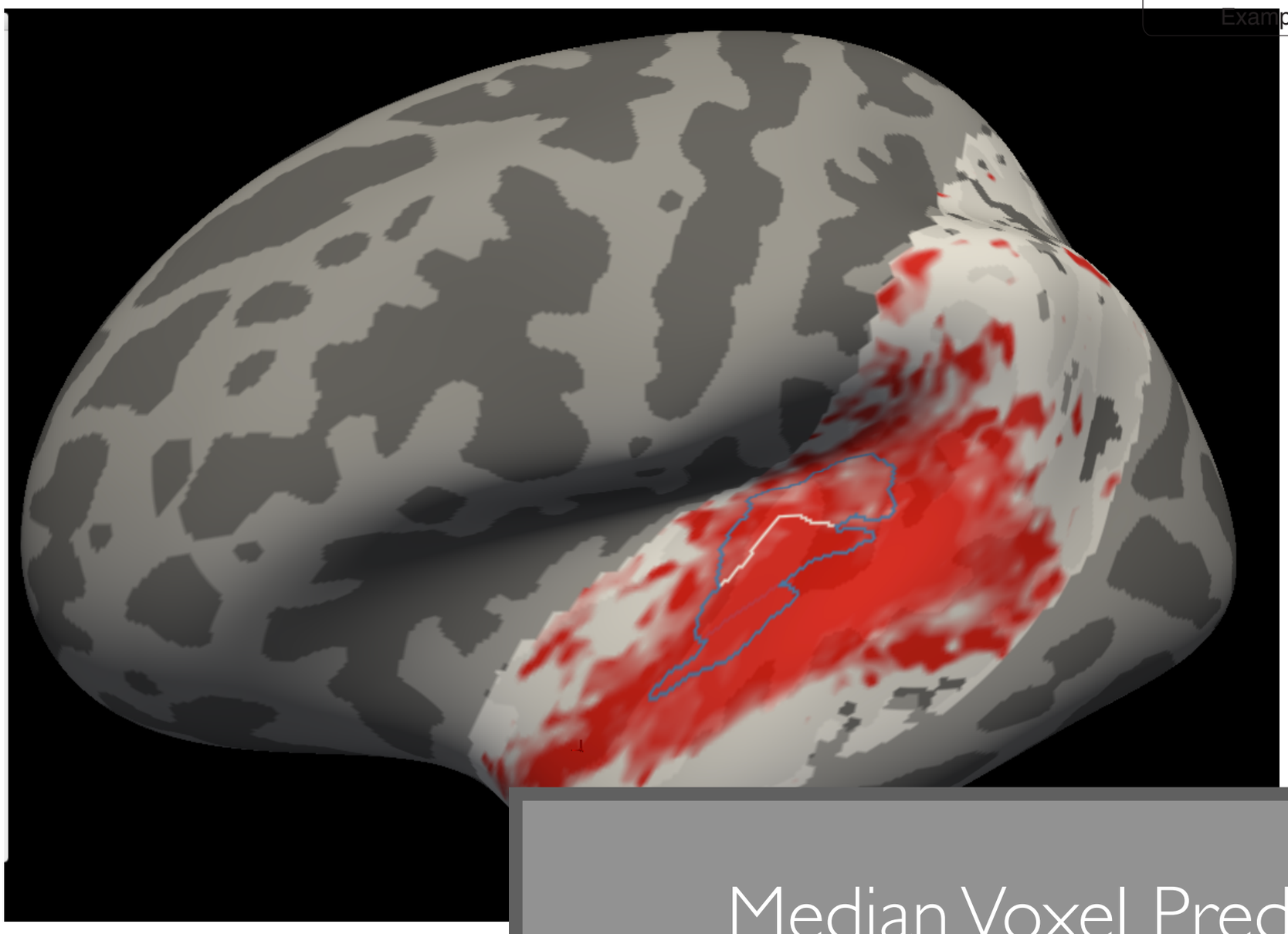
Example first-layer filters

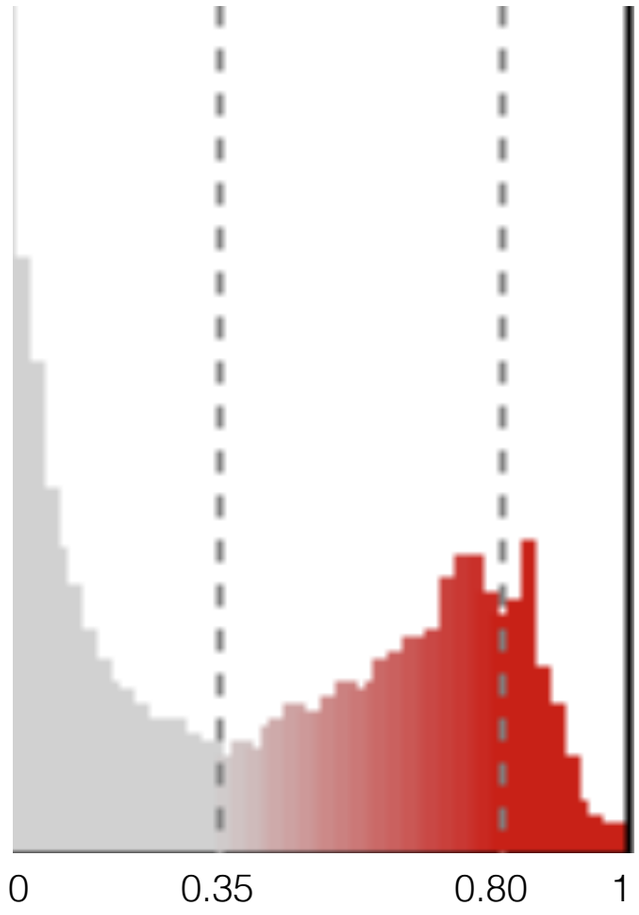
(Remember: spectrotemporal model (Shamma, 2005):)



Model Productivity at Best Layer


Example first-layer filters

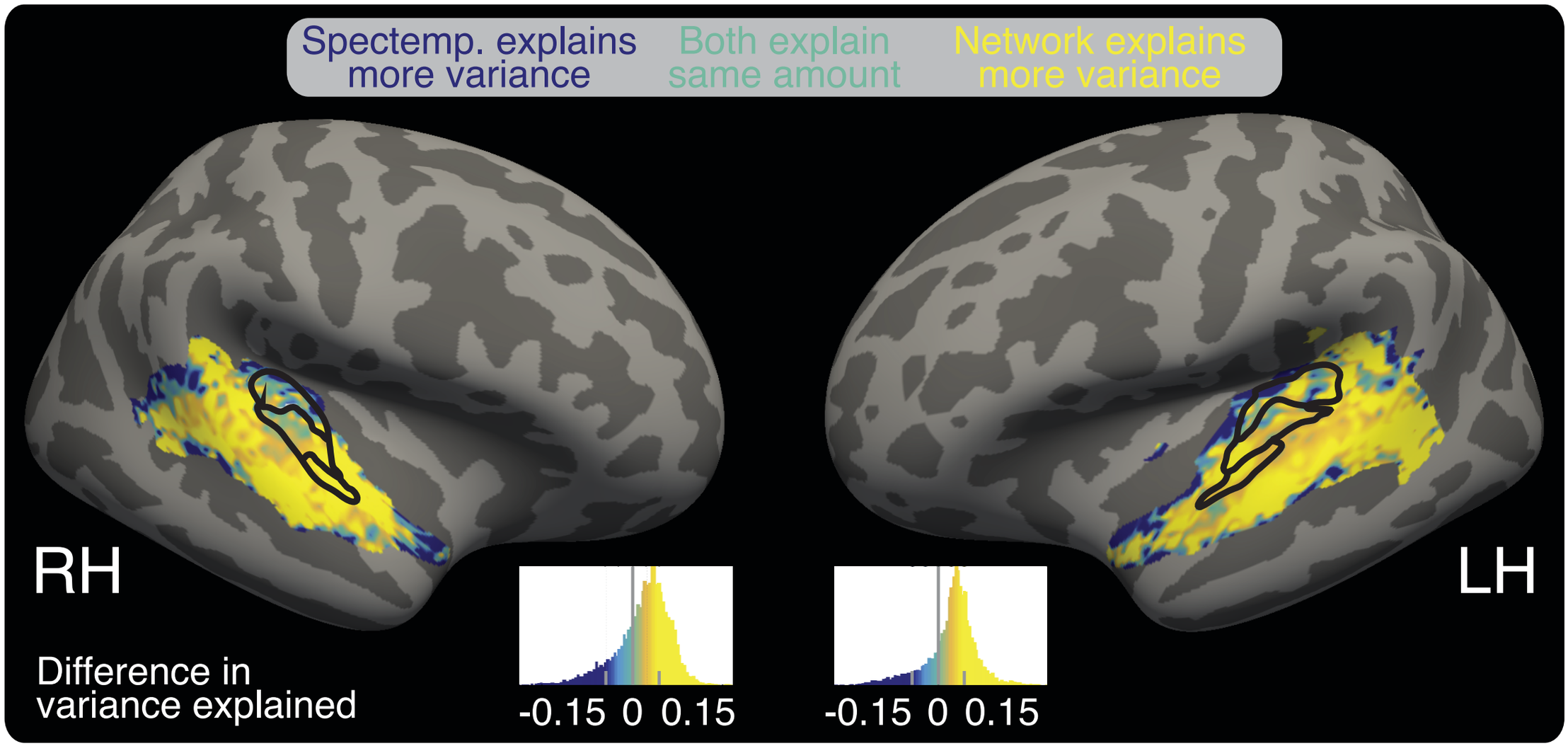




Median Voxel Predictivity

~80%.

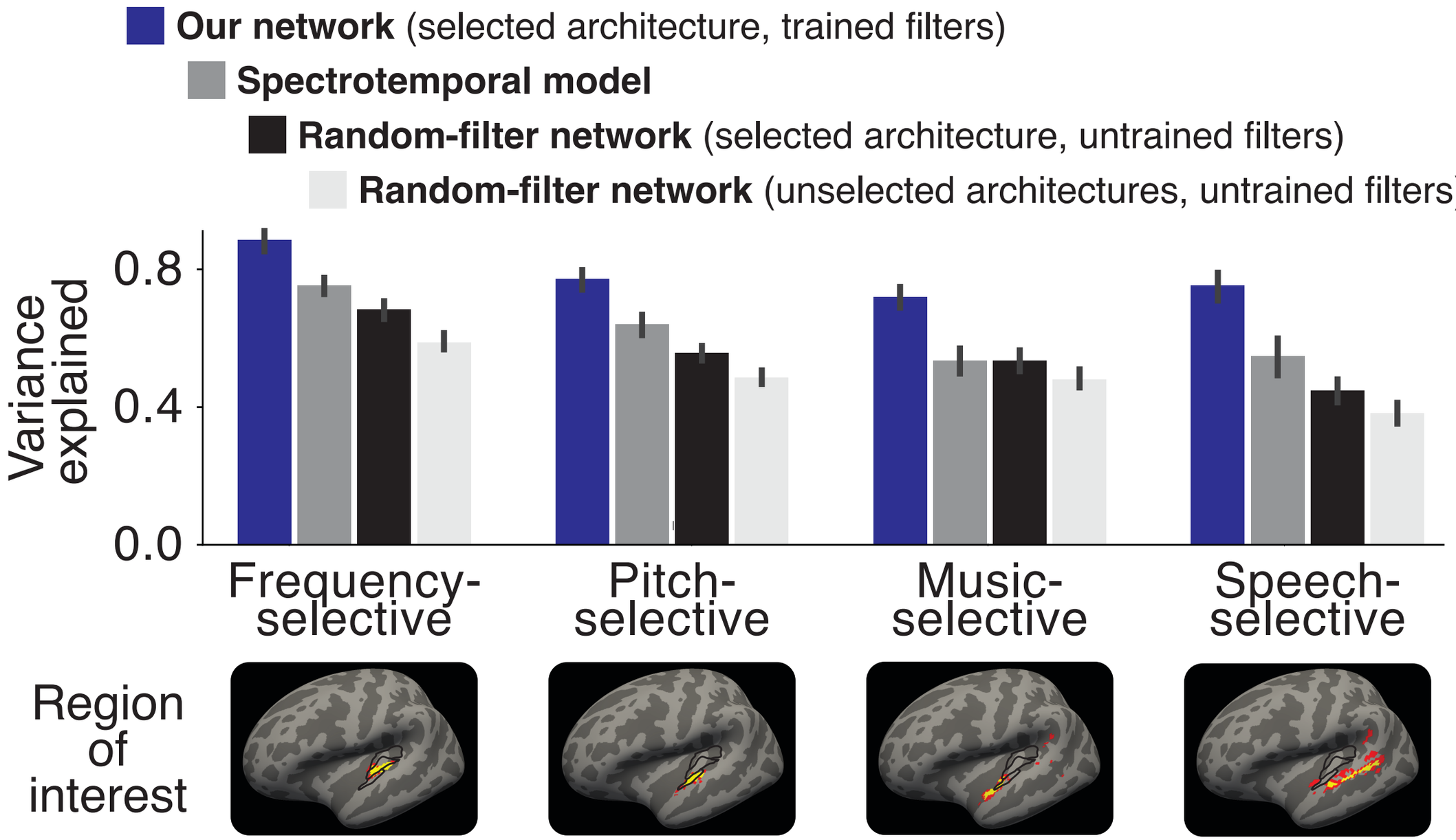
Comparison to Spectrotemporal Filtering Model



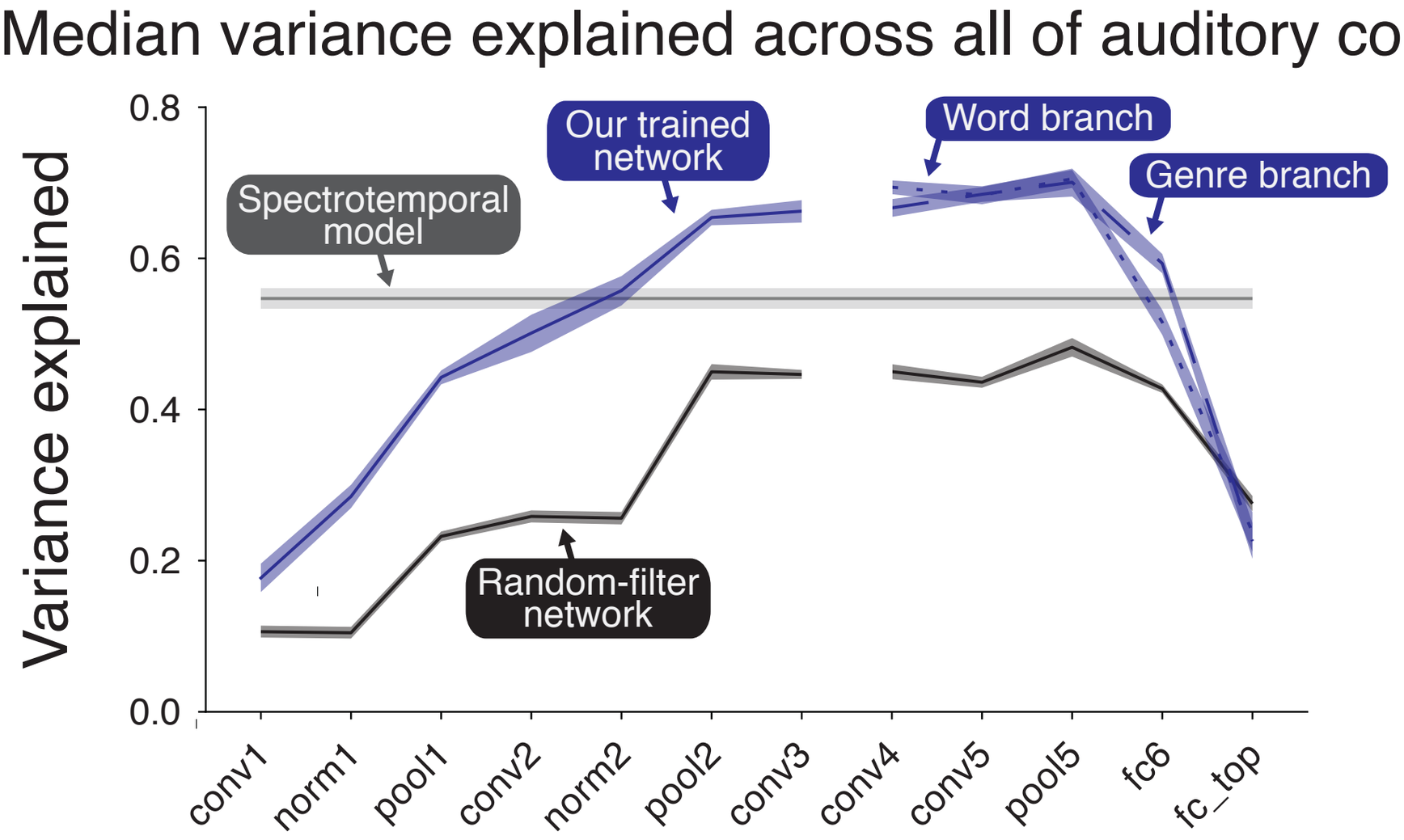
Significant improvement relative to existing models,

but especially in non-primary areas.

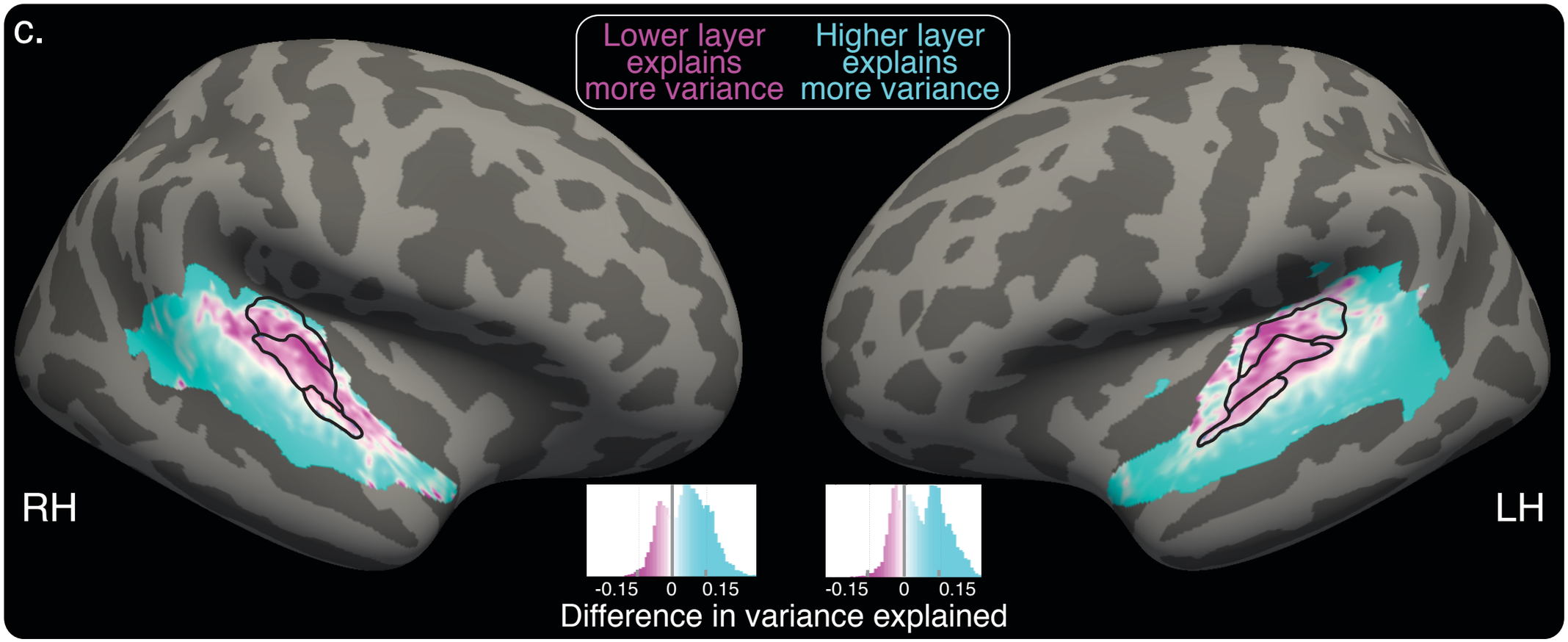
Comparison of Predictivity by RoI



Median Predictivity as a Function of Model Layer



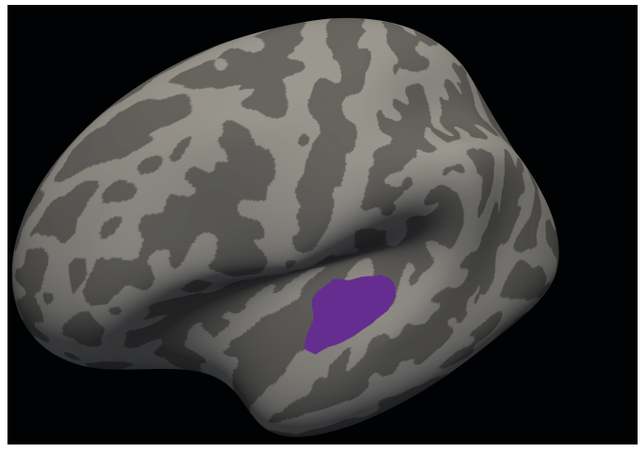
Predictivity Difference Between High and Low Model Layers

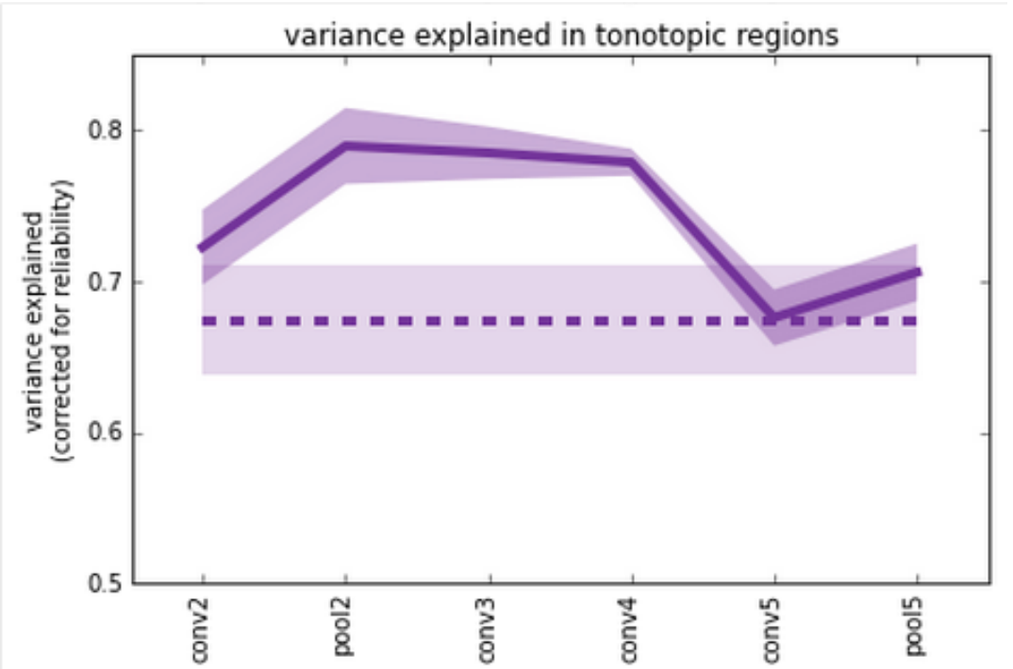


Early layers better explanation of primary cortex, higher layers better explanation of non-primary cortex.

