ICMPC17-APSCOM7, Tokyo, Japan | August 26, 2023

Emotion-relevant Representations of Music Extracted by Convolutional Neural Networks Are Encoded in Medial Prefrontal Cortex

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

How does music evoke emotions *via* the brain?

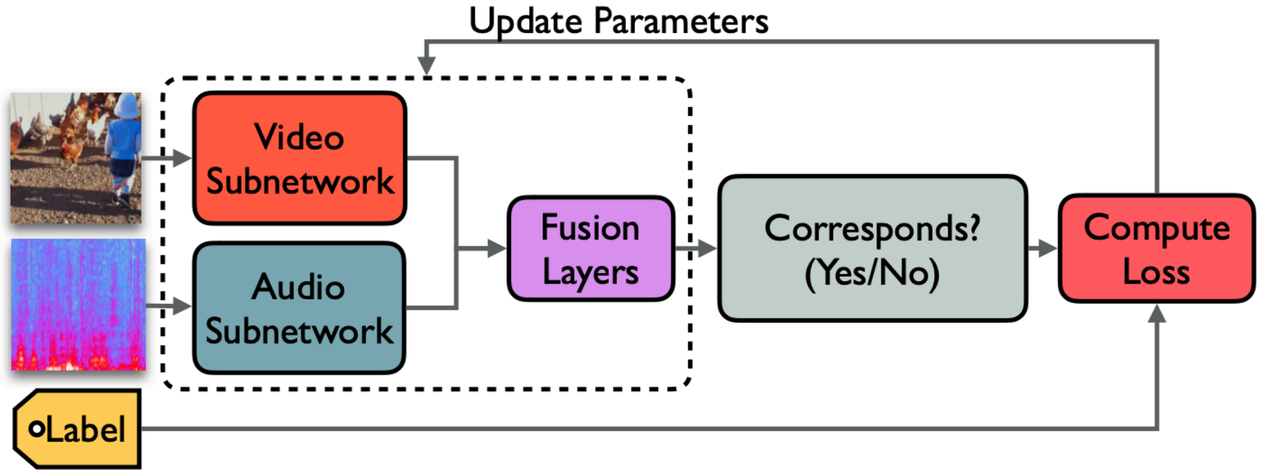
Neuroscientific view

Music → ? → Brain → ? → Emotion

How is auditory information of music transformed to contribute to the emergence of emotions?

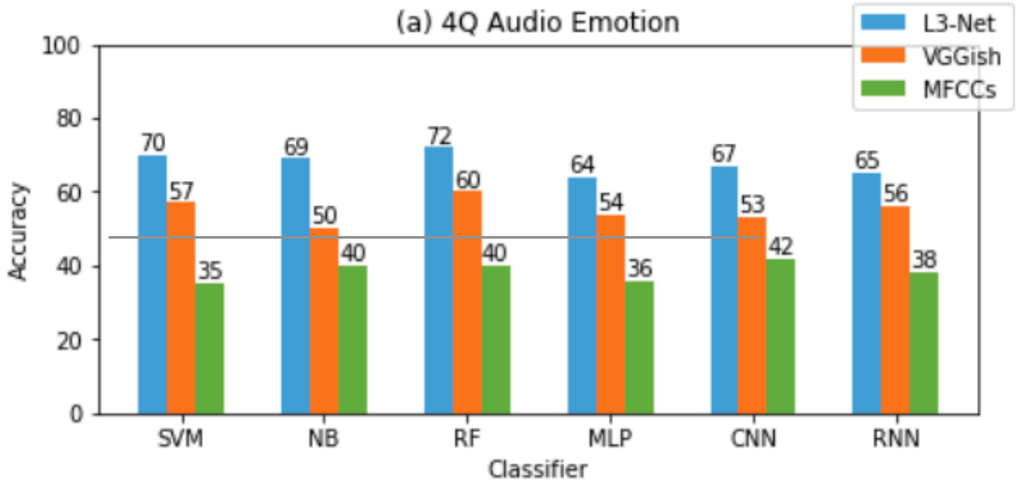
CNN embedding for music emotion *recognition*

