Music Wellness Center, Ewha University, South Korea | June 26, 2024

Neural encoding of musical emotion

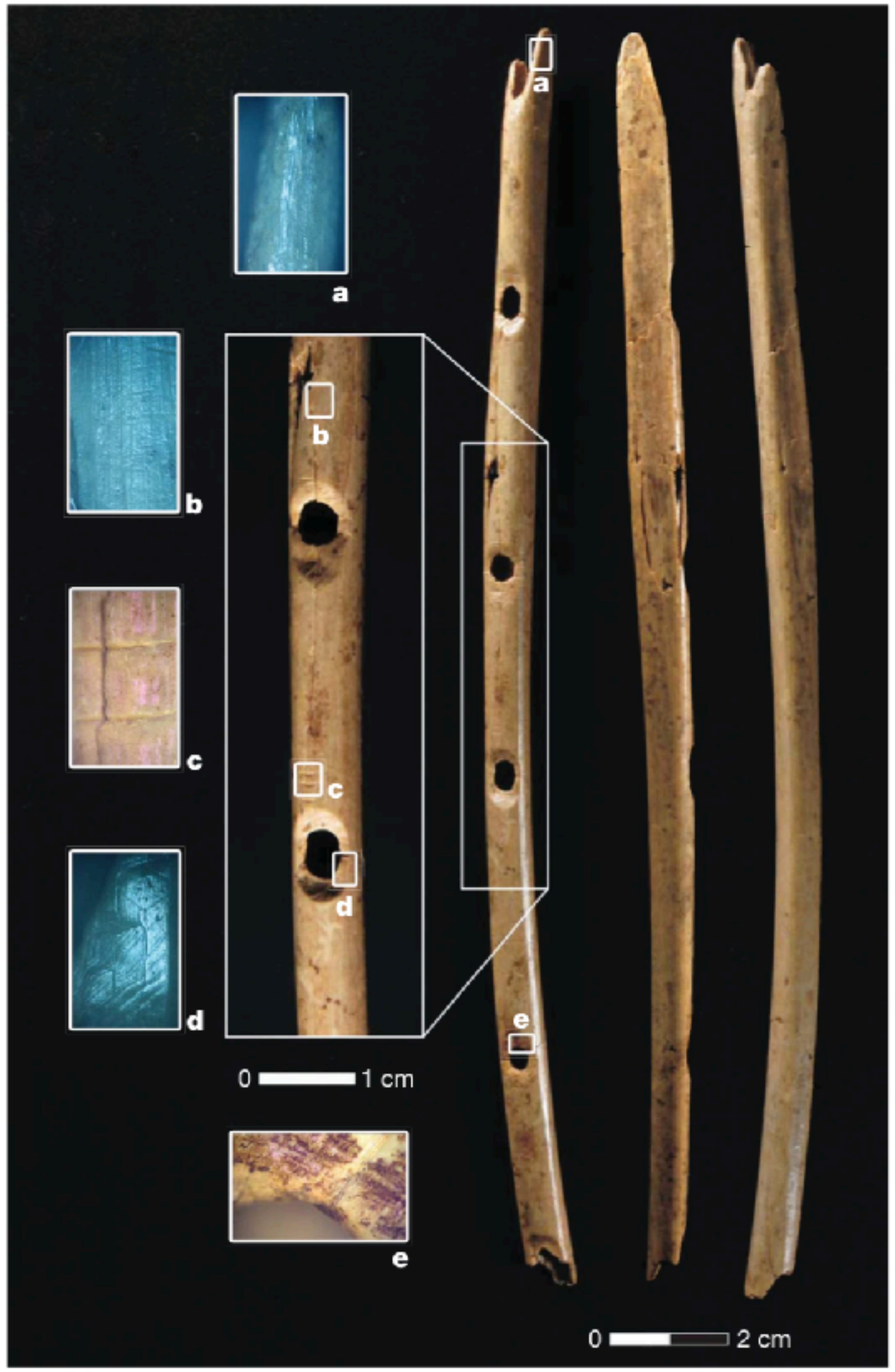
Seung-Goo Kim[1], Tobias Overath[2], & Daniela Sammler[1]

[1]**Research Group Neurocognition of Music and Language**
Max Planck Institute for Empirical Aesthetics, **Frankfurt am Main, Germany**
[2]Department of Psychology & Neuroscience, Duke University, NC, USA

https://voxpopulisphere.com/2015/11/01/audio-what-ancient-greek-music-actually-sounded-like/

Music in culture

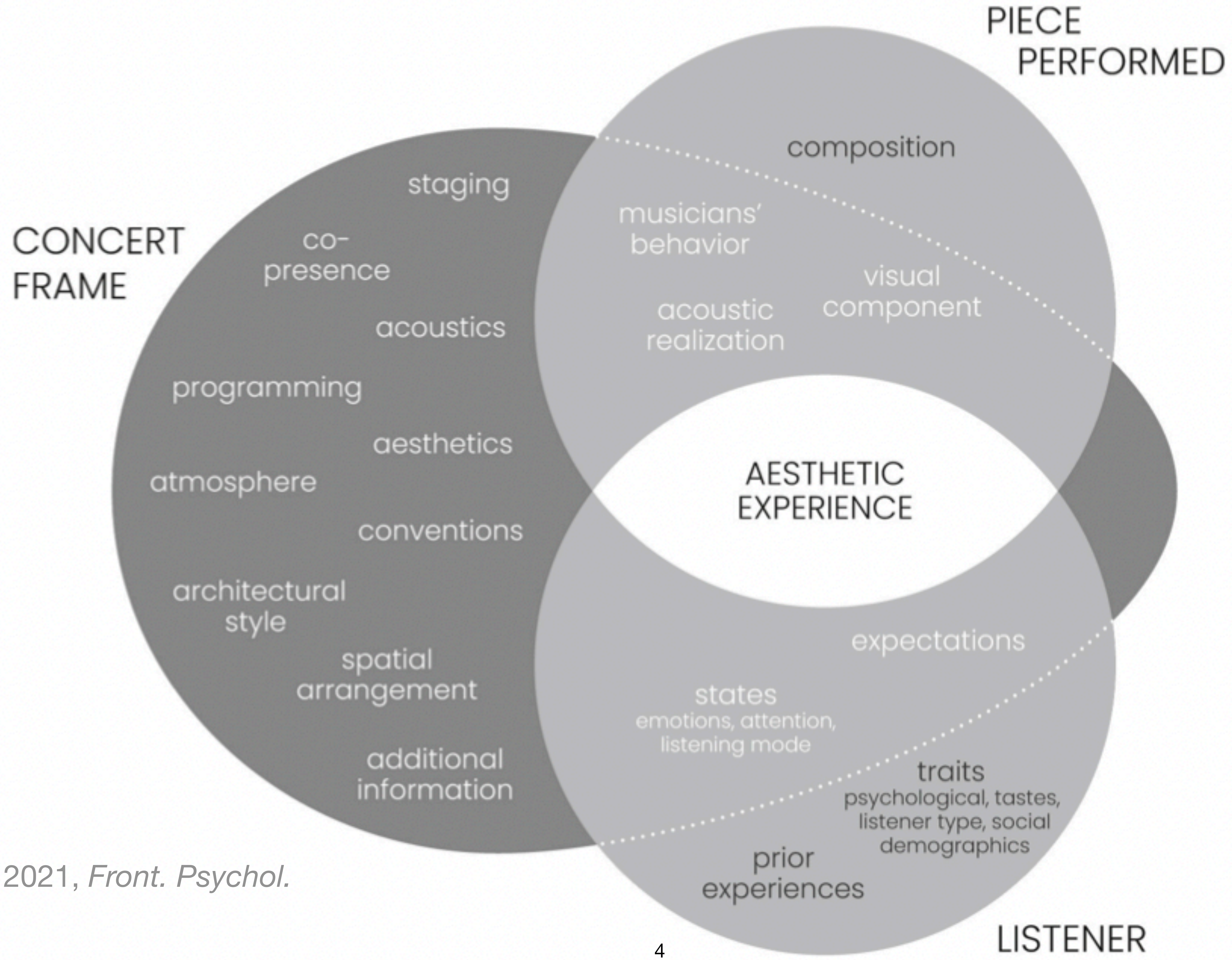
- Current archeological evidence:
  - Ancient origin **of music**: > **35k** yr ago [1]
  - Human & **speech**: **200k** yr ago [2,3]
  - Writing: 5k yr ago [4]
- Ubiquitous in every human culture [5]



35,000 year-old vulture bone flute from Danube, Germany. [1]

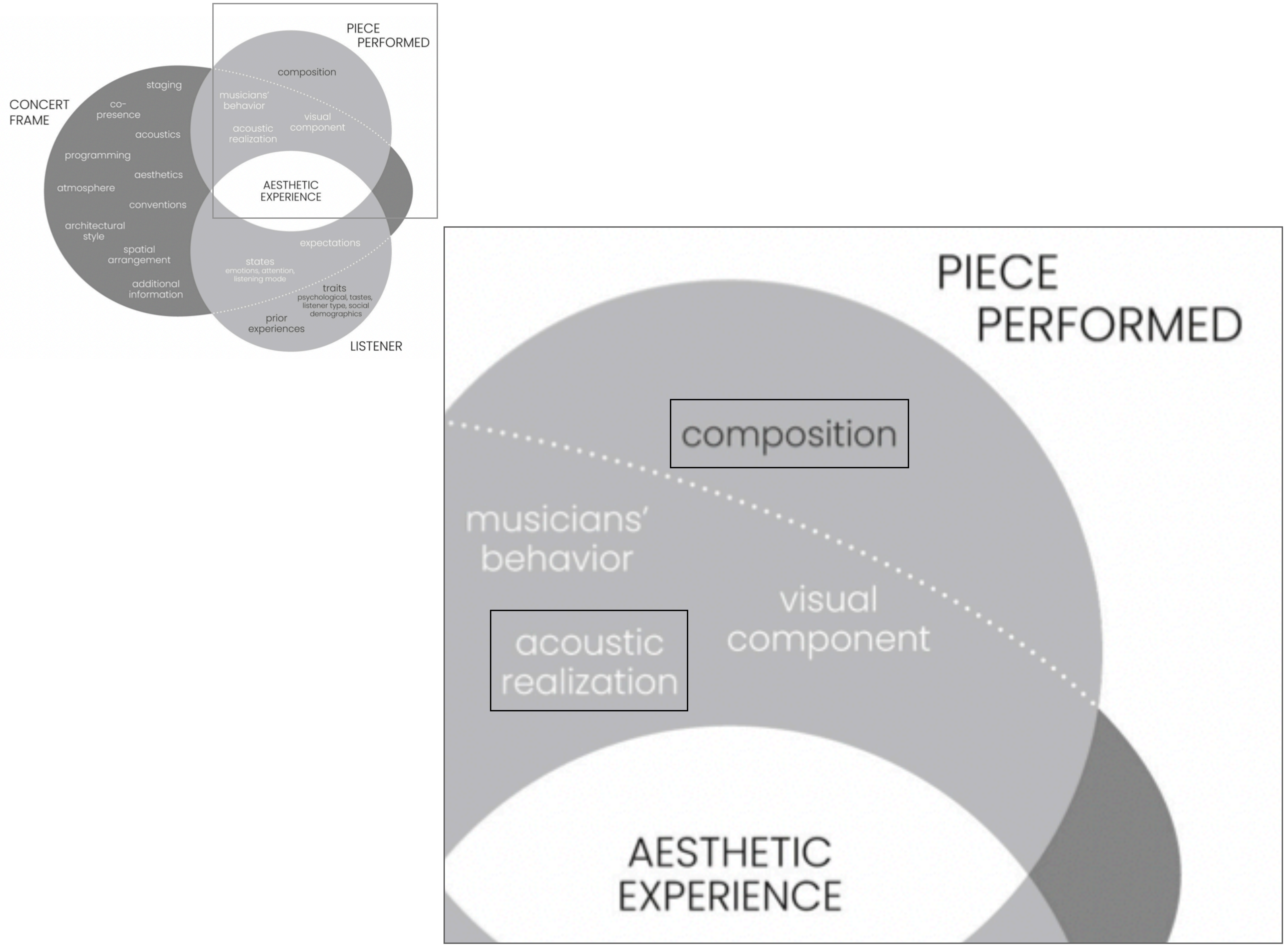
[1] Conard et al. (2009). Nature. [2] Mounier & Lahr. (2019). Nat Comm.
[3] Perreault & Mathew. (2012). PLOS ONE. [4] Senner. (1991). [5] Honing et al. (2015). Phil Trans B.

Frameworks of music-evoked emotion

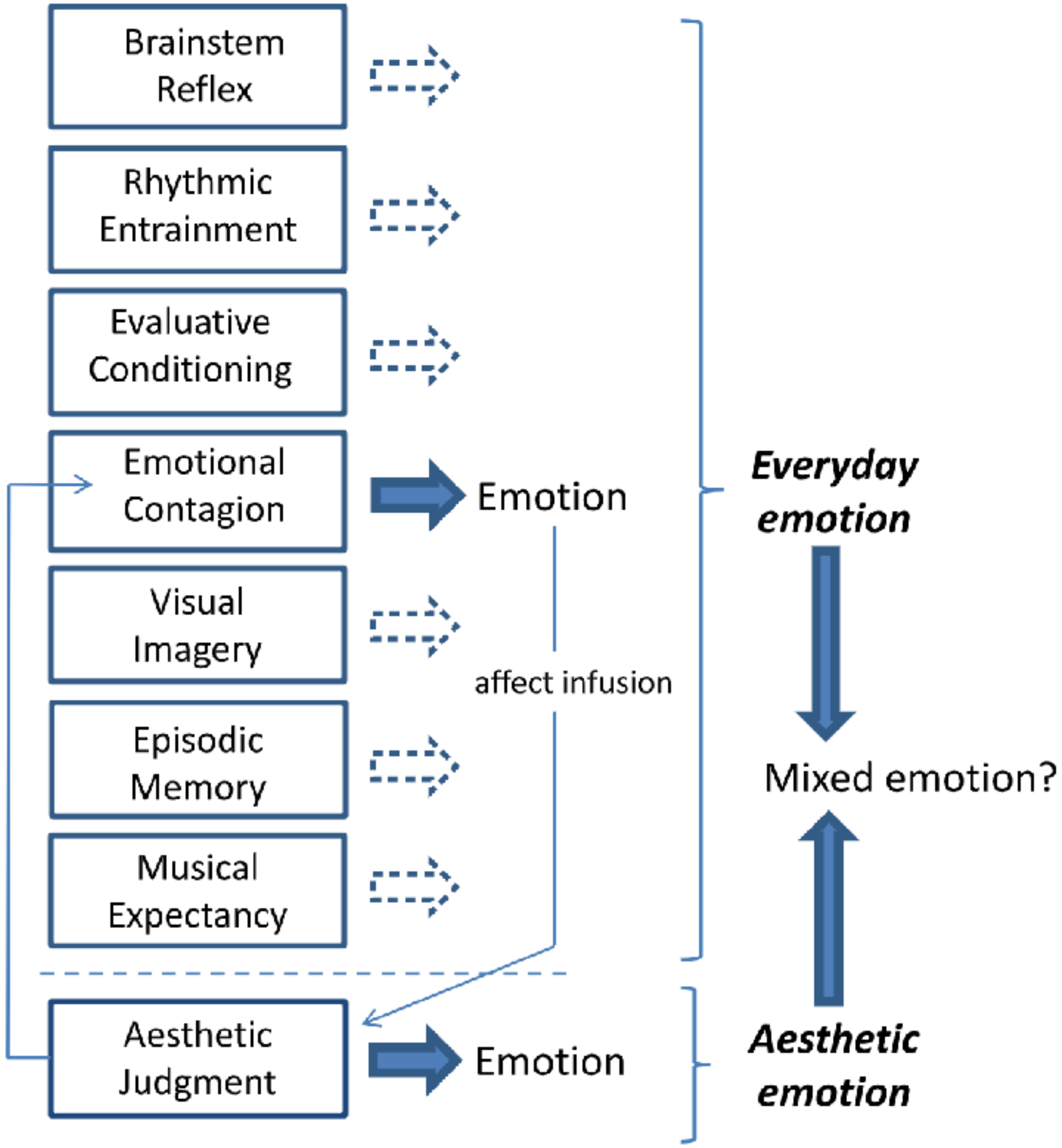


Wald-Fuhrmann et al., 2021, *Front. Psychol.*

Frameworks of music-evoked emotion



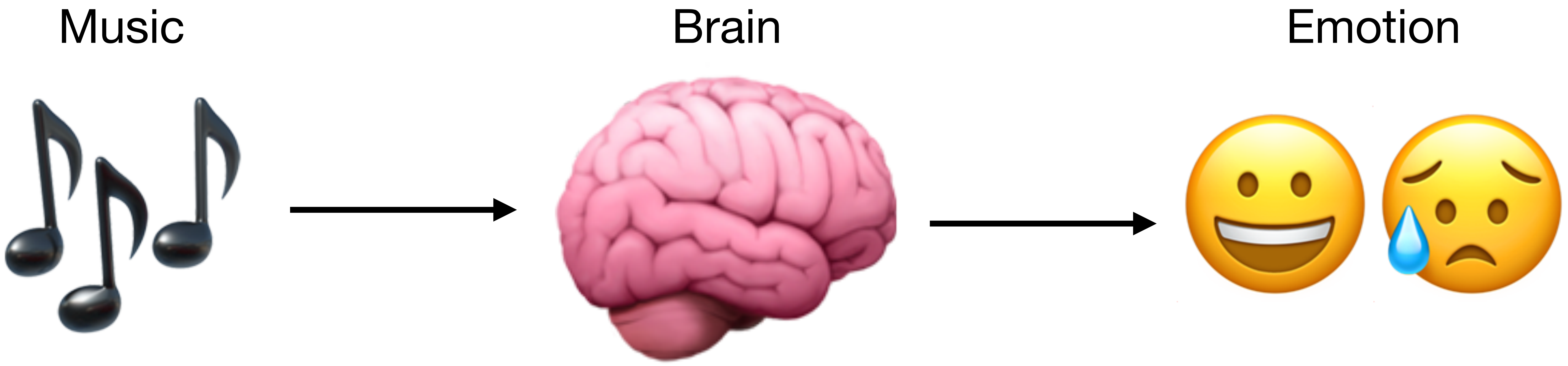
Wald-Fuhrmann et al., 2021, *Front. Psychol.*



"BRECVEMA" (Juslin, 2013, *Physics of Life Rev.*)

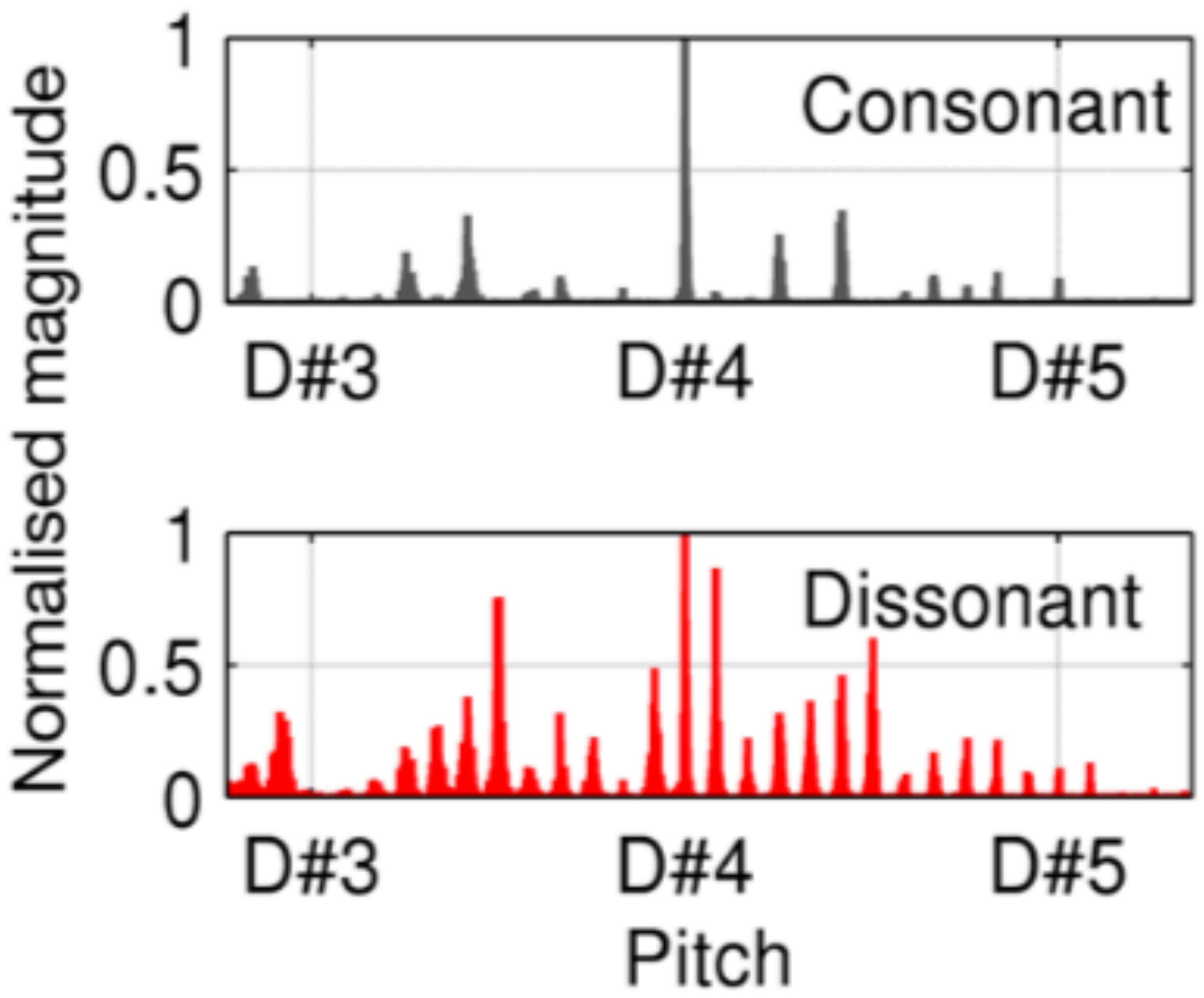
How does music evoke emotions *via* the brain?