Differentiation by Region of Interest

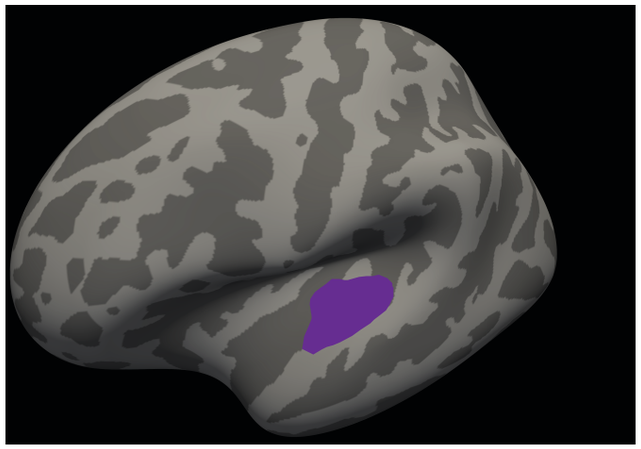
Tonotopic
(⊂ Primary)

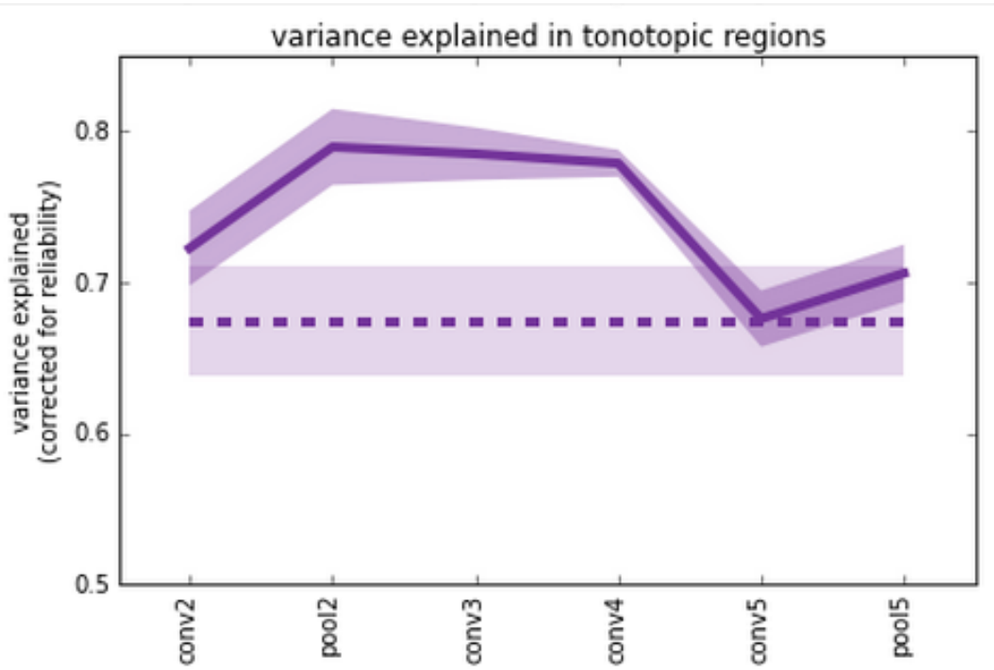




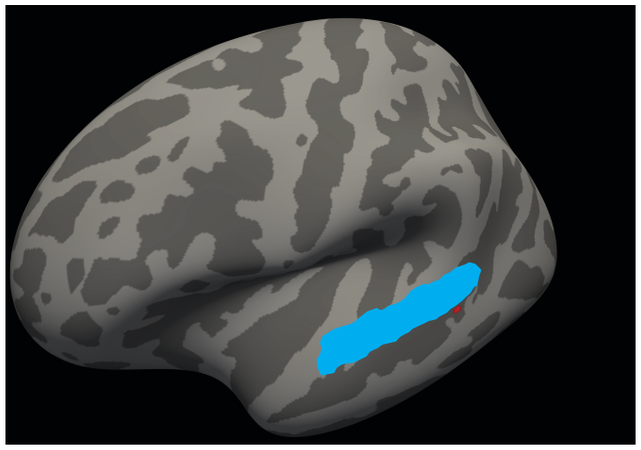
Differentiation by Region of Interest

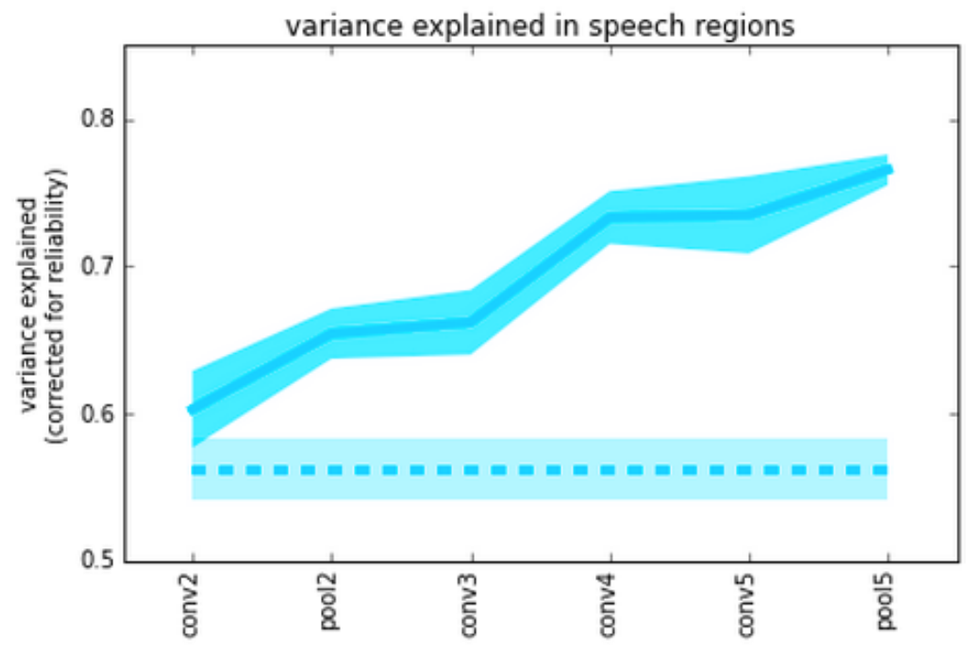
Tonotopic
(⊂ Primary)





Speech-selective
(⊂ Non-primary)





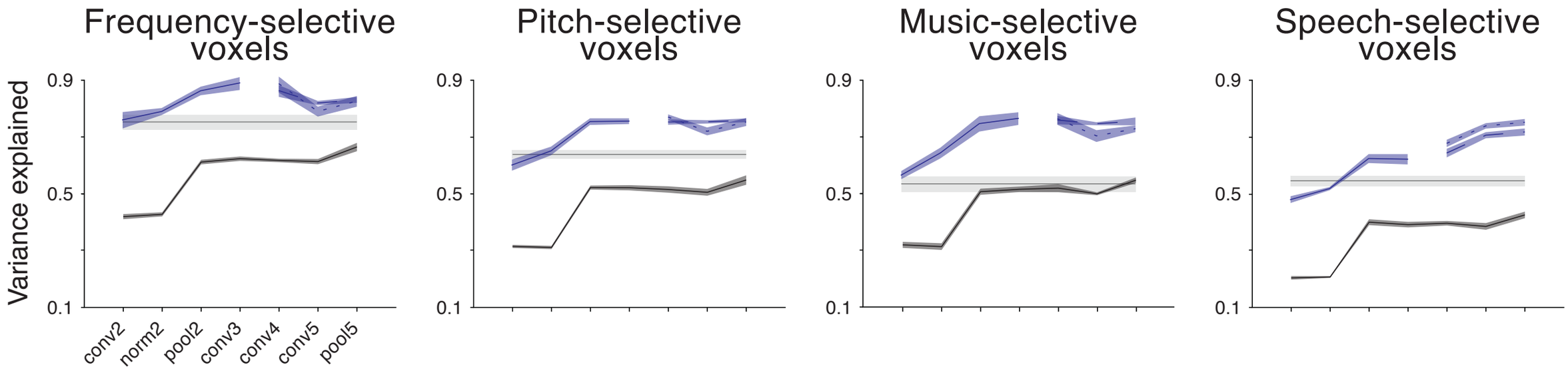
Differentiation by Region of Interest

Our network (selected architecture, trained filters)

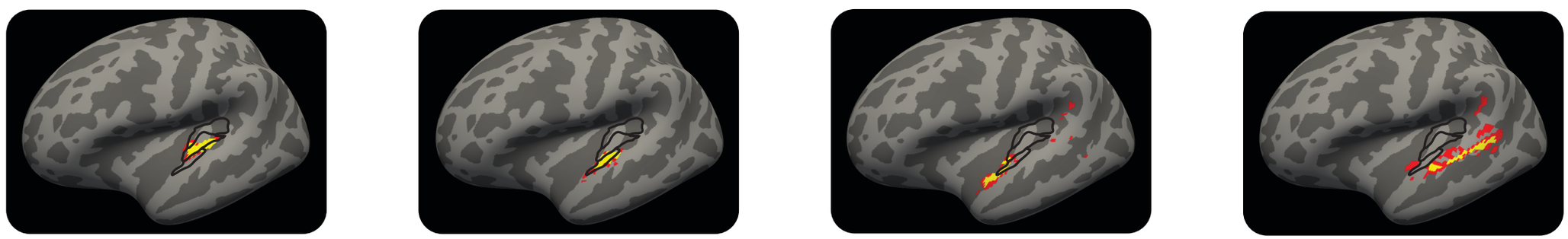
Spectrotemporal model

Random-filter network (selected architecture, untrained filters)

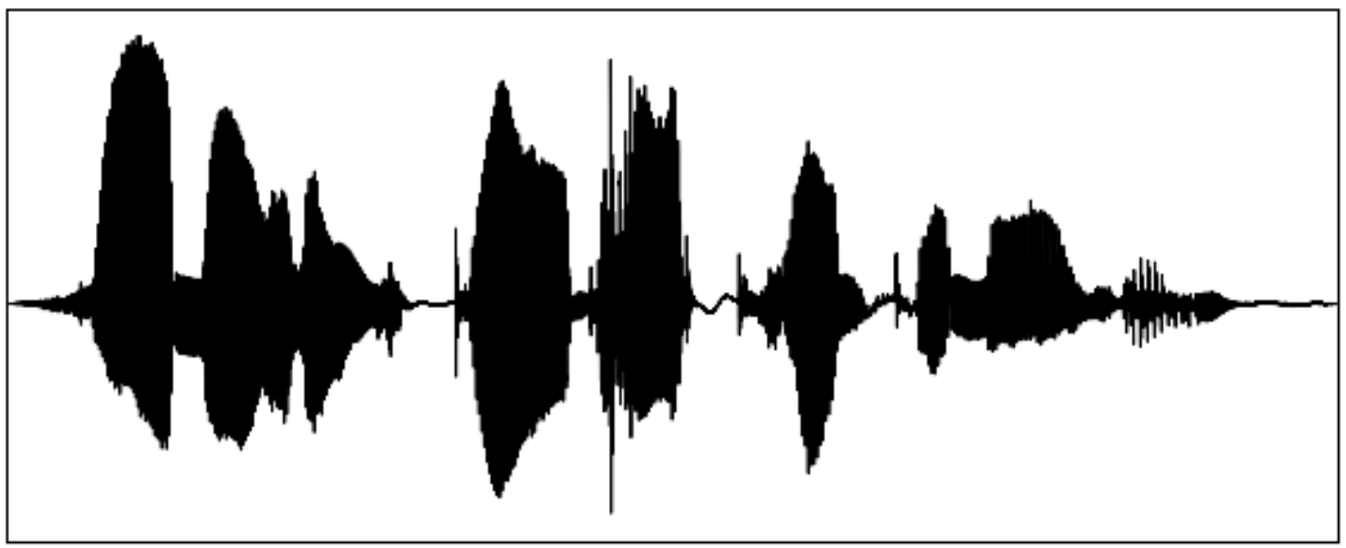
Random-filter network (unselected architectures, untrained filters)

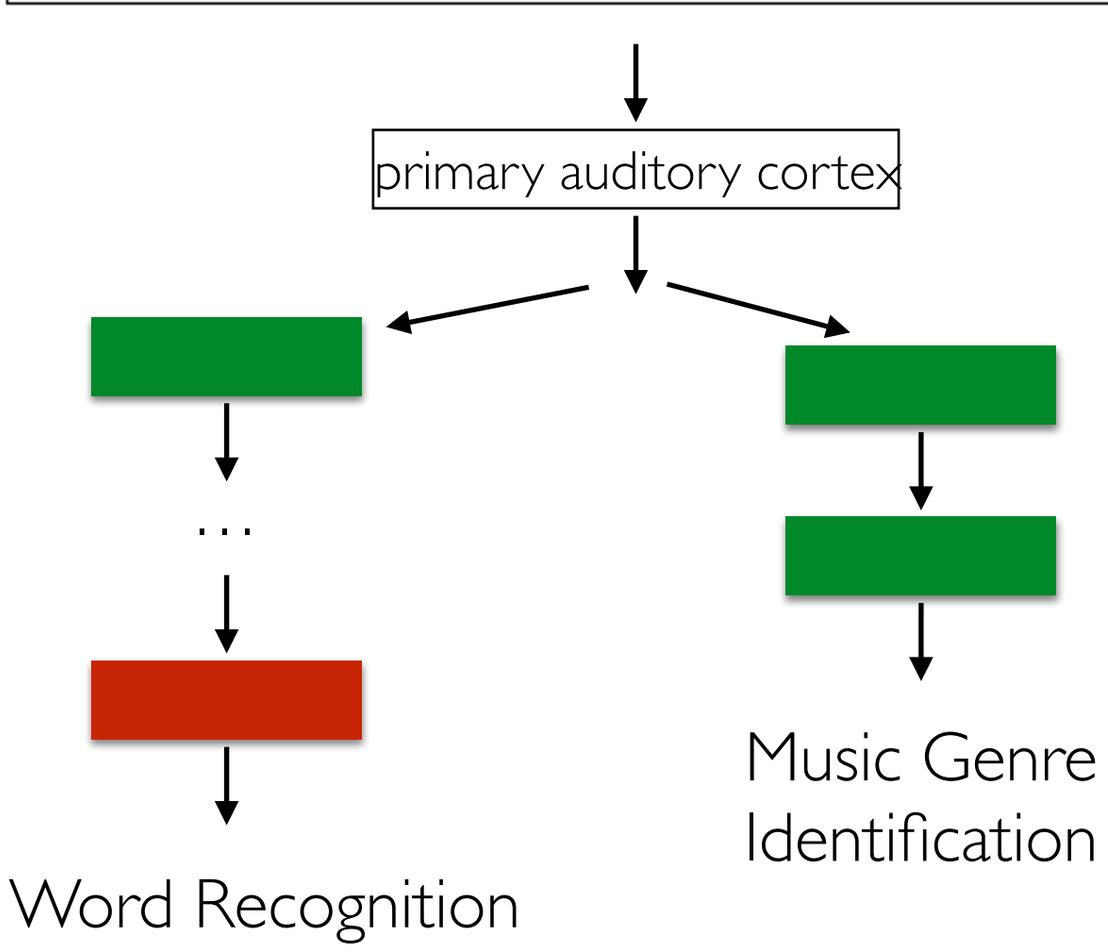


Region of interest

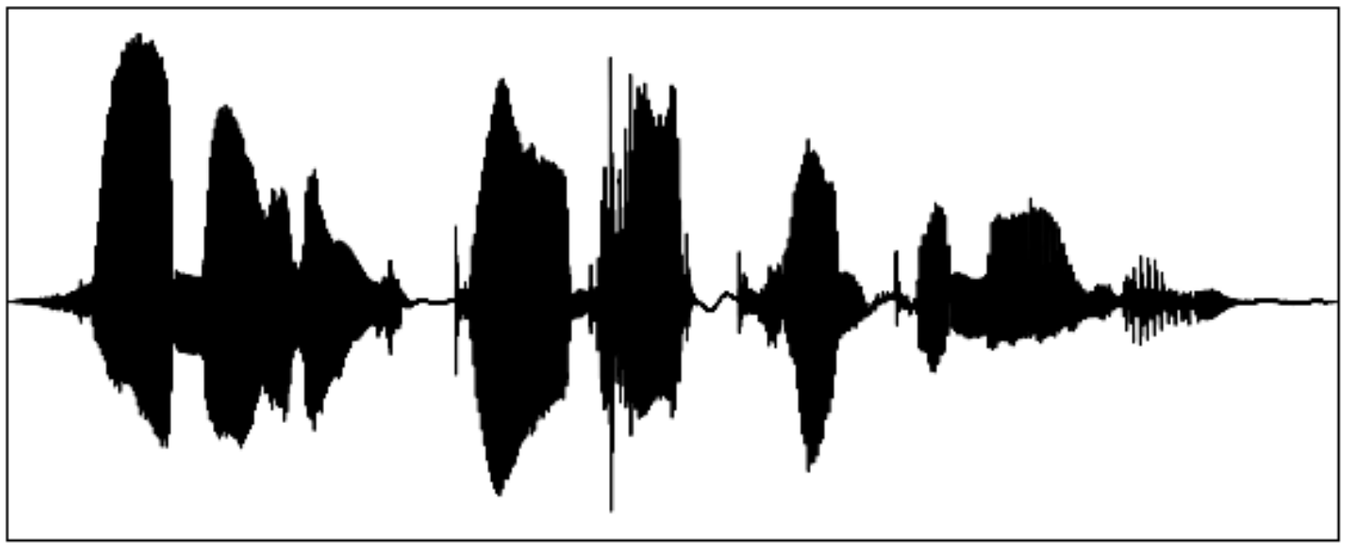


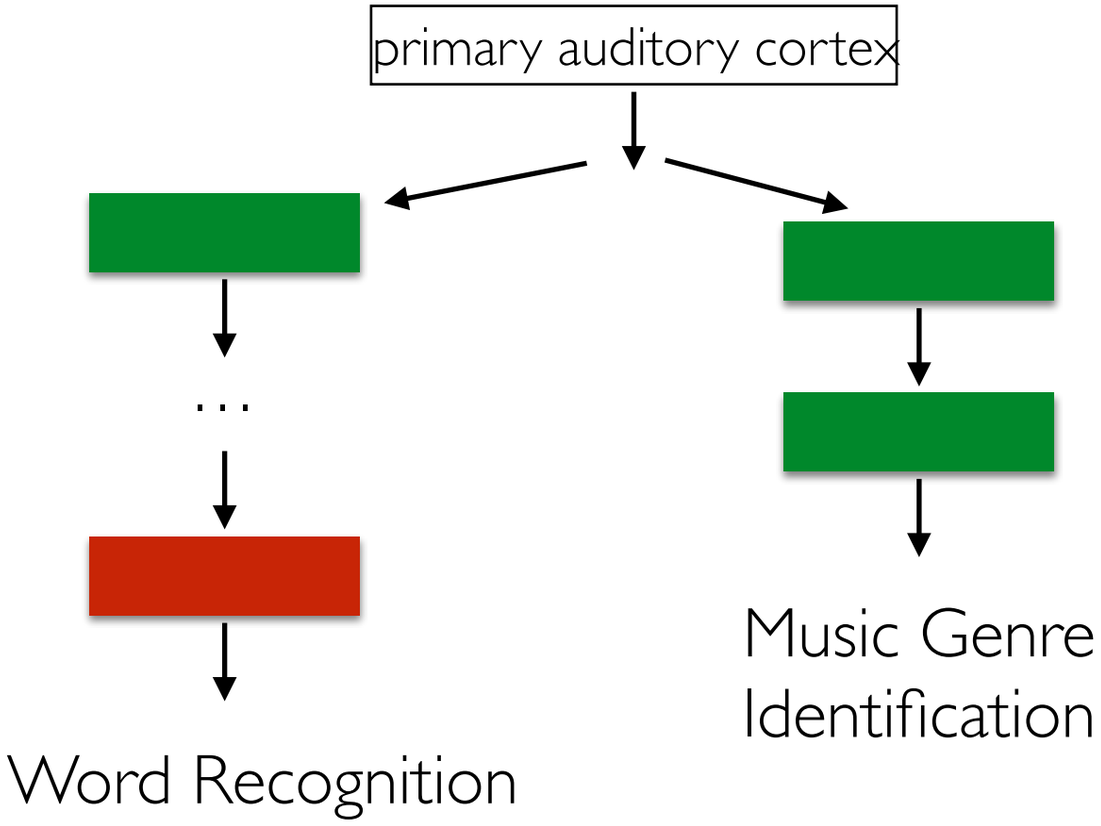
Ongoing: Functionality Organization by Task

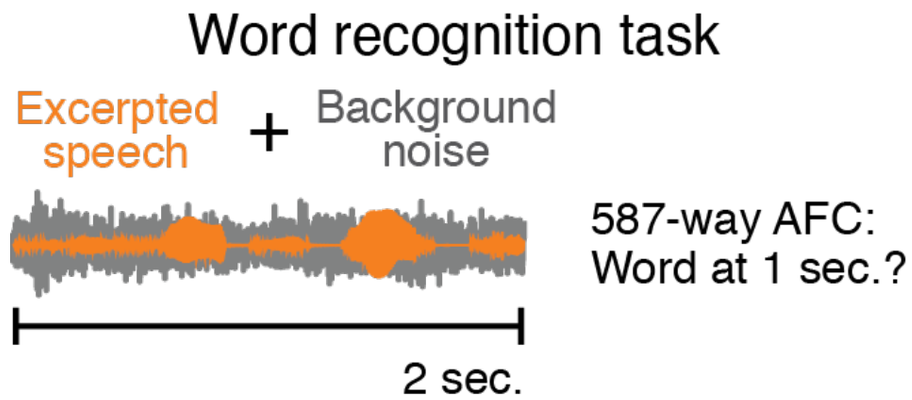




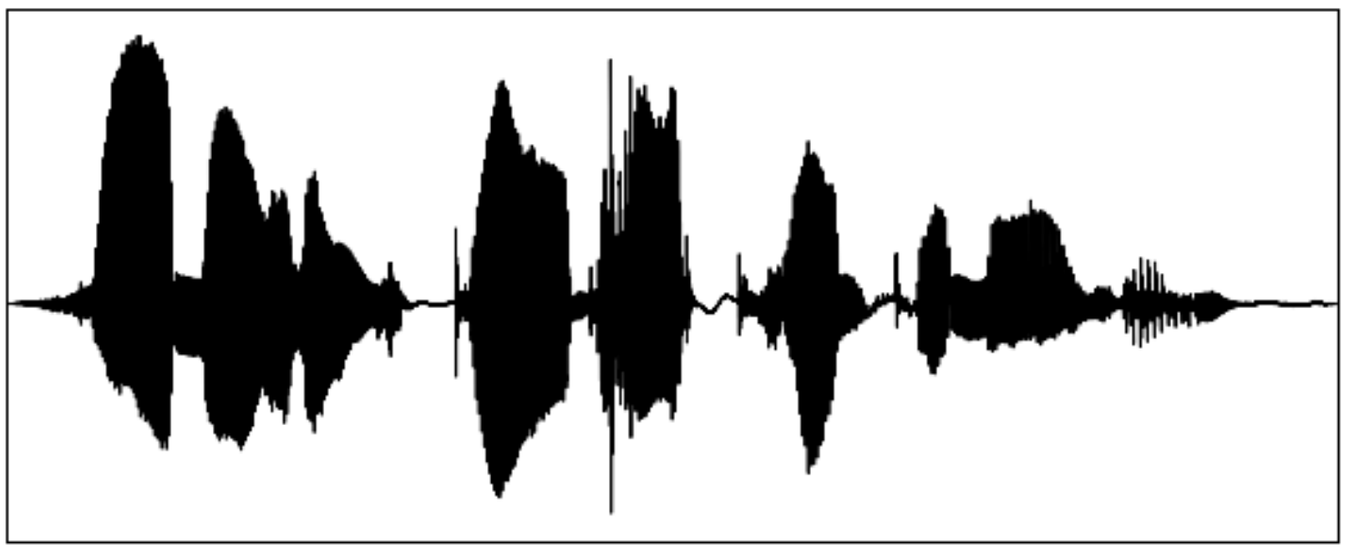
Ongoing: Functionality Organization by Task