Potentially mid/high-level representation of music signal



Open-L3: Cramer et al., 2019, *ICASSP*

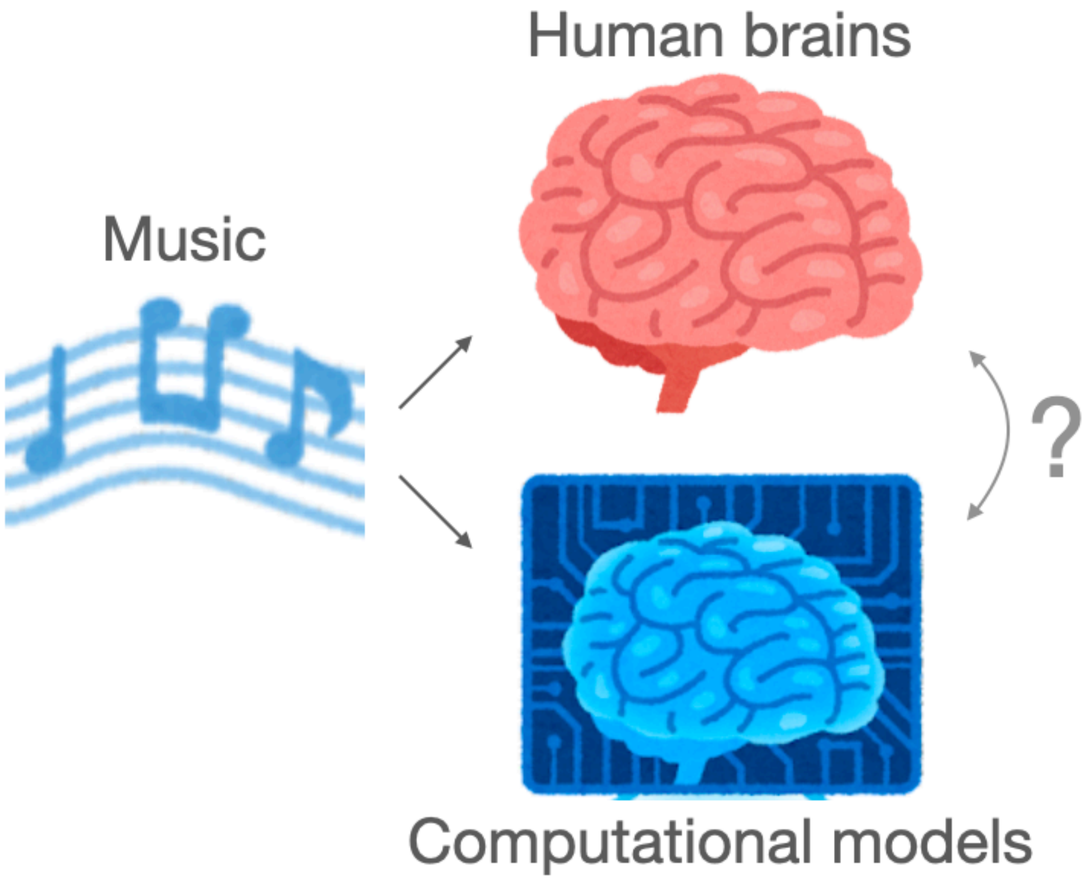
4Q Audio Emotion Dataset: 255 music clips (30 s) for **Arousal-Valence quadrants**



Koh & Dubnov, 2021, *ACA*

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

Research Questions



- **Q1**: Are the embeddings of pre-trained **CNN models** predictive of ***felt* emotions** and **neural activity**, in comparison to low-level audio features?
- **Q2**: How are ***felt* emotions** and **musical enjoyment** associated with neural activity over time?