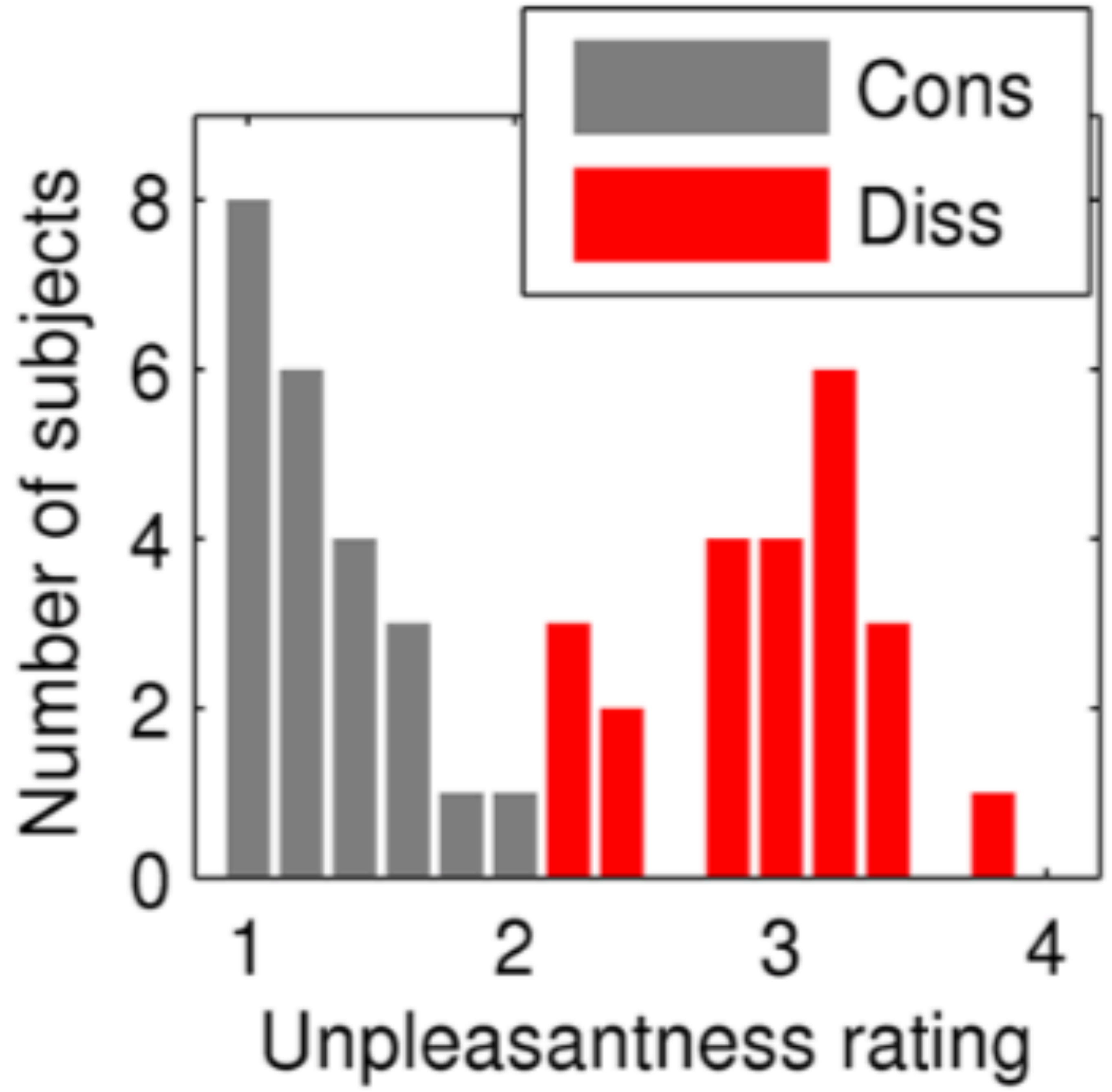
Neuroscientific view



Study I: Consonant vs. Dissonant music

Music 🎶 → 😀😥 Emotions





"Dissonant" stimuli were rated more unpleasant.

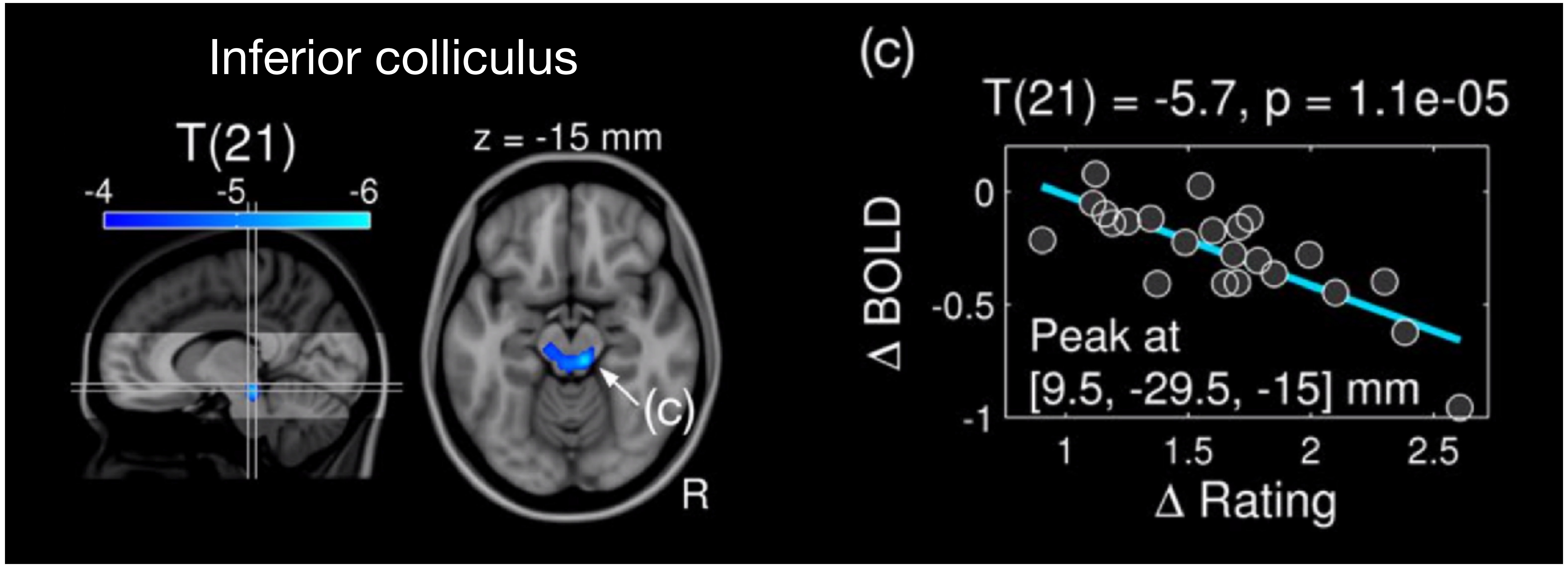
stimuli per condition = 20, stimulus duration = 30 s, # subjects = 23 (25.9 yo; 13 F)

Kim et al., 2017, *Sci Rep.*

Study I: Consonant vs. Dissonant music

🧠 → 😀😥

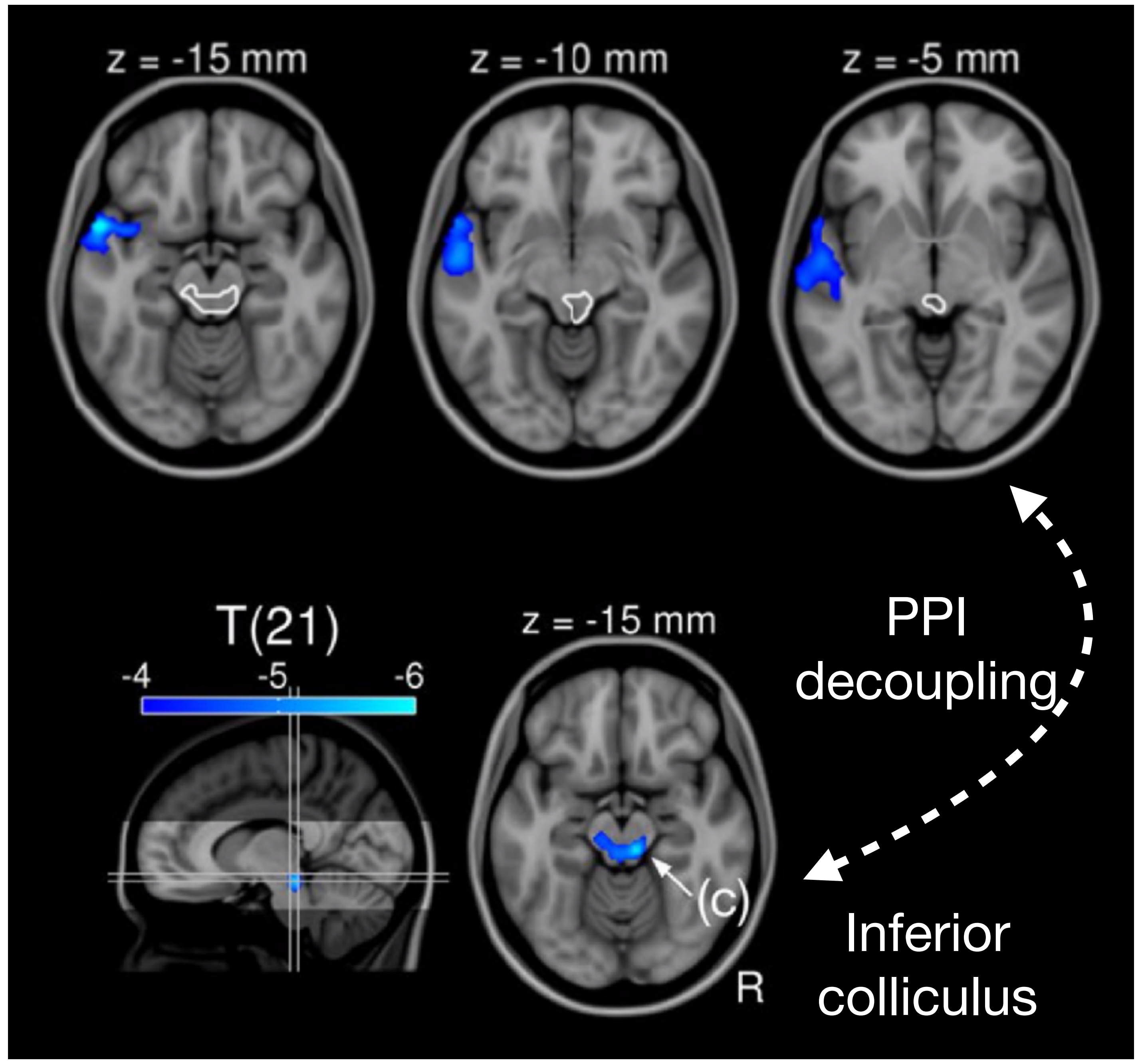
$$\Delta BOLD = \beta_0 + \beta_1 \Delta Rating + error$$



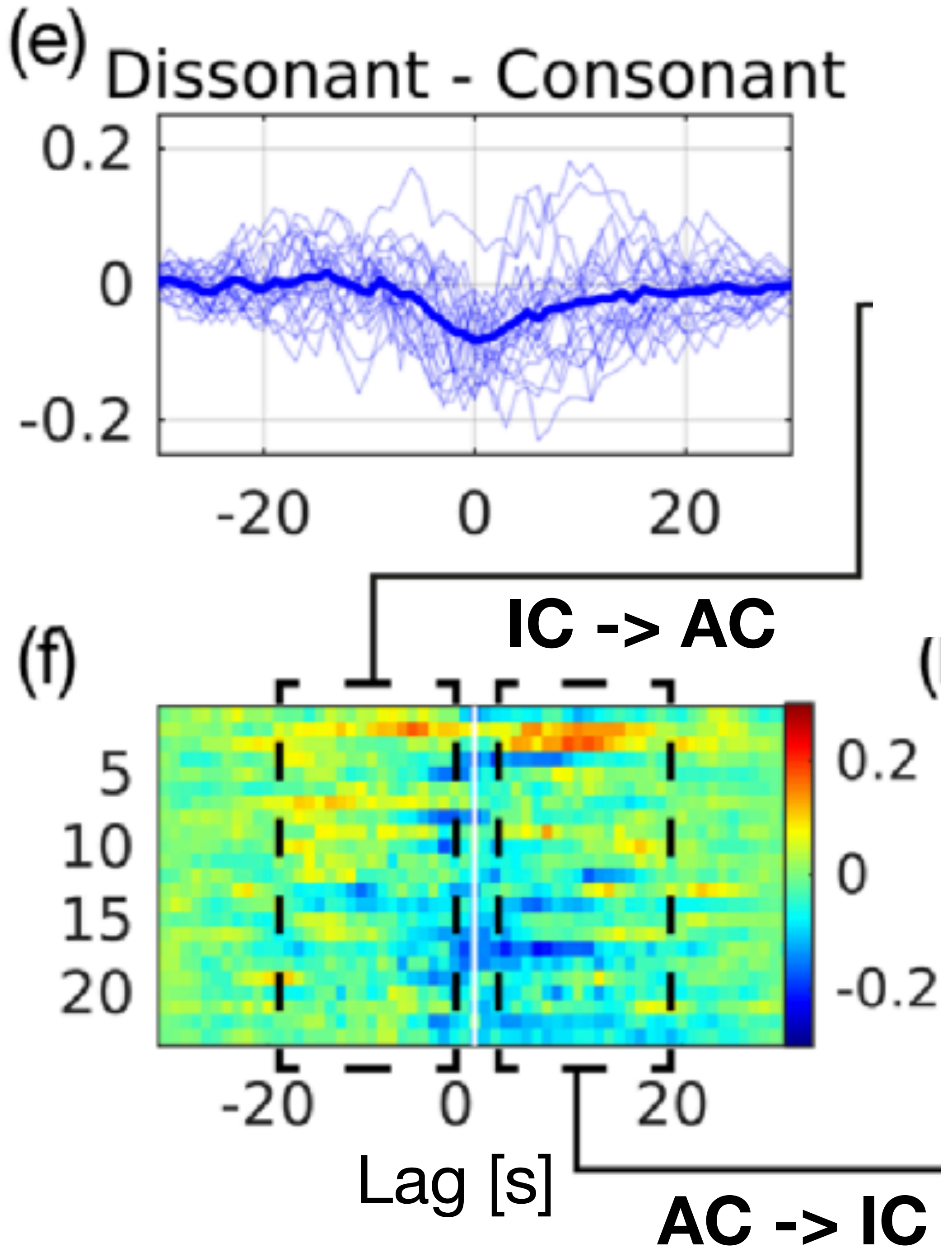
BOLD activation in the inferior colliculus (IC) decreased more in individuals who rated dissonant versions worse.

Study I: Consonant vs. Dissonant music

Psycho-Physiological Interaction (PPI)



Cross-correlation



The IC <u>decoupled</u> more from the left auditory cortex (top-down) in individuals who rated dissonant versions worse.

Kim et al., 2017, *Sci Rep.*

Limitations of controlled stimuli

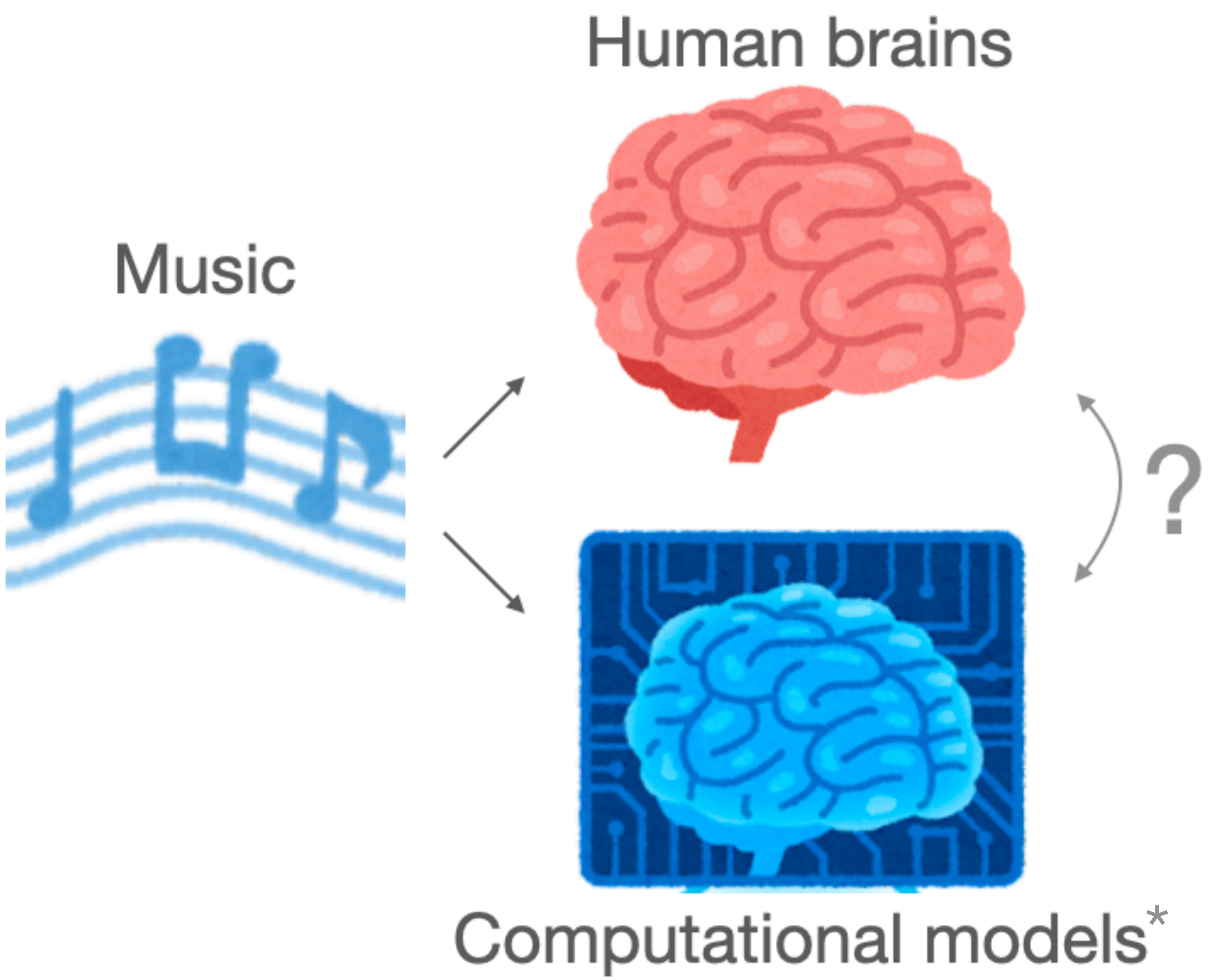
Ceteris paribus... and the subject is not listening.

Disruptive manipulations: validity, generalizability, & specificity?

Solution: Use of natural(istic) stimuli to increase the external validity of neuroscientific studies.

Study II: Approach

Computational models to extract information from natural music



*Possible mid/high-level representations that are relevant to emotions

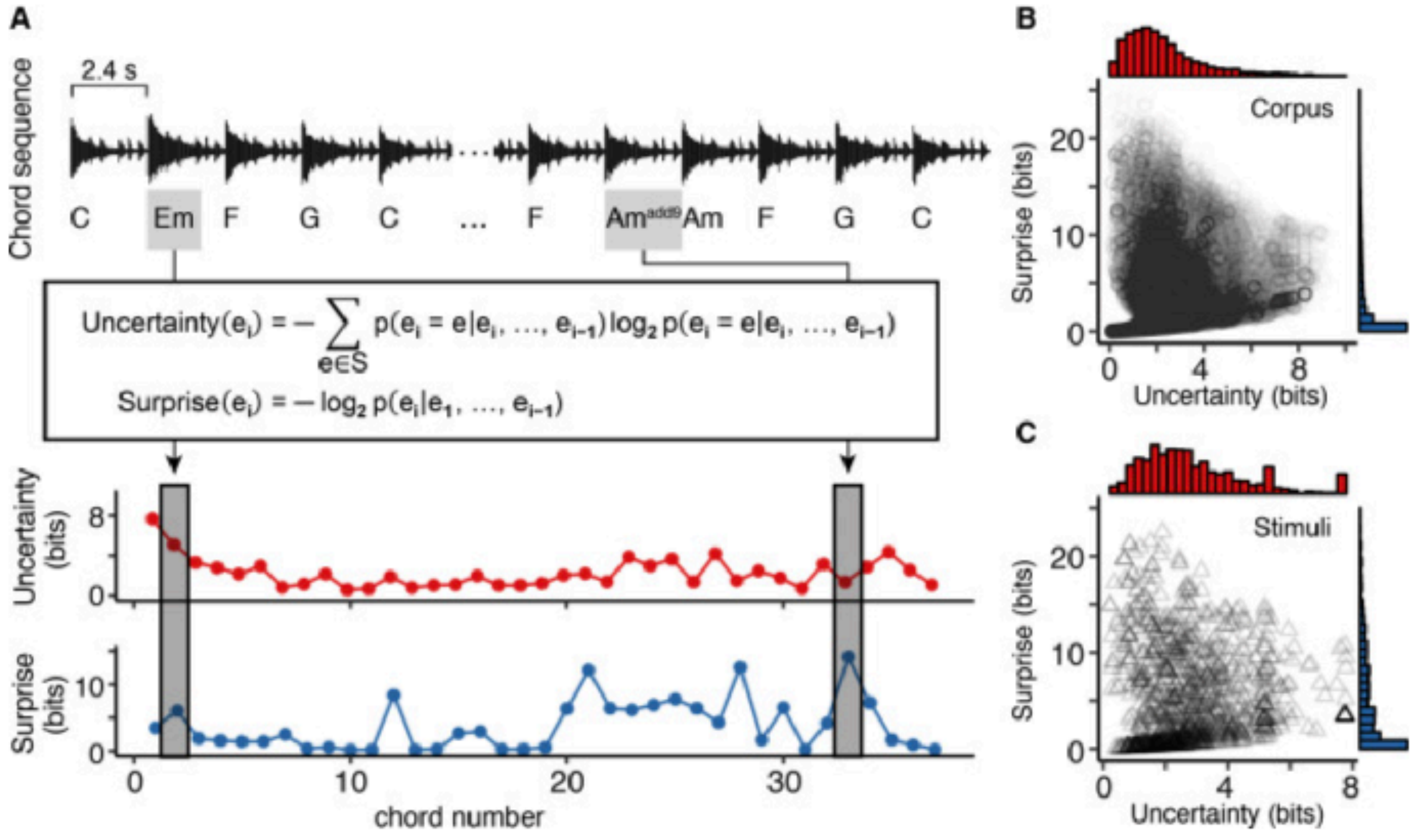
https://www.irasutoya.com/

How can we extract "meaningful" features from music signals?