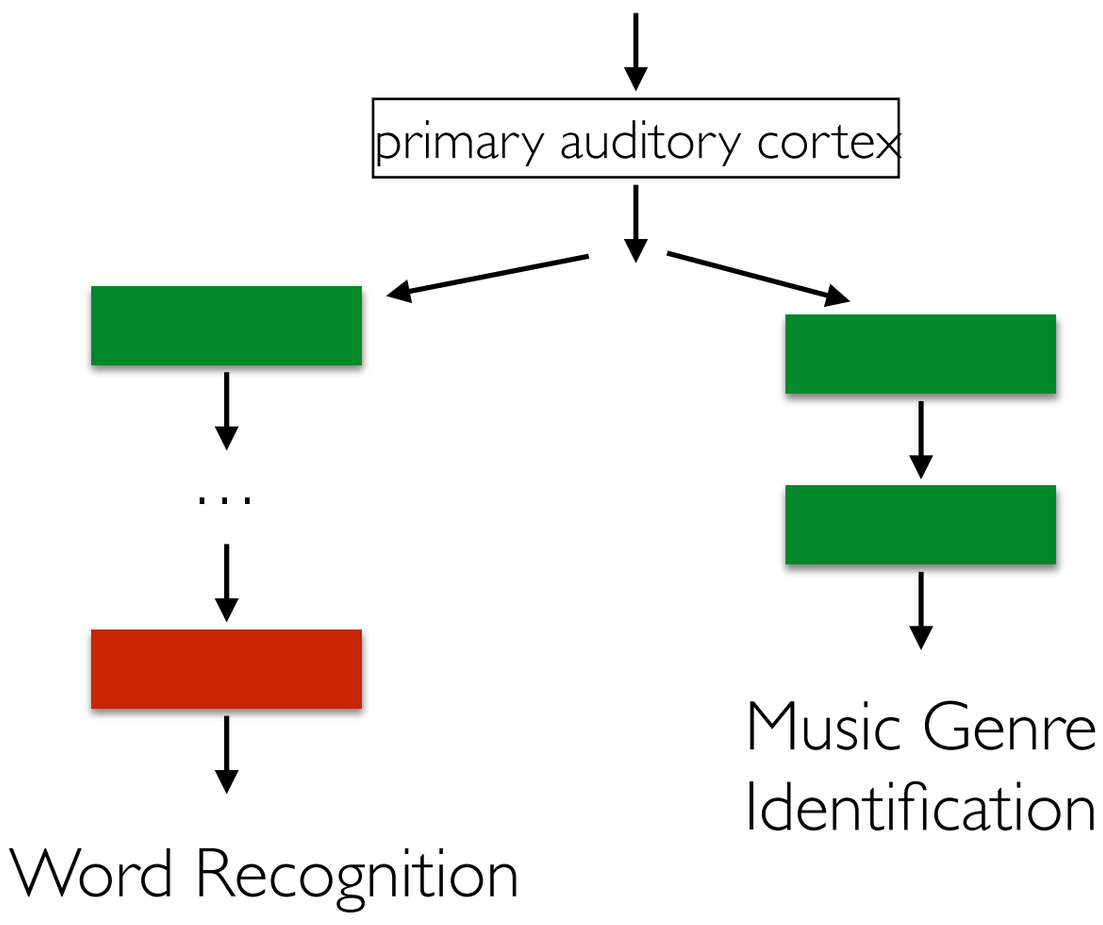


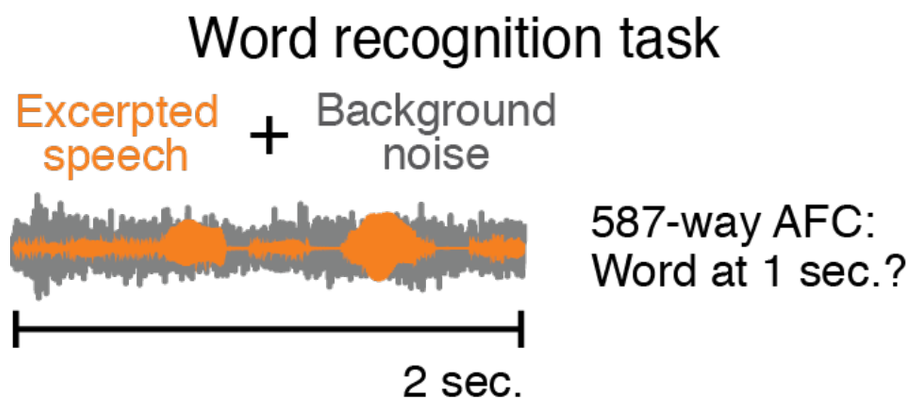


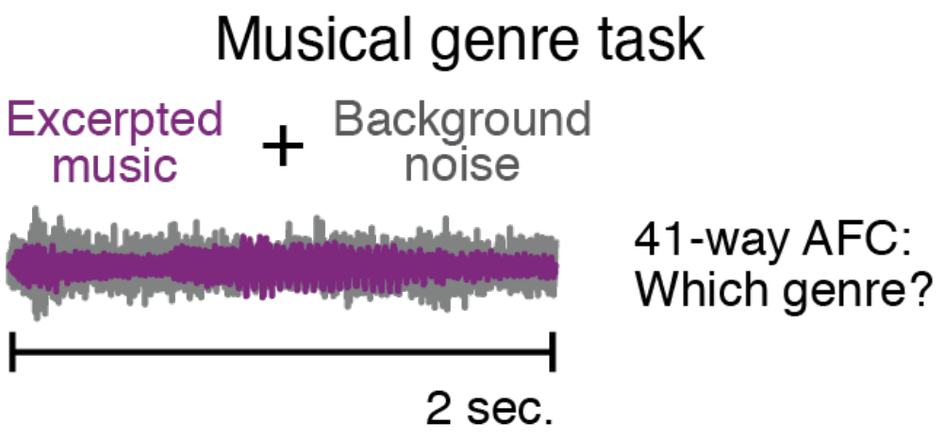


Ongoing: Functionality Organization by Task

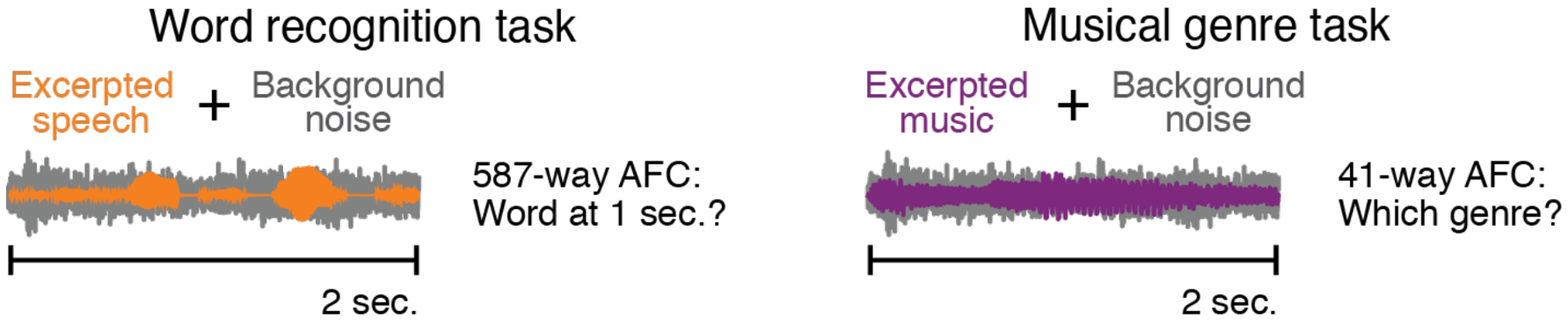






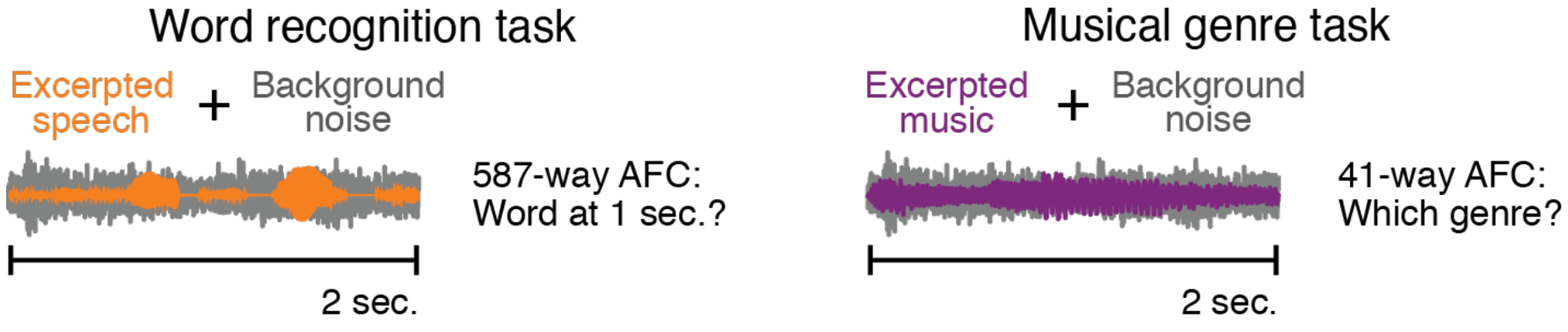


Ongoing: Functionality Organization by Task



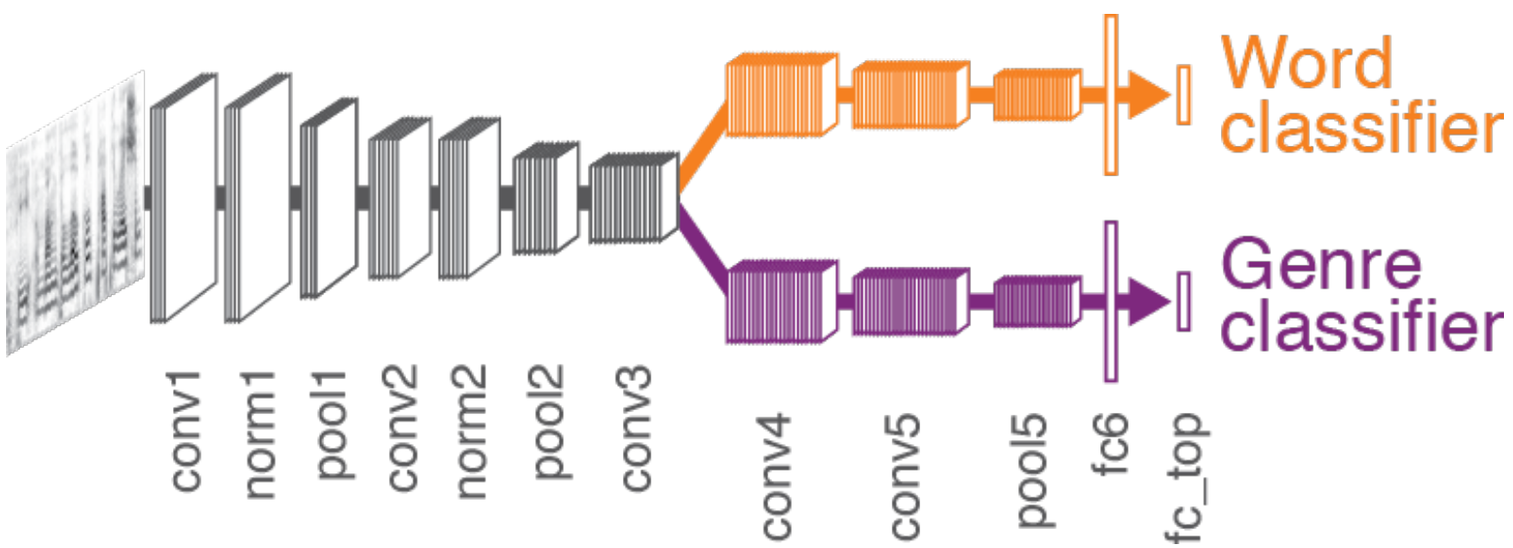
Variety of architectures with different stream branching points:

Ongoing: Functionality Organization by Task



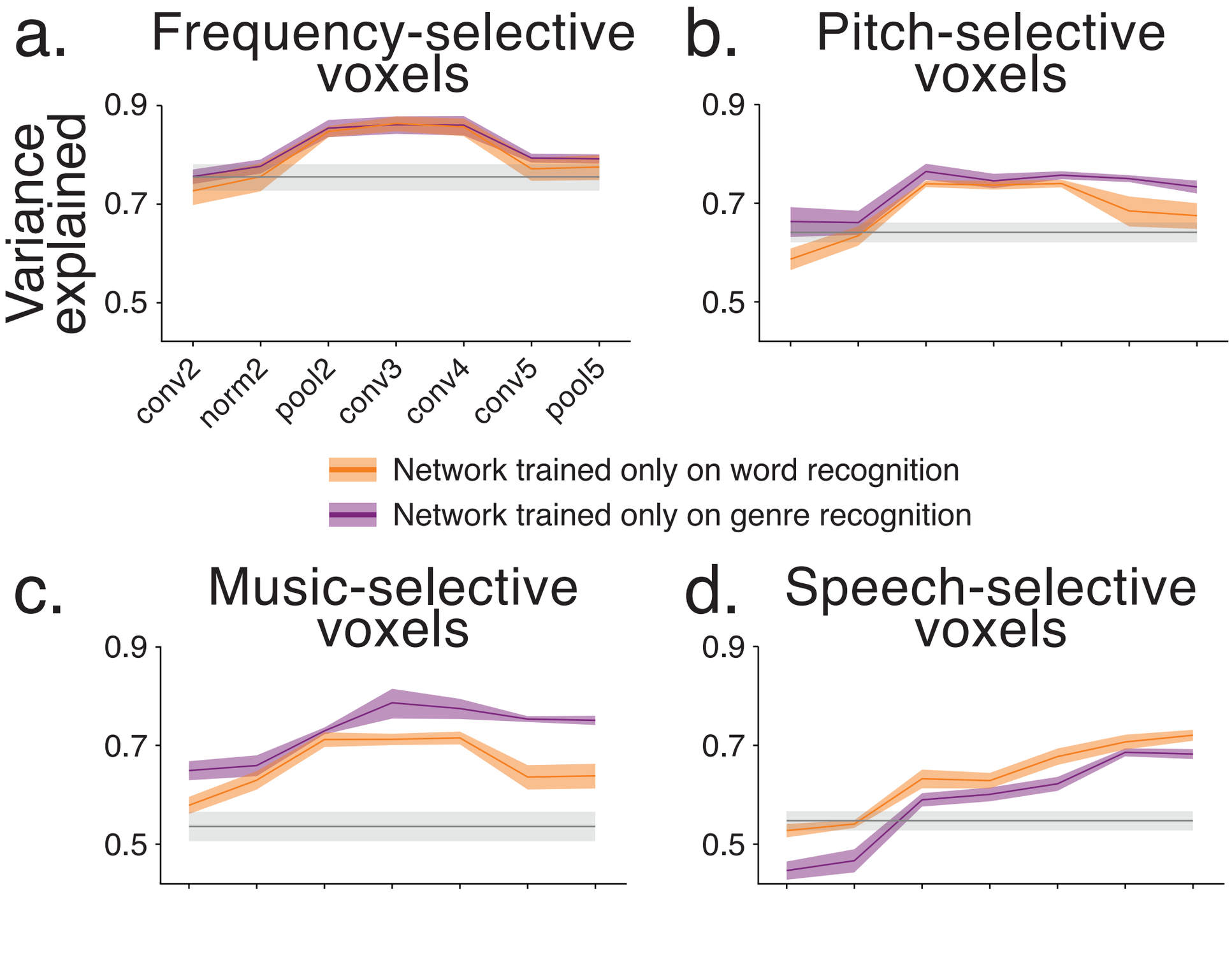
Variety of architectures with different stream branching points:

Architectural meta-parameter optimization yields specific branched model:

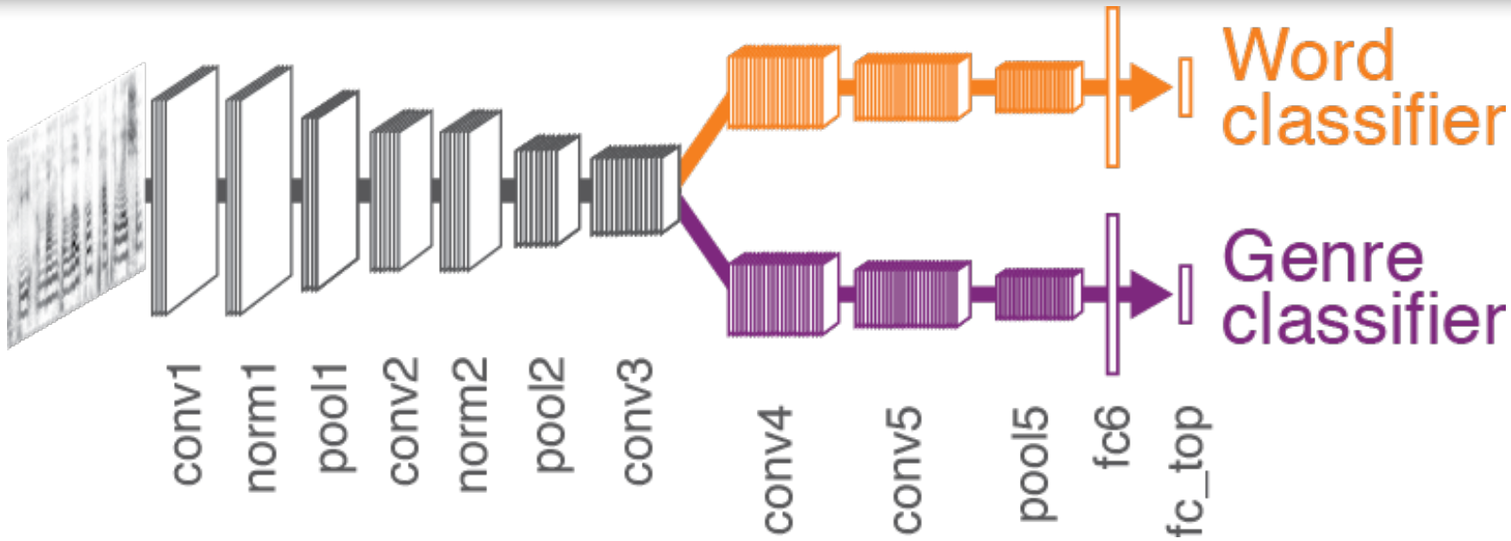


Analysis of Model Architectures

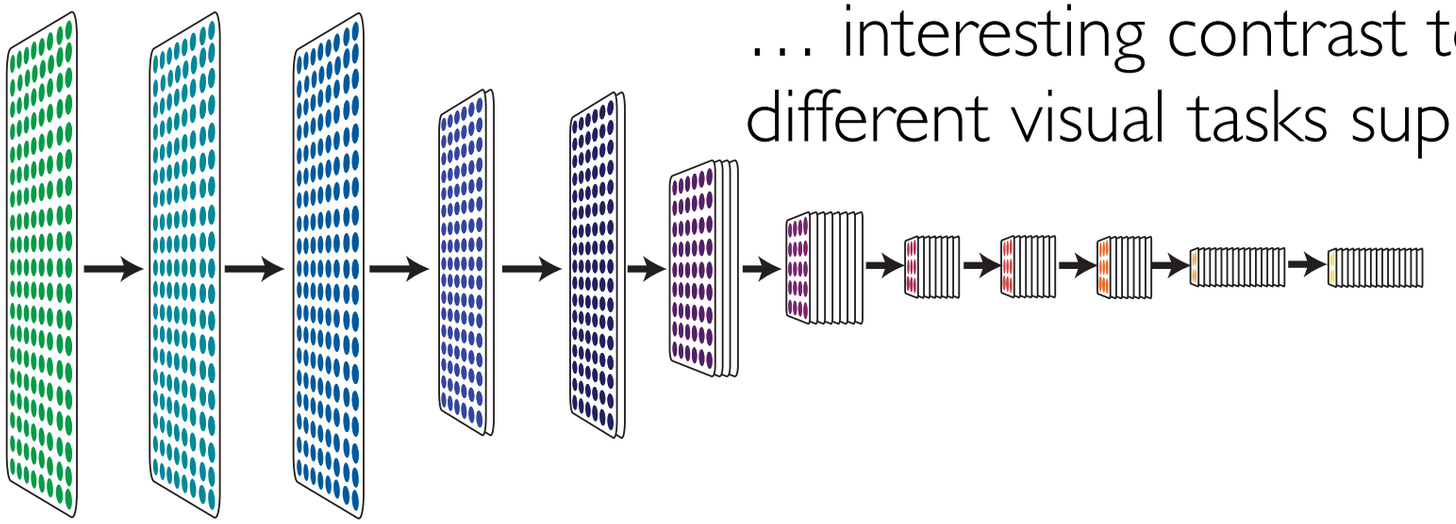
Differentiation of processing streams into different subsets of brain voxels:



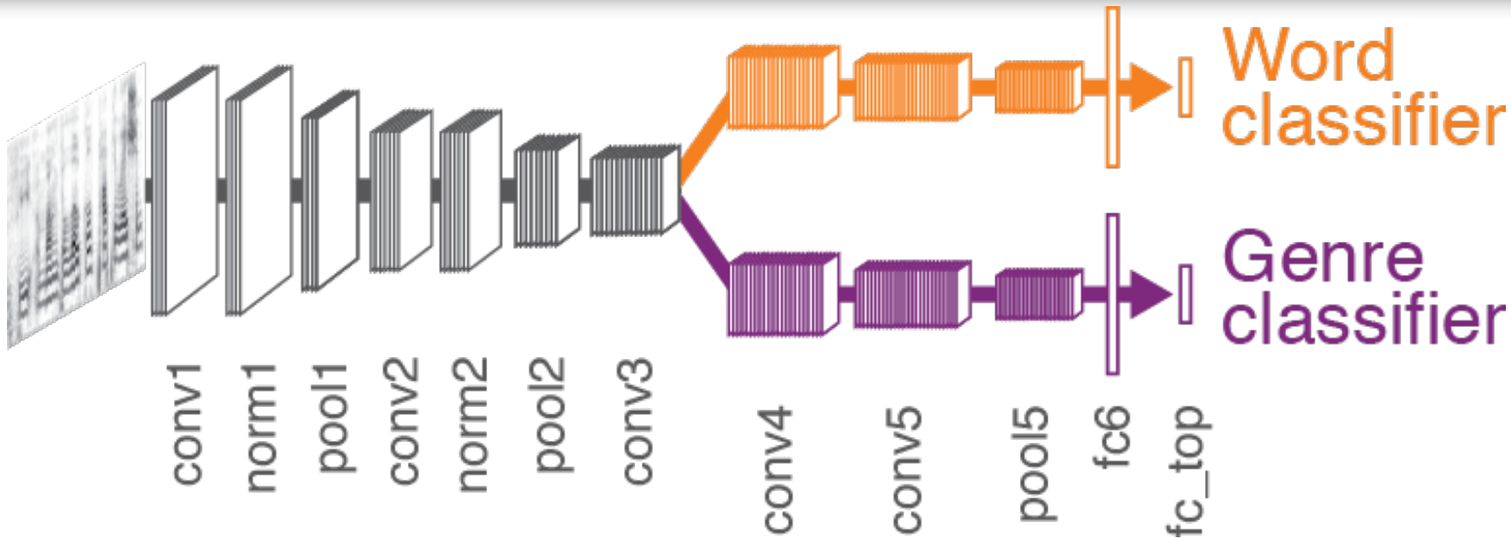
Analysis of Model Architectures



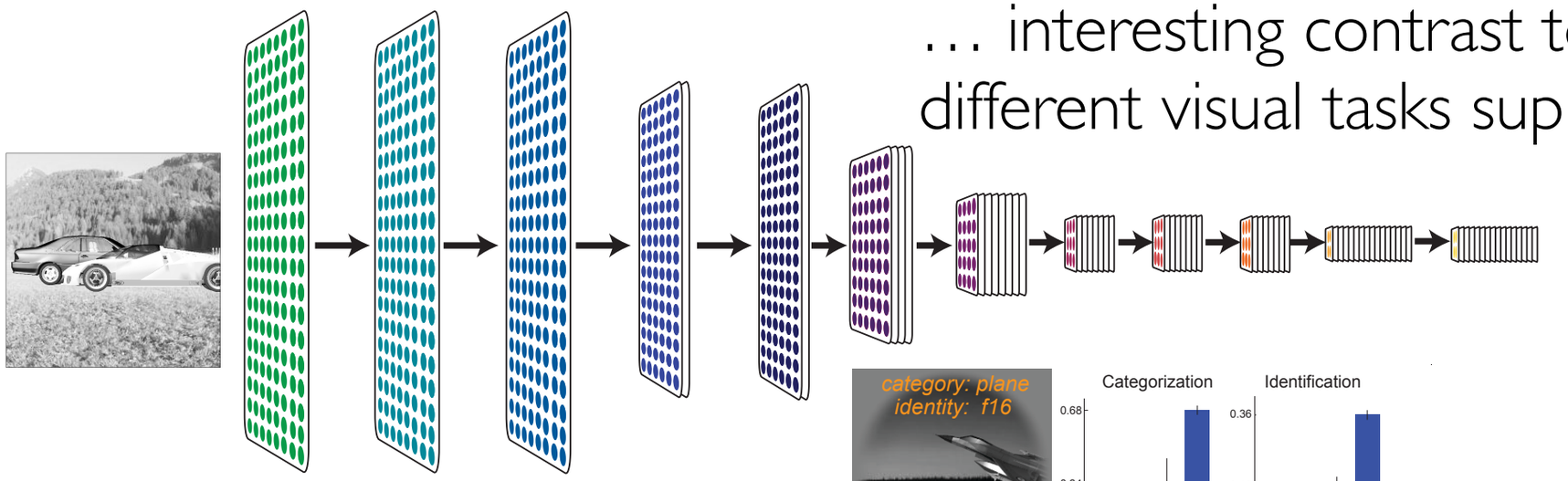
… interesting contrast to ventral stream, many different visual tasks supported by single stream…



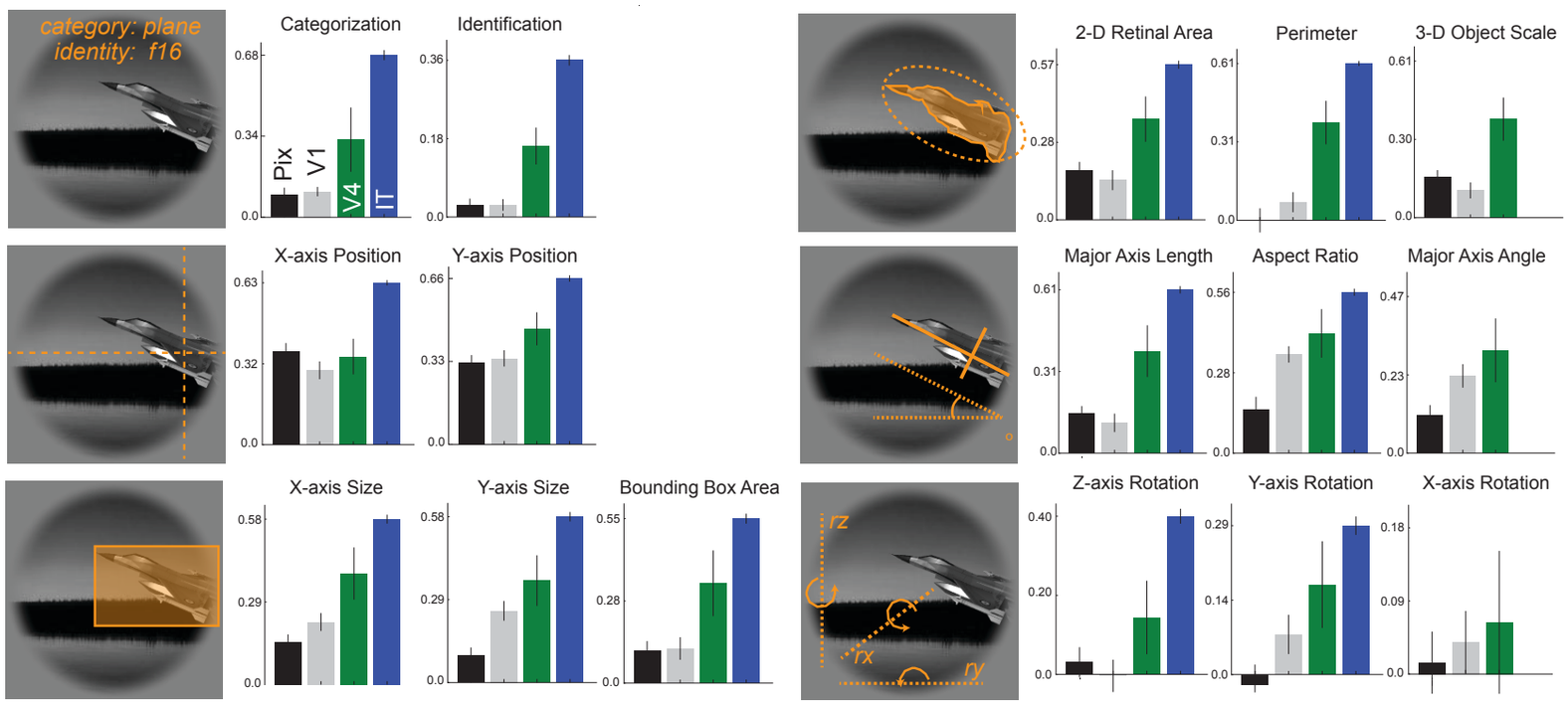
Analysis of Model Architectures



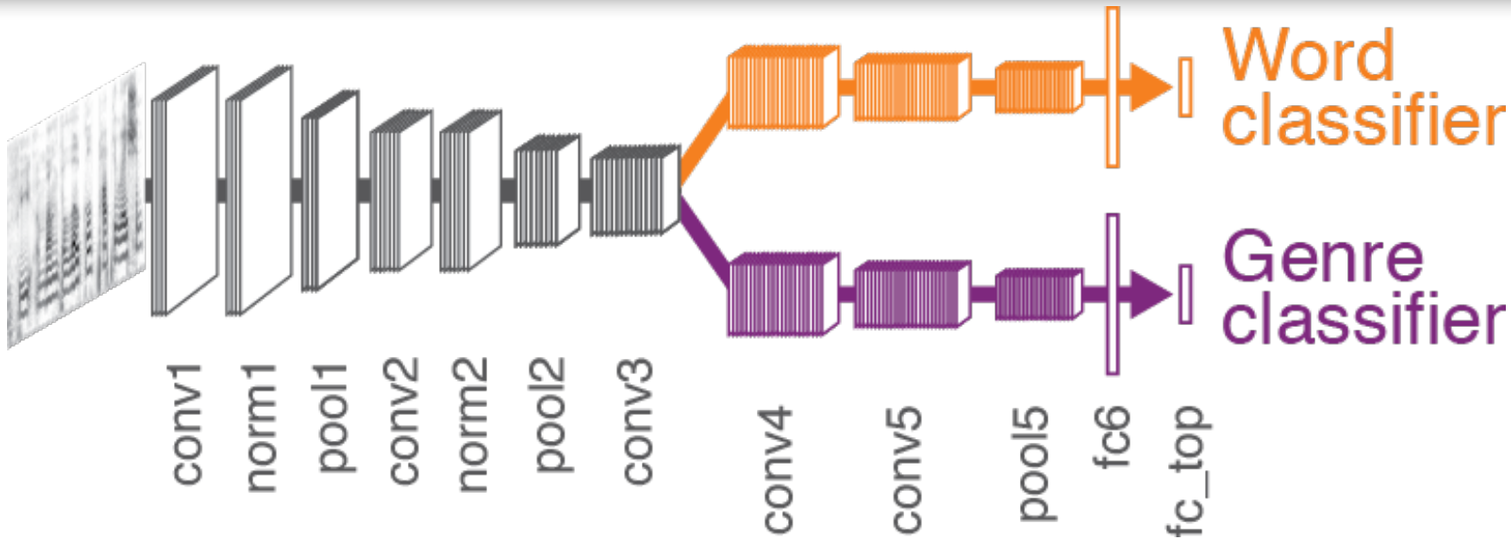
… interesting contrast to ventral stream, many different visual tasks supported by single stream…



IT > V4 > V1 across many tested visual tasks (see lect. 3)