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

Methods

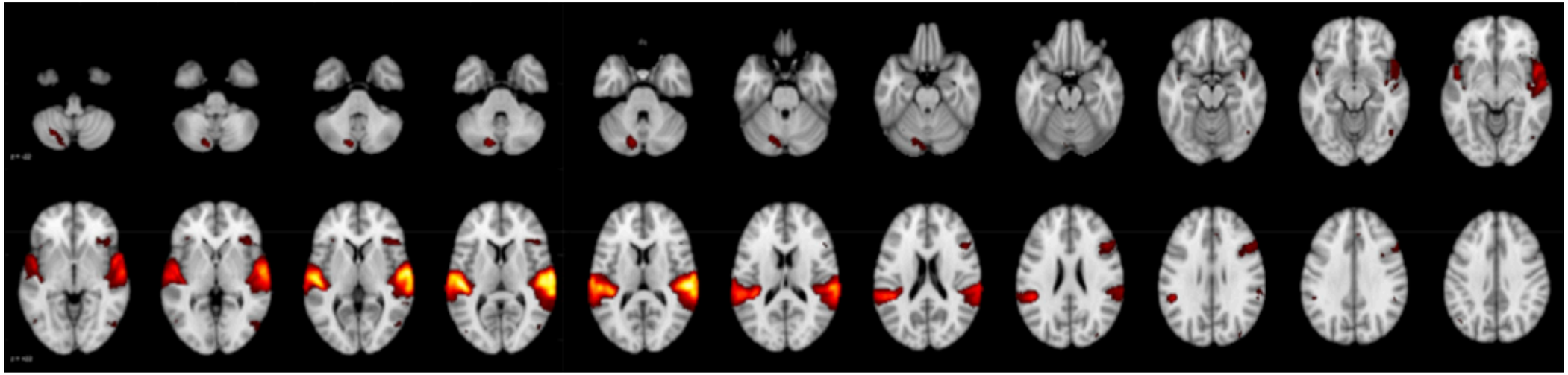
Original study

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