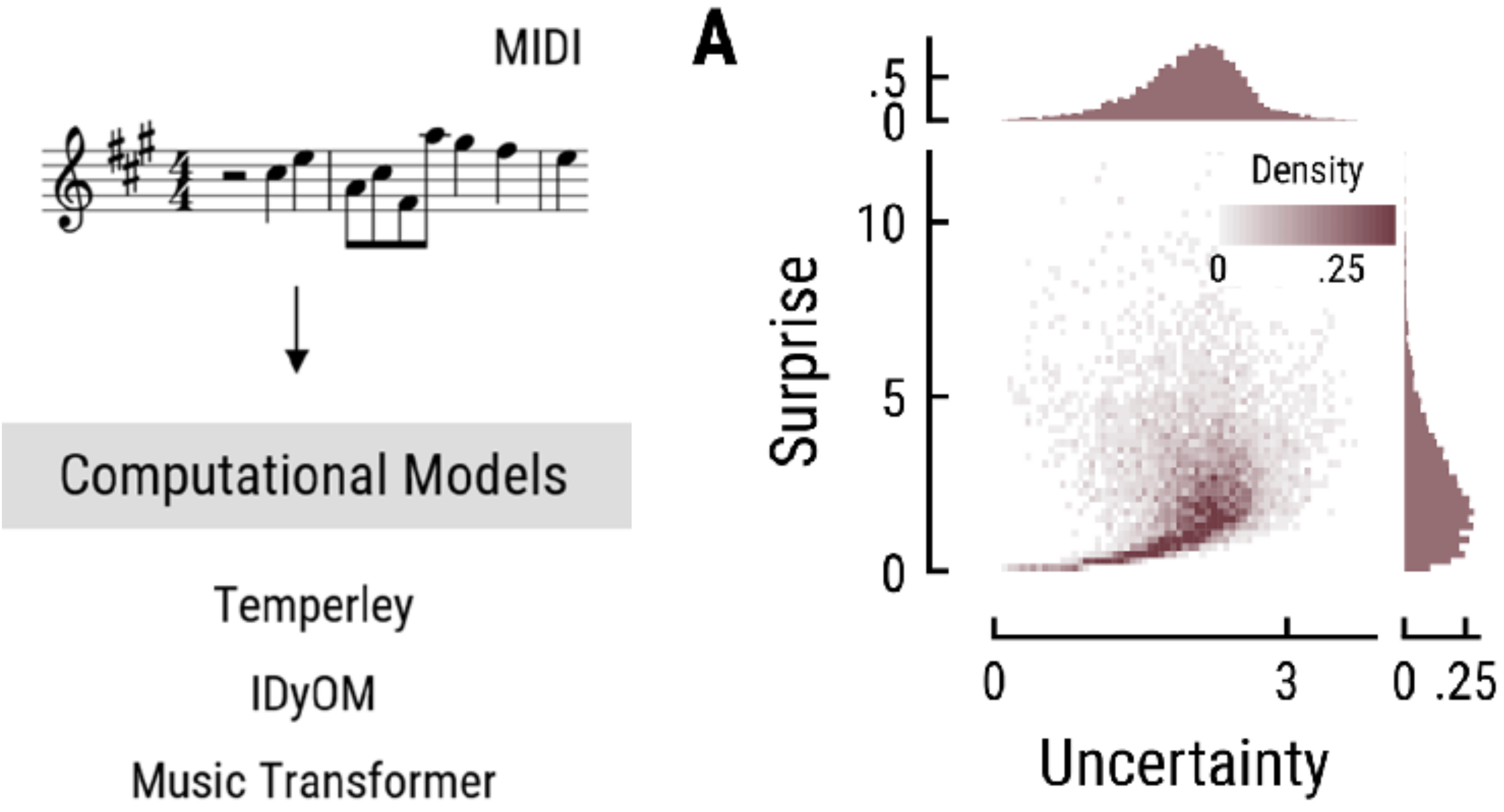
🎶

Symbol-domain models

Modeling on MIDI notes: **n-gram modeling** (e.g., PPM, HMM),
deep neural networks (e.g., Melody RNN, MusicRNN, Music Transformer)



Cheung et al., 2019, *Current Biology.*

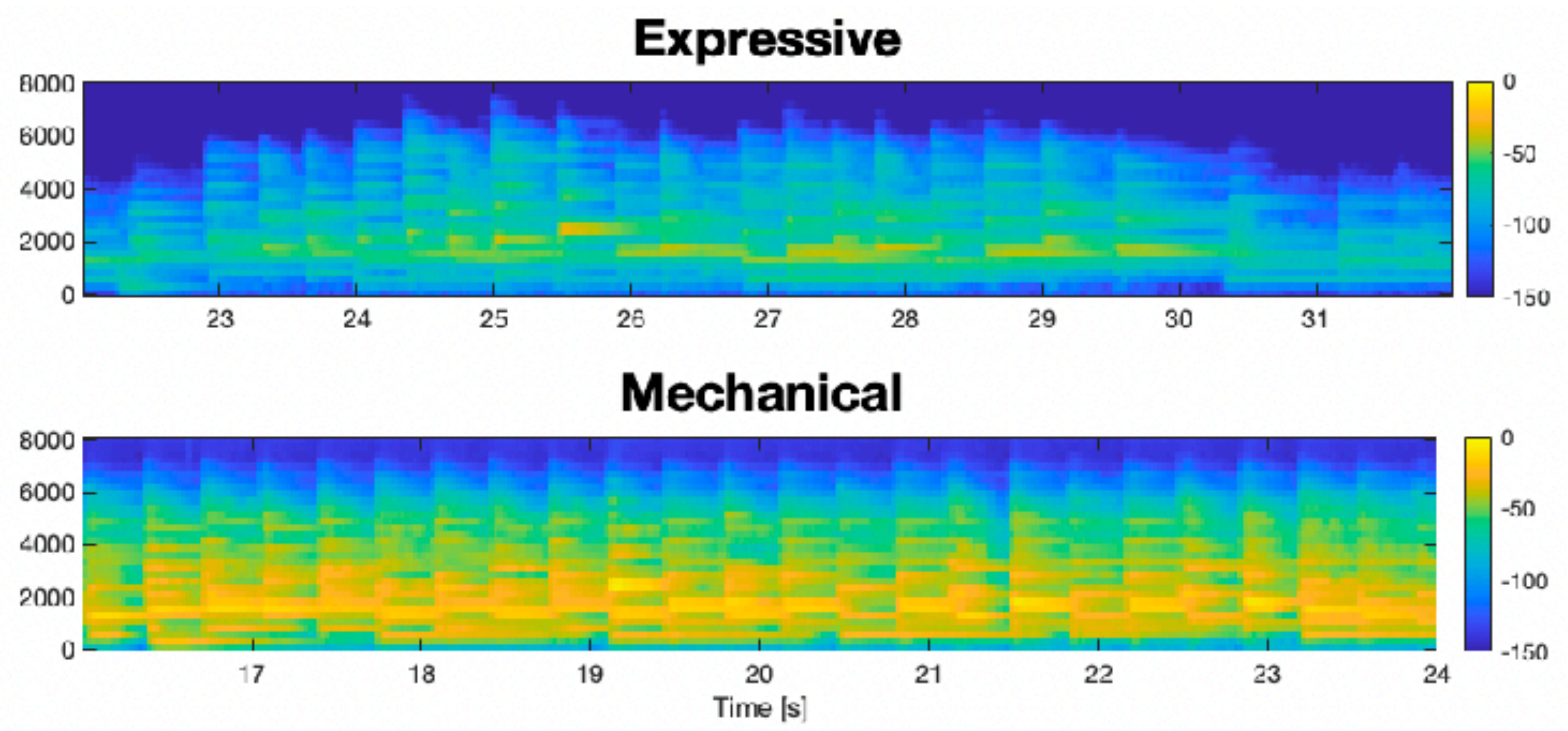


Kern et al., 2022, *eLife*.

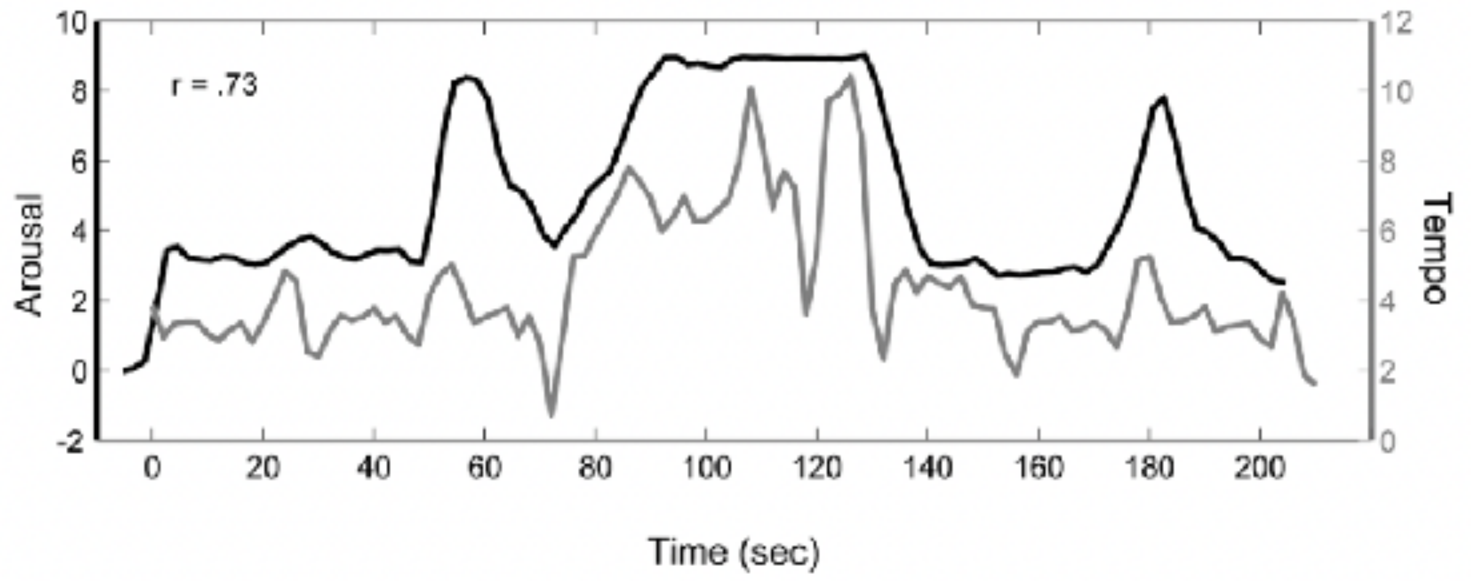
But notes/chords should be transcribed correctly.

Effects of expressions (tempi & dynamics)

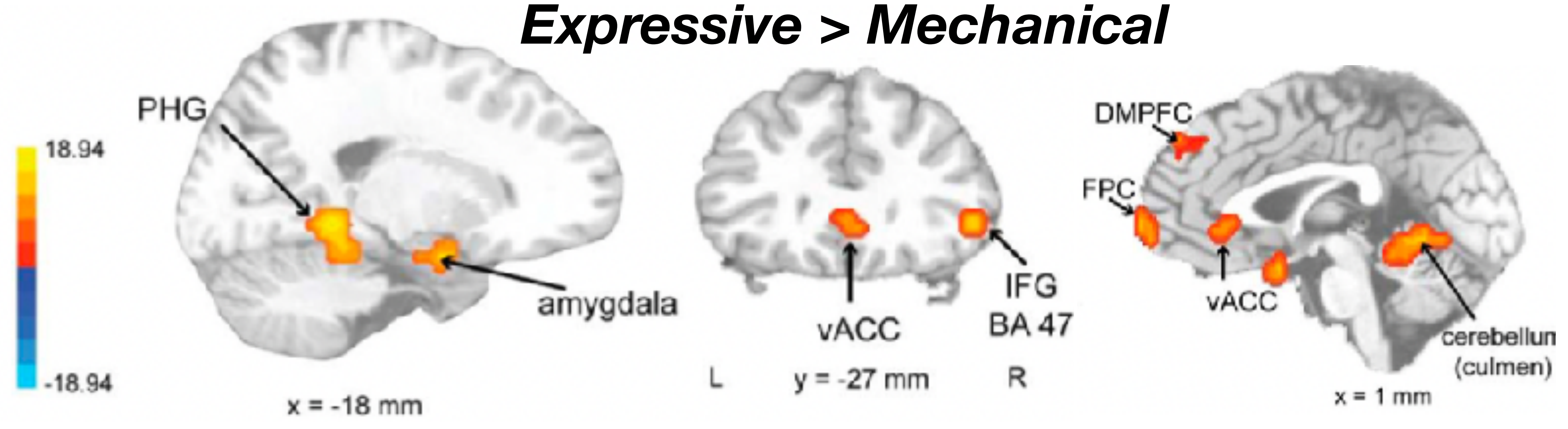
Chapin et al., 2010, *PLOS One*.



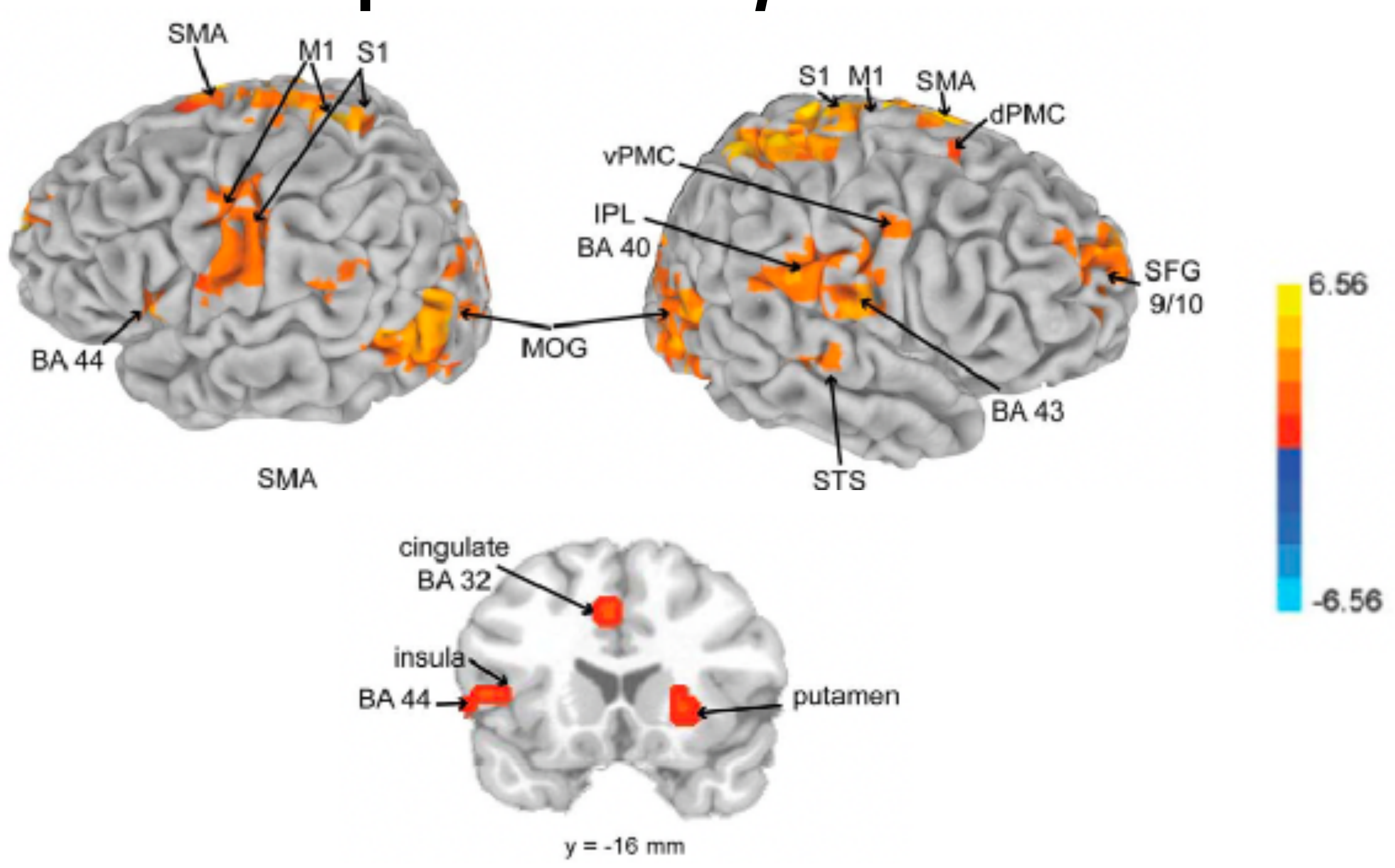
An exemplary participant (lag = 5.4 s)



Expressive > Mechanical



Tempo within *Expressive*

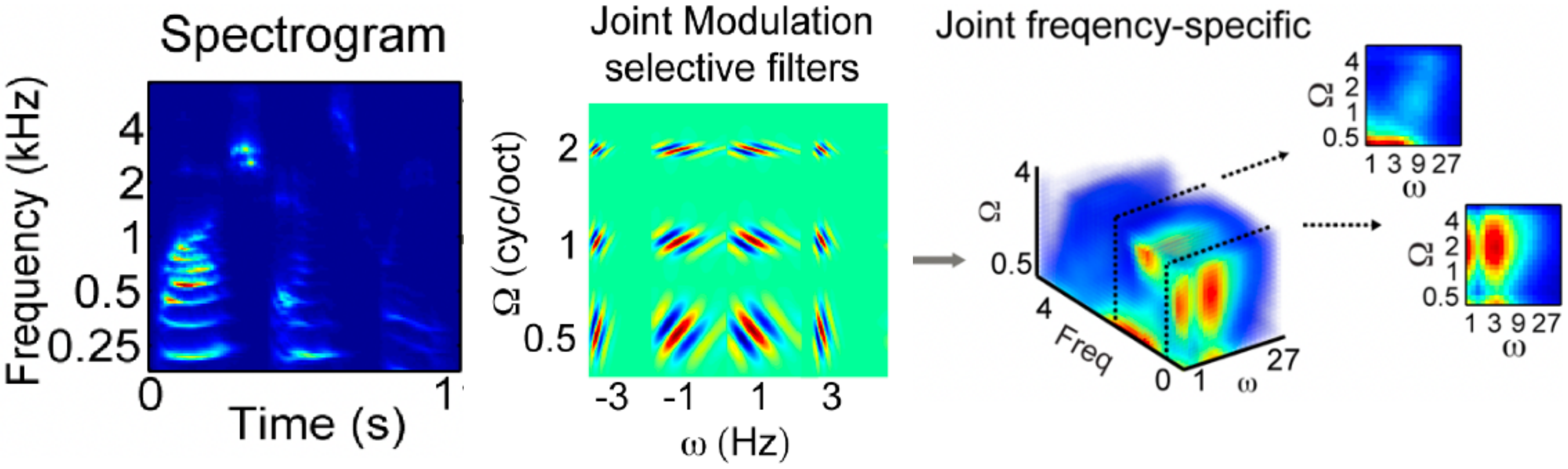


Chapin et al., 2010, *PLOS One*.

Symbolic models do not capture musical expressions!

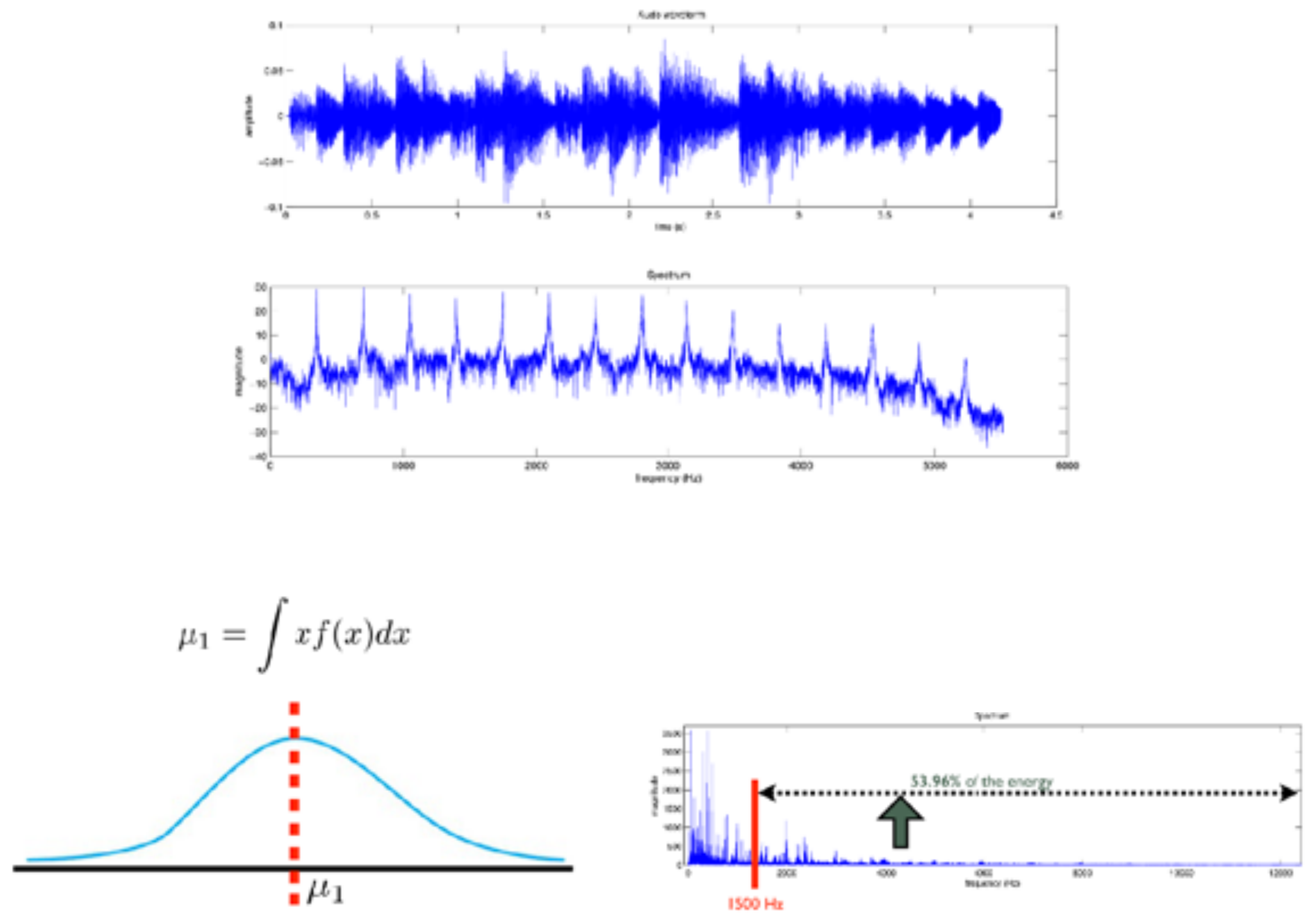
Audio-domain models

Auditory models: simulated activity of peripheral and central auditory neurons



Santoro et al., 2014, *PLOS Comp Biol*.

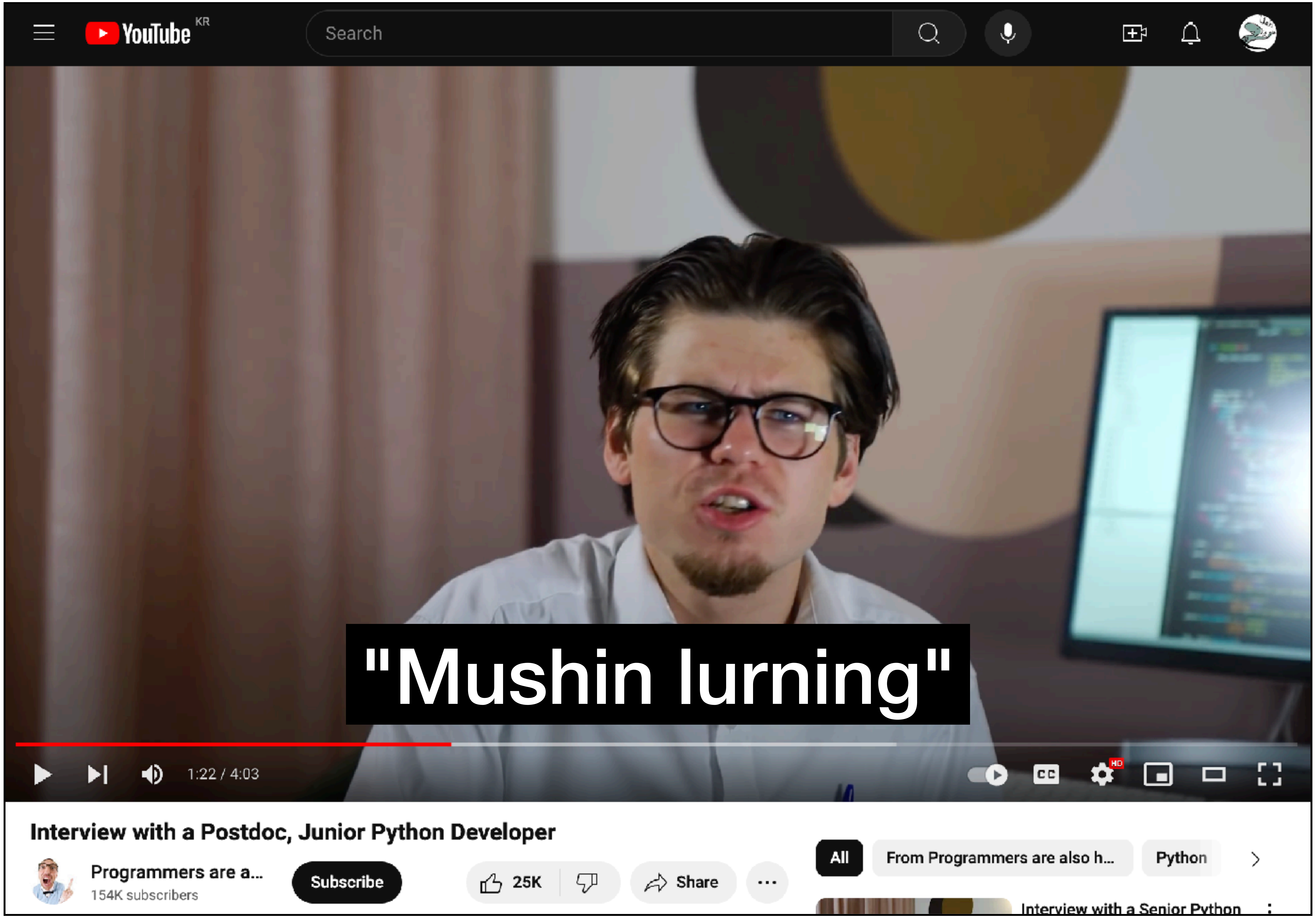
Music information retrieval (MIR) models: acoustic features relevant to music database



Lartillot, 2017, *MIRtoolbox User's Manual.*

Traditional models mostly simulate cochlear and primary auditory cortex activity.