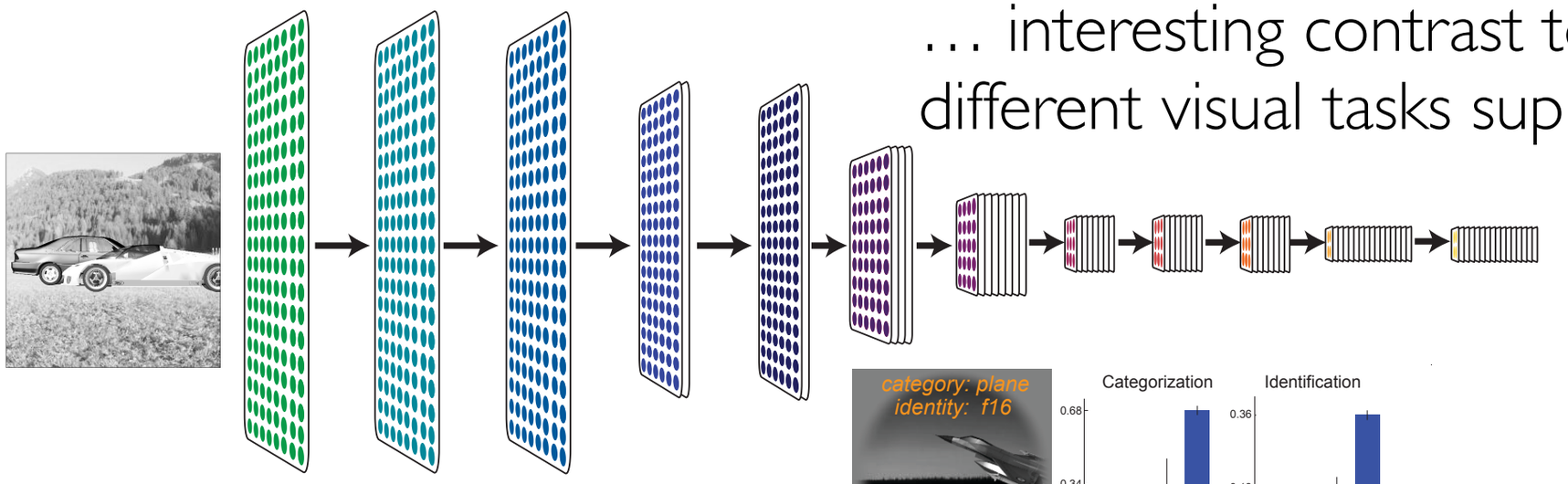


Analysis of Model Architectures



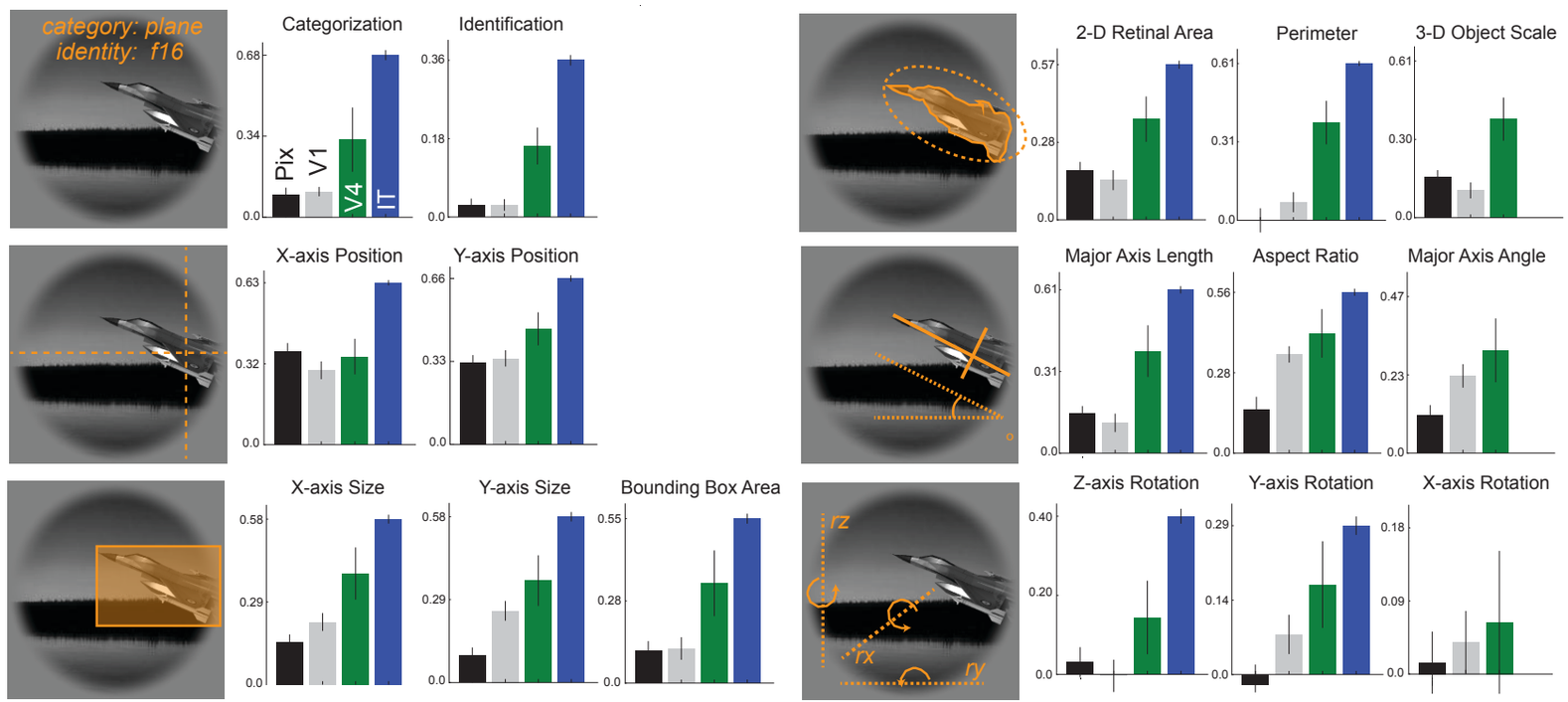
Auditory & Visual Cortex:

similar but different

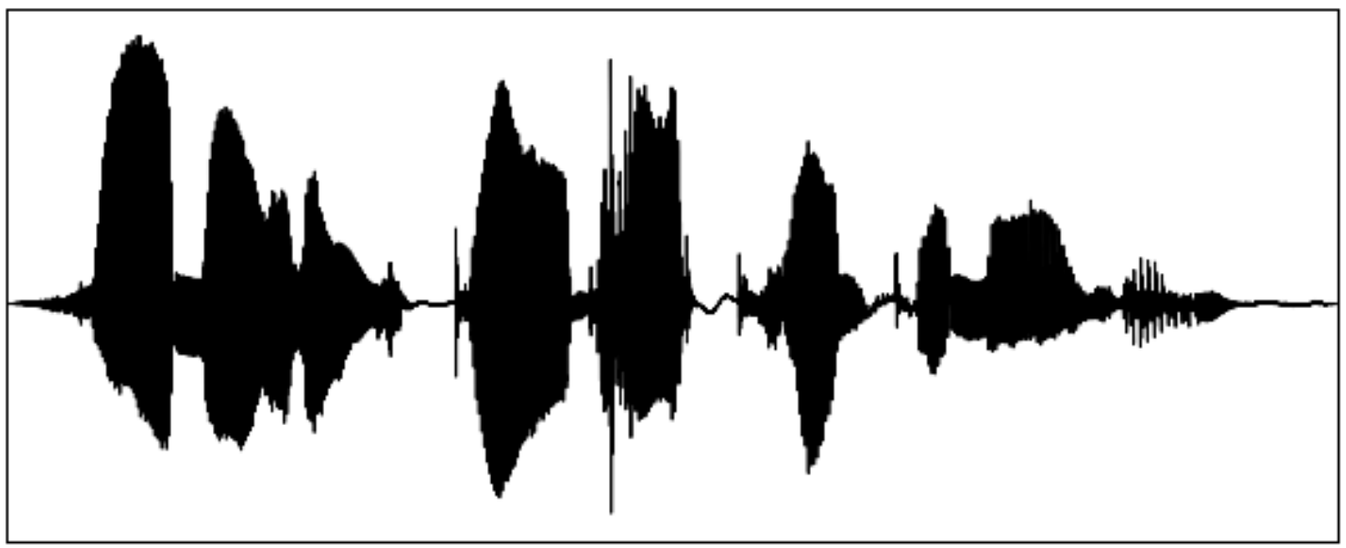


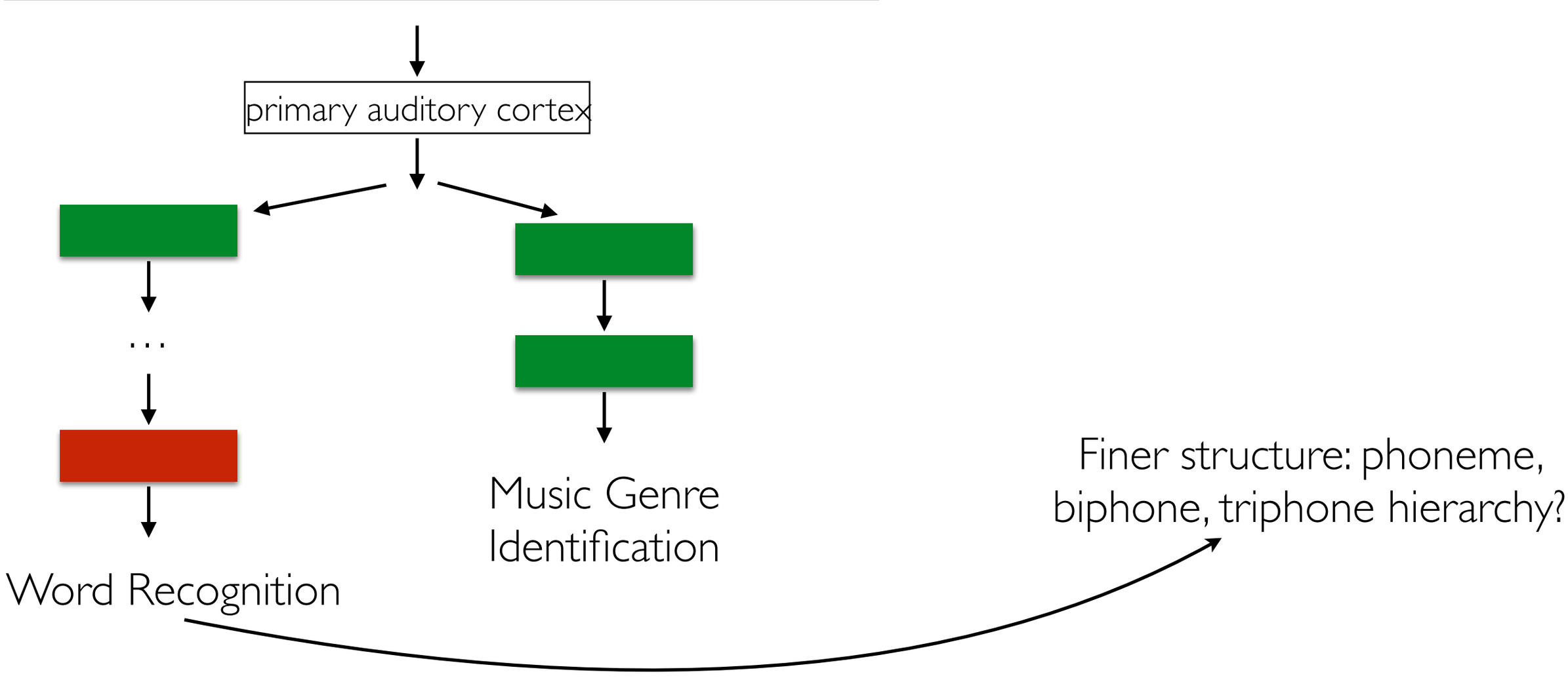
… interesting contrast to ventral stream, many different visual tasks supported by single stream…

IT > V4 > V1 across many tested visual tasks (see lect. 3)



Ongoing: Functionality Organization by Task





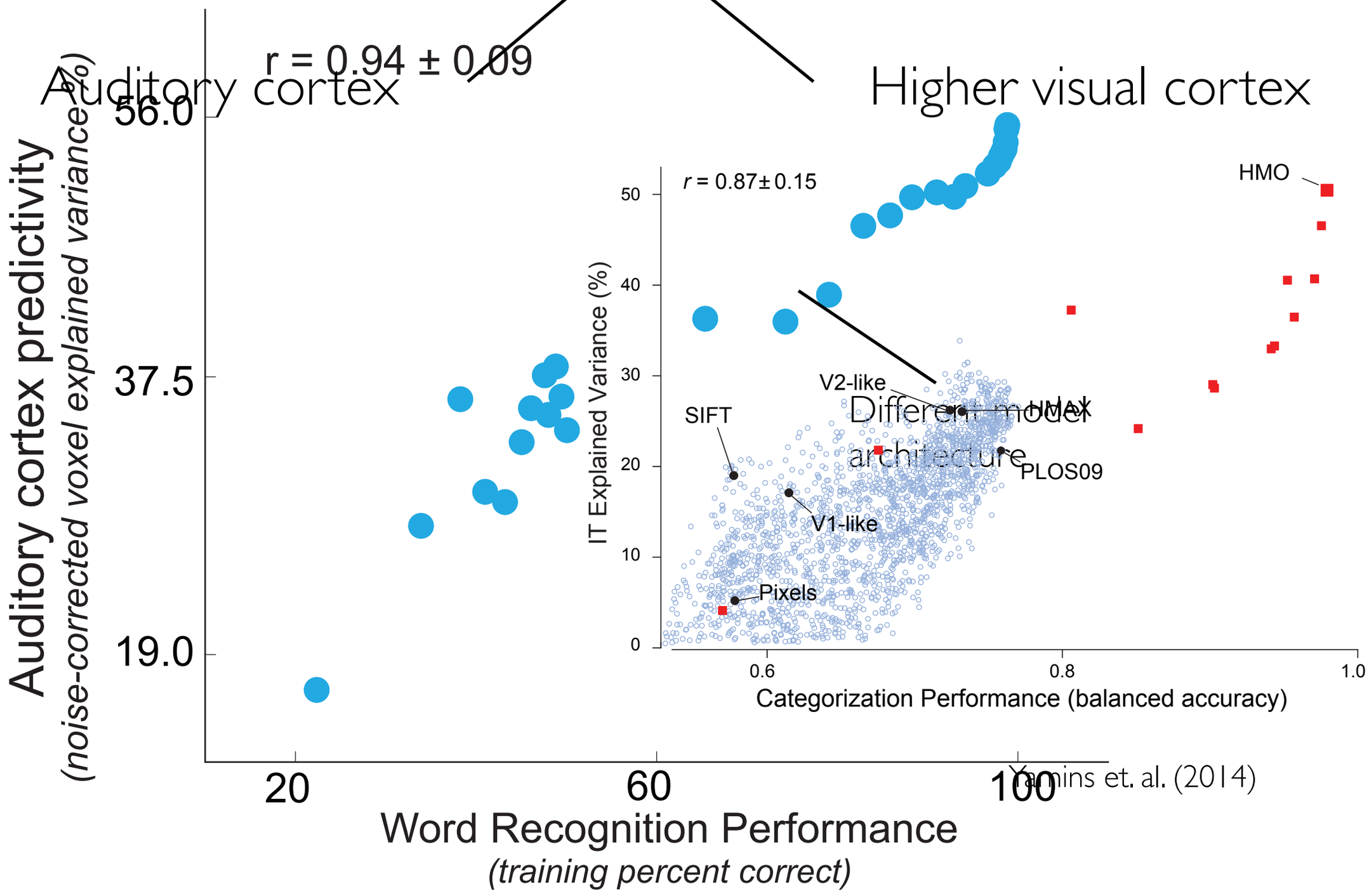
Goal-Driven Modeling Principle

High-variation task performance vs:

r = 0.94 ± 0.09

Auditory cortex

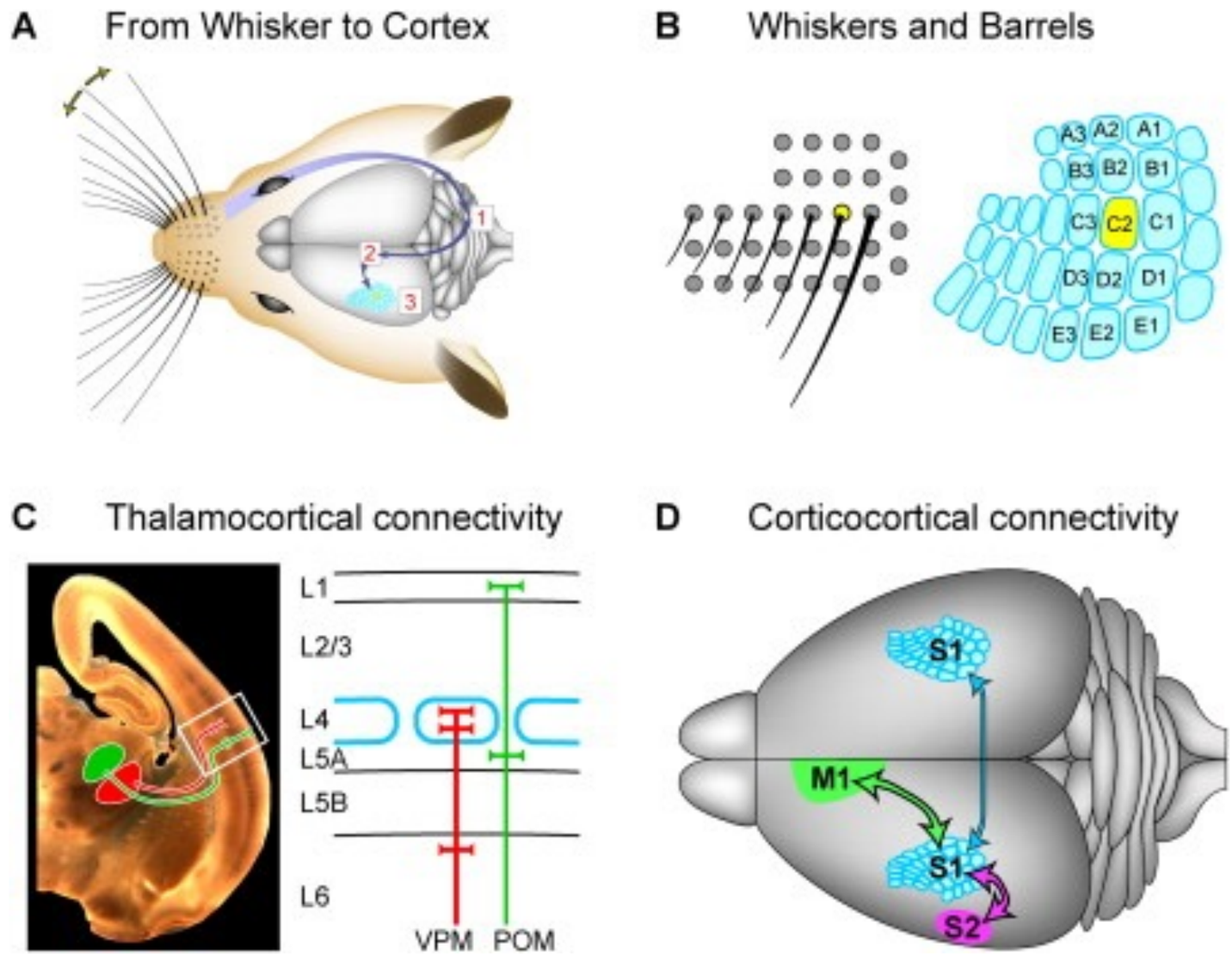
Higher visual cortex

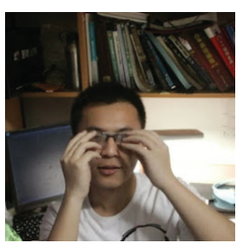


Yamins et. al. (2014)

Rodent Somatosensory Cortex

Petersen, 2007



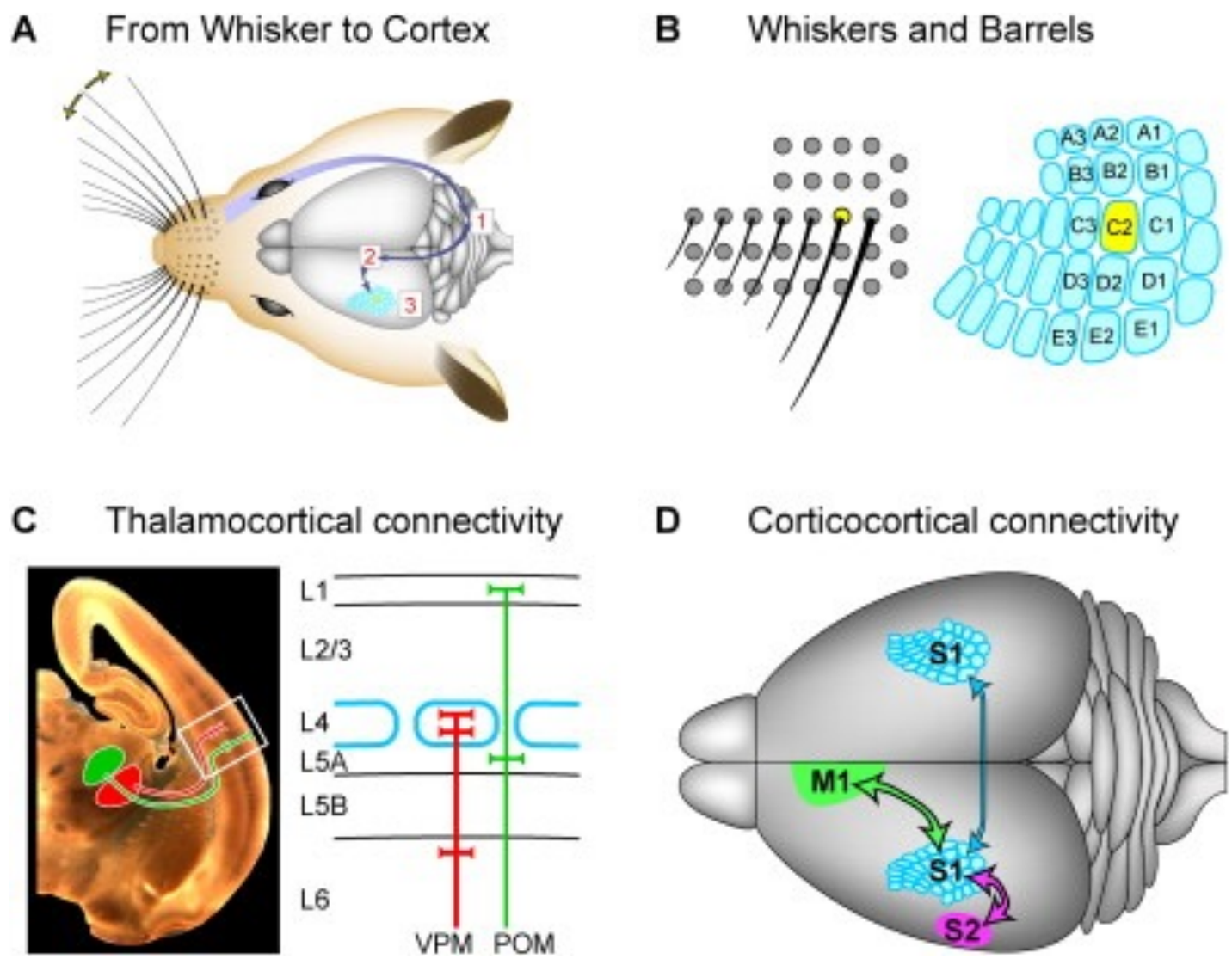


Chengxu Zhuang

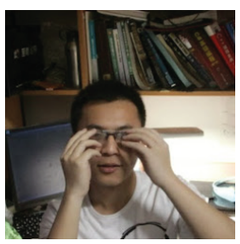
Mitra Hartmann & Lab

Rodent Somatosensory Cortex

Petersen, 2007



- Spatially-structured input data

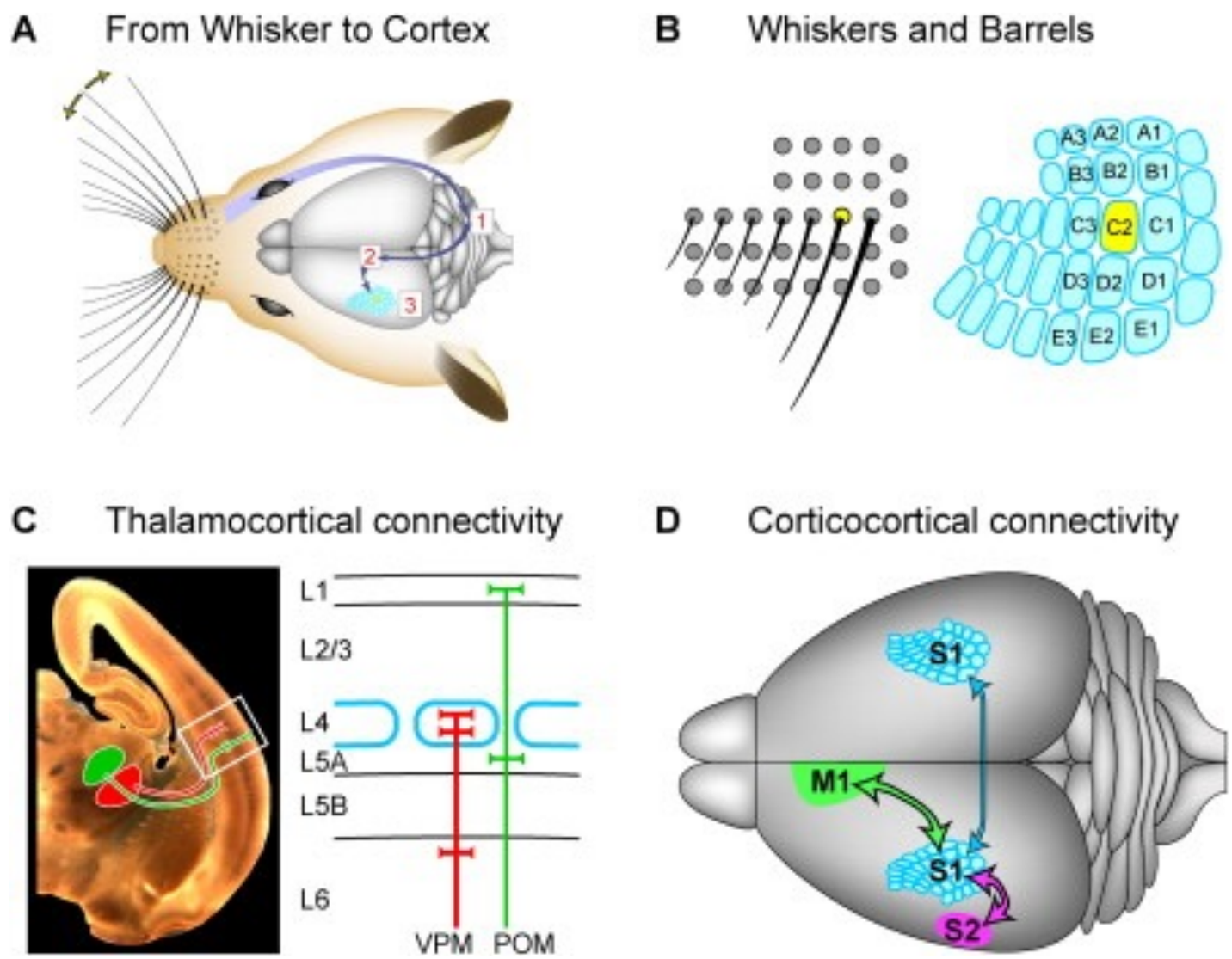


Chengxu Zhuang

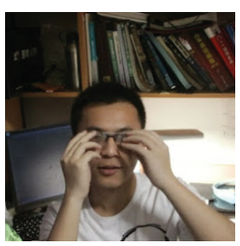
Mitra Hartmann & Lab

Rodent Somatosensory Cortex

Petersen, 2007



- Spatially-structured input data
- Spatiotopic sensor

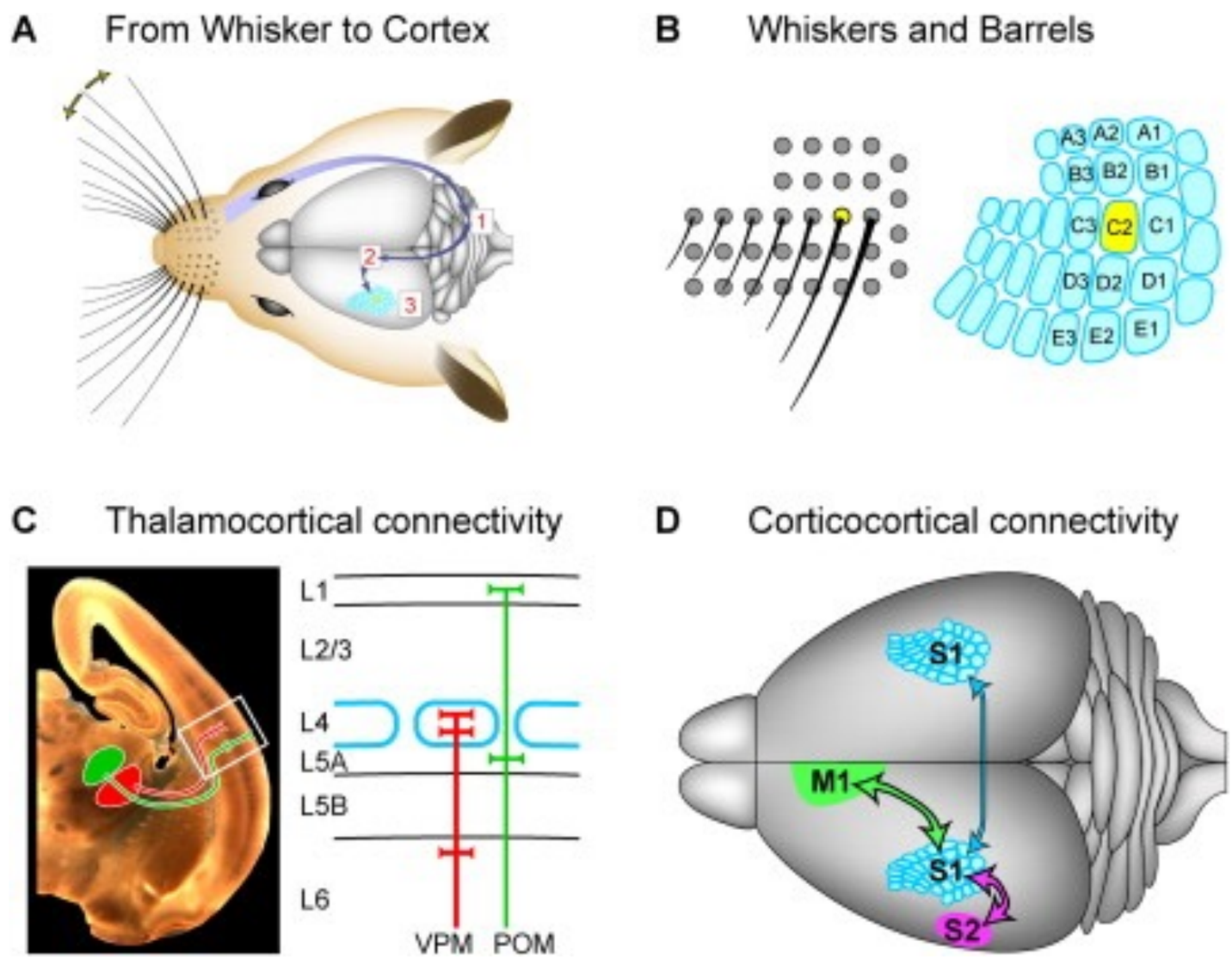


Chengxu Zhuang

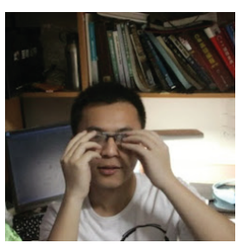
Mitra Hartmann & Lab

Rodent Somatosensory Cortex

Petersen, 2007



- Spatially-structured input data
- Spatiotopic sensor
- Potentially hierarchical structure