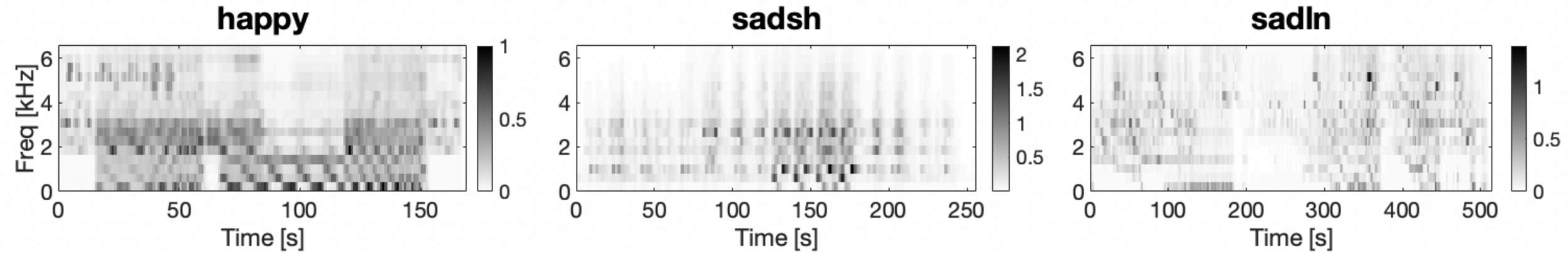


Sachs et al., 2020, *NeuroImage.*

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

Music

{Envelope, |Envelope'|, |VGGish'|, |OpenL3'| }

Encoding