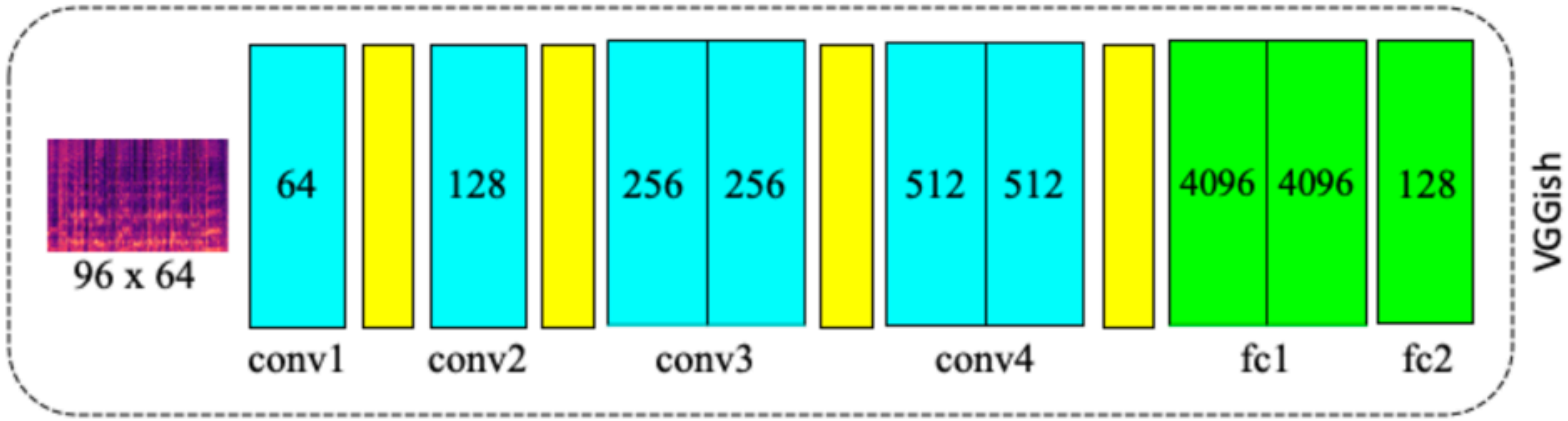
https://youtu.be/YnL9vAFphmE?si=RyR8PnvE2y9cCijl

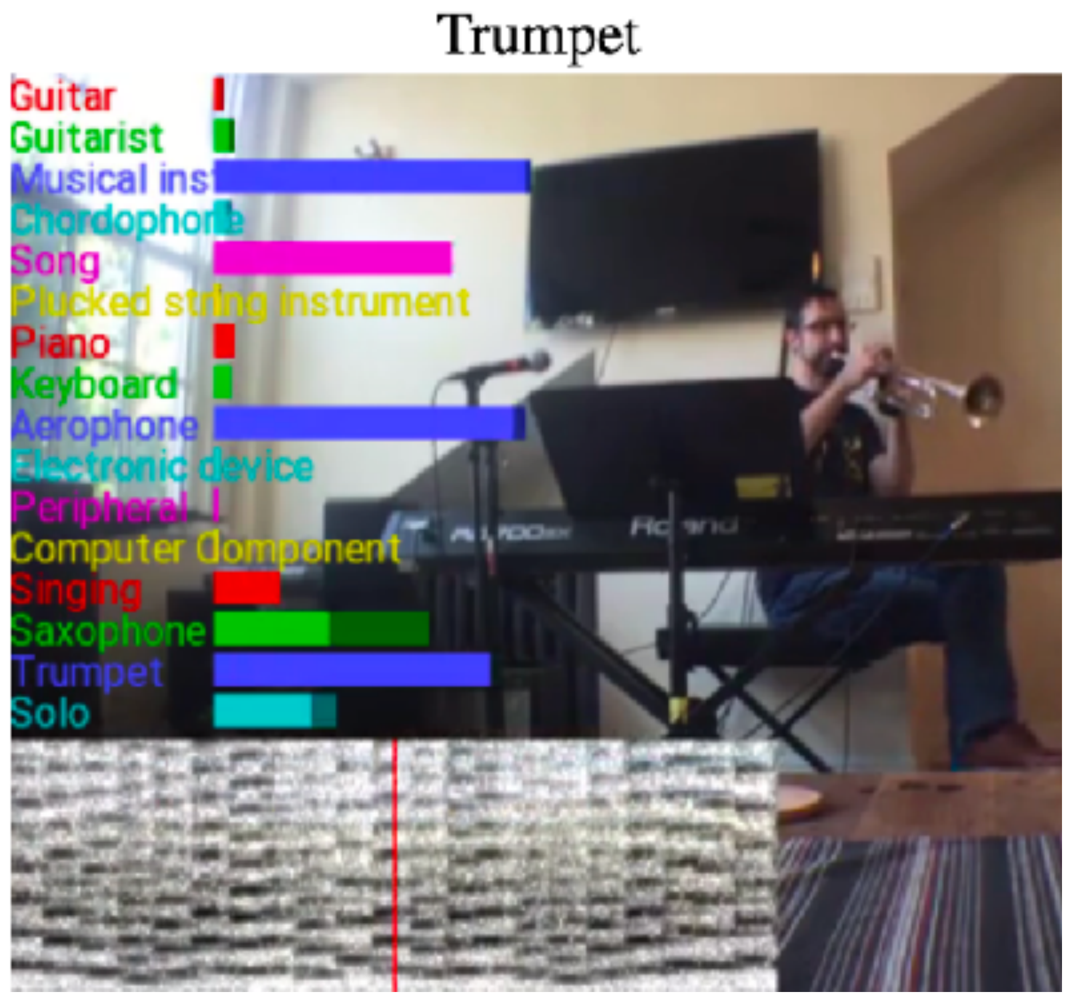


"Audio semantic" models

- **Convolutional neural networks (CNNs)** applied to short [~1 s] spectrograms to generate text labels of audio data (e.g., `'babyCry'`, `'dogBark'`)

- Trained on **video-audio correspondence** to learn **the second-order acoustics** (`"this kind of spectrogram often goes with that video"`).





VGGish: Hershey et al., 2017, ICASSP

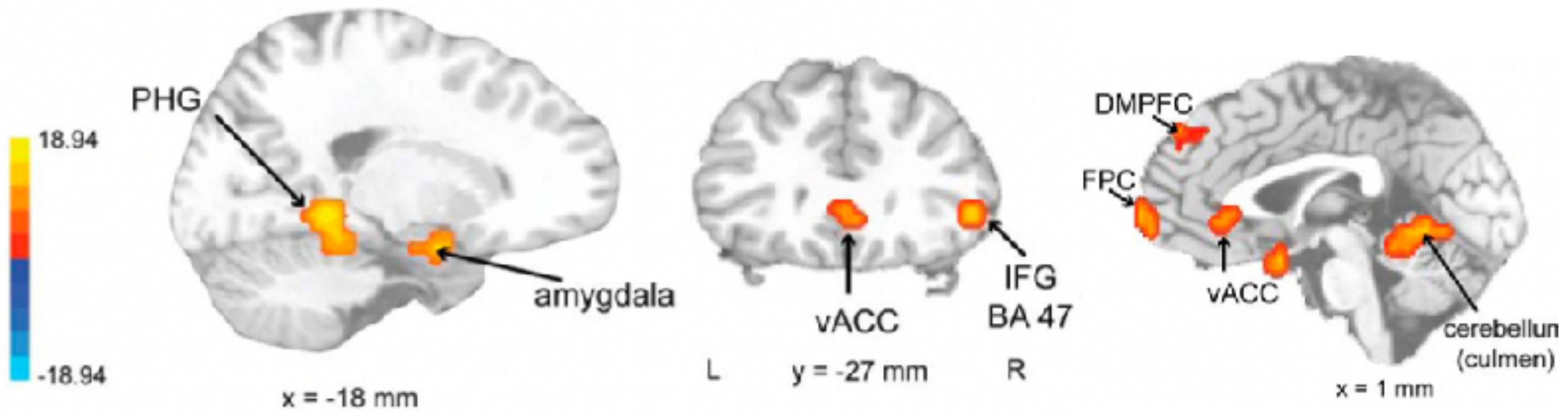
A possible mid/high-level representation of sounds

But really? Isn't it just timbre?

Example: Expressive vs. Mechanical

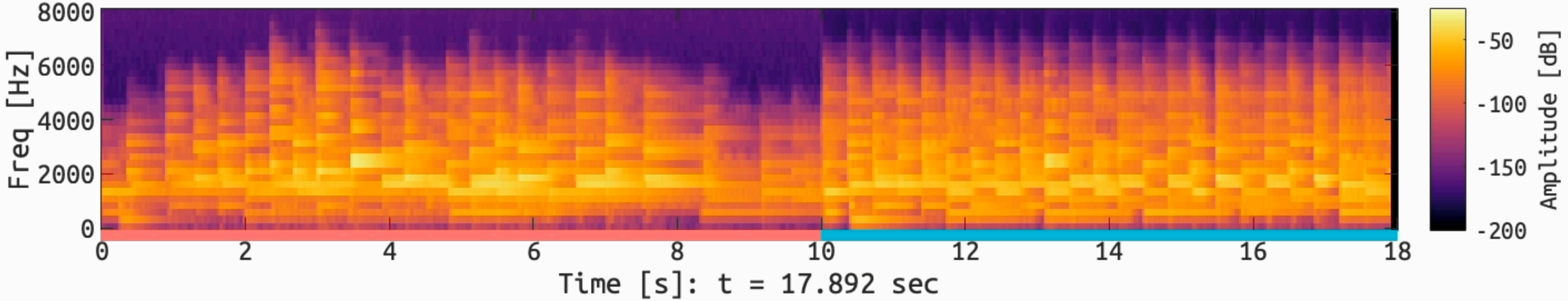
Chapin et al., 2010, PLOS ONE.

Expressive > Mechanical



Chapin et al., 2010, *PLOS One*.

[EXPRESSIVE]: "Frédéric Chopin's Etude in E major, Op.10, No. 3 was performed by an undergraduate piano major (female, 22 years old) on a Kawai CA 950 digital piano"

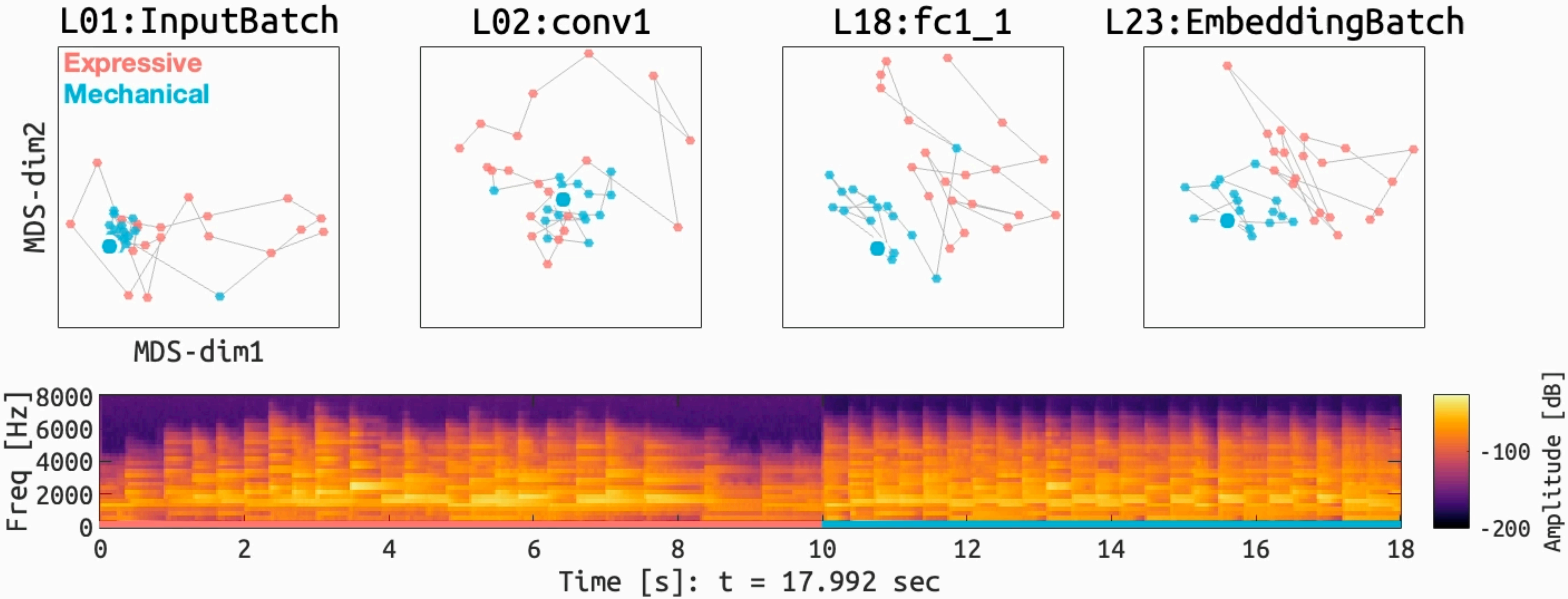


[MECHANICAL]: "The MIDI (Musical Instrument Digital Interface) onset velocity (key pressure) of each note (correlating with sound level) was set to 64 (range 0–128), and pedal information was eliminated."

Audio source: https://doi.org/10.1371/journal.pone.0013812

Example: Expressive vs. Mechanical

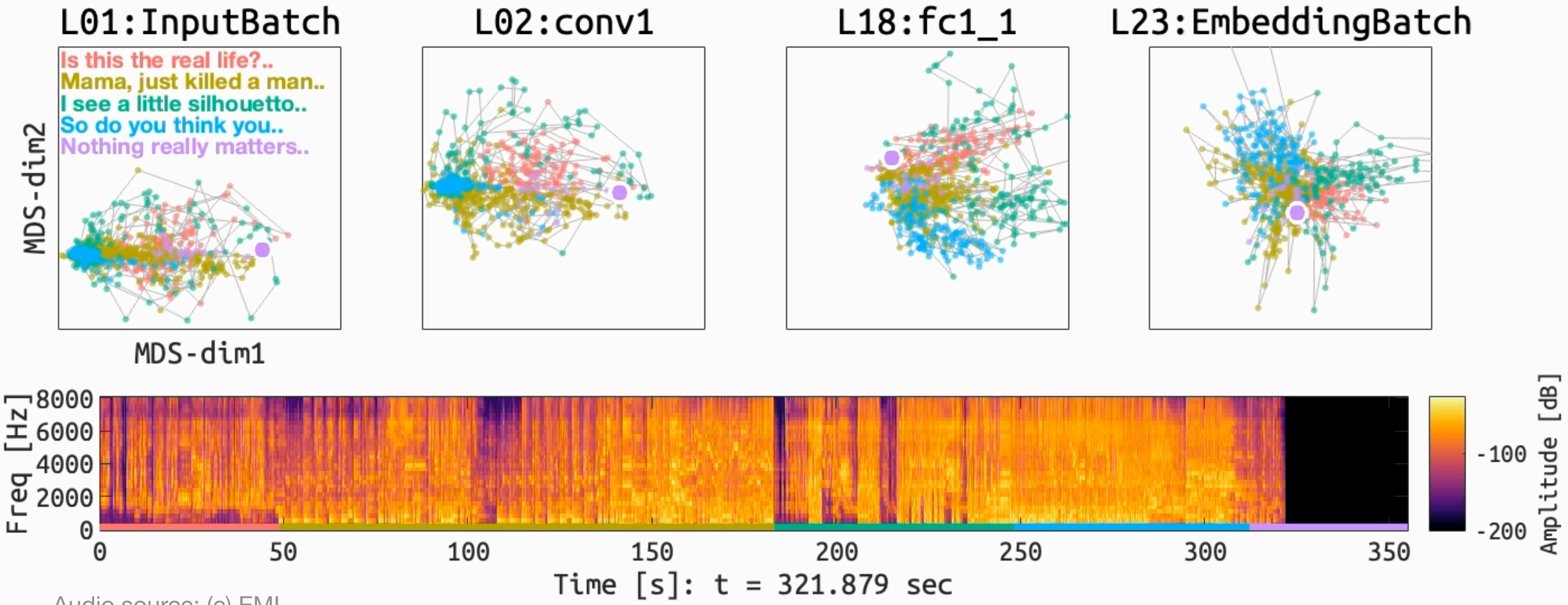
VGGish representations at various layers (from 1 to 23)



Audio source: https://doi.org/10.1371/journal.pone.0013812

Example: Queen. (1975). Bohemian Rhapsody

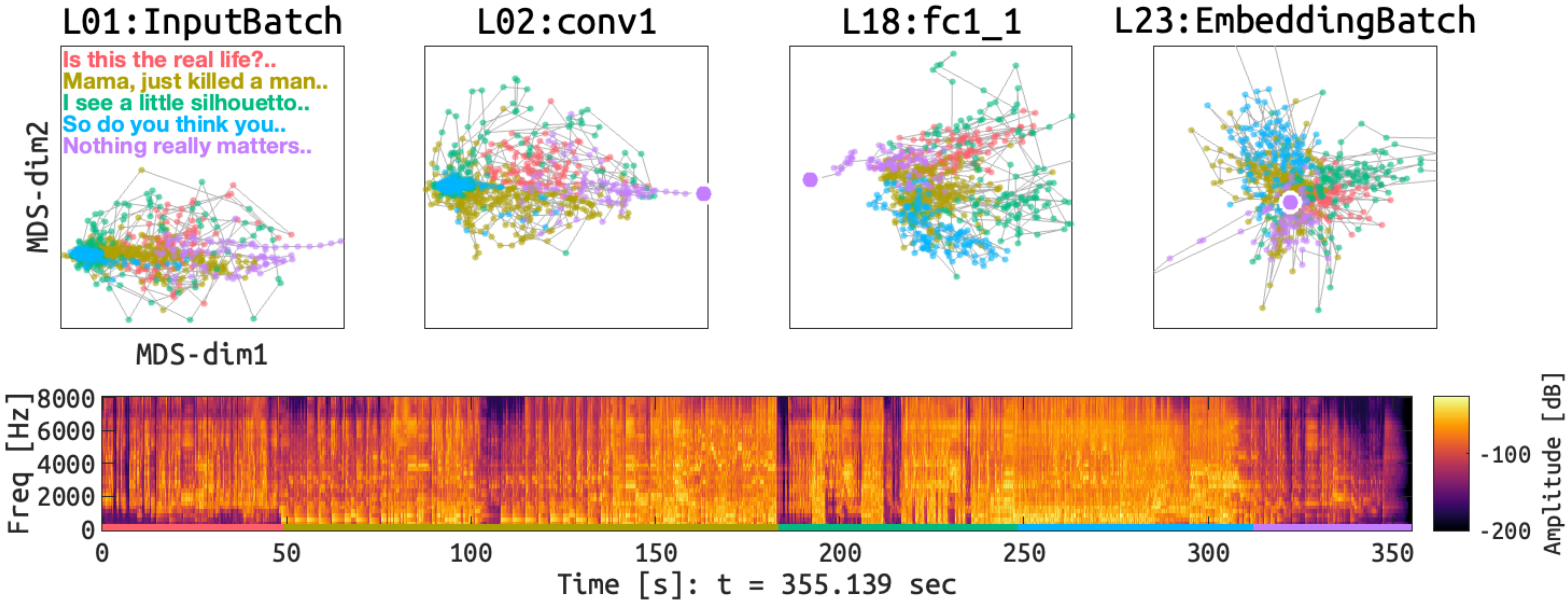
VGGish representations at various layers (from 1 to 23)



Audio source: (c) EMI

Example: Queen. (1975). Bohemian Rhapsody

VGGish representations at various layers (from 1 to 23)



CNN embedding for music emotion *recognition*

Potentially mid/high-level representation of music signal

An "deep audio semantic" model called "VGGish"

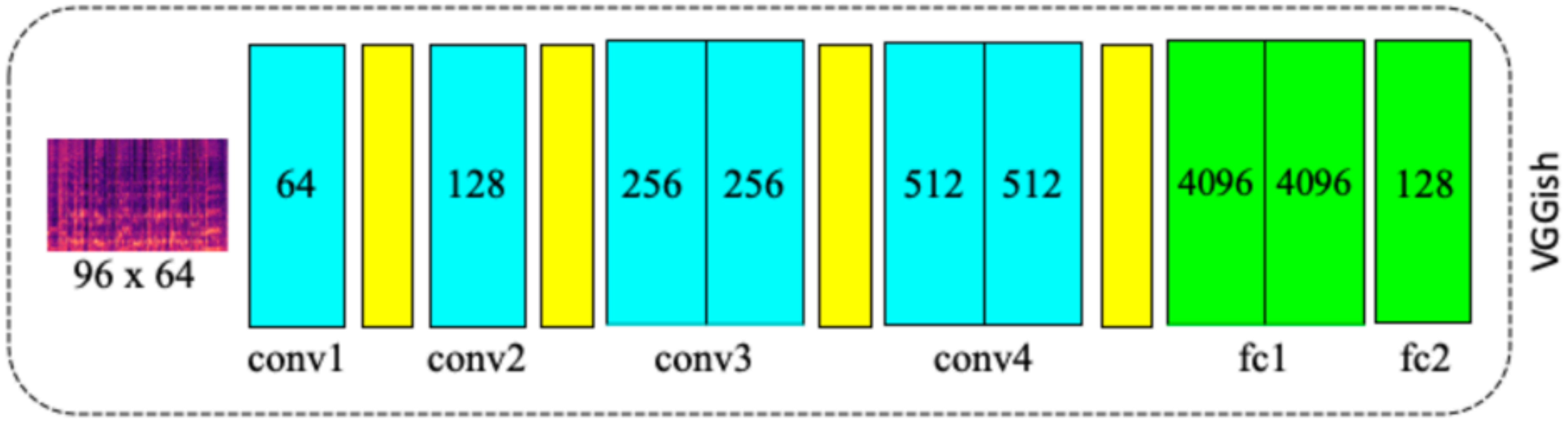
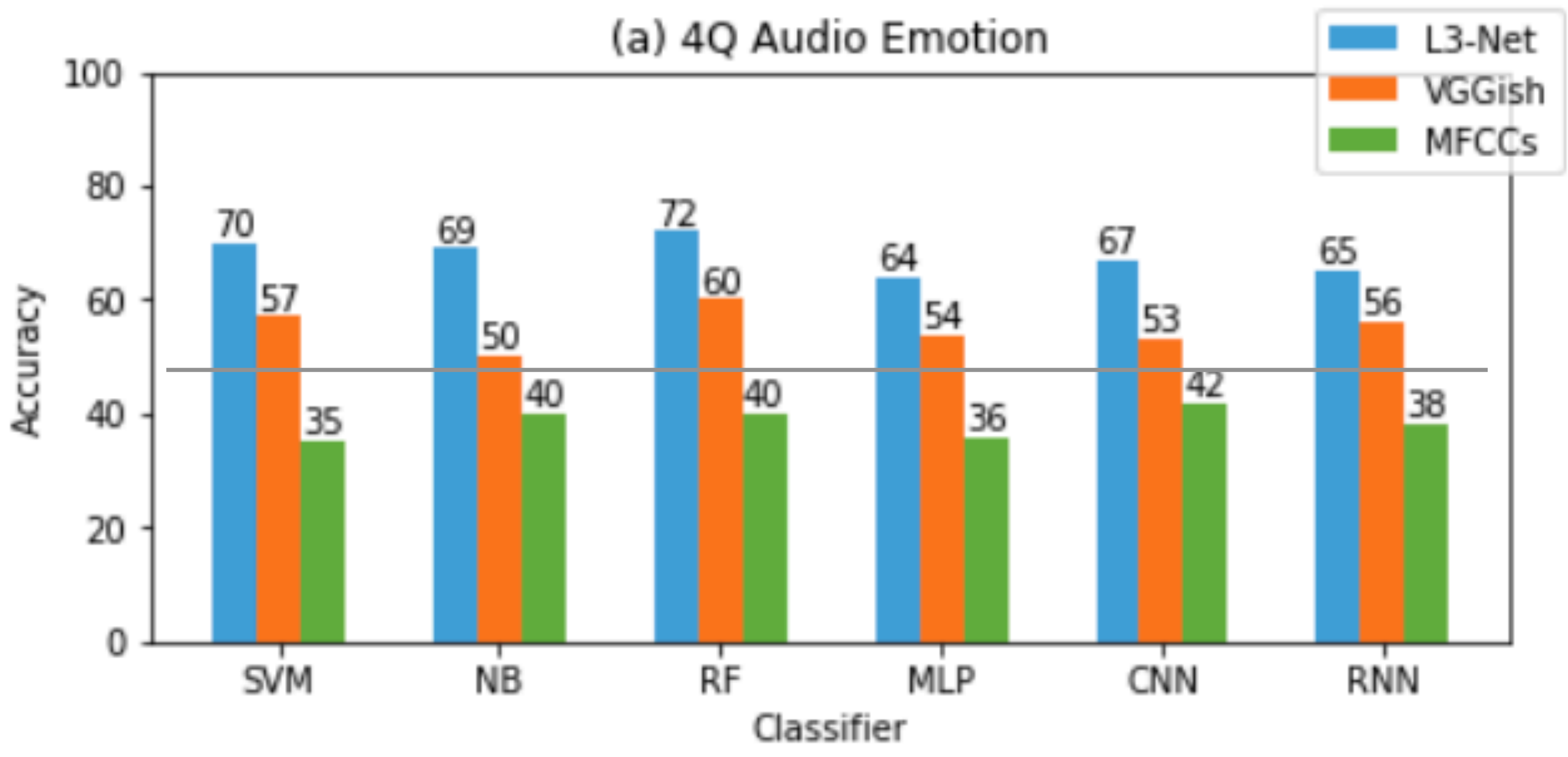


Diagram from Koh & Dubnov, 2021, *ACA*

4Q Audio Emotion Dataset: 255 music clips (30 s) for **Arousal-Valence quadrants**



Koh & Dubnov, 2021, *ACA*

Deep audio semantic models carry more information related to expressed emotions than a traditional audio descriptor.