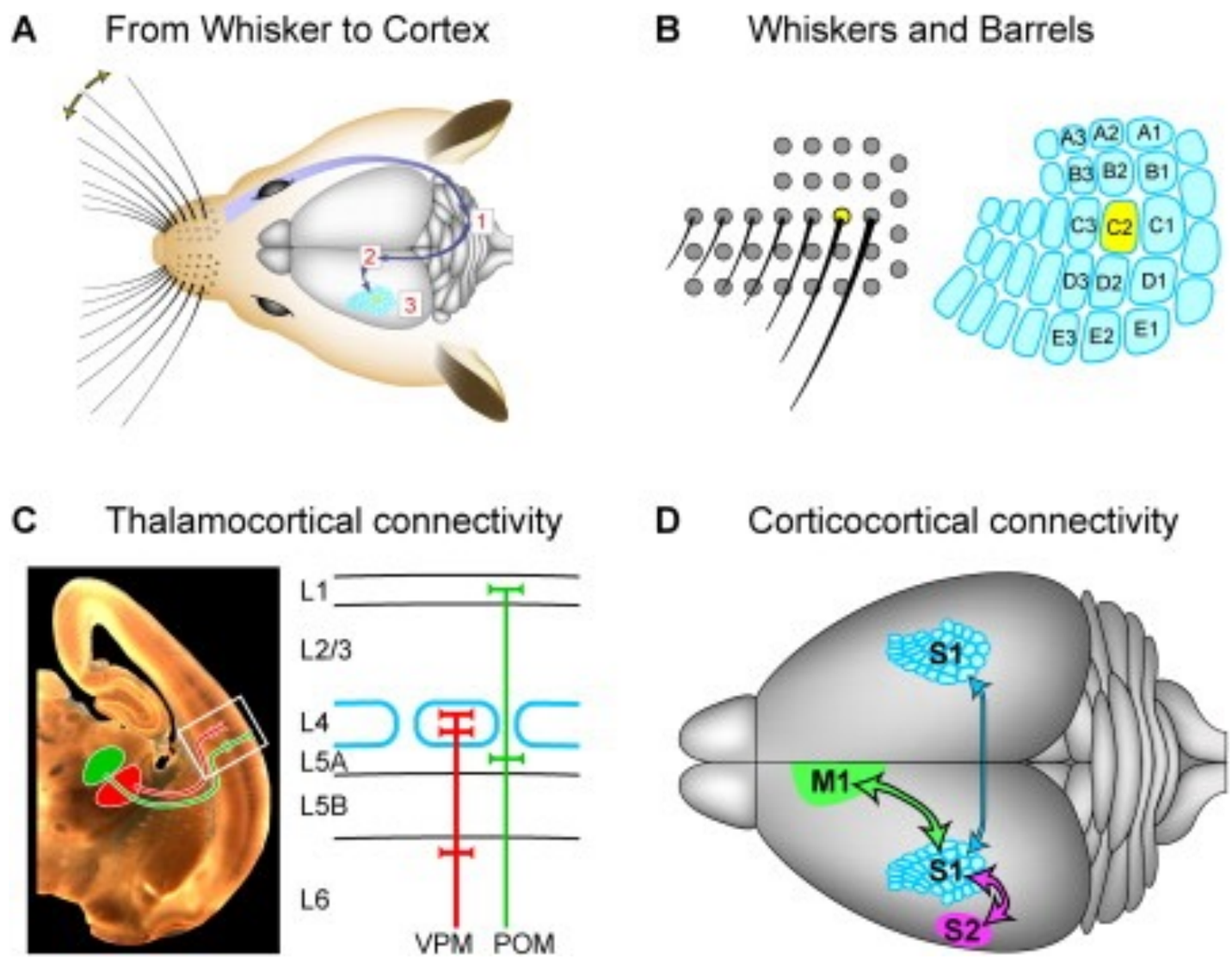


Chengxu Zhuang

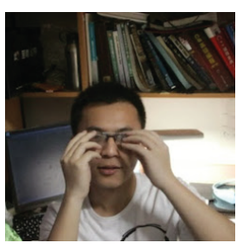
Mitra Hartmann & Lab

Rodent Somatosensory Cortex

Petersen, 2007



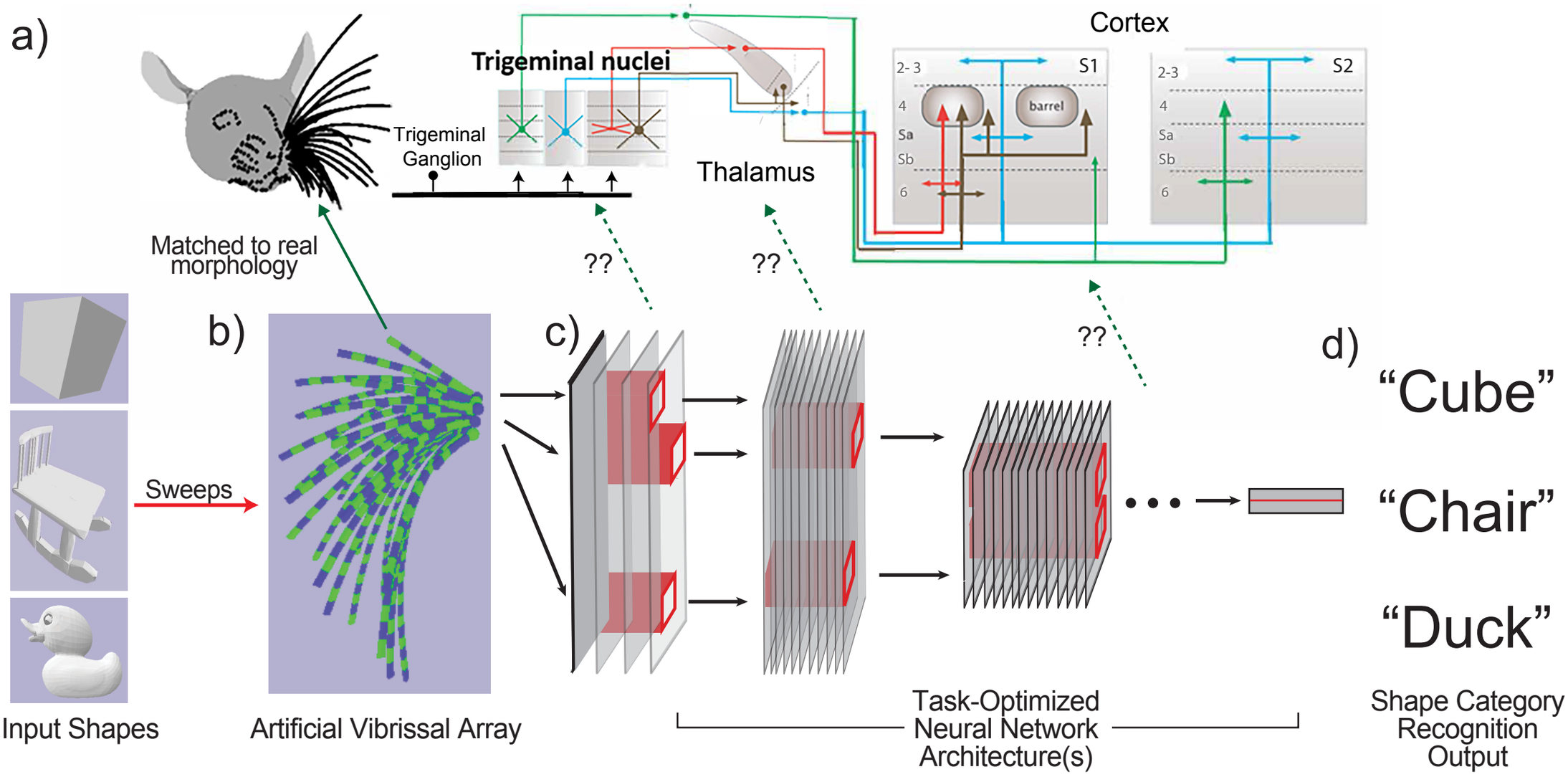
- Spatially-structured input data
- Spatiotopic sensor
- Potentially hierarchical structure
- Poorly understood higher cortical areas

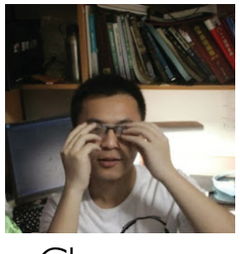


Chengxu Zhuang

Mitra Hartmann & Lab

Rodent Somatosensory Cortex





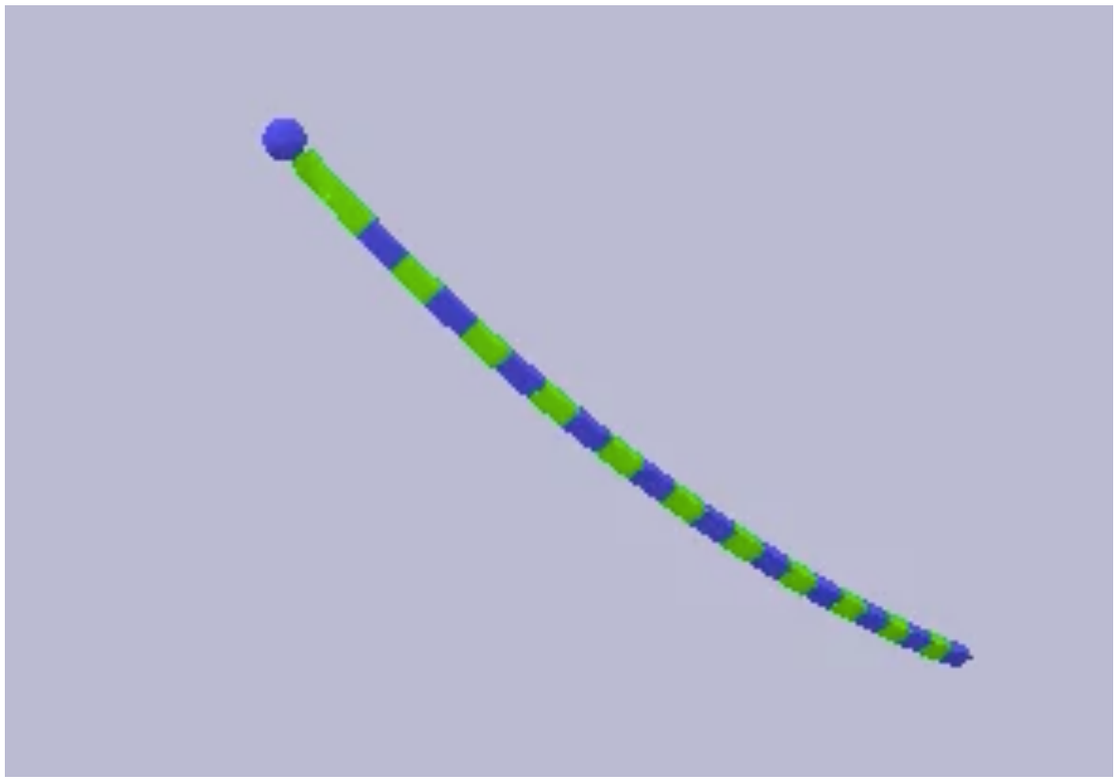
Chengxu Zhuang

Mitra Hartmann & Lab

Hypothesis: can get a model for this cortical cascade by optimizing properly-sized CNN with whisker-like sensor input for some ethologically relevant somatosensory task.

Rodent Somatosensory Cortex

First have to build a model of the sensory to gather data.



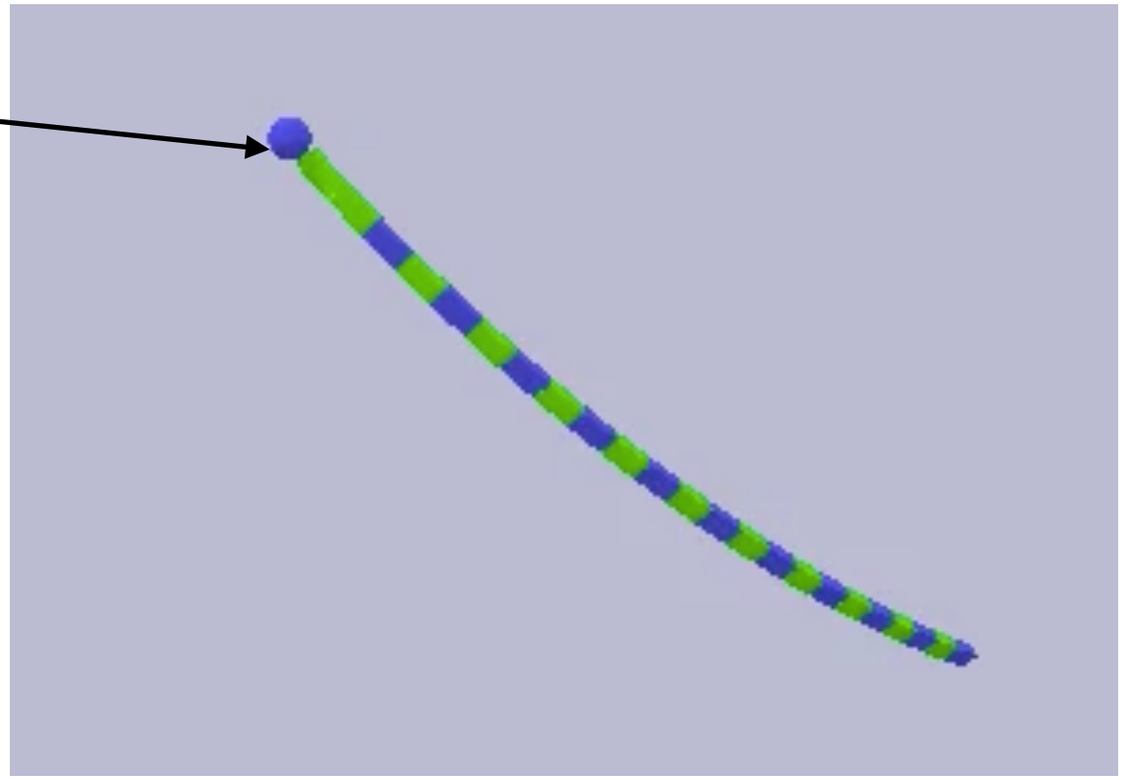
Using published data from Mitra Hartmann's group

Chengxu Zhuang

Mitra Hartmann & Lab

Rodent Somatosensory Cortex

First have to build a model of the sensory to gather data.

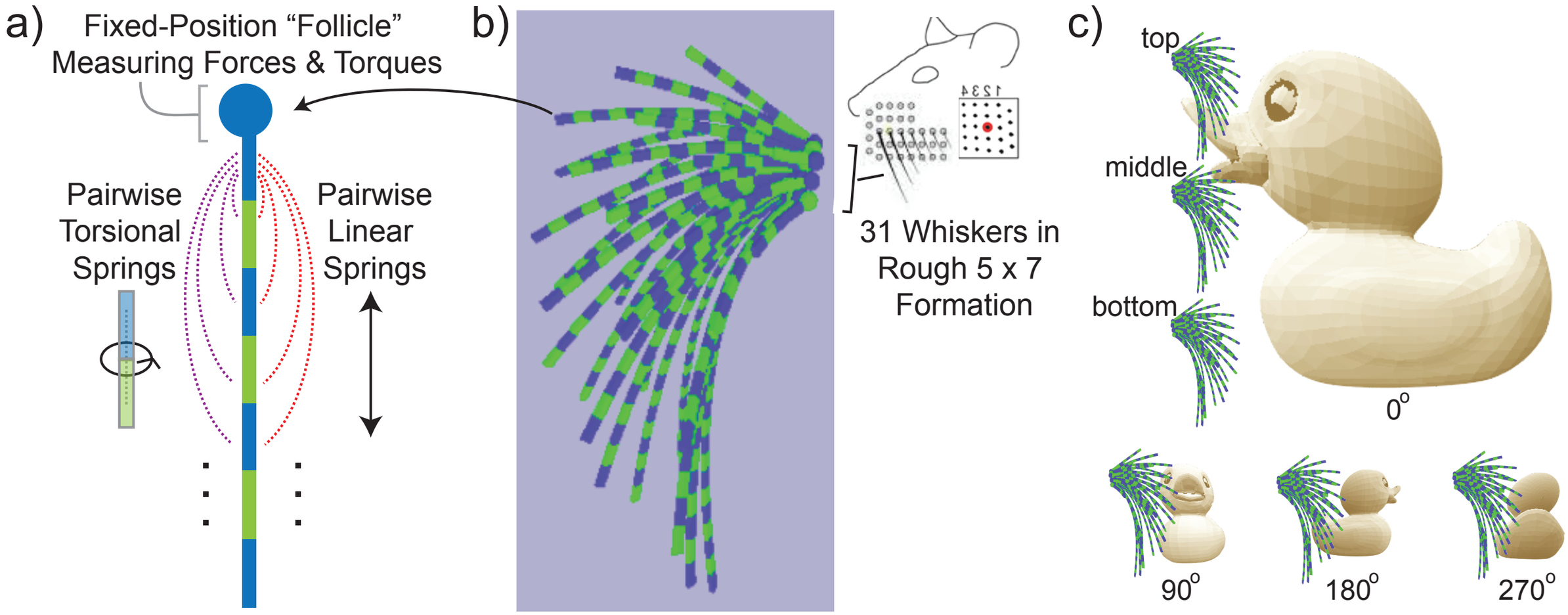


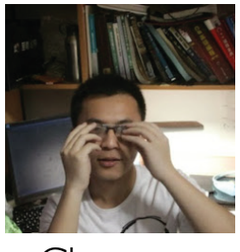
Using published data from Mitra Hartmann's group

Chengxu Zhuang

Mitra Hartmann & Lab

Rodent Somatosensory Cortex

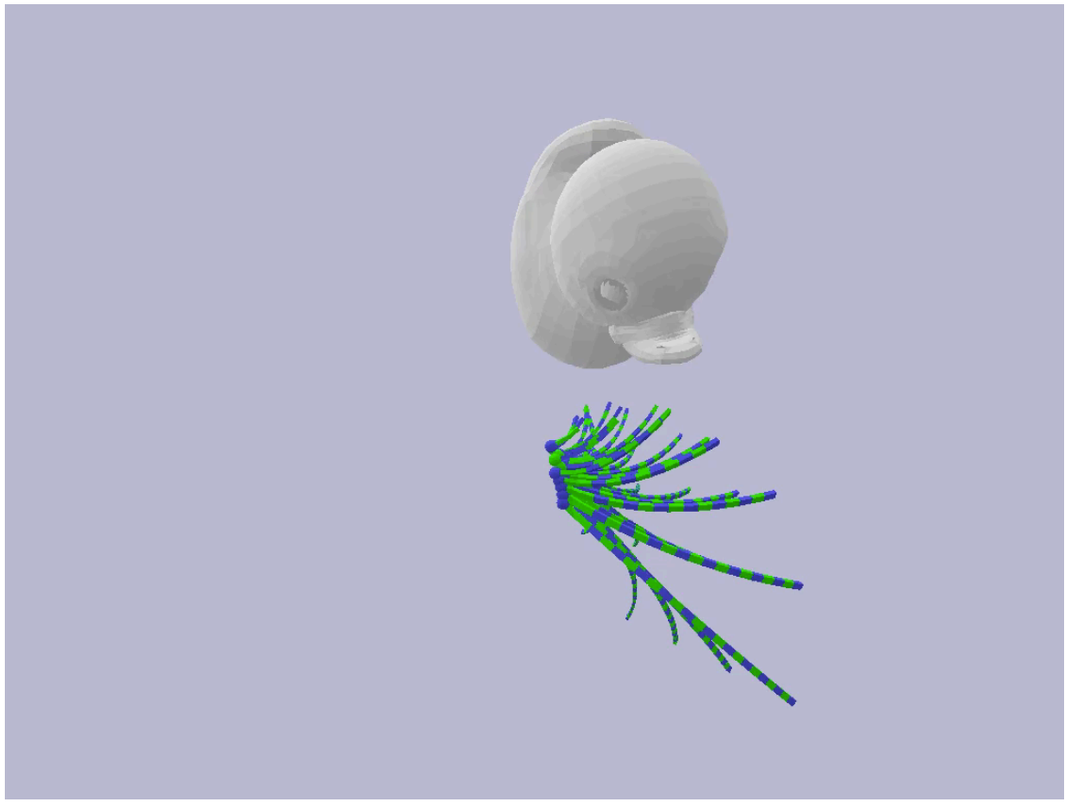


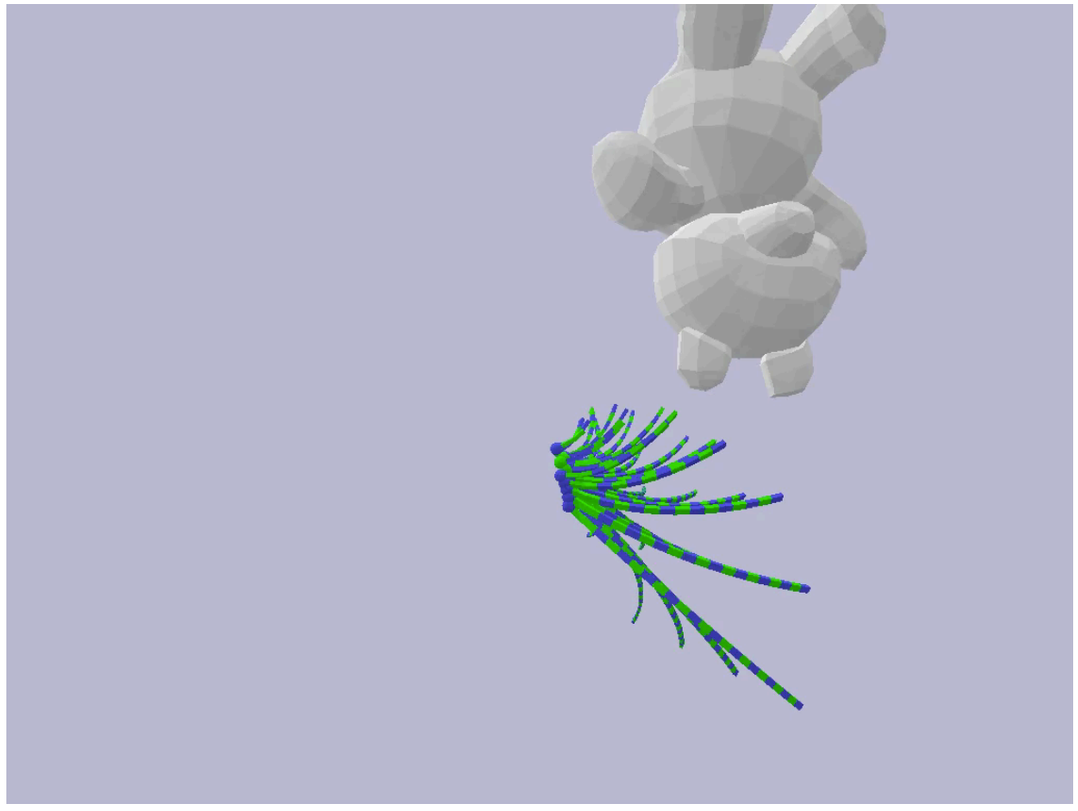


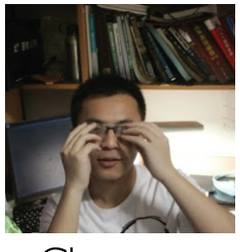
Chengxu Zhuang

Mitra Hartmann & Lab

Rodent Somatosensory Cortex





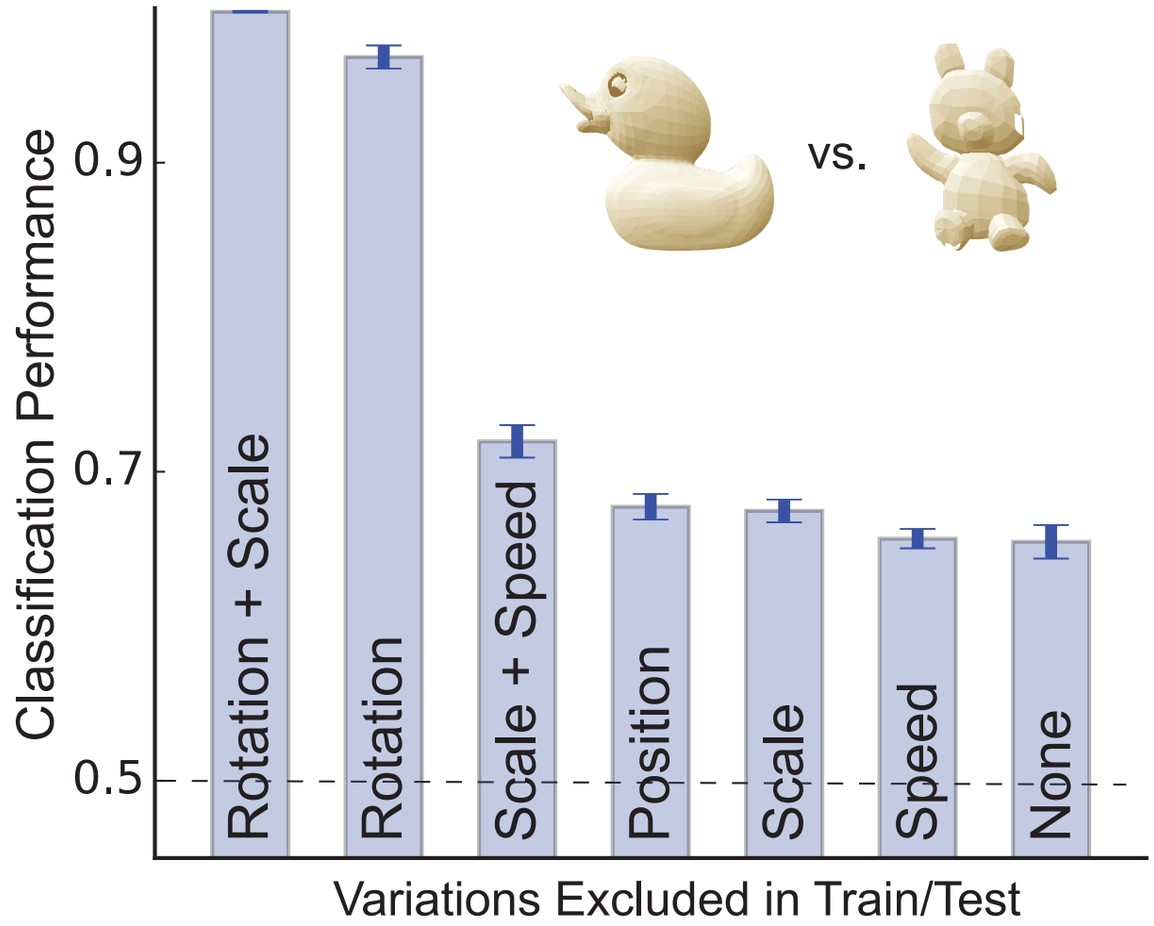


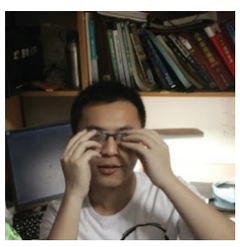
Chengxu Zhuang

Mitra Hartmann & Lab

Rodent Somatosensory Cortex

Exactly the "right" case for a deep cortical cascade ….