Brain

{fMRI time-series}

Encoding

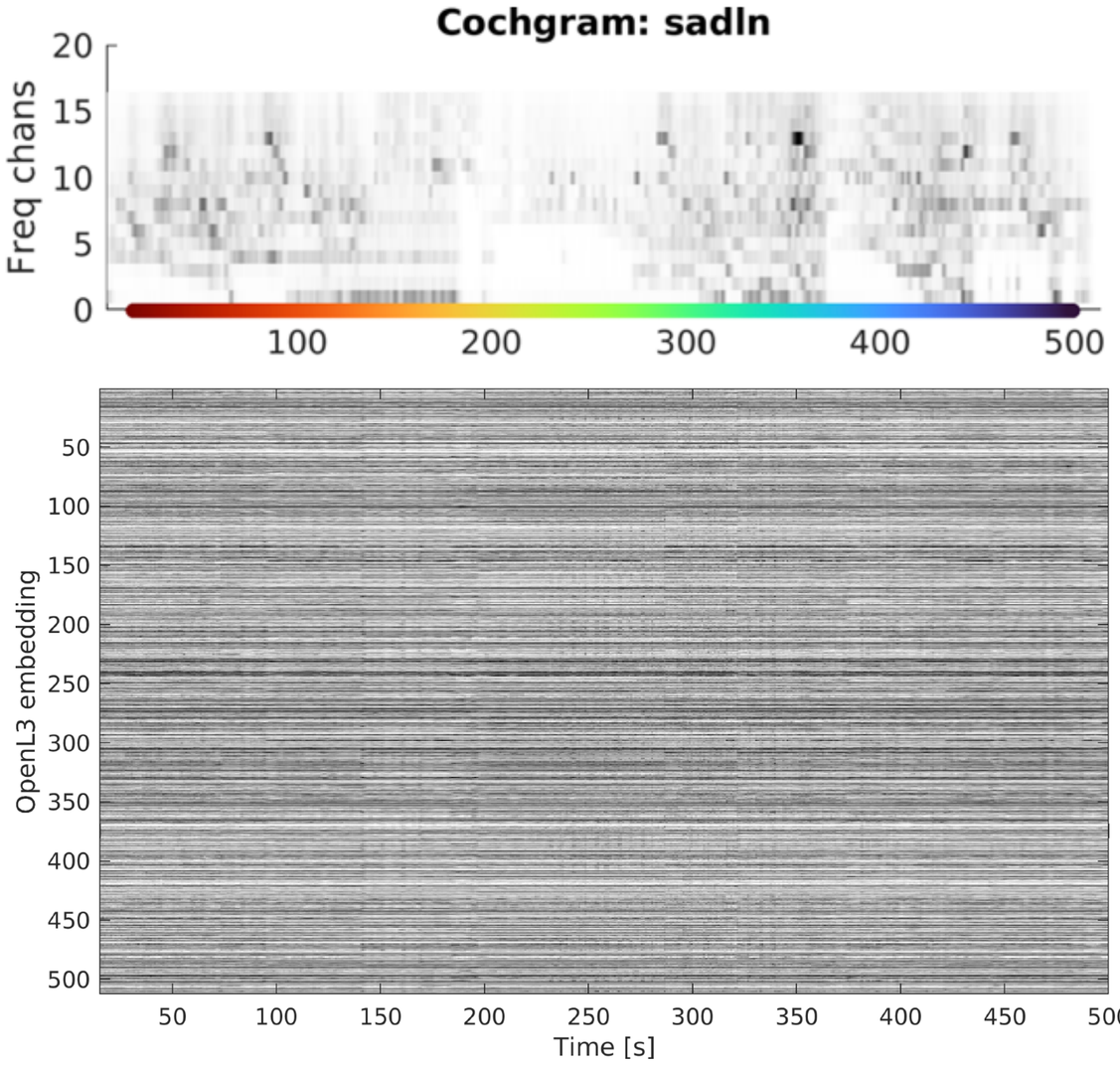
Emotion

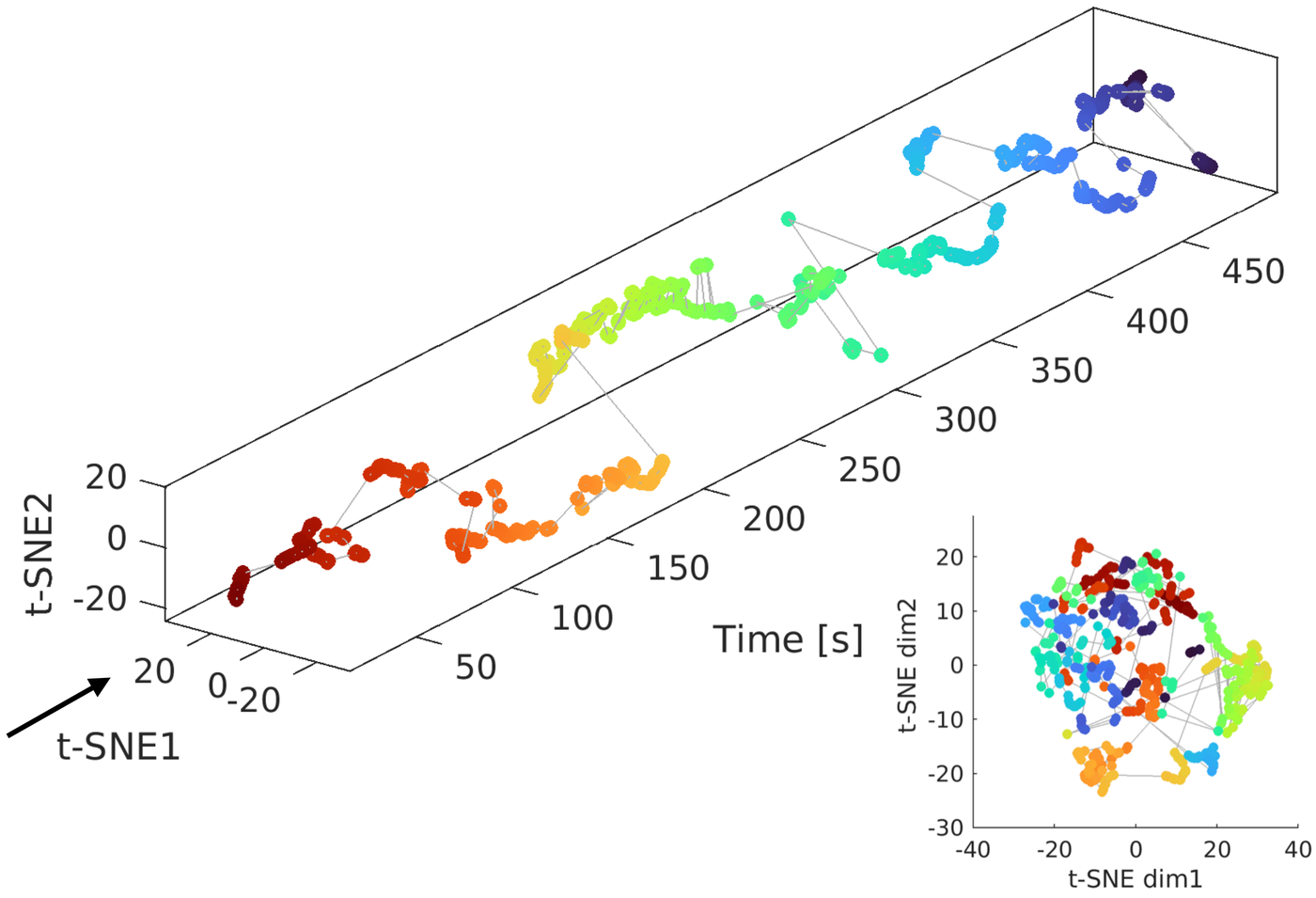
{|Emotionality'|, |Enjoyment'|}

max(Cross-correlation)