How do we measure brain activity?

Functional magnetic resonance imaging

Still the state-of-the-art non-invasive functional imaging

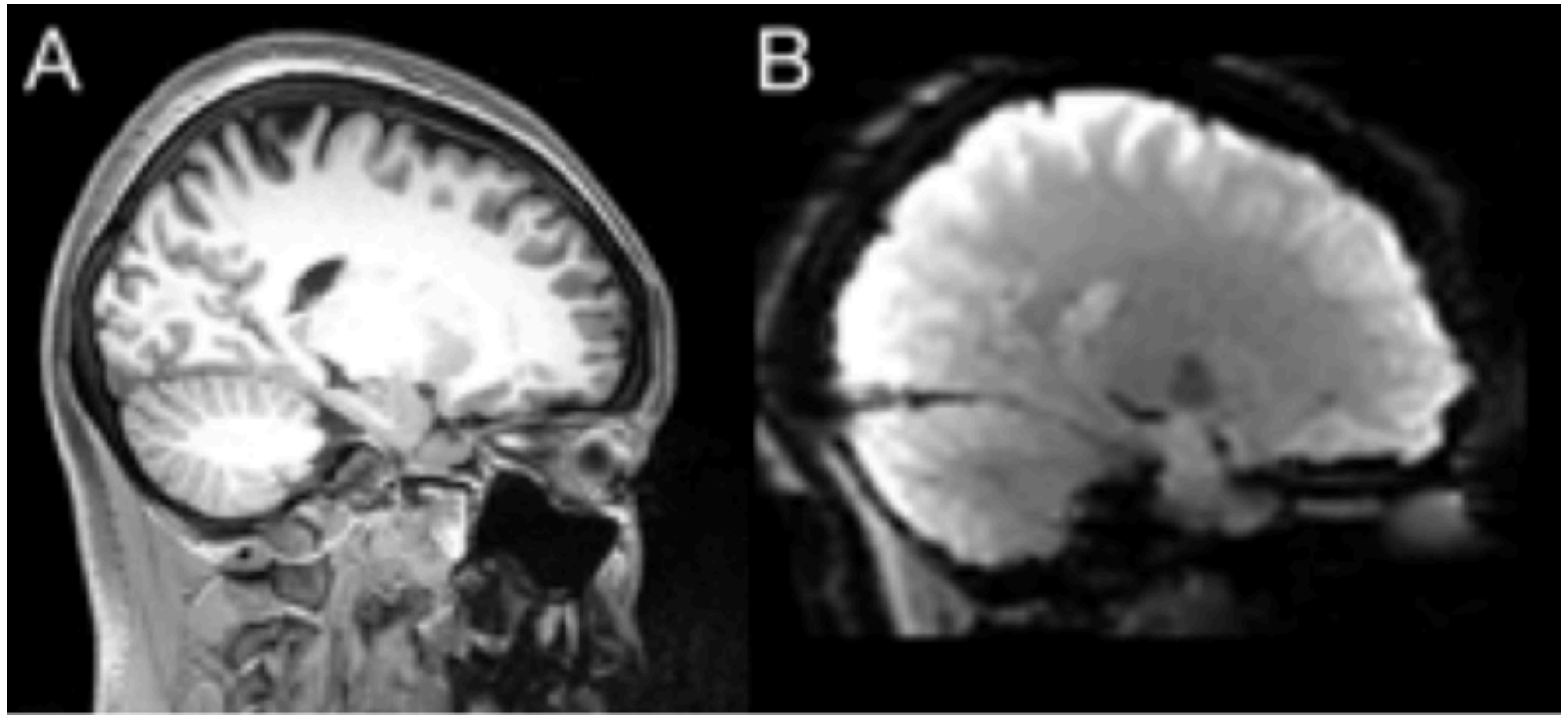
- **Hardware**: the same MRI at hospitals (but with a different program)

- **Mechanism**: Blood-oxygenation-level-dependent (BOLD) effect

- **Strengths**: the highest spatial (3-mm-iso) and temporal resolution (1 sec) than any other non-invasive imaging (e.g., PET)

- **Weaknesses**: indirect neural activity (unlike M/EEG, ECoG), ~120 dB acoustic scanning noise (but there are active noise cancelling headphones for MRI)

Structural MRI **Functional (BOLD) MRI**



Kim, 2017, *Dissertation*.

https://www.irasutoya.com/

How do we measure emotions?

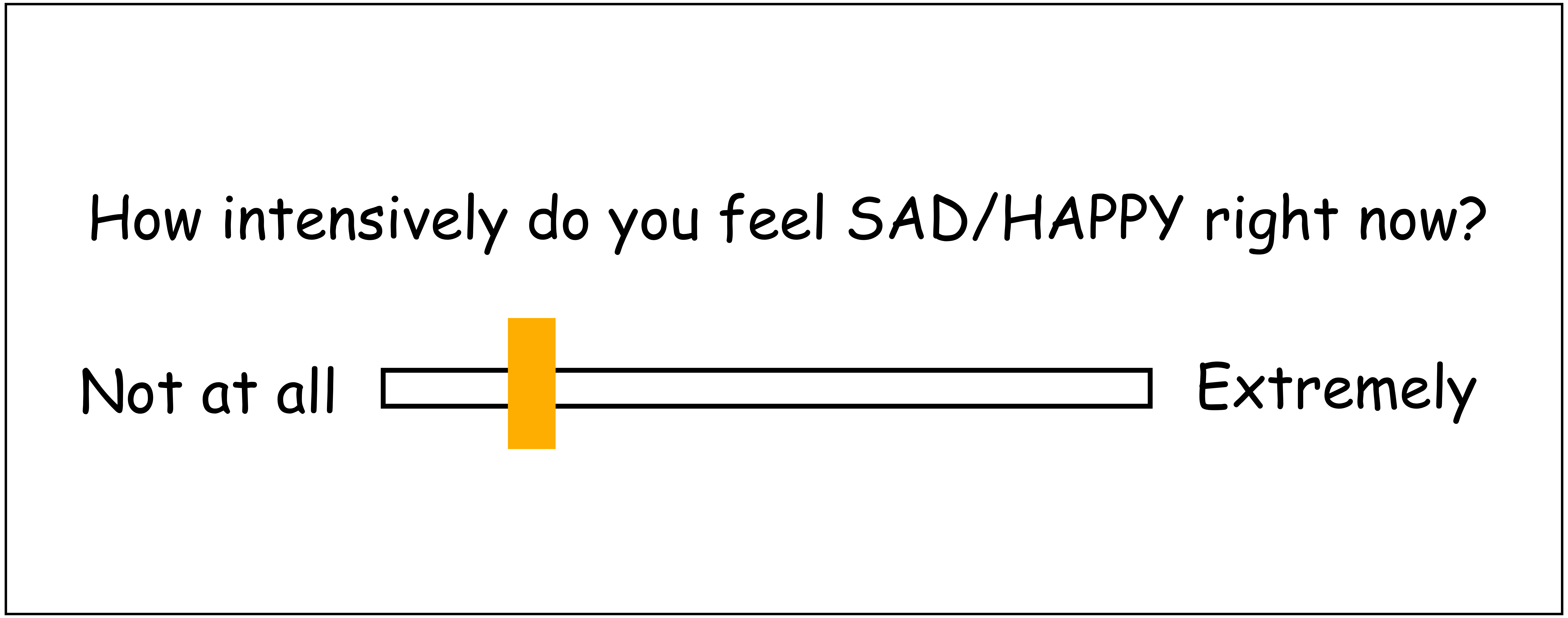
😀😥

Felt vs. perceived emotions

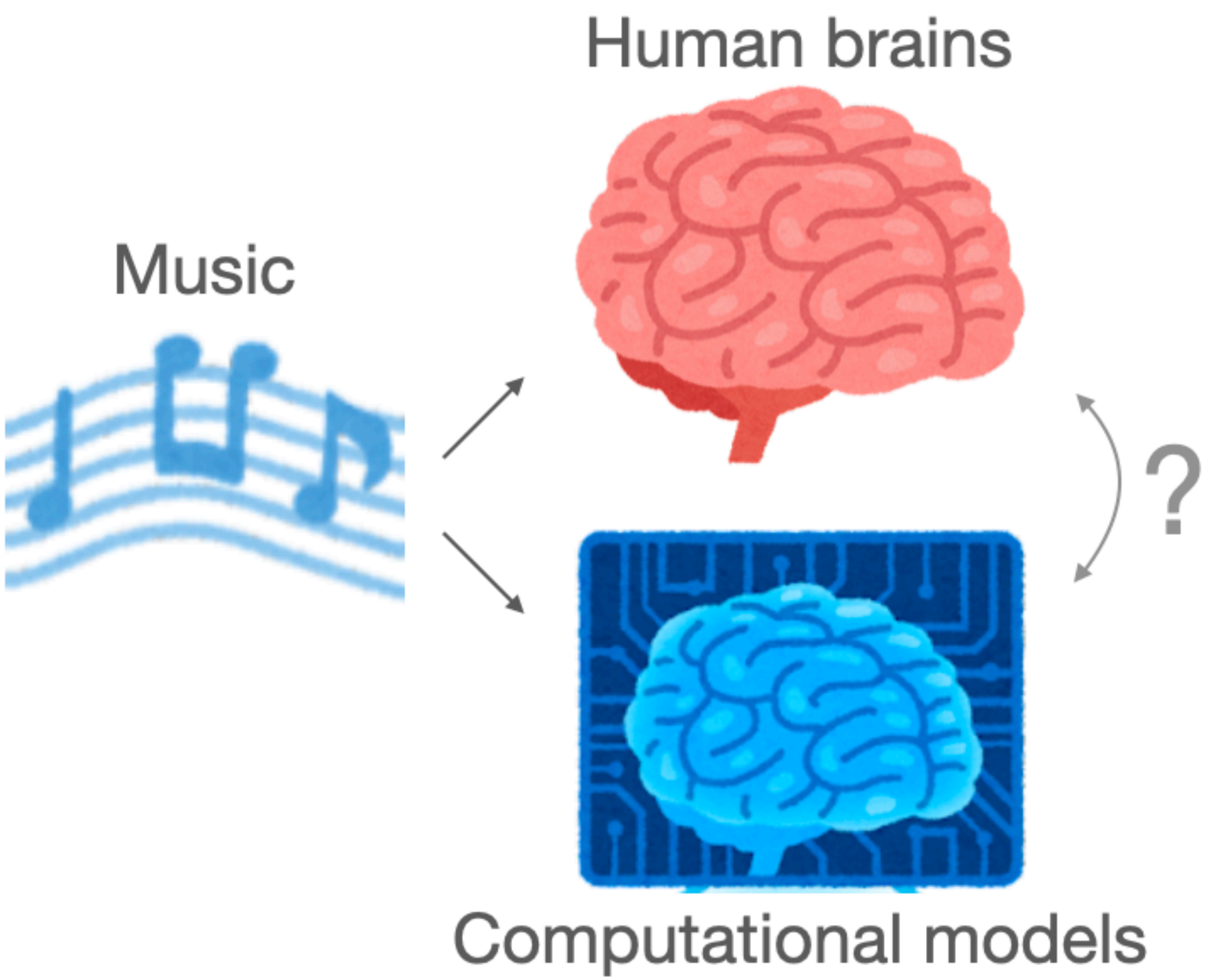
We just ask people how they feel...

Rating with a slider

https://www.irasutoya.com/



Research Questions



- **Q1**: How are ***felt* emotions** and **musical enjoyment** associated with neural activity over time?
- **Q2:** Would **increasingly abstract representations of music** in different layers of the CNN be encoded along the cortical gradient axis of abstraction?
- **Q3**: How do **layer-specific CNN embeddings** predict human behavioral ratings of musical emotions?

Kim, 2022. *Front. Neurosci.*
https://www.irasutoya.com/

Methods

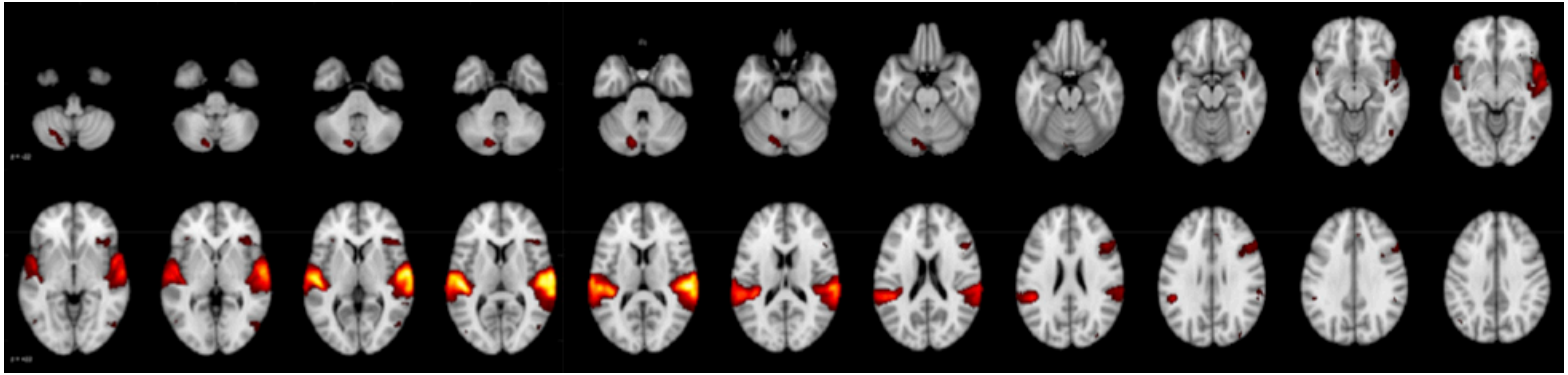
Original study

Sachs et al., 2020, *NeuroImage*.

OpenNEURO

https://openneuro.org/datasets/ds003085

Inter-subject correlation during a "sad" piece of music: r ~ [0, 0.16], cluster-$P < 0.05$

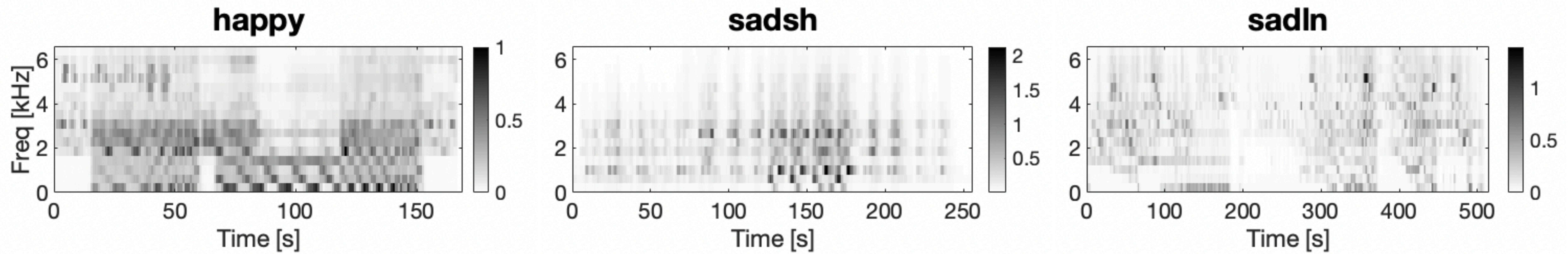


Sachs et al., 2020, *NeuroImage.*

Stimuli

Sachs et al., 2020, *NeuroImage*.

- **Happy** [2 min 48 sec]: Lullatone's "Race against the Sunset"
- **Sad-short** [4 min 16 sec]: Olafur Arnalds's "Frysta"
- **Sad-long** [8 min 35 sec]: Michael Kamen's "Discovery of the Camp"



Kim et al., *In prep*.

Participants & protocol

Sachs et al., 2020, *NeuroImage*.

- N = 40 (21 female, mean age = 24.1 ± 6.24 from LA)
  - Unfamiliar with 3 stimuli and reported "intended" emotions from 60-s excerpts

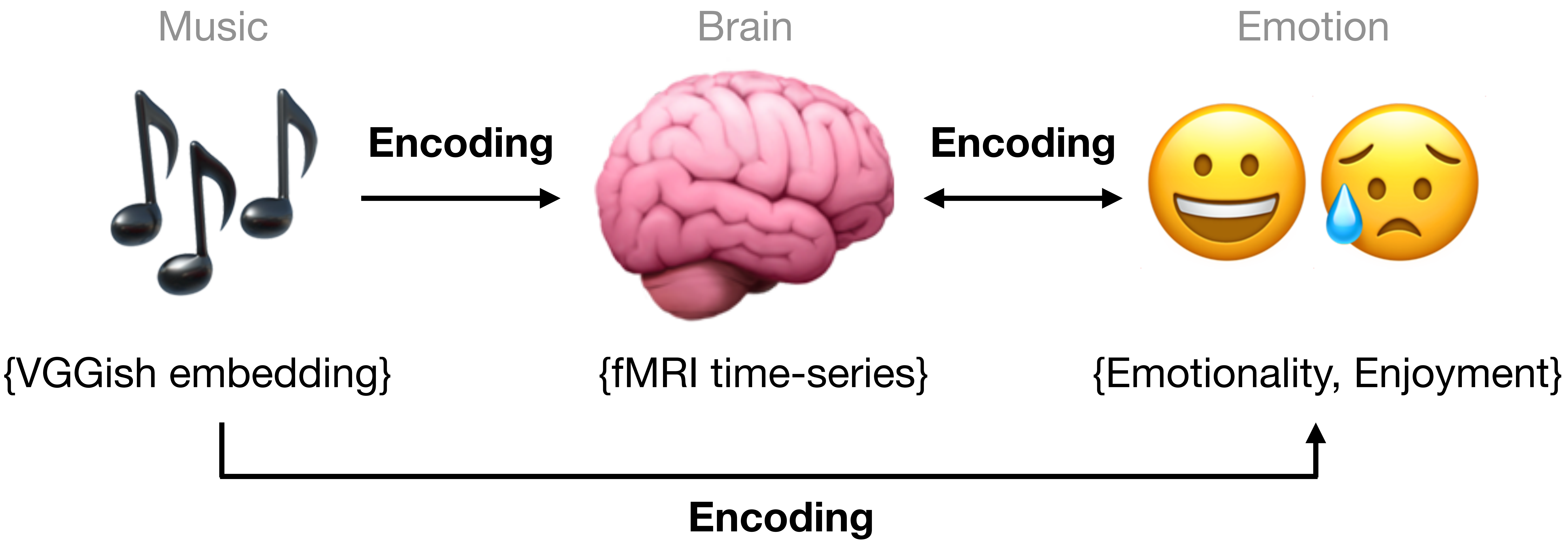
Passive listening with eyes open

Rating with a slider

- "The intensity of felt sadness or happiness" (***Emotionality***)
- "The intensity of enjoyment" (***Enjoyment***)

https://www.irasutoya.com/

Analysis overview



Encoding analysis

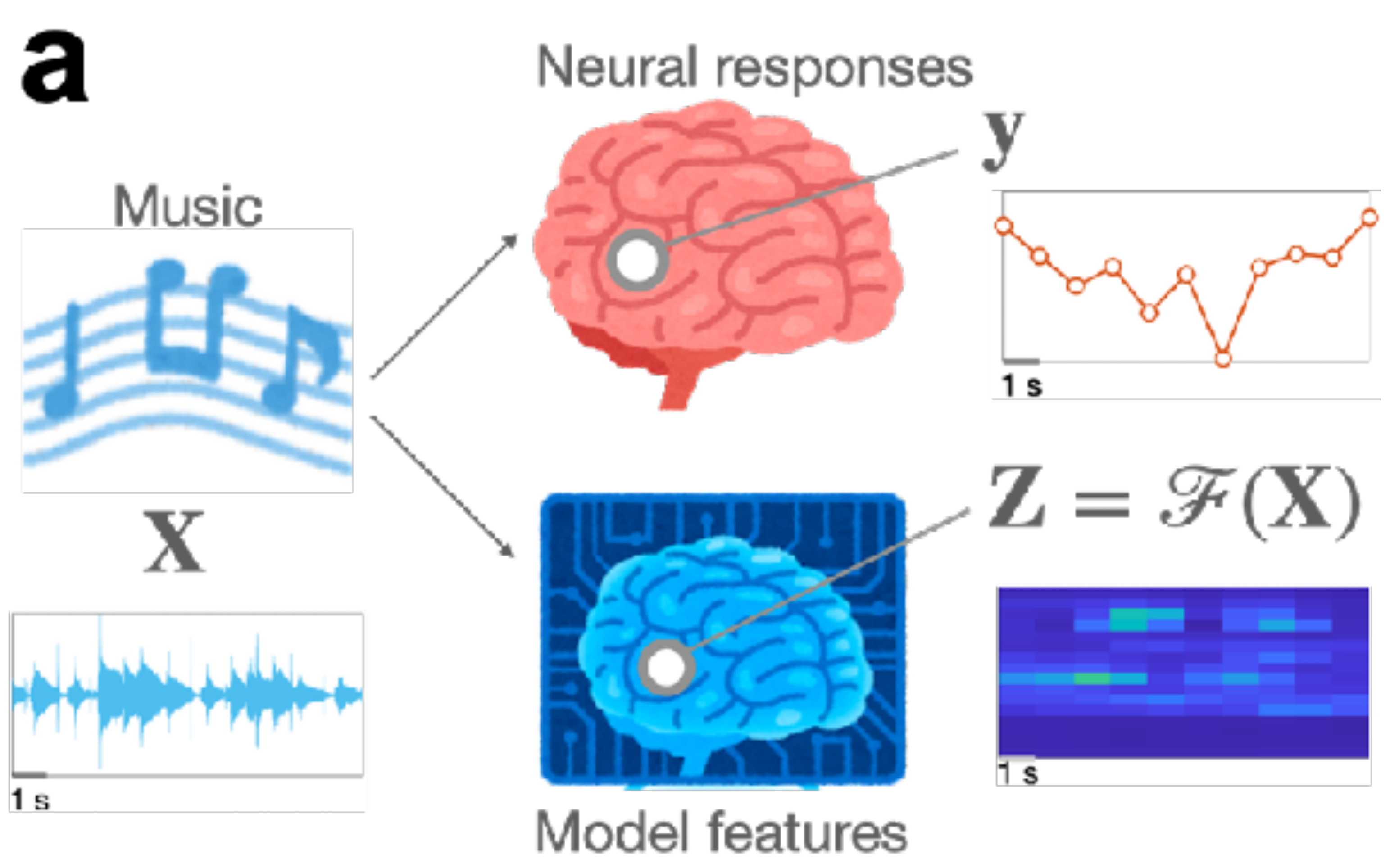
Kim, 2022. *Front. Neurosci.*

Layer-specific profile

How do we map VGGish-layer-specific encoding on the cortex?