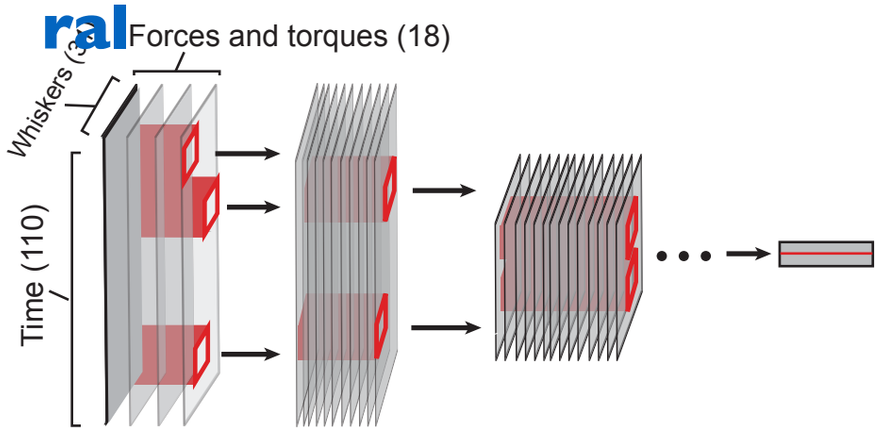
Chengxu Zhuang

Mitra Hartmann & Lab

Rodent Somatosensory Cortex

Four distinct architecture families with different hypotheses about how temporal and spatial information is integrated.

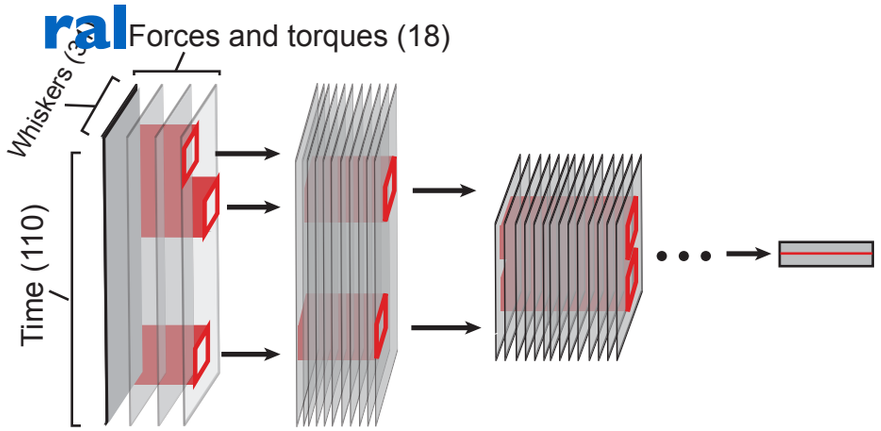
Spatiotempo
ral



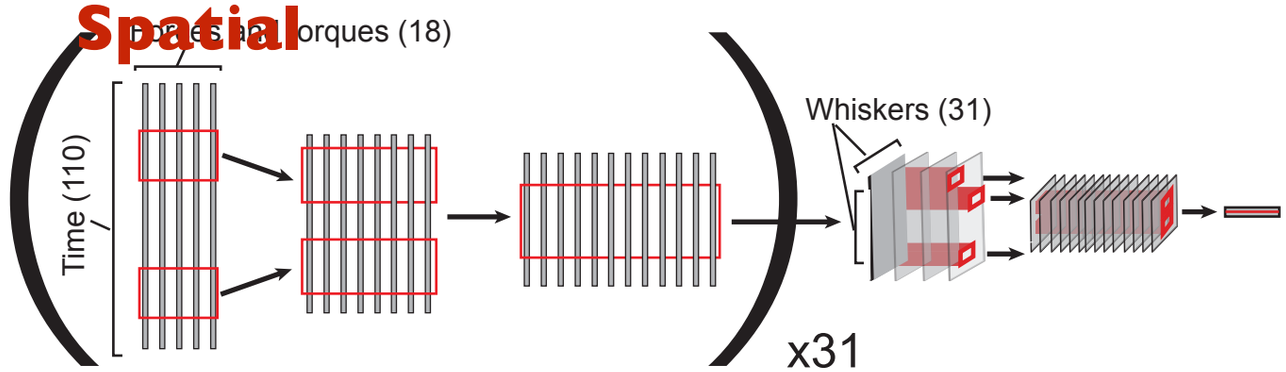
Rodent Somatosensory Cortex

Four distinct architecture families with different hypotheses about how temporal and spatial information is integrated.

Spatiotemporal



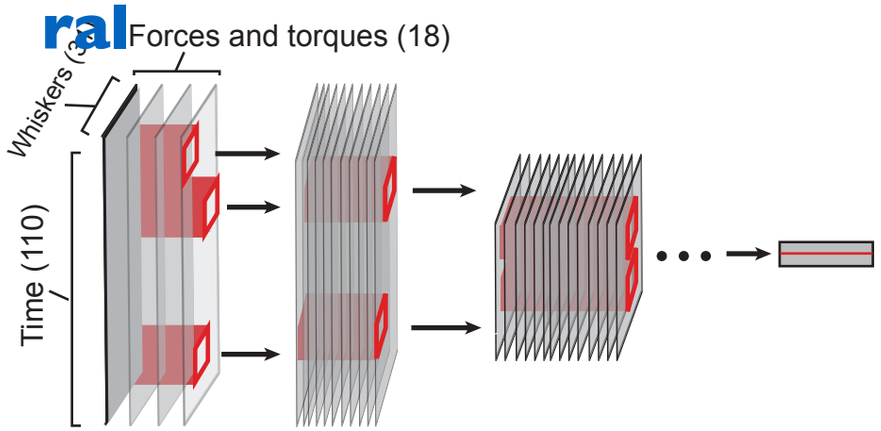
Temporal-Spatial



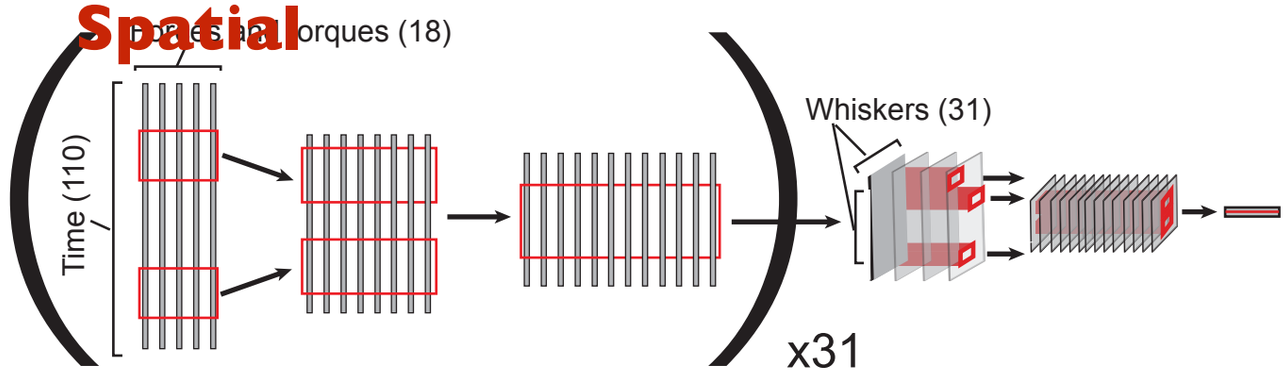
Rodent Somatosensory Cortex

Four distinct architecture families with different hypotheses about how temporal and spatial information is integrated.

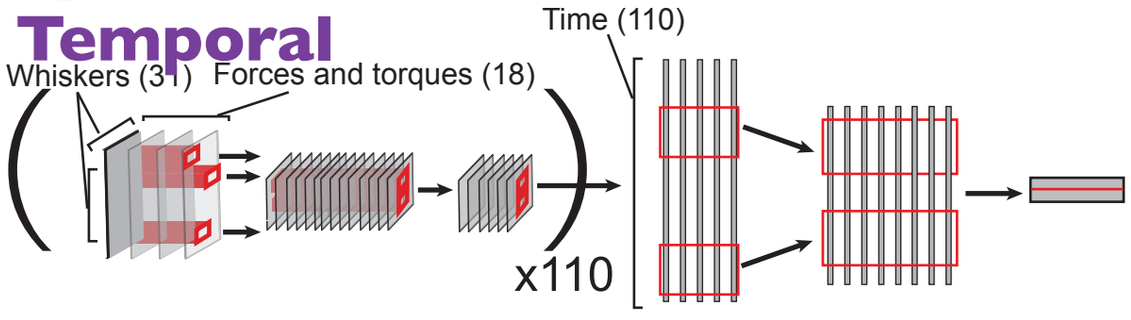
Spatiotemporal



Temporal-Spatial



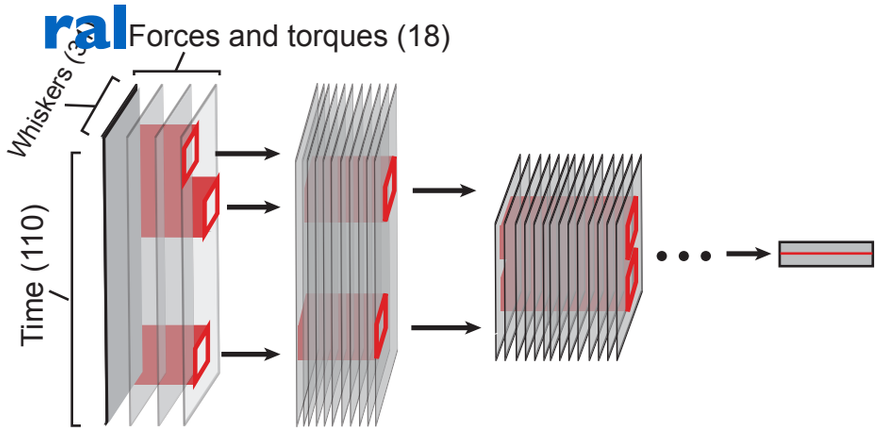
Spatial-Temporal



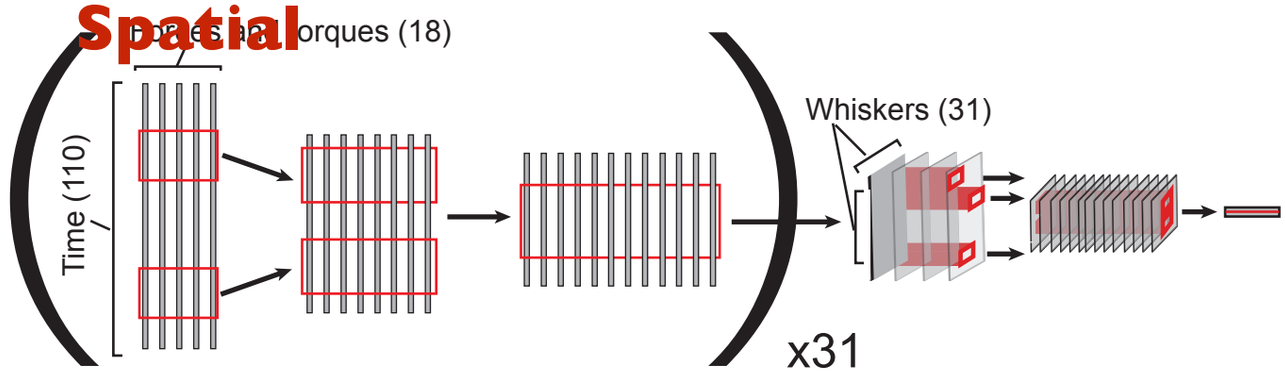
Rodent Somatosensory Cortex

Four distinct architecture families with different hypotheses about how temporal and spatial information is integrated.

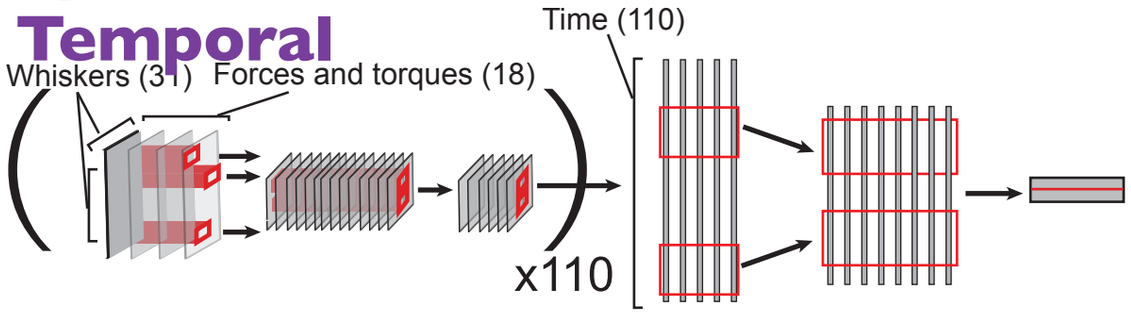
Spatiotemporal



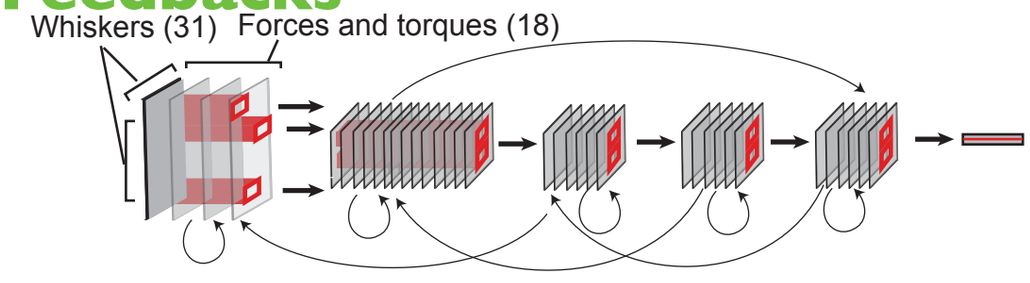
Temporal-Spatial



Spatial-Temporal

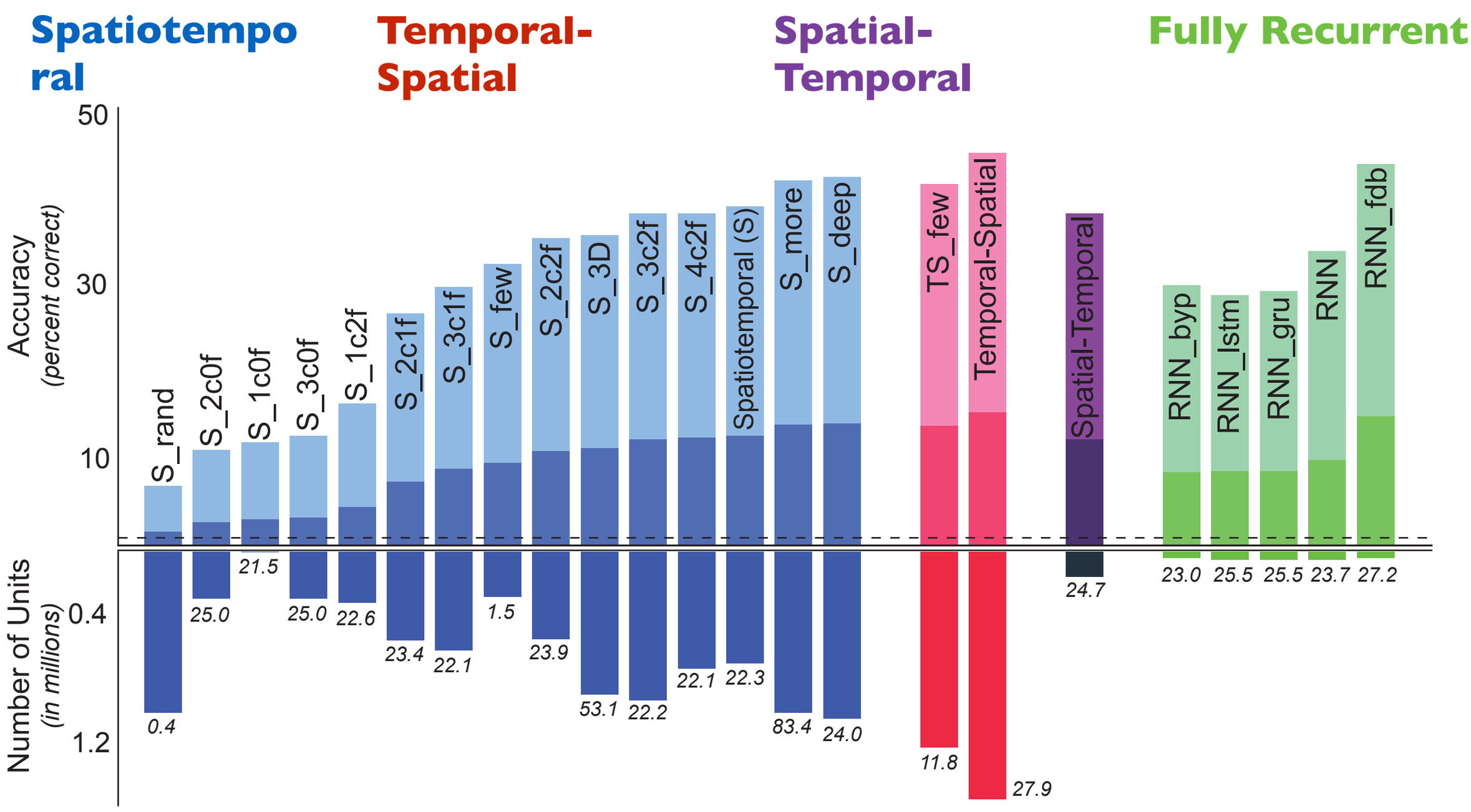


Fully Recurrent w/ Deep Feedbacks



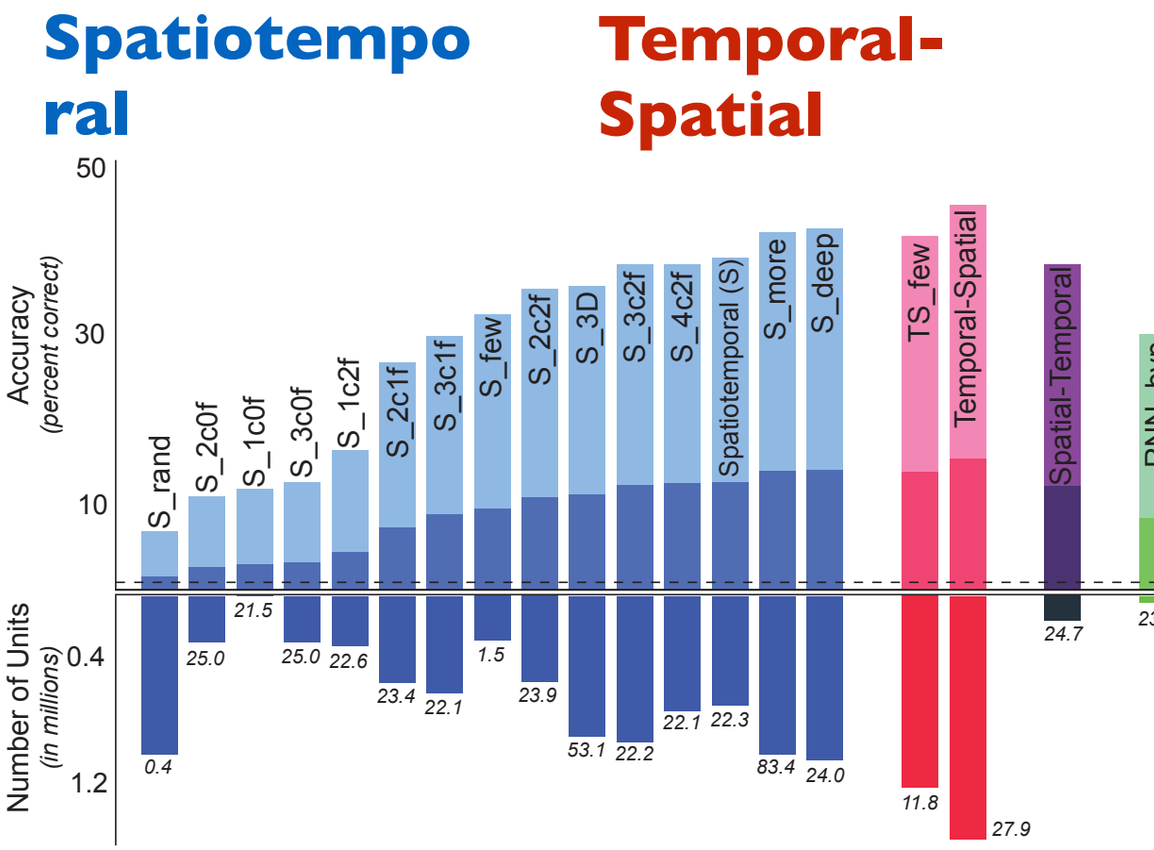
Optimization goal: object shape recognition.

Rodent Somatosensory Cortex



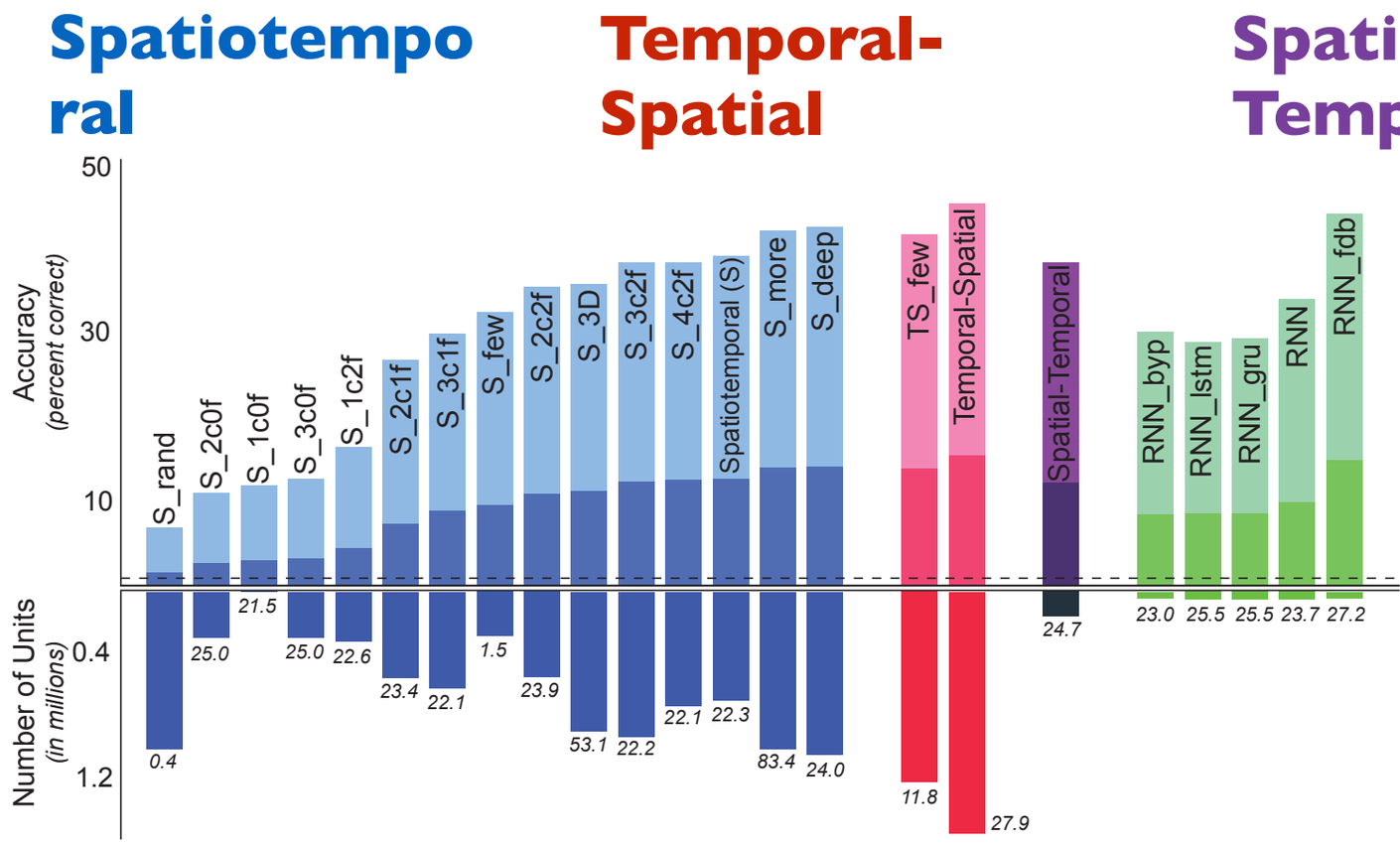
Optimization goal: object shape recognition.

Rodent Somatosensory Cortex



Optimization goal: object shape recognition.