CNN representation of music

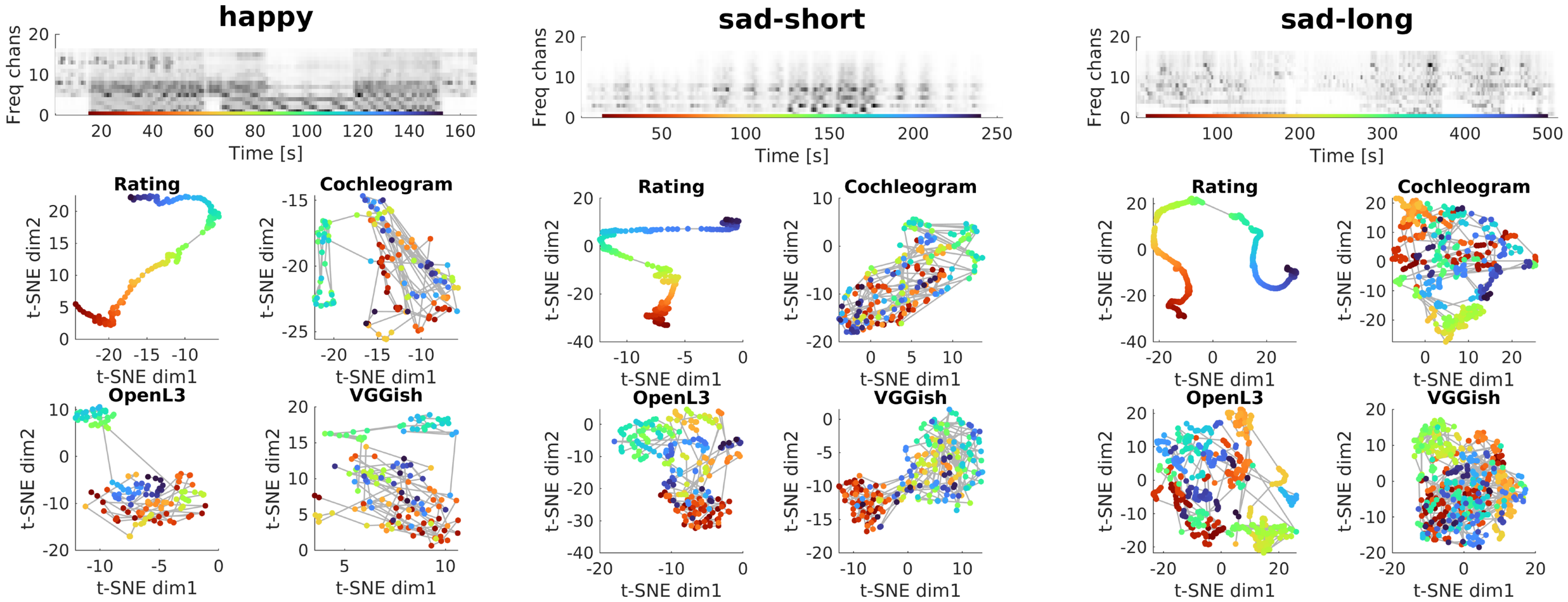
t-SNE trajectory





Emotional ratings vs. Audio features

t-SNE trajectories

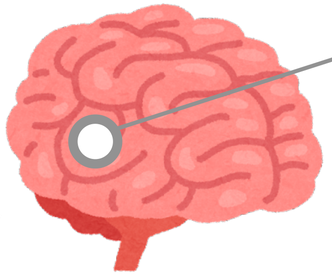


Kim et al., *In prep*.