How good the "Layer X" is for this brain area?

A "profile" of layer-specific prediction accuracies



b

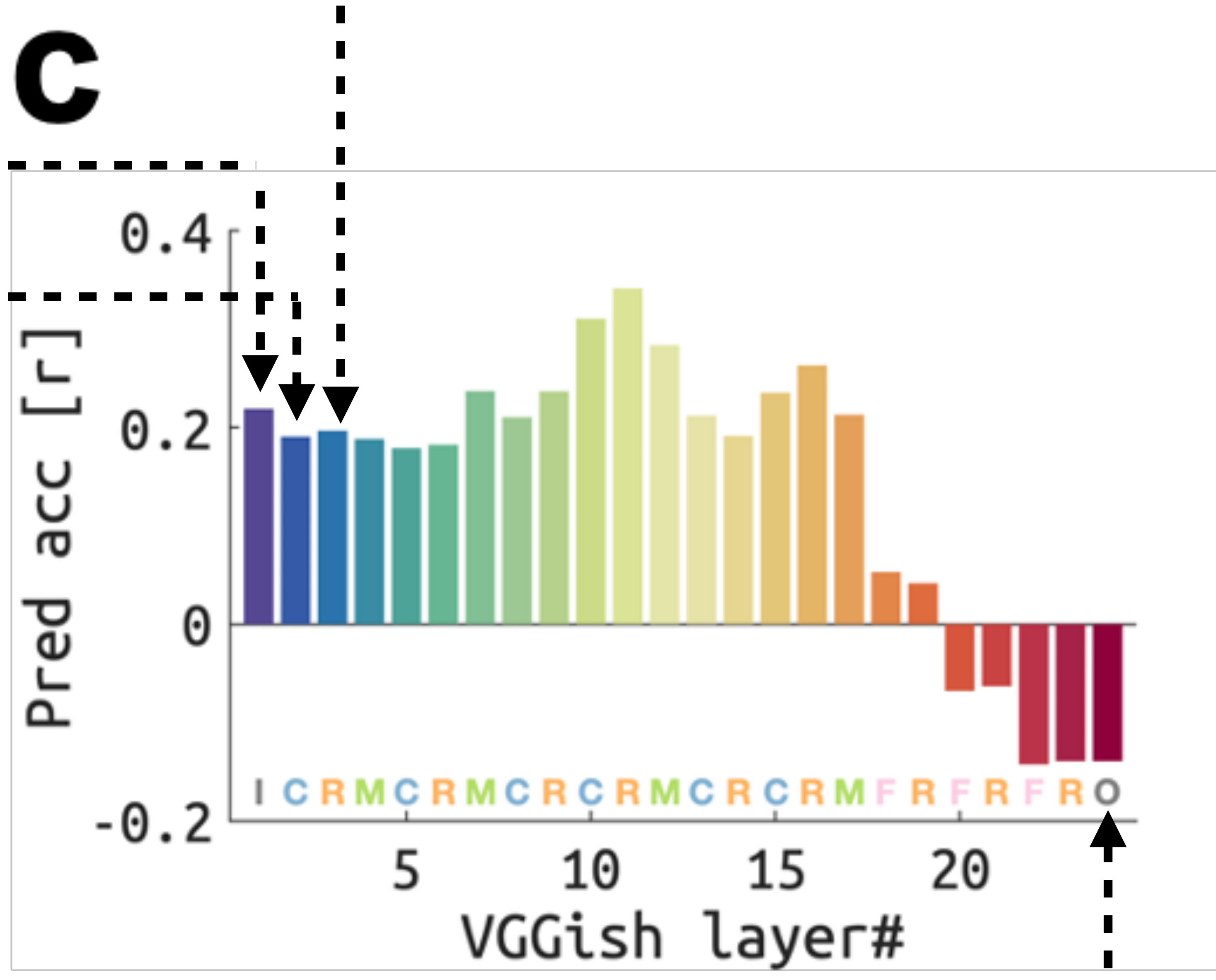
$$\boldsymbol{y}^{(1)} = \mathcal{F}_1(\mathbf{X}^{(1)})\mathbf{b}_1 + \varepsilon \rightarrow \boldsymbol{r}_1 = \mathrm{corr}(\hat{\boldsymbol{y}}_1^{(2)}, \boldsymbol{y}^{(2)})$$

$$\boldsymbol{y}^{(1)} = \mathcal{F}_2(\mathbf{X}^{(1)})\mathbf{b}_2 + \varepsilon \rightarrow \boldsymbol{r}_2 = \mathrm{corr}(\hat{\boldsymbol{y}}_2^{(2)}, \boldsymbol{y}^{(2)})$$

$$\boldsymbol{y}^{(1)} = \mathcal{F}_3(\mathbf{X}^{(1)})\mathbf{b}_3 + \varepsilon \rightarrow \boldsymbol{r}_3 = \mathrm{corr}(\hat{\boldsymbol{y}}_3^{(2)}, \boldsymbol{y}^{(2)})$$

$$\vdots$$

$$\boldsymbol{y}^{(1)} = \mathcal{F}_{24}(\mathbf{X}^{(1)})\mathbf{b}_{24} + \varepsilon \rightarrow \boldsymbol{r}_{24} = \mathrm{corr}(\hat{\boldsymbol{y}}_{24}^{(2)}, \boldsymbol{y}^{(2)})$$



Kim et al., In prep.

Results

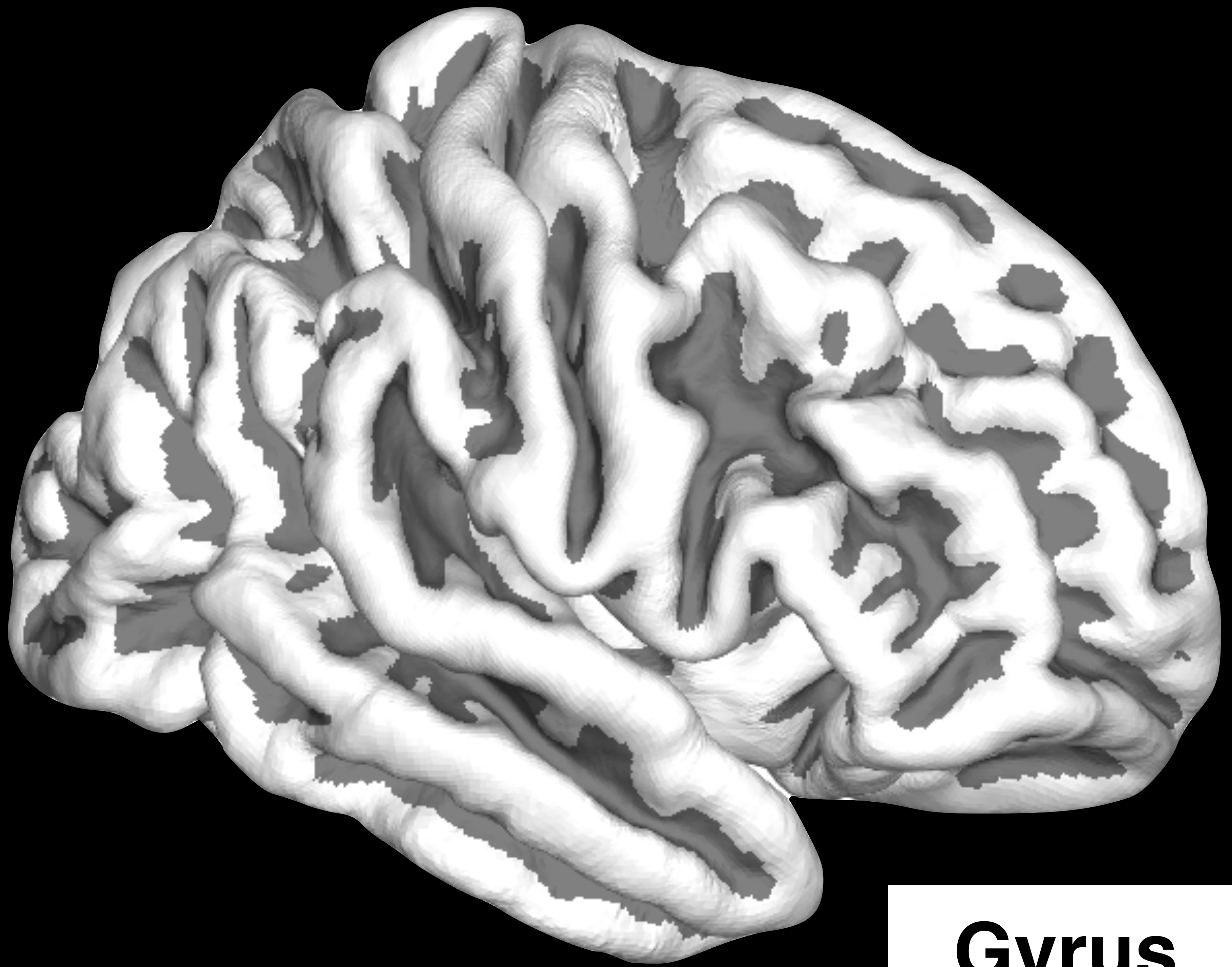
Inflated cortical surface and anatomy

Just to see better the buried parts of the cortex

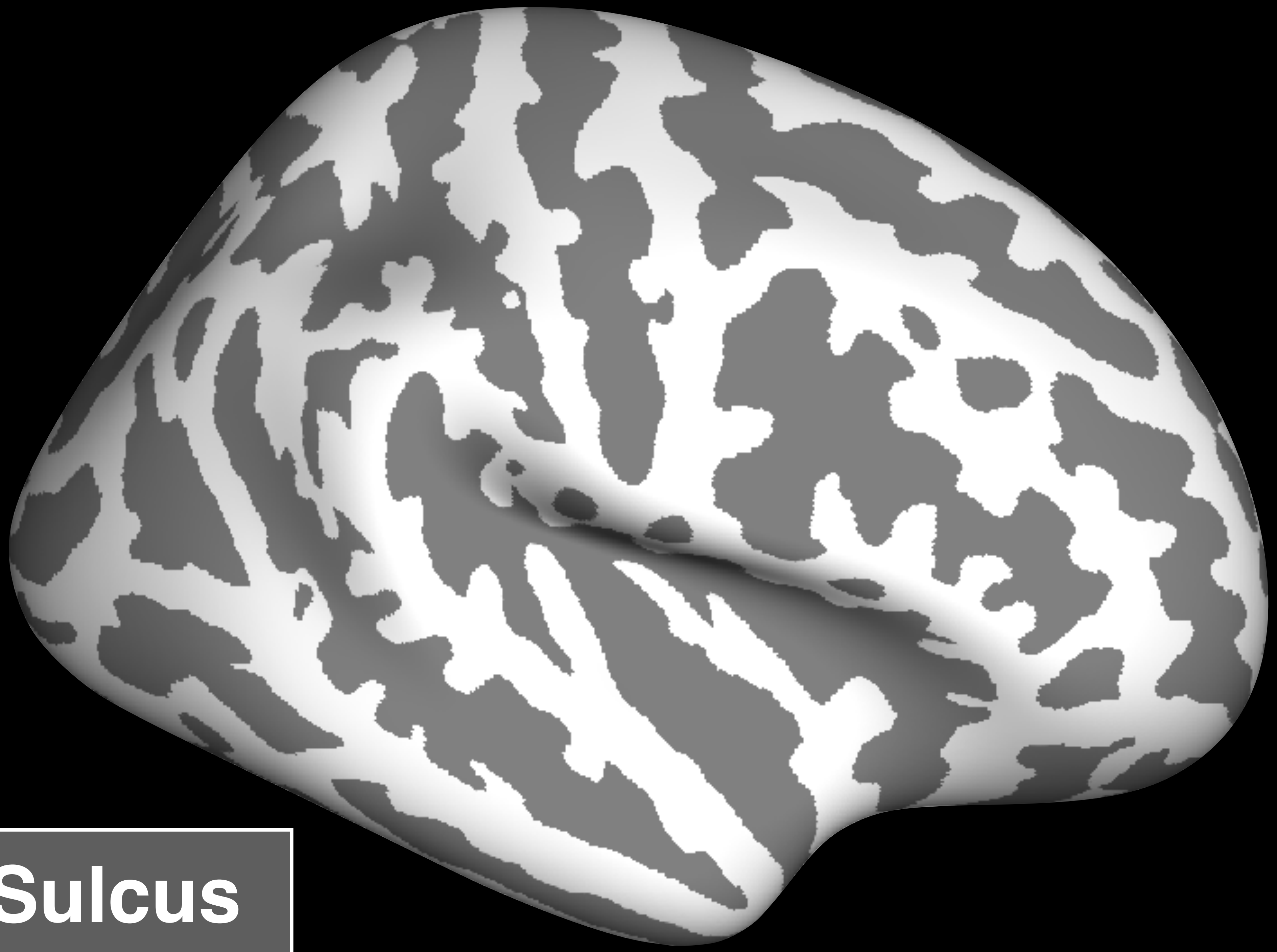
Actual brain

(cc) Pr. M. Forest @pixel.com

"Averaged brain"

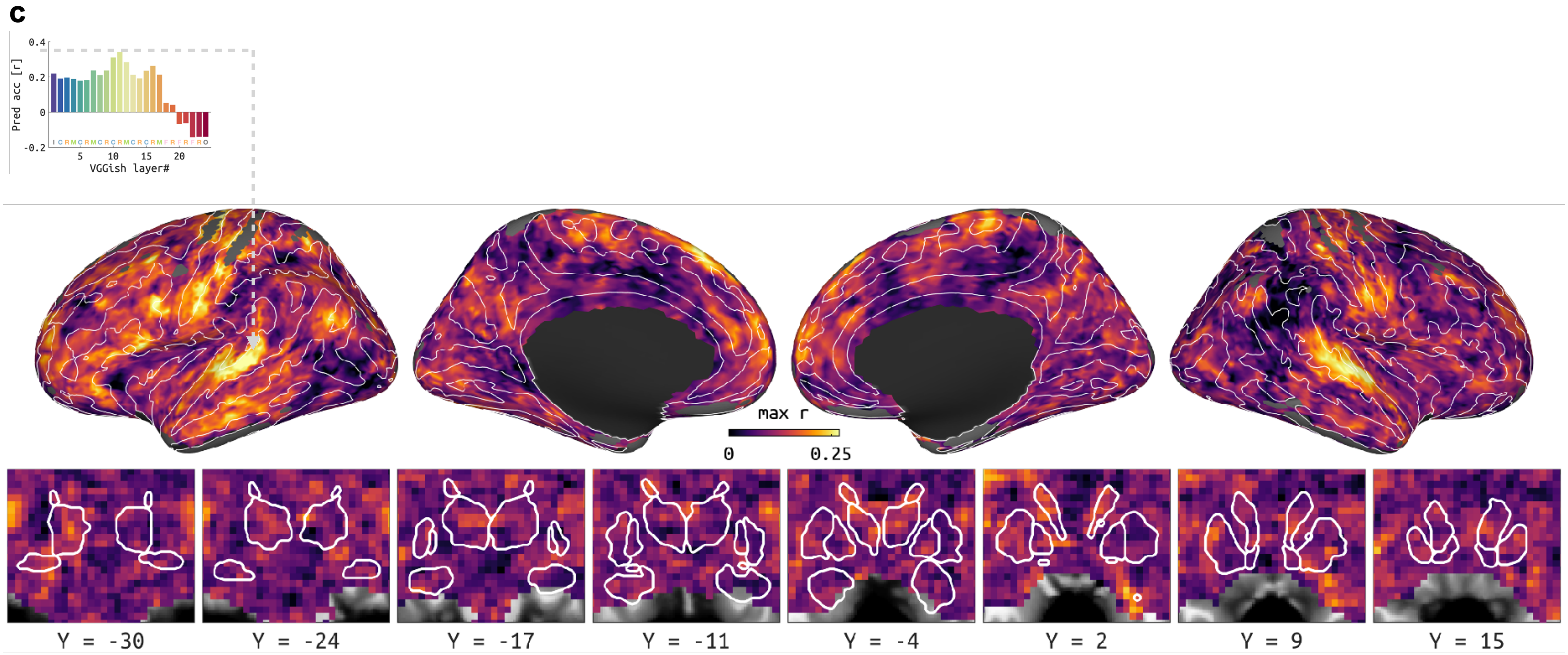


"Inflated brain"

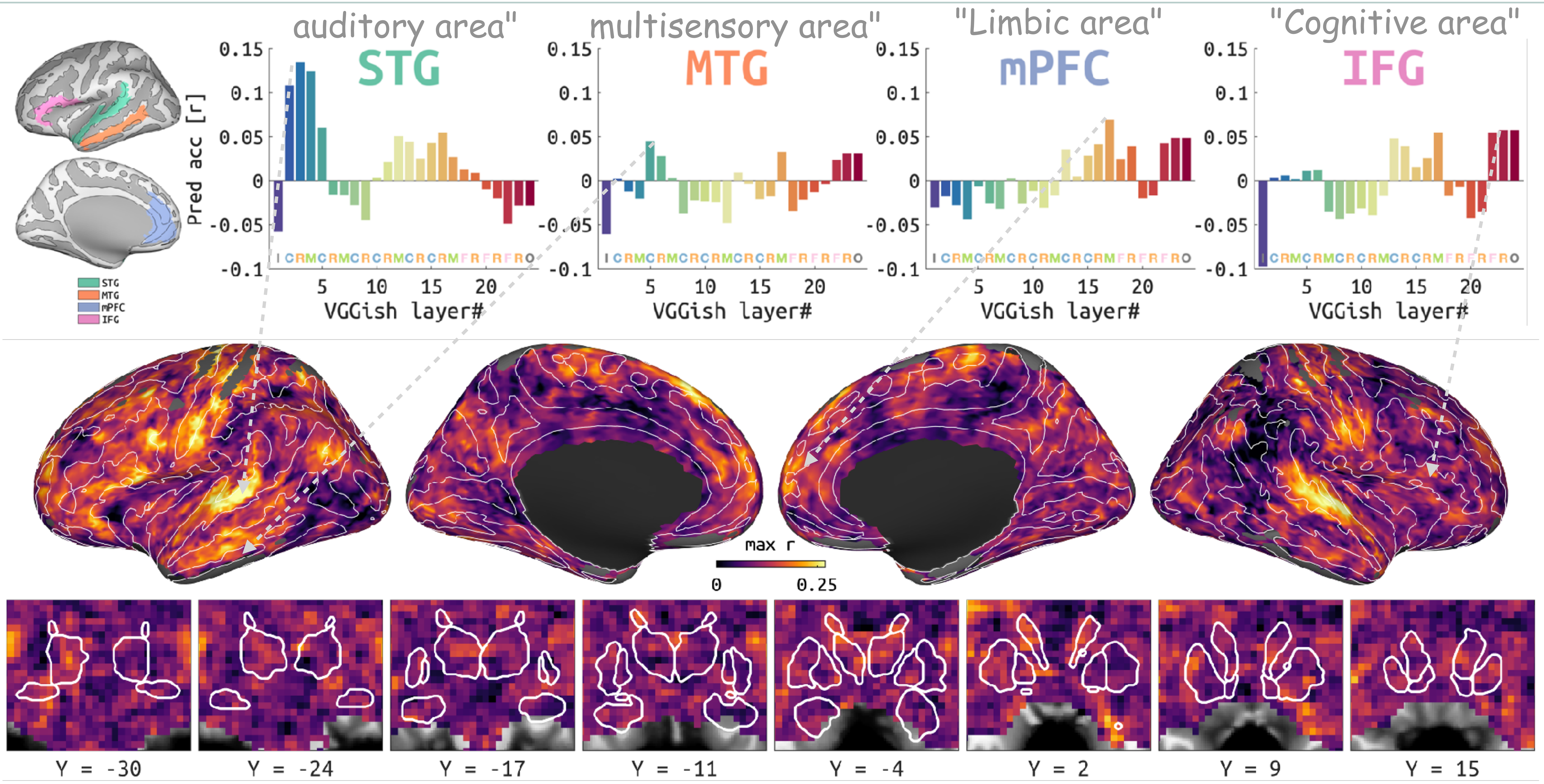


Gyrus | Sulcus

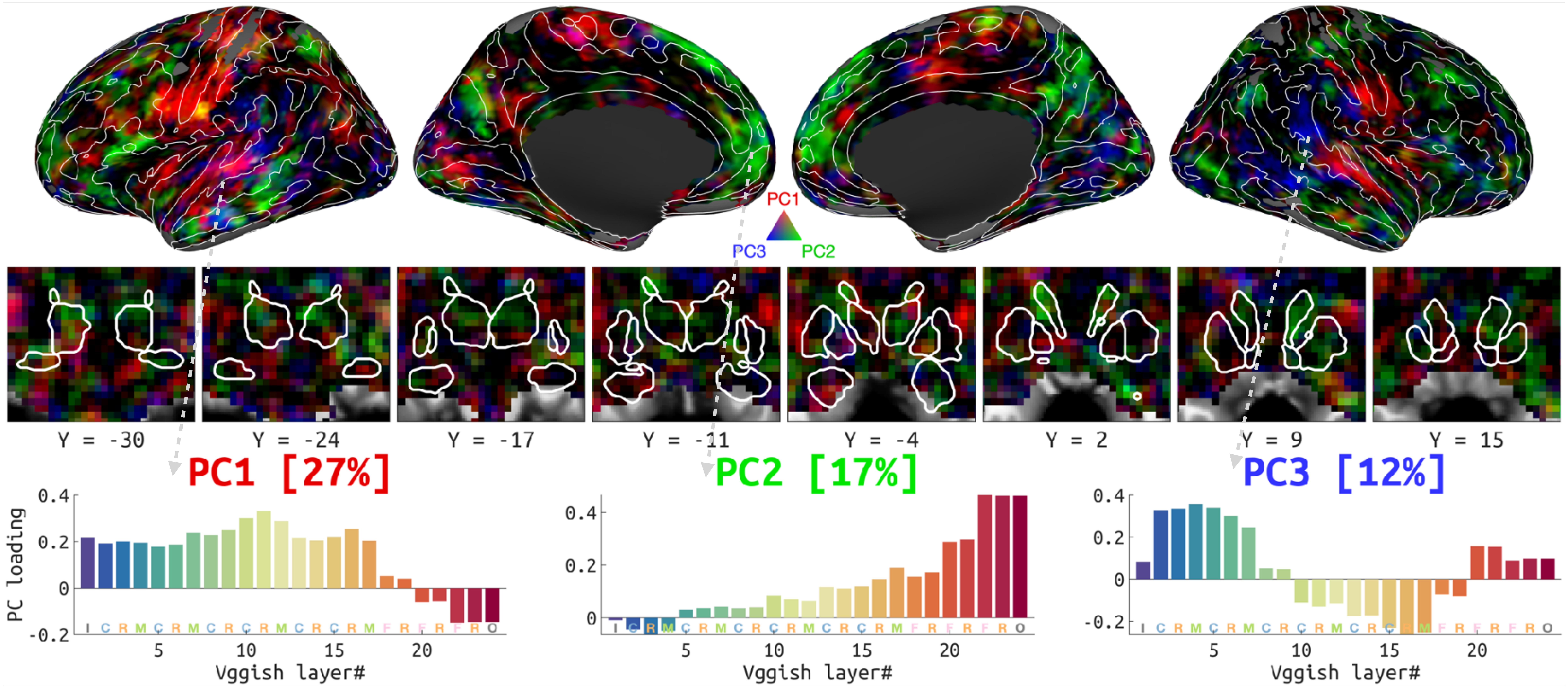
Maximal prediction accuracy 🎶→🧠



*Simplified labels:

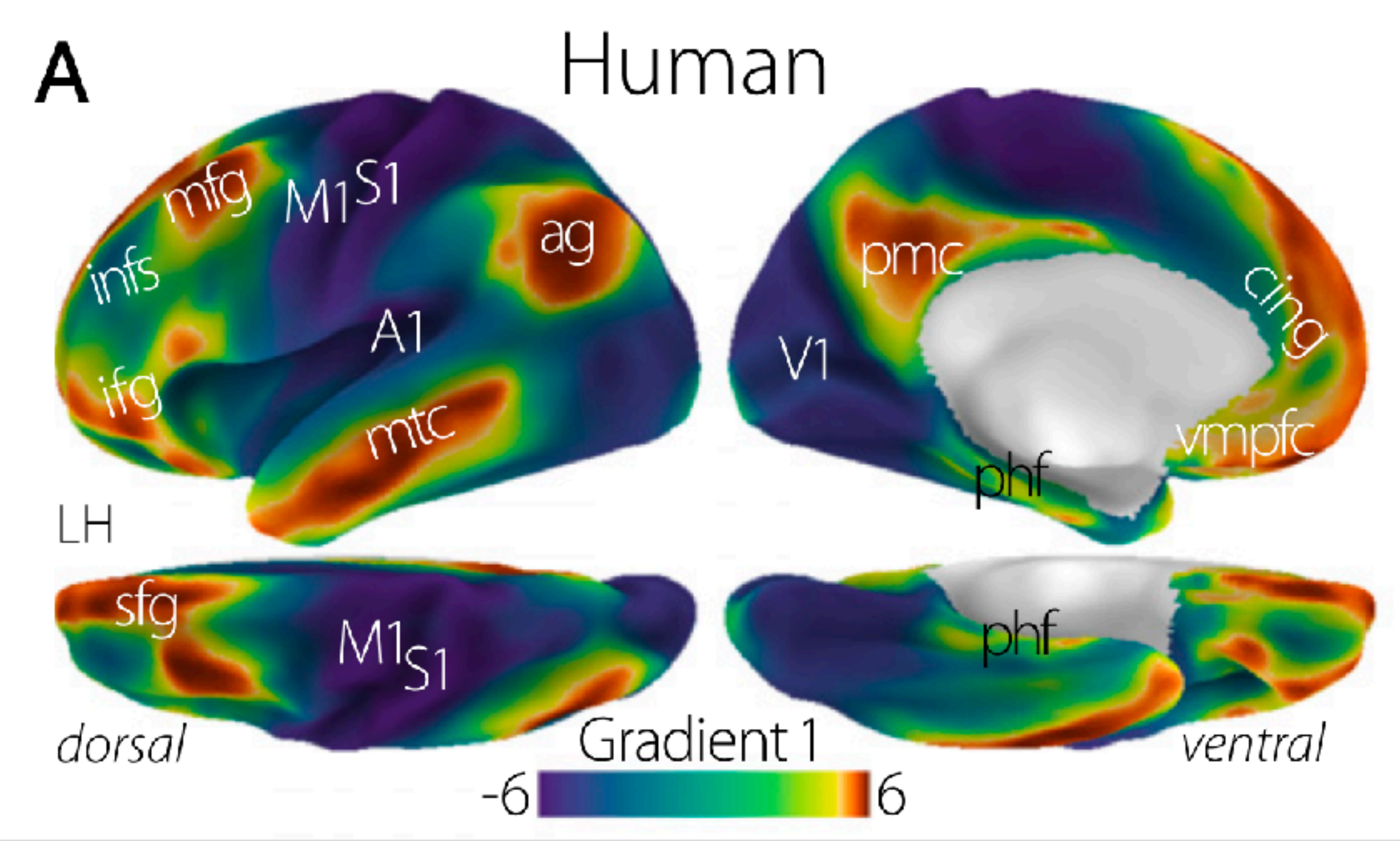


Topography of CNN-layer-specific encoding 🎶→🧠

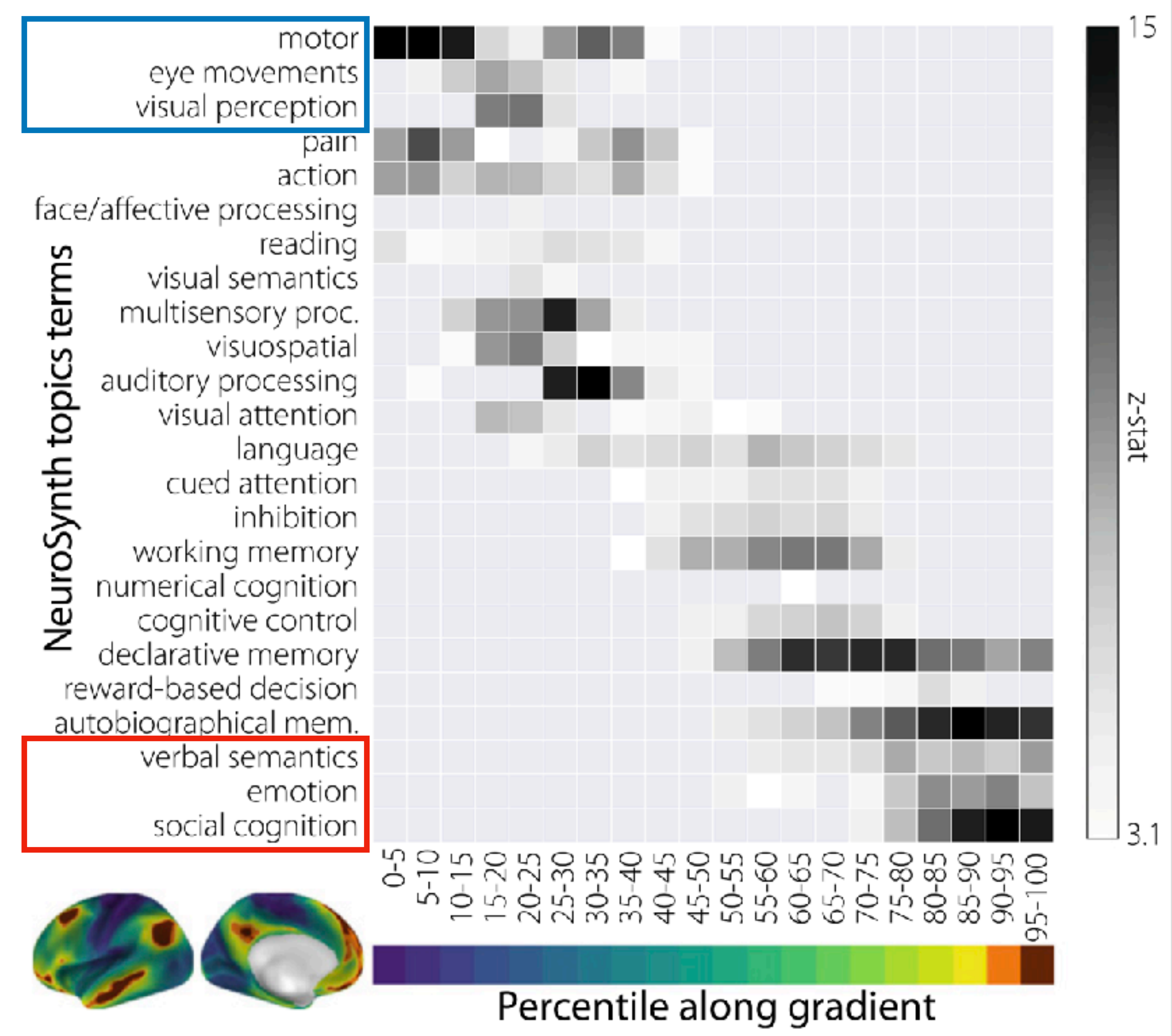


What is already known about the cortical topography of information abstraction?

Margulies et al., 2016, PNAS

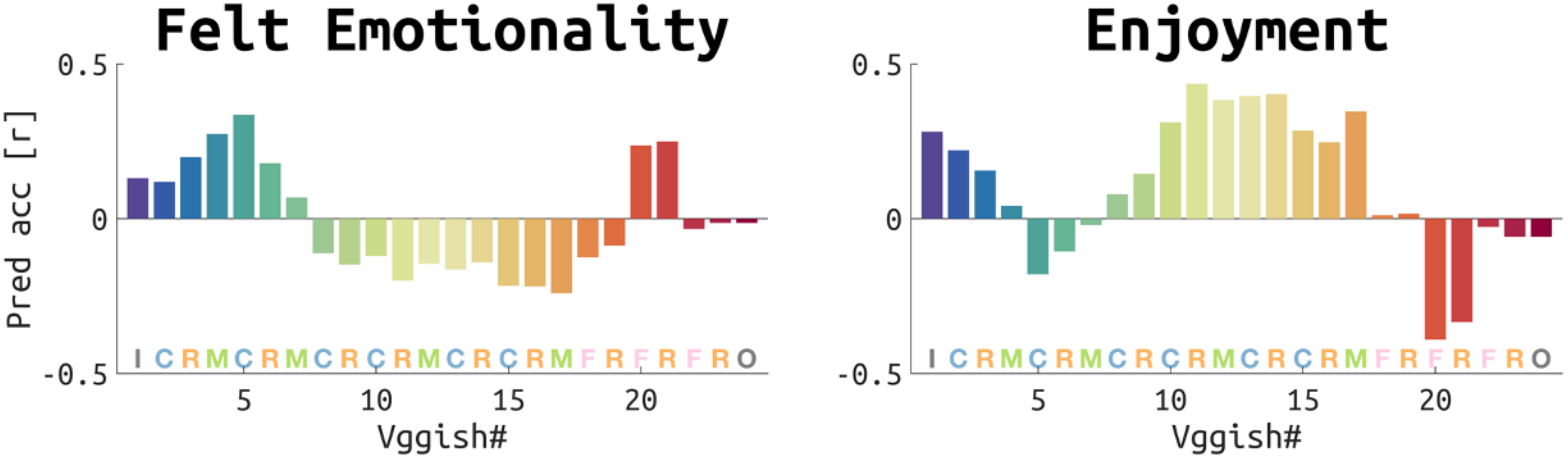


Topography found from resting-state functional connectivity



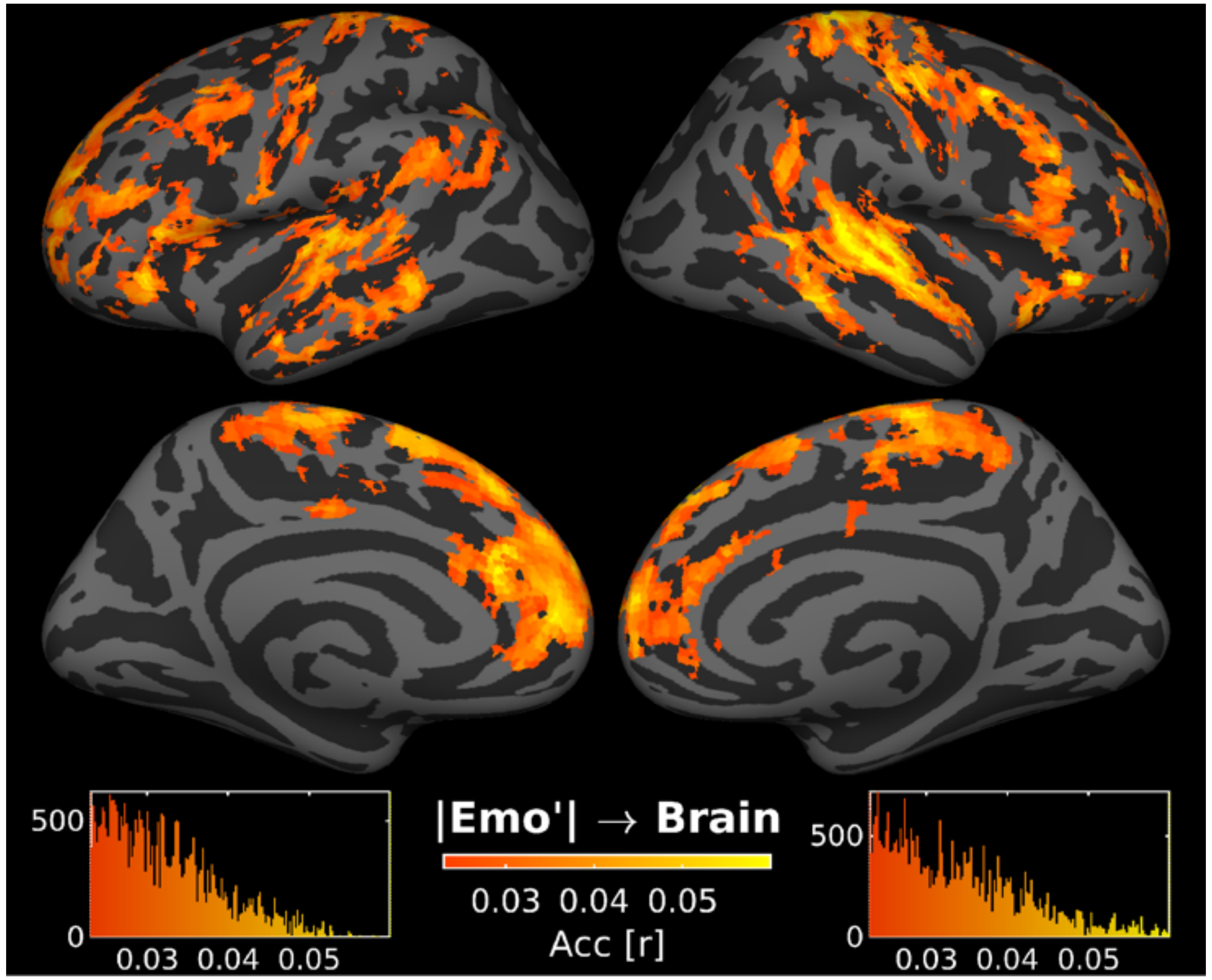
vs. the known topography of abstraction

Distinctive patterns of Emotionality and Enjoyment 😀😥→🧠

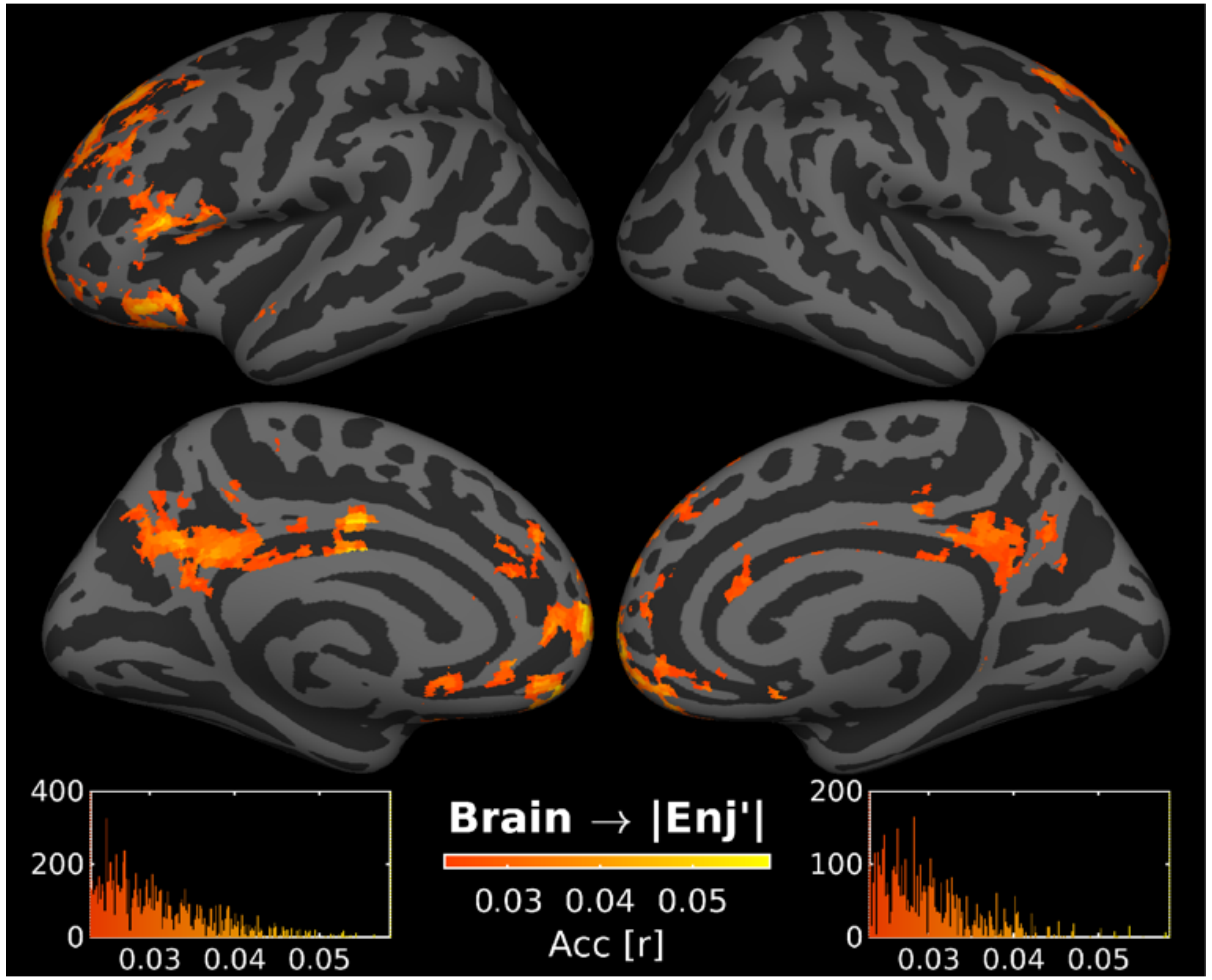


Brain-emotion correlation 😀😥↔🧠

fMRI = |Emotionality’| + |Enjoyment’| + error



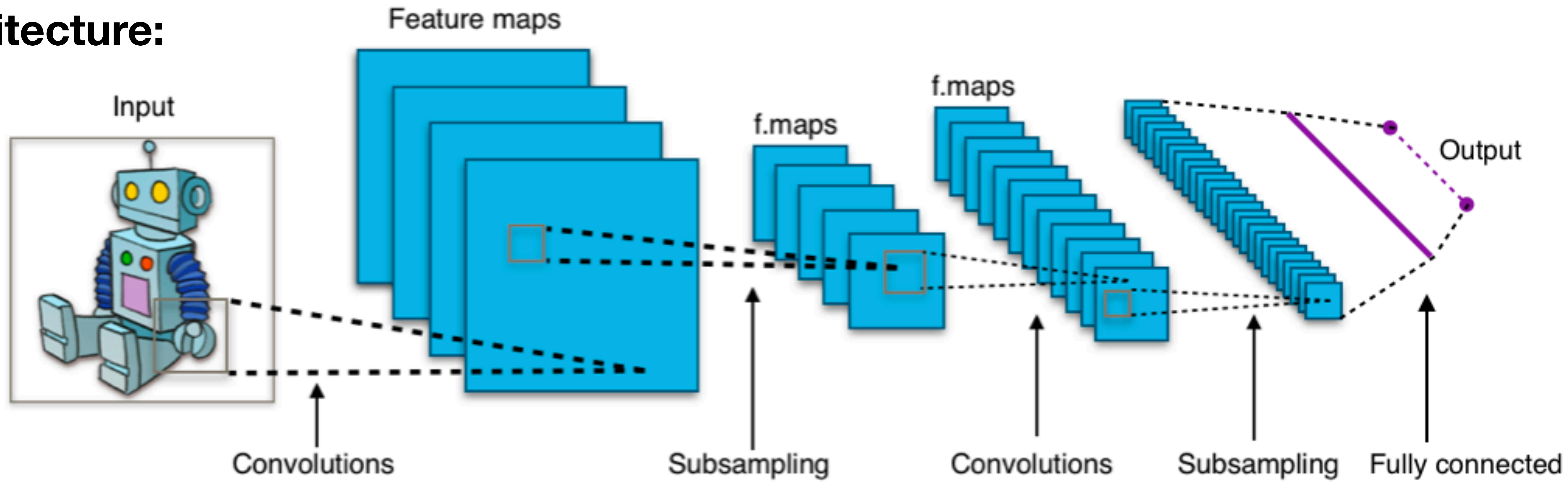
Kim et al., *In prep*.



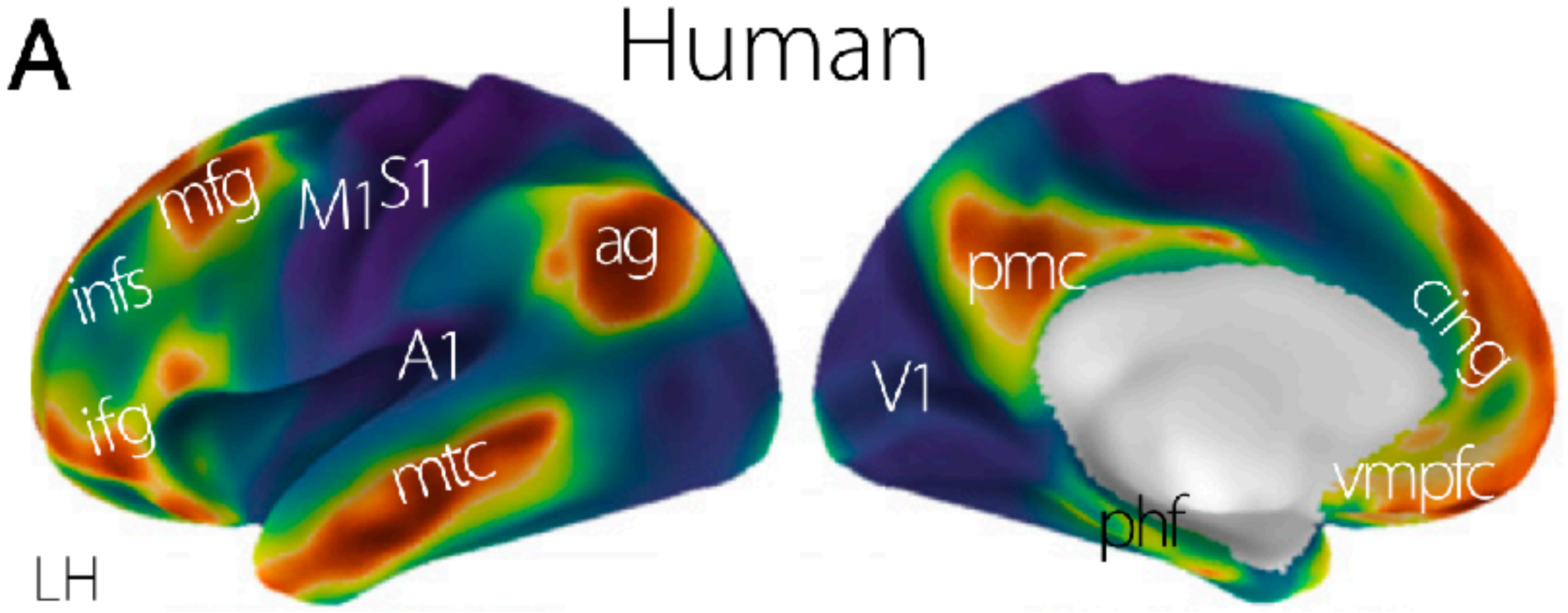
Discussion

Abstraction of sounds

A CNN architecture:

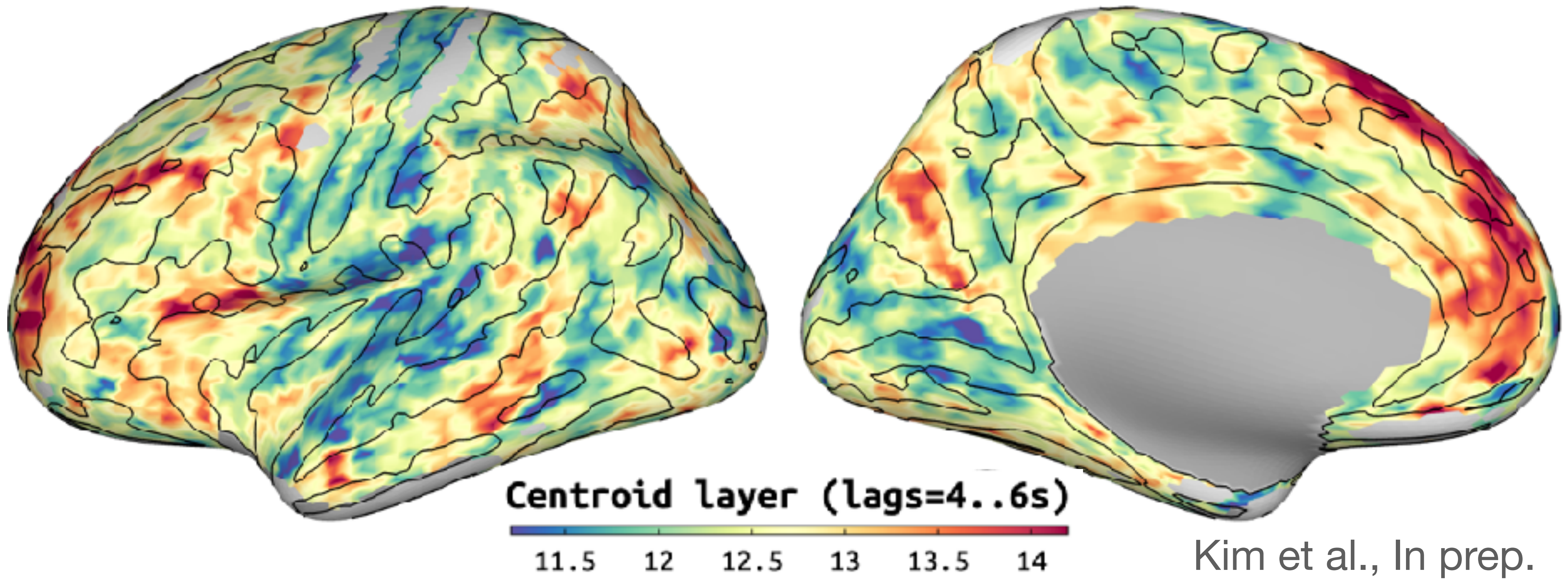


Abstraction in the human cerebral cortex:



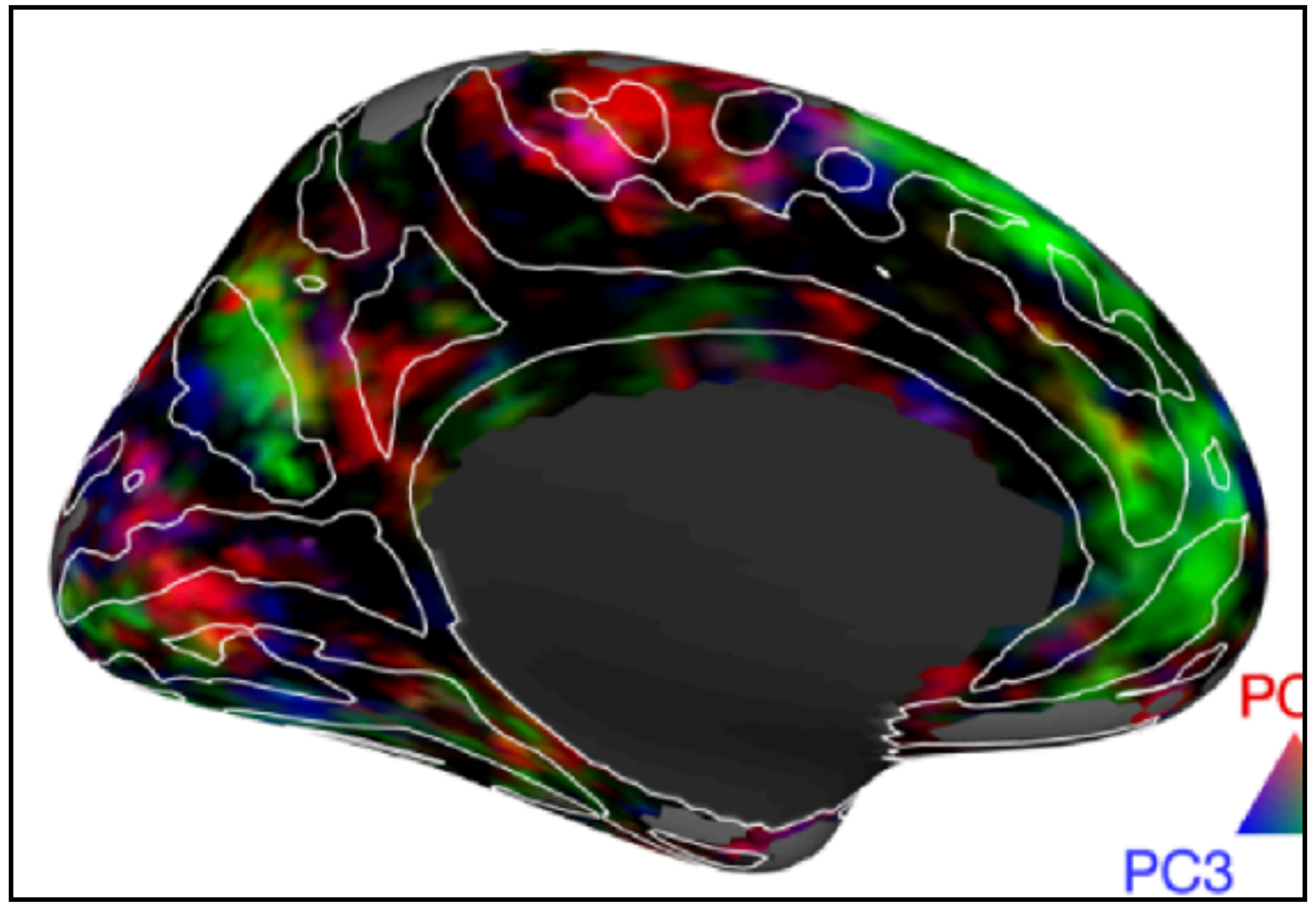
Marguleius et al., 2016, PNAS

Musical abstraction in the human cerebral cortex:

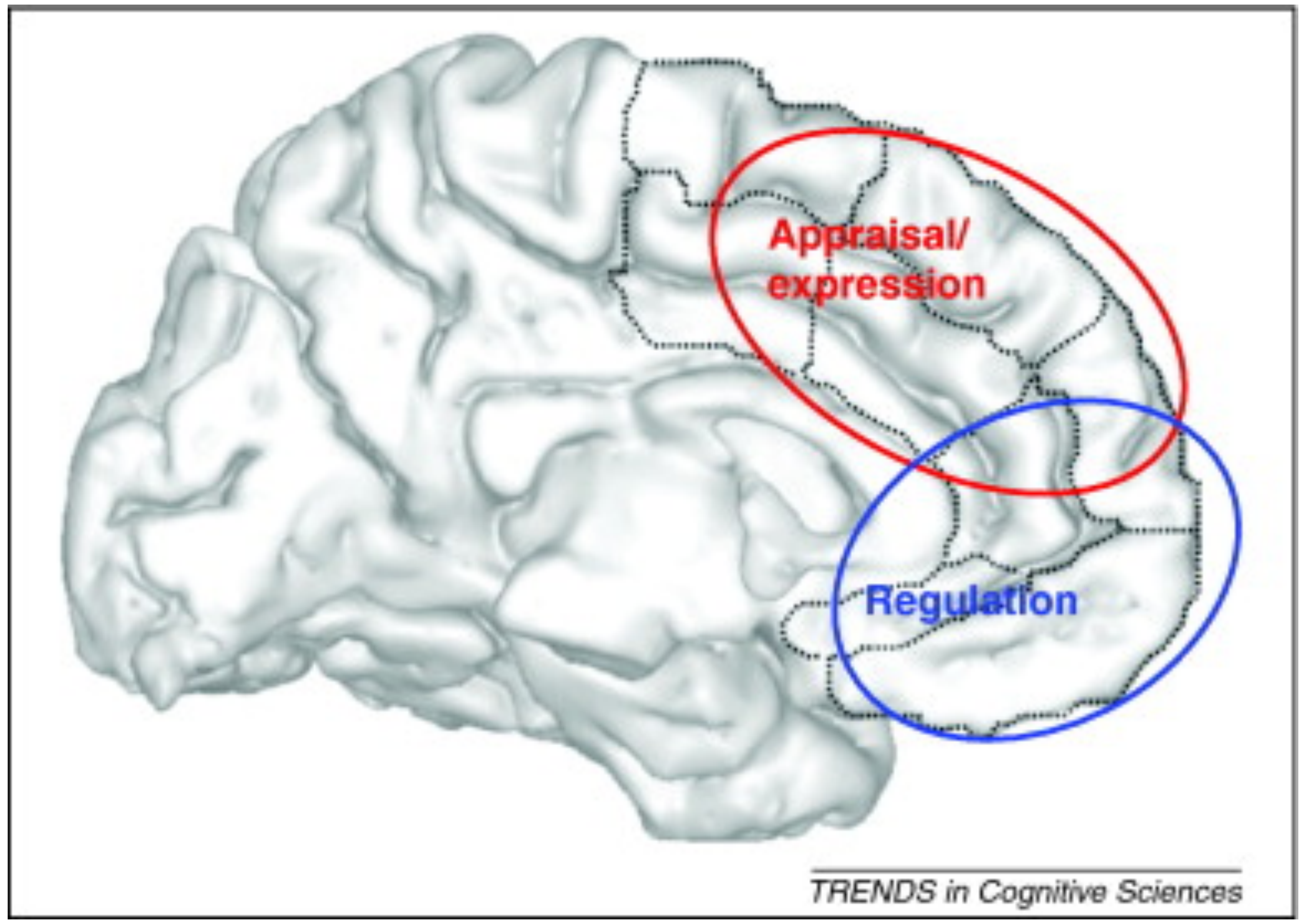


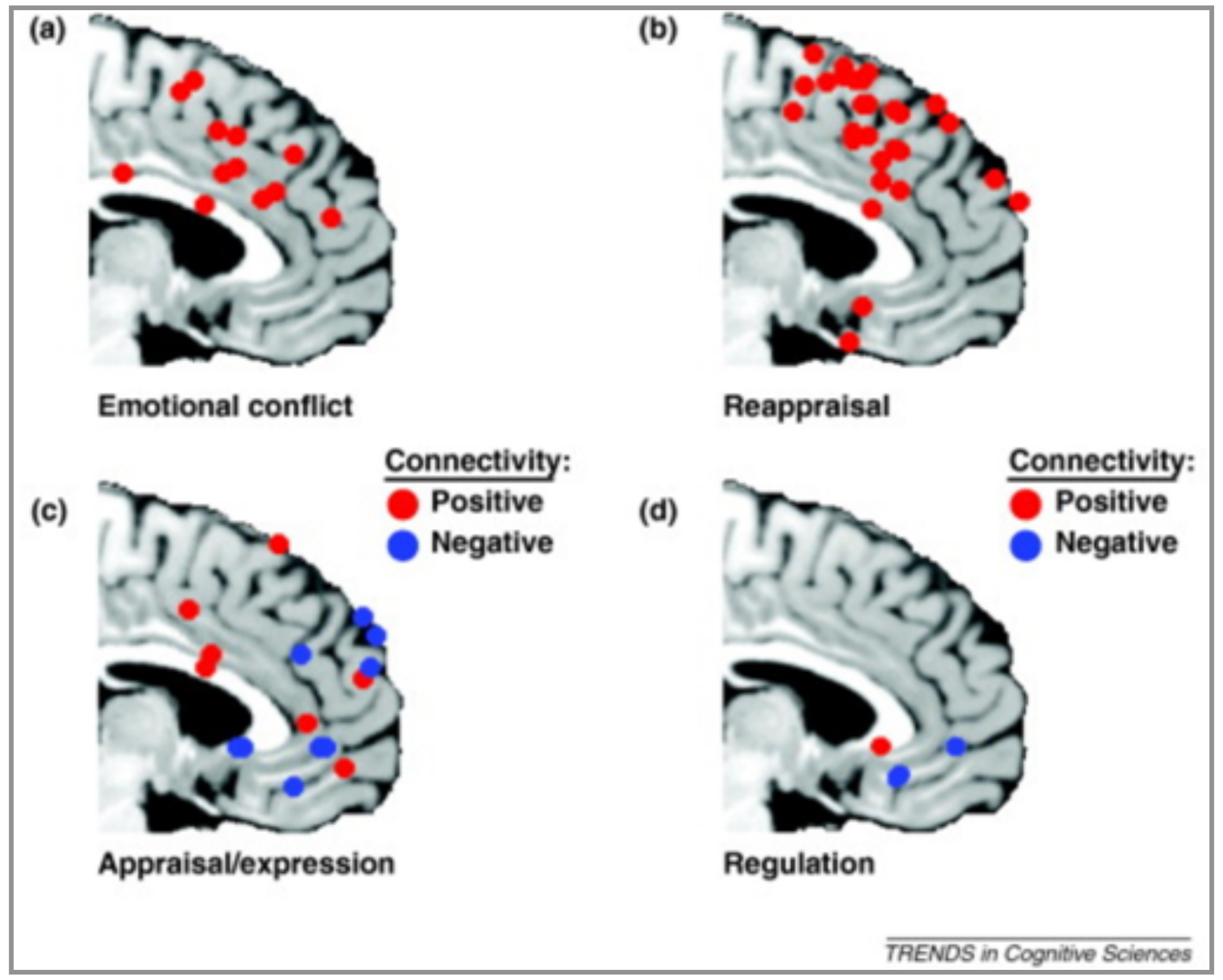
Kim et al., In prep.

mPFC and emotional processing



Kim et al., In prep.

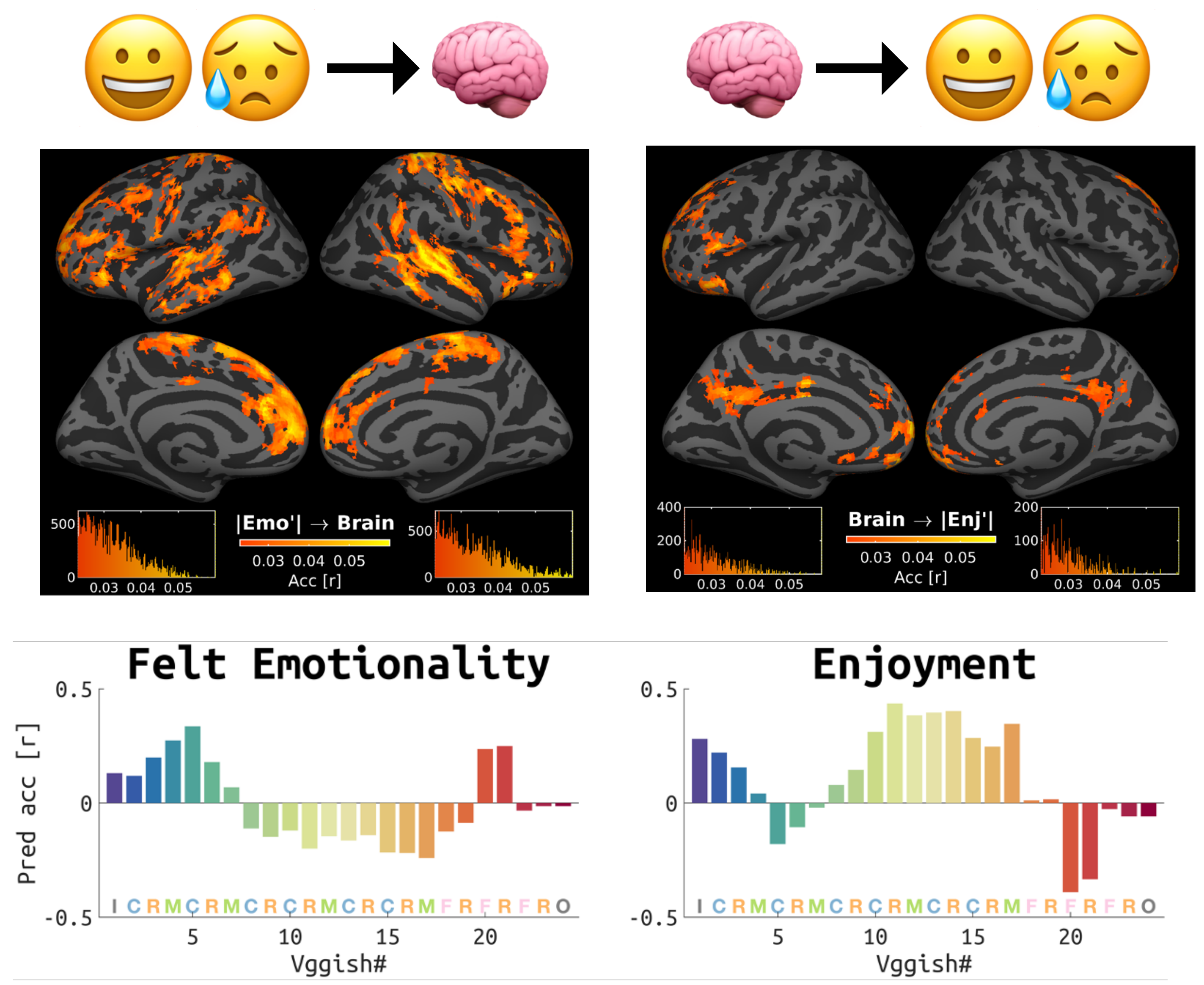




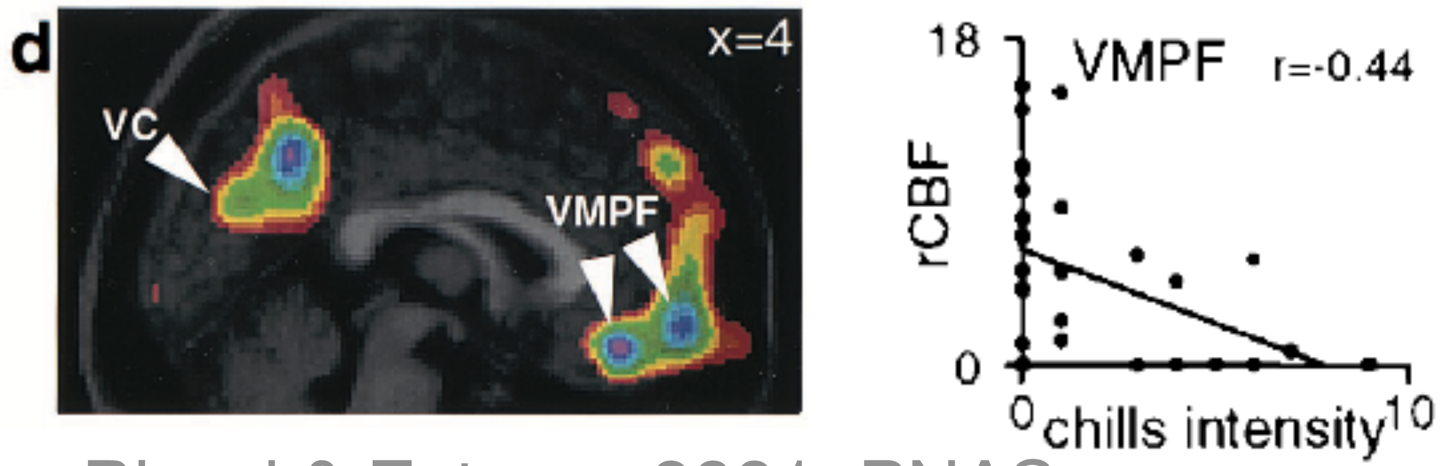
Etkin, Egner, Kalisch, 2010, *Trends in Cognitive Sciences.*

Audio semantic model changes were encoded in the mPFC, which showed a sensitivity to musical structures ("boundaries").

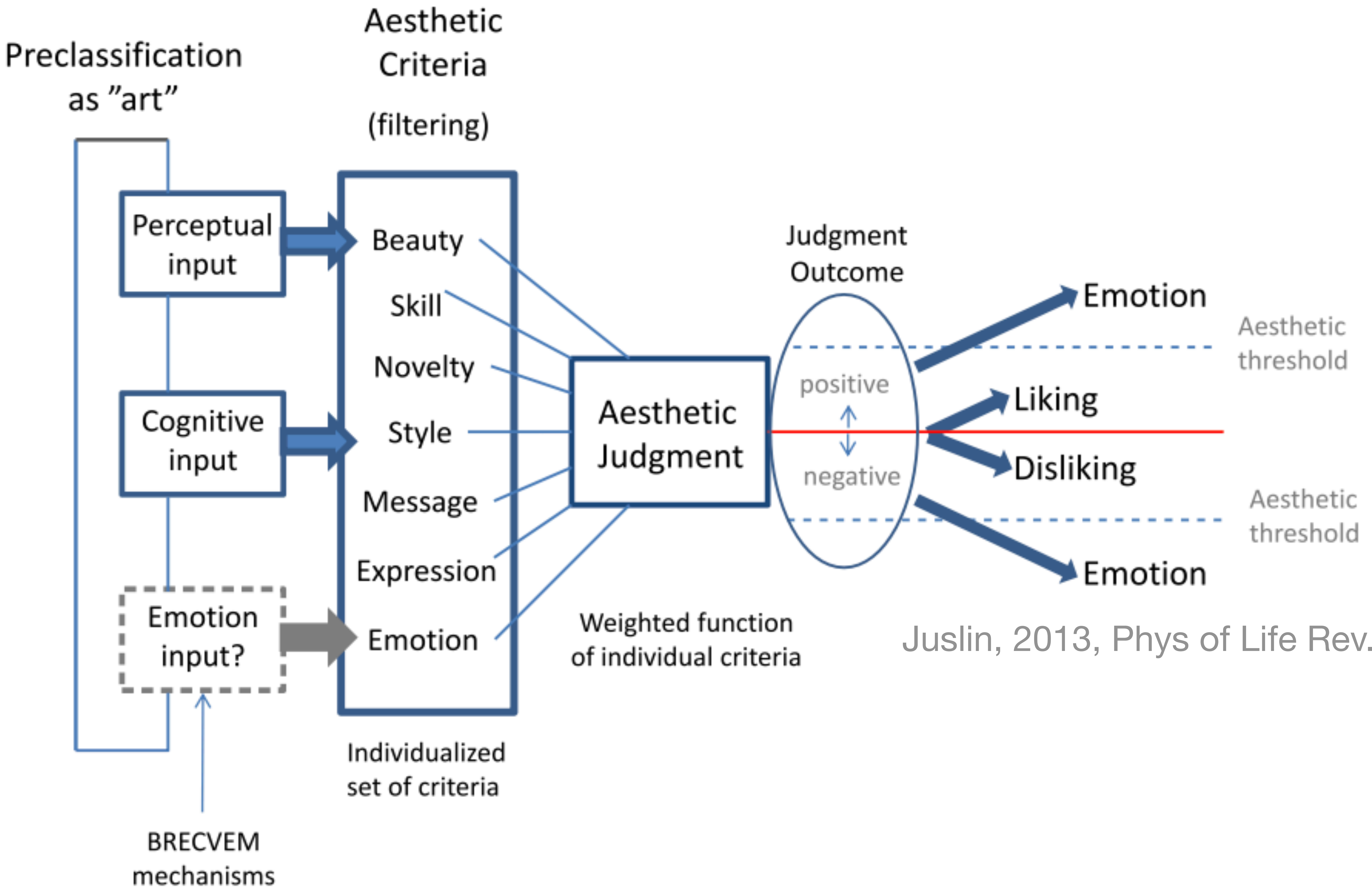
Different encoding of emotionality & enjoyment



Kim et al., In prep.



Blood & Zatorre, 2001, PNAS



Juslin, 2013, Phys of Life Rev.

vmPFC activity was followed by Enjoyment rating changes.

Conclusions

- **Naturalistic stimuli** with computational models allow ecologically valid investigation of evoked emotions.
- CNN embeddings are sensitive to information that is relevant for emotional responses, **beyond low-level audio features.** In particular, the abstraction in CNN shows high similarity to **the known cortical topography of information abstraction**, as well as relevance to behavioral ratings of emotional responses.
- Two continuous ratings (*Emotionality* and *Enjoyment*) were differentially encoded in the brain and predicted by different CNN layers, potentially reflecting **distinct mechanisms** of ***felt* emotions** and **aesthetic judgements**.

Thank you for your attention! (and time for discussion🤓)

https://seunggookim.github.io/

Dr. Daniela Sammler
MPI-EA, Frankfurt,
Germany

Dr. Tobias Overath
Duke University,
NC, USA

Dr. Tom H. Fritz
MPI-CBS, Leipzig,
Germany

https://www.aesthetics.mpg.de/en/research/research-group-neurocognition-of-music-and-language.html