Rodent Somatosensory Cortex



- Most specific networks don't work, but within each family, some much better than others

- Filter training and depth very important factors, but number of parameters less so

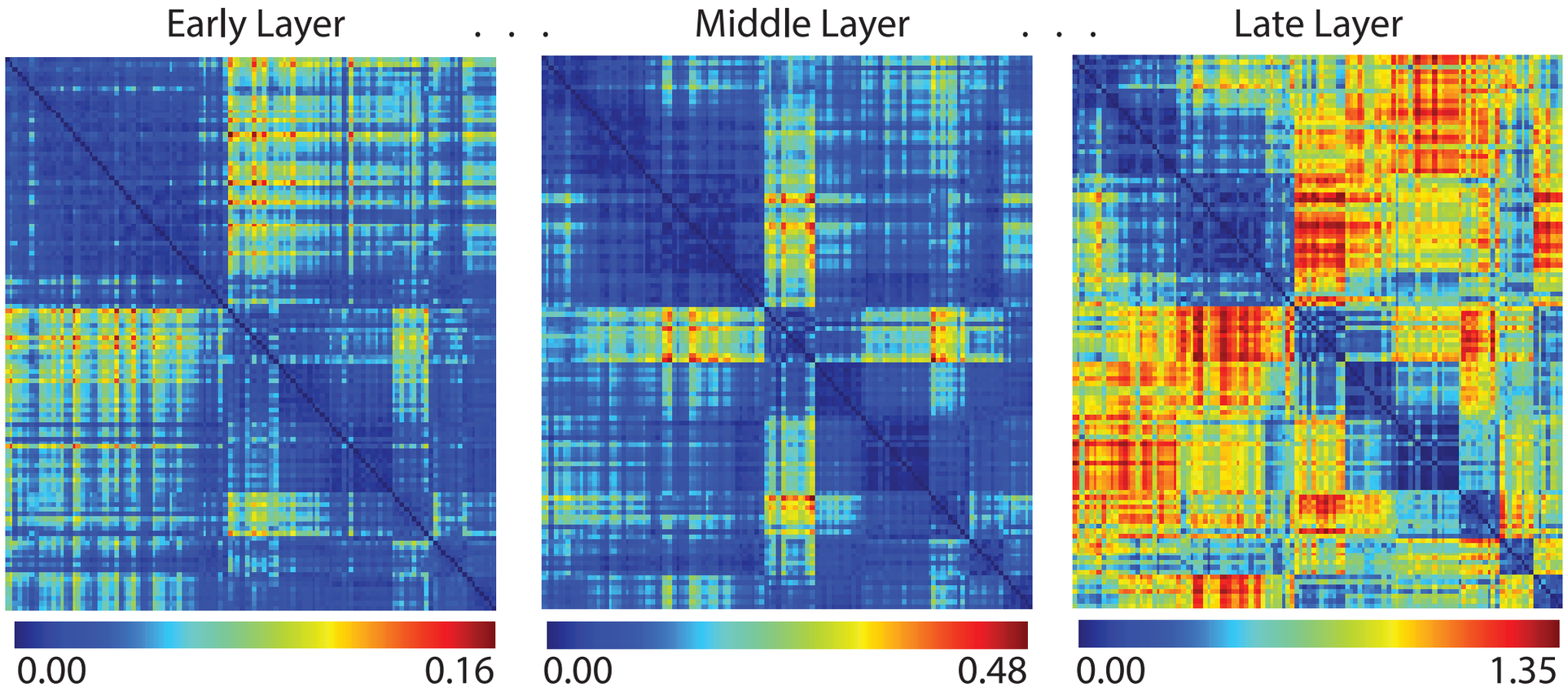
- Recurrent networks with long-range feedbacks achieve highest performance with comparatively few parameters and small (neurally reasonable) numbers of units

Rodent Somatosensory Cortex

Representational Dissimilarity Matrices (RDMs) capture signatures of different stages of the computation

v_i = vector of responses to i-th stimuli

$$RDM[i, j] = corr(v\_i, v\_j)$$

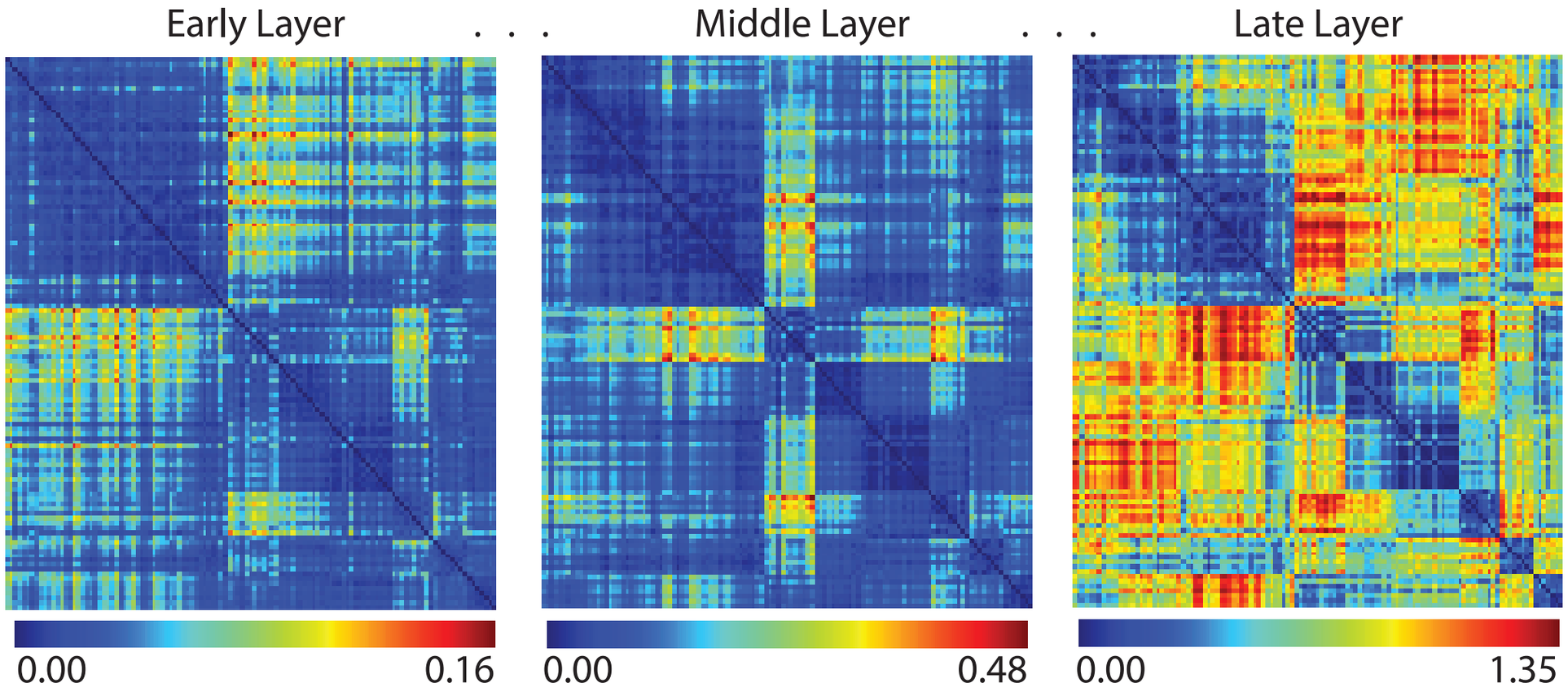


Rodent Somatosensory Cortex

Representational Dissimilarity Matrices (RDMs) capture signatures of different stages of the computation

v_i = vector of responses to i-th stimuli

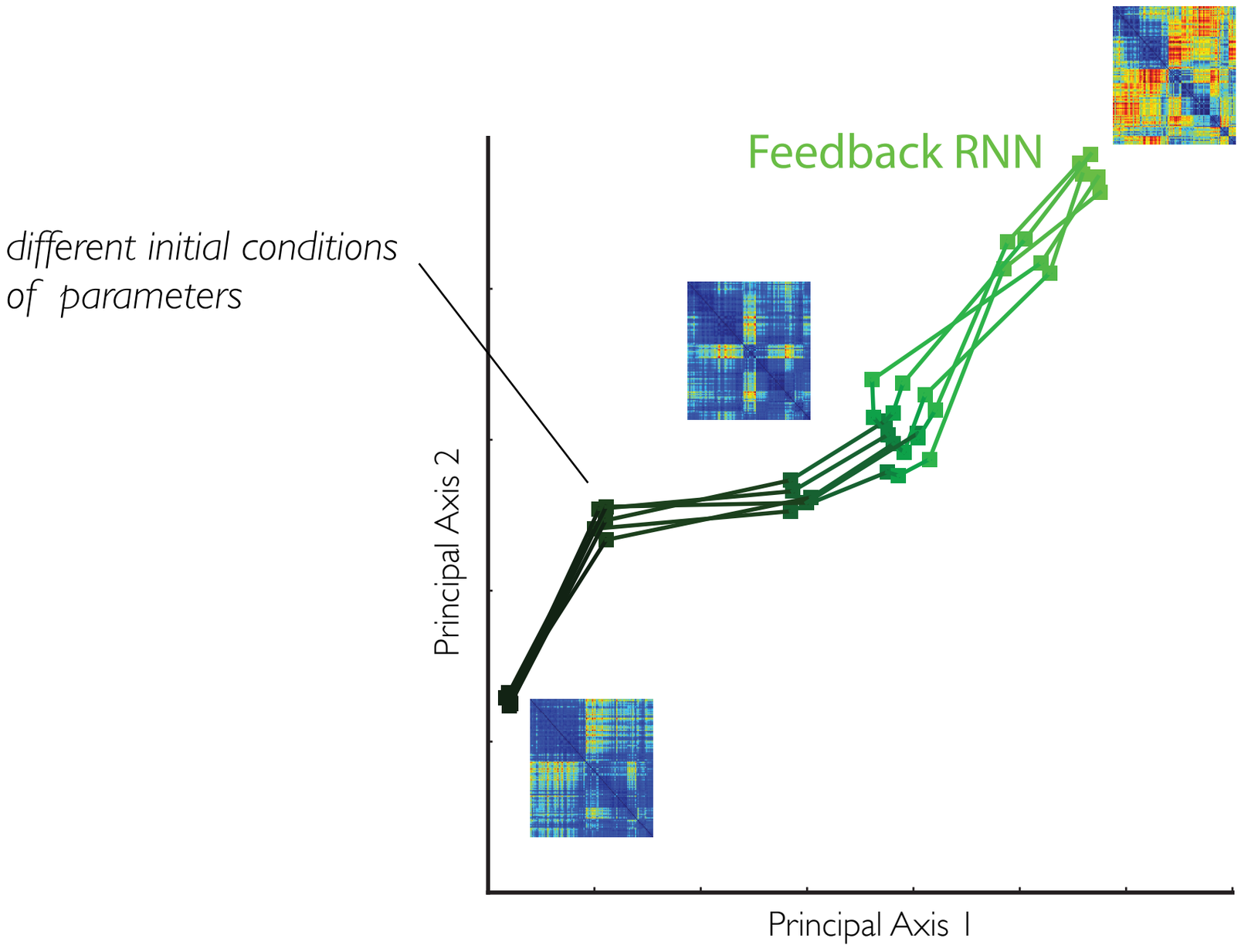
$$RDM[i, j] = corr(v\_i, v\_j)$$



RDMs have been successfully used to compare models to real neural data in visual cortex (Kriegeskorte)

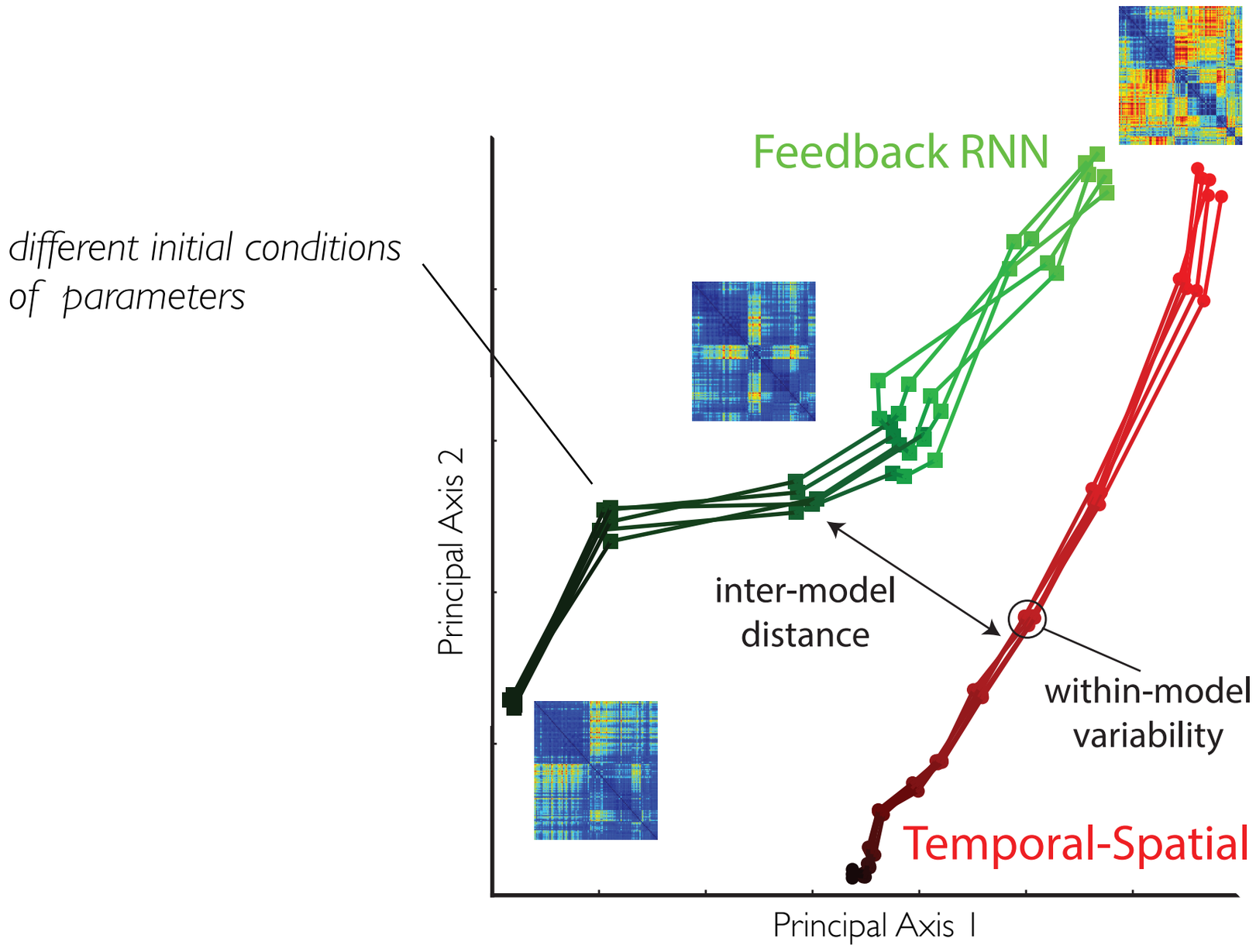
Rodent Somatosensory Cortex

Dimension reduction on RDMs can be used to "visualize" model differences.



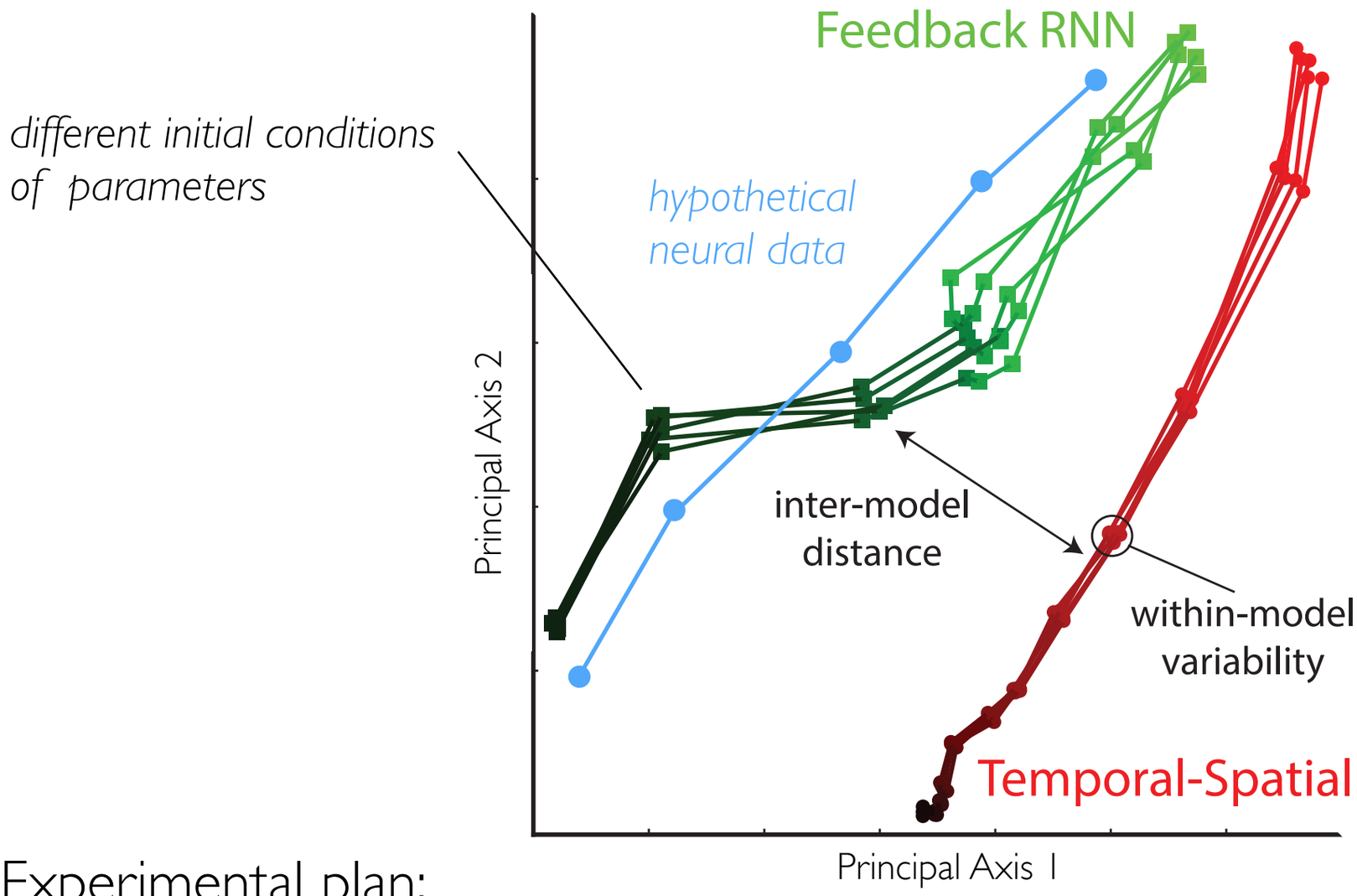
Rodent Somatosensory Cortex

Dimension reduction on RDMs can be used to "visualize" model differences.



Rodent Somatosensory Cortex

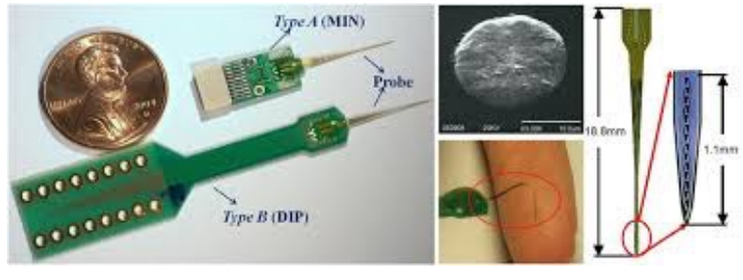
Dimension reduction on RDMs can be used to "visualize" model differences.



Mitra Hartmann & Lab

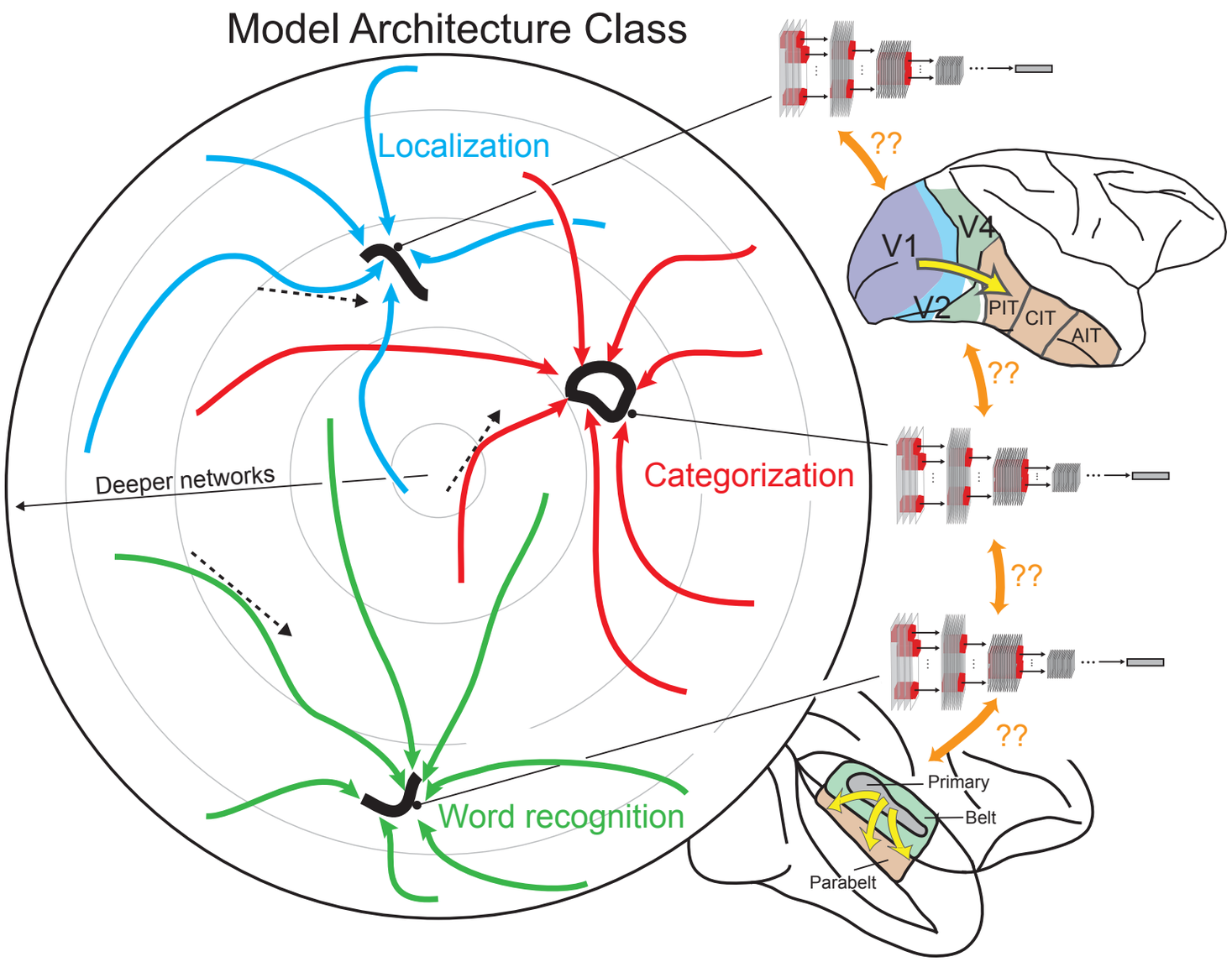
Experimental plan:

(i) collect neural responses in an intermediate area S1
(ii) compute RDMs,
(iii) compare to model families



Where it might work …

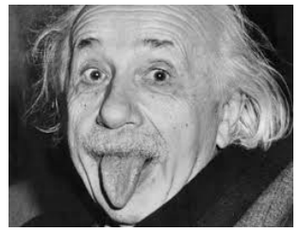
If *successful:*



somatosensation

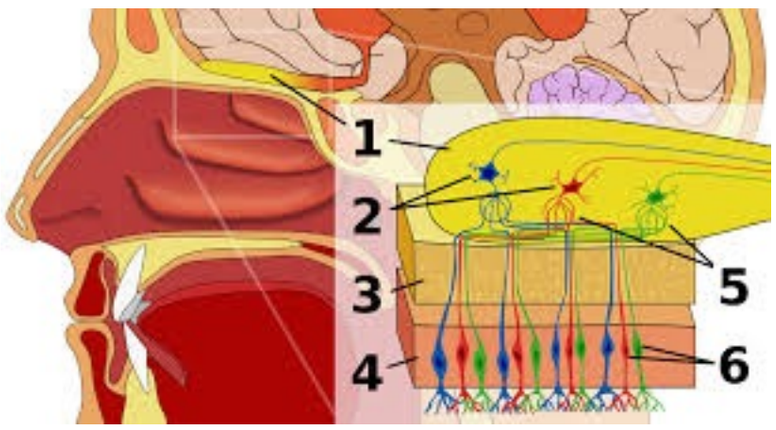
vision

audition

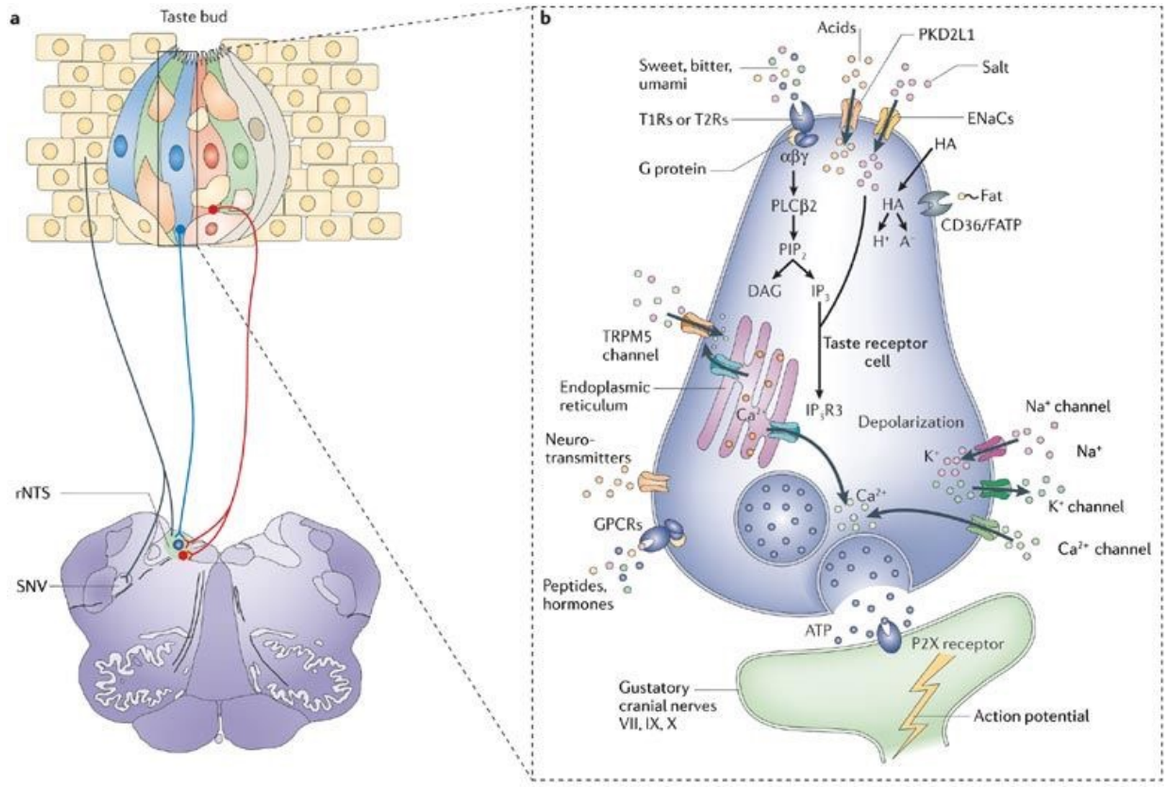


Where it might NOT work …

Unlikely *to yield to this form of analysis:*



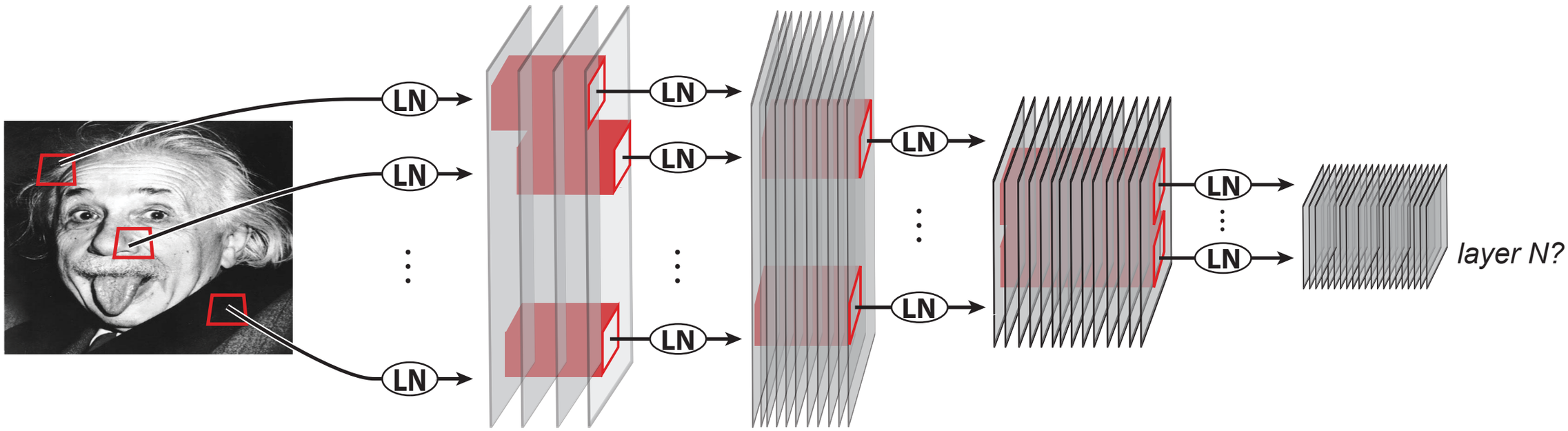
Olfaction (smell)