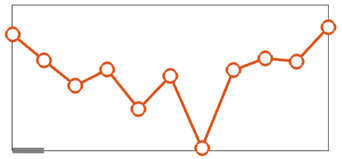
Encoding analysis

Training set

Neural responses



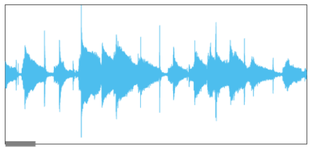
$\mathbf{y}_1$



1 s

Music #1

$\mathbf{X}_1$



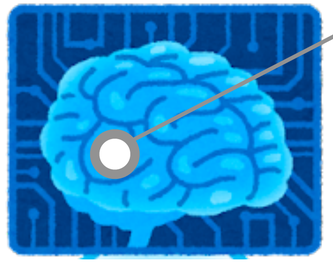
1 s

Weight fitting: $\hat{\mathbf{w}}_1 = \mathbf{Z}_1^{+}\mathbf{y}_1$

$\mathbf{y}_1 = \mathbf{Z}_1\mathbf{w}_1 + \epsilon$

Delays

$\mathbf{z}_1 = \mathscr{J}'(\mathbf{X}_1)$

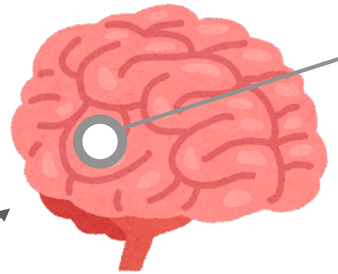


1 s

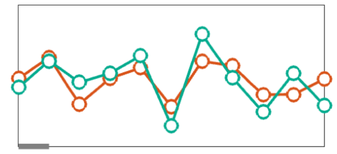
Model features

Test set

Neural responses



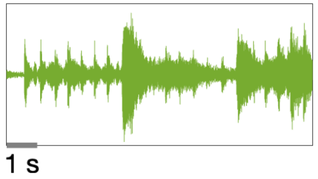
$\mathbf{y}_2$



1 s

Music #2

$\mathbf{X}_2$

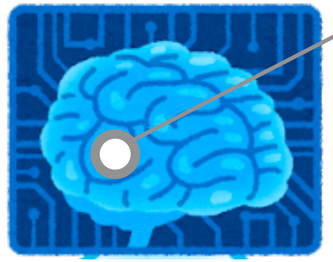


1 s

Prediction: $\hat{\mathbf{y}}_2 = \mathbf{Z}_2\hat{\mathbf{w}}_1$

Delays

$\mathbf{z}_2 = \mathscr{J}'(\mathbf{X}_2)$



1 s

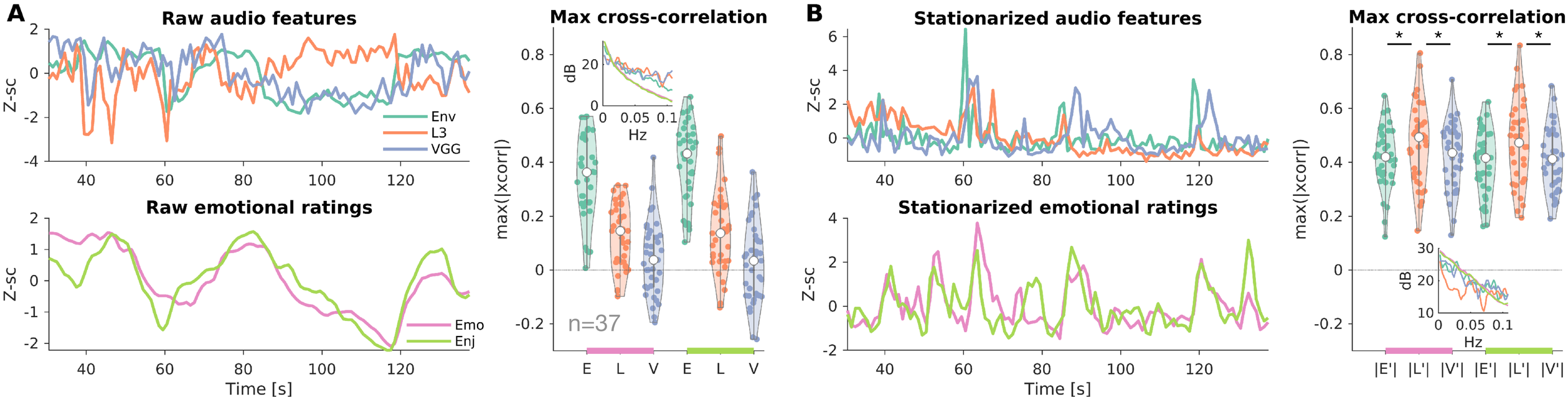
Model features

Kim, 2022. *Front. Neurosci.*

Results

Audio-emotion correlation 🎶→😀😢

Stationarization to match modulation spectra