Gustation (taste)

Where it might NOT work …

So, recall, in vision:



at sensor:

- wide spatial layout
- few channels (RGB = 3 channels)

audition:
spectrotemporal layout, 1 channel

somatosensation:
spectrotemporal layout, ~6 channels

<— *many subcortical and cortical processing layers* —>

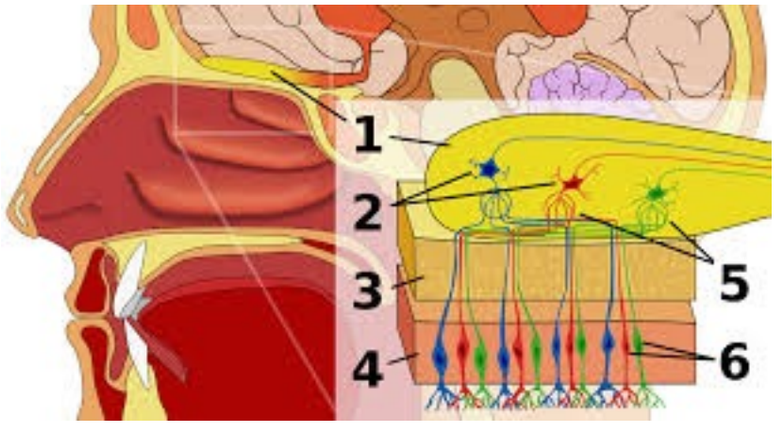
at higher sensory cortex:

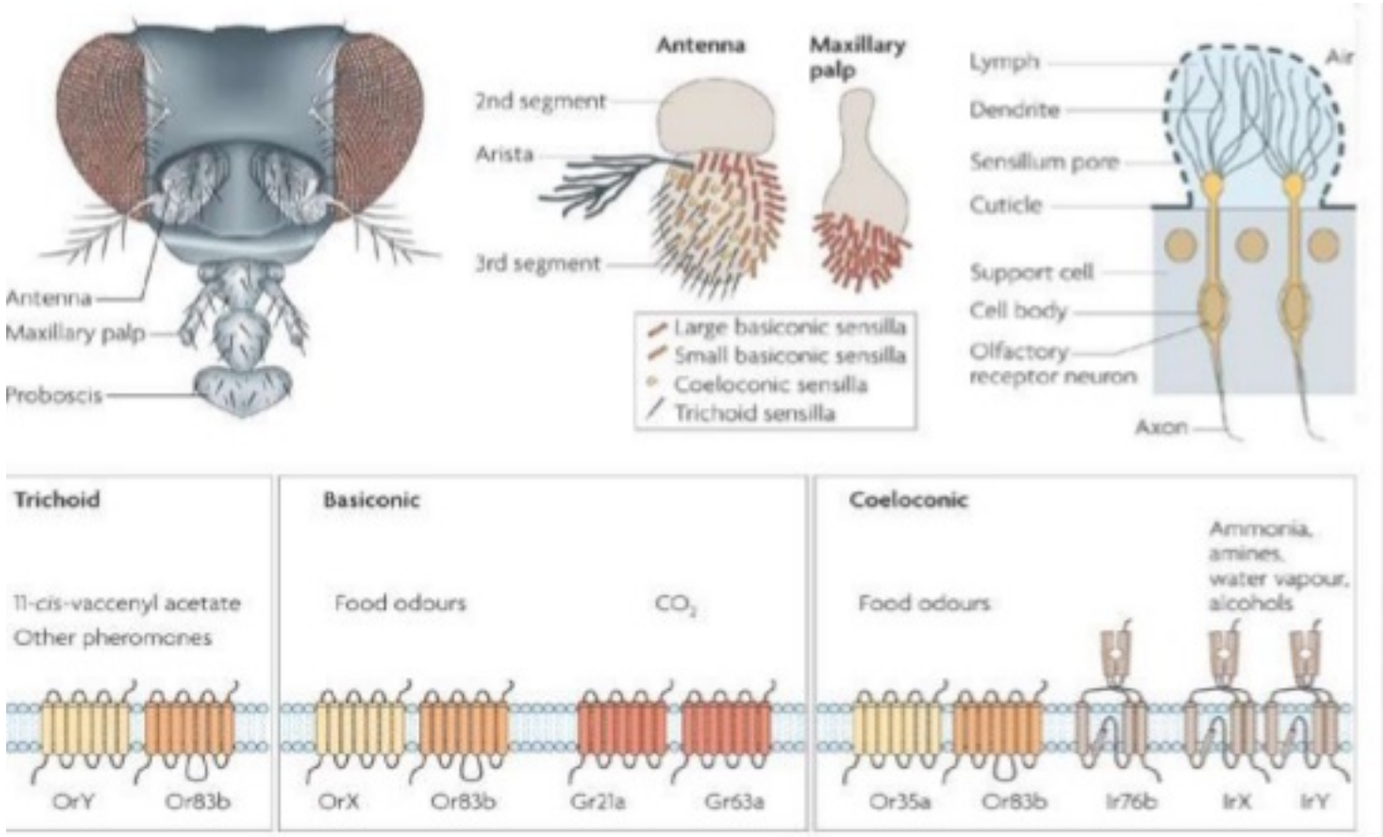
- less spatiotopy
- many channels (~1000)

. . . so there's something to explain.

Where it might NOT work …

Olfaction (smell)

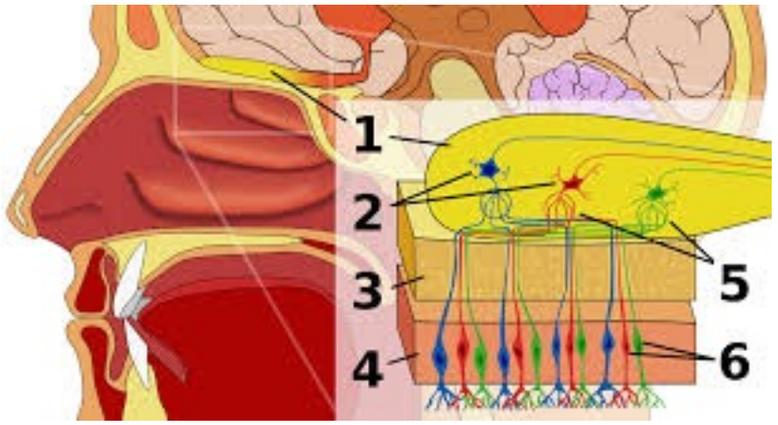


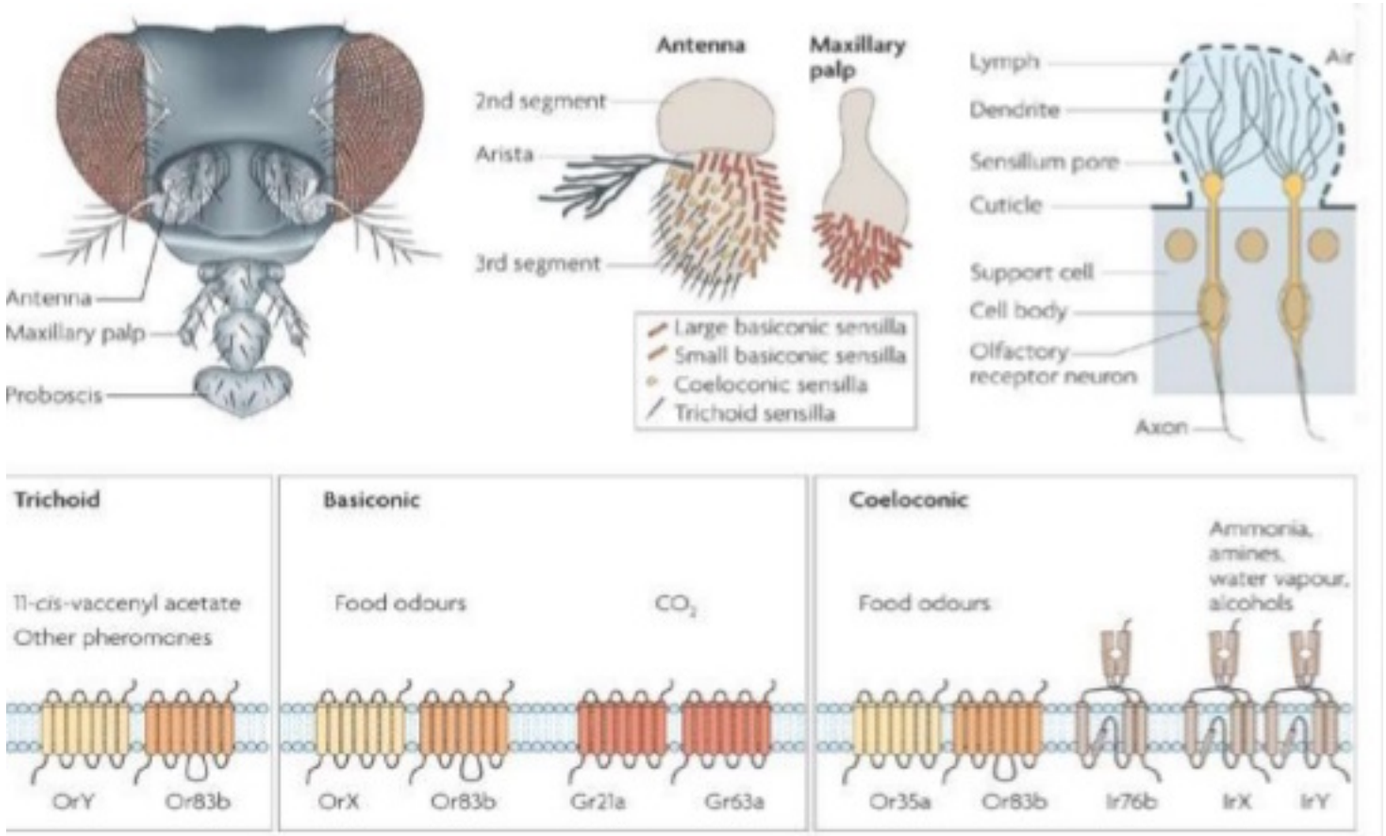


Thousands of (genetically encoded) input channels … no obvious spatial structuring … simple behaviors …

Where it might NOT work …

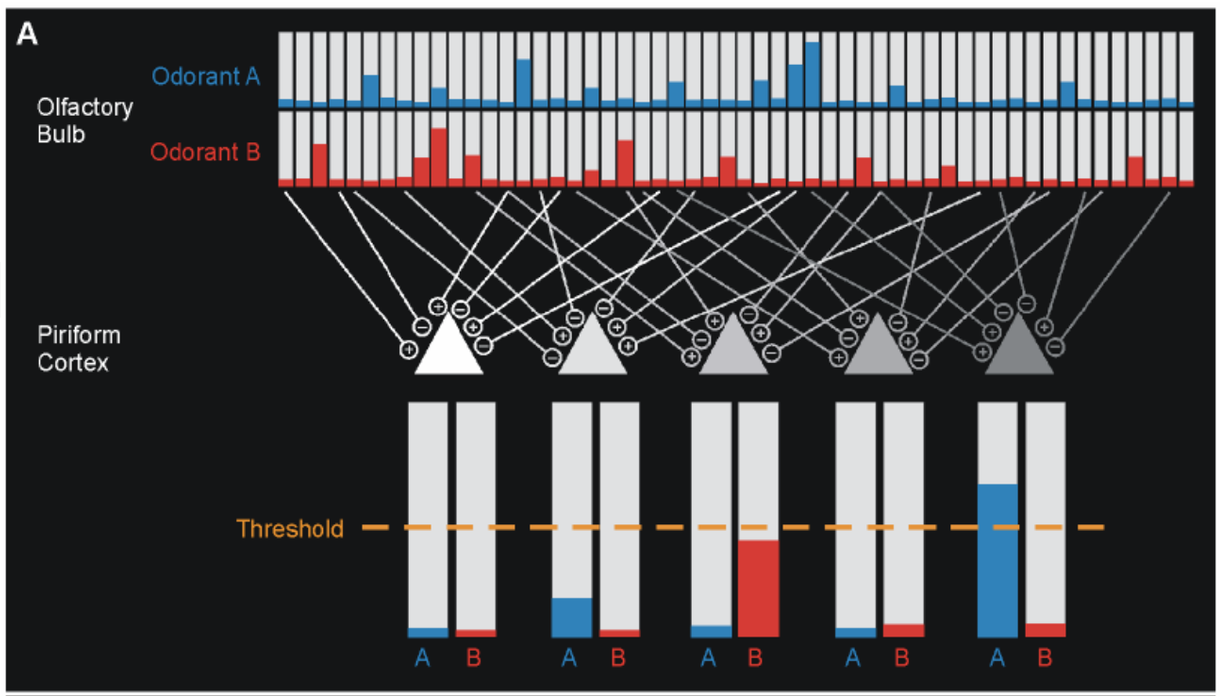
Olfaction (smell)





Thousands of (genetically encoded) input channels … no obvious spatial structuring … simple behaviors …

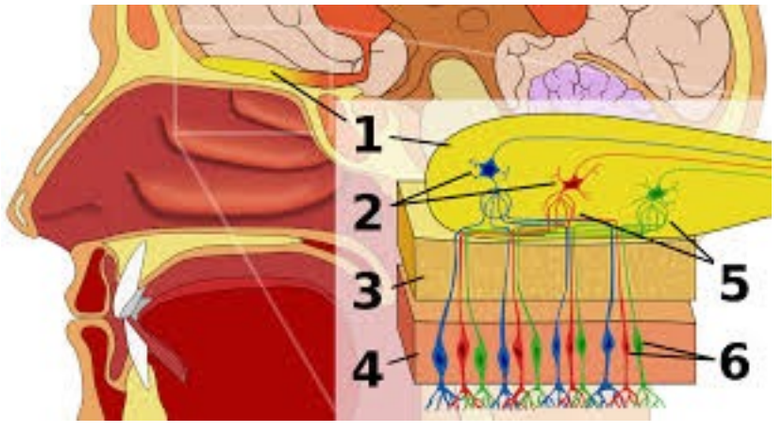
… so no need for deep networks.

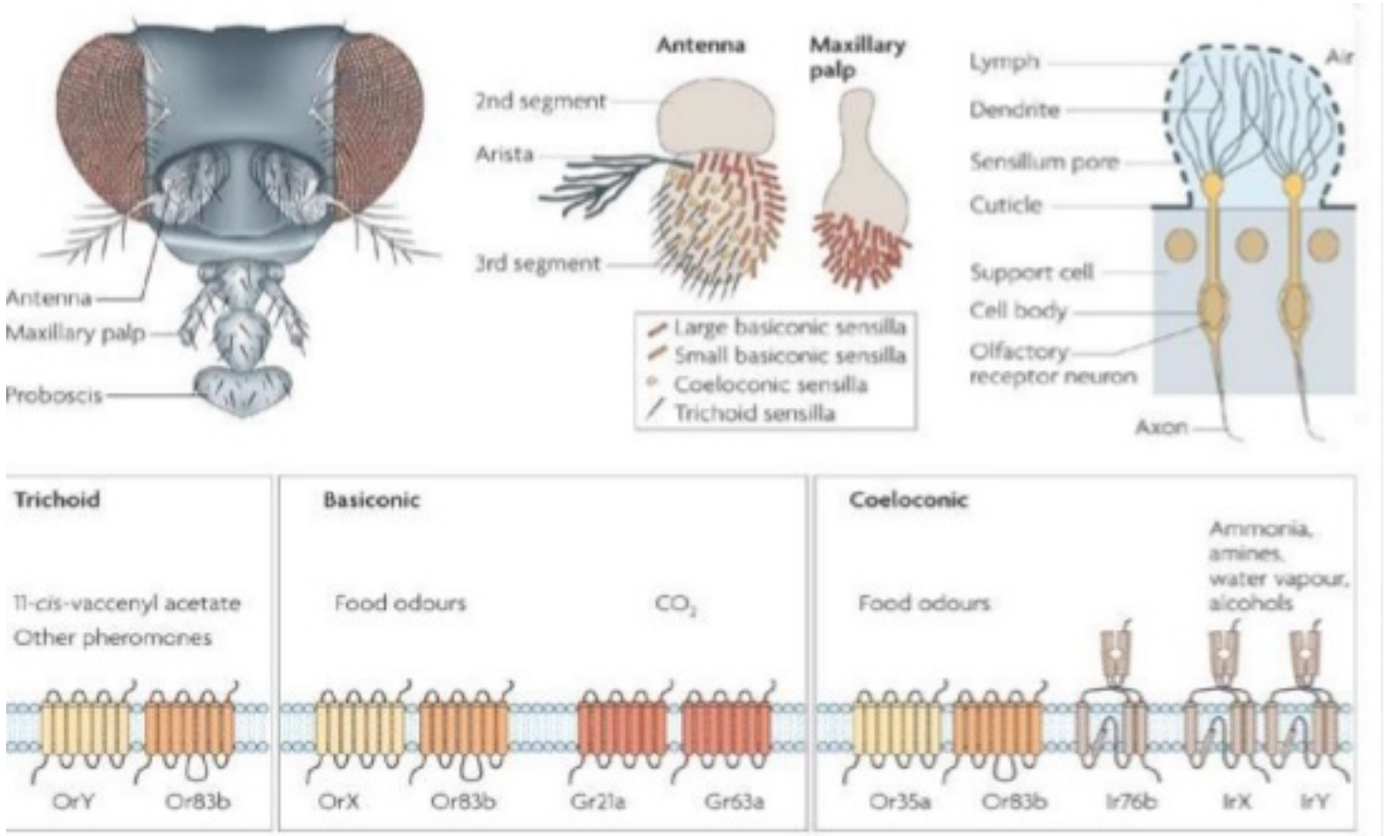


Two layer random association model of piriform cortex. (Stettler & Axel 2009)

Where it might NOT work …

Olfaction (smell)

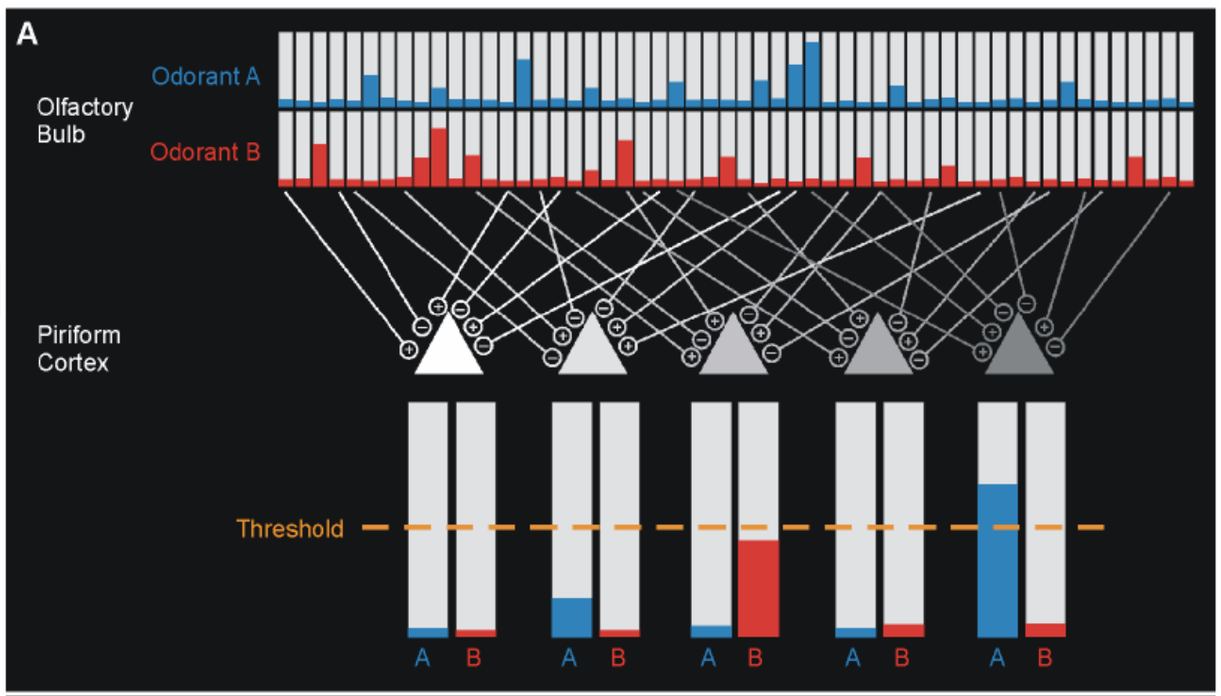




Thousands of (genetically encoded) input channels … no obvious spatial structuring … simple behaviors …

… so no need for deep networks.

Can task-trained model beat this—> ??? maybe not …



Two layer random association model of piriform cortex. (Stettler & Axel 2009)

But see…

› Neuron. 2021 Dec 1;109(23):3879-3892.e5. doi: 10.1016/j.neuron.2021.09.010. Epub 2021 Oct 7.

Evolving the olfactory system with machine learning

Peter Y Wang [1], Yi Sun [2], Richard Axel [3], L F Abbott [1], Guangyu Robert Yang [4]

Affiliations + expand

PMID: 34619093 DOI: 10.1016/j.neuron.2021.09.010

Free article

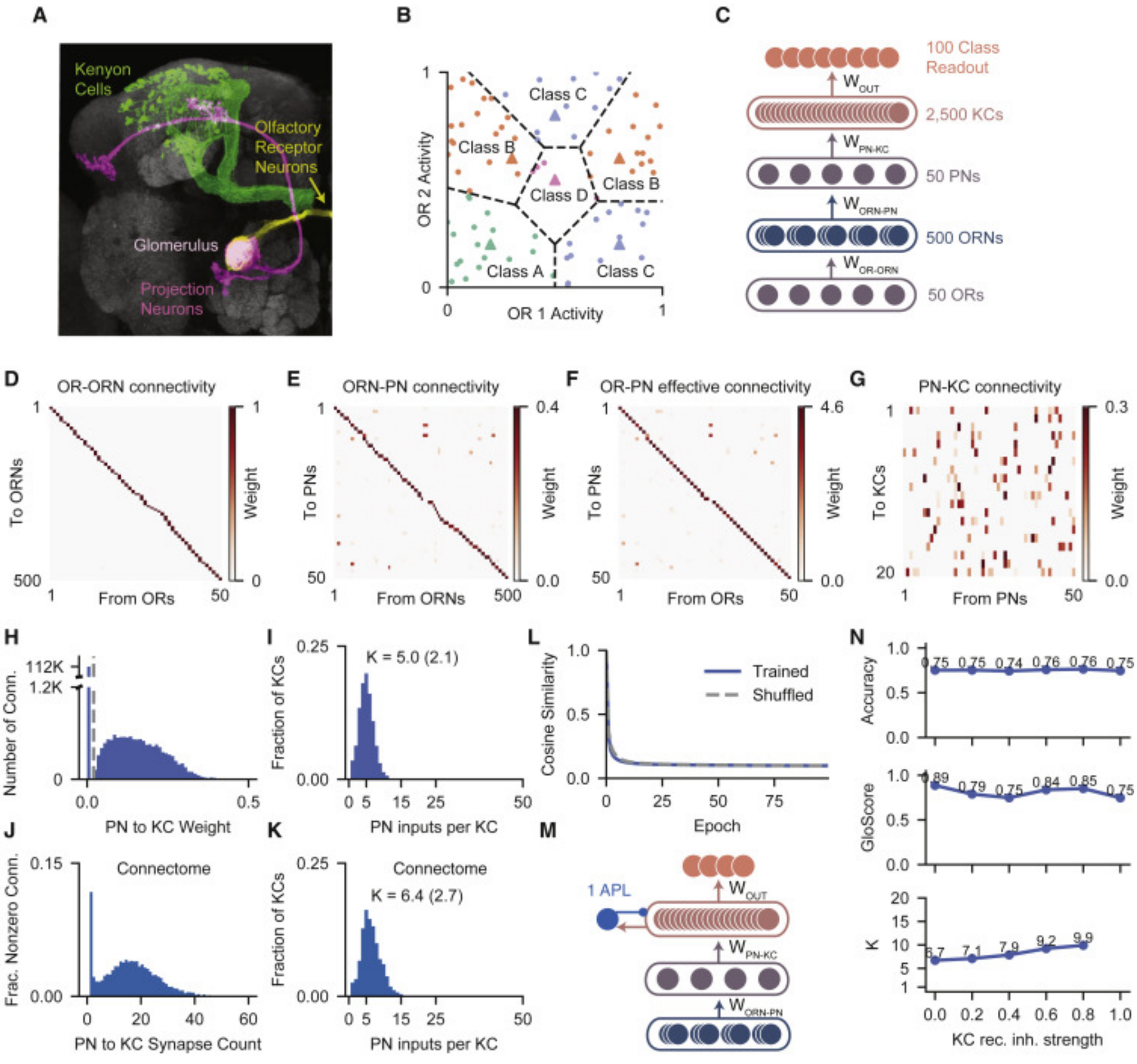


Figure 1. Artificial neural network evolves with the connectivity of the fly olfactory system

(A) The fly olfactory system.

(B) Illustration of the task. Every odor (a million in total; 100 shown) is a point in the space of ORN activity (50 dimensions; two dimensions shown) and is classified based on the closest prototype odor (triangles, 100 in total; four shown). Each class is defined by two prototype odors.

(C) Architecture of the artificial neural network. The expression profile of ORs in every ORN, as well as all other connection weights, is trained.

(D) OR-ORN expression profile after training. ORNs are sorted by the strongest projecting OR.

(E) ORN-PN mapping after training. Each PN type is sorted by the strongest projecting ORN.

(F) Effective connectivity from OR to PN type, produced by multiplying the matrices in (D) and (E).

Figure 3. Recurrent neural networks converge with biological structures

(A) Schematic of a recurrent neural network using recurrent connections (W_{REC}) (left) and the equivalent "unrolled" network diagram (right).

(B and C) Network connectivity between neurons whose activity, when averaged across all odors, exceeds a threshold at different steps. (B) Connectivity from neurons active at step 1 to neurons active at step 2. Connections are sorted. (C) Connectivity from neurons active at step 2 to neurons active at step 3, showing only the first 20 active neurons at step 3.

(D) Number of active neurons at each step of computation. At step 1, only the first 500 units in the recurrent network are activated by odors. Classification performance is assessed after step 3.

(E–I) Similar to (A)–(D), but for networks unrolled for four steps, instead of 3. Classification readout occurs at step 4. Effective step 2–4 connectivity is the matrix product of step 2–3 (G) and step 3–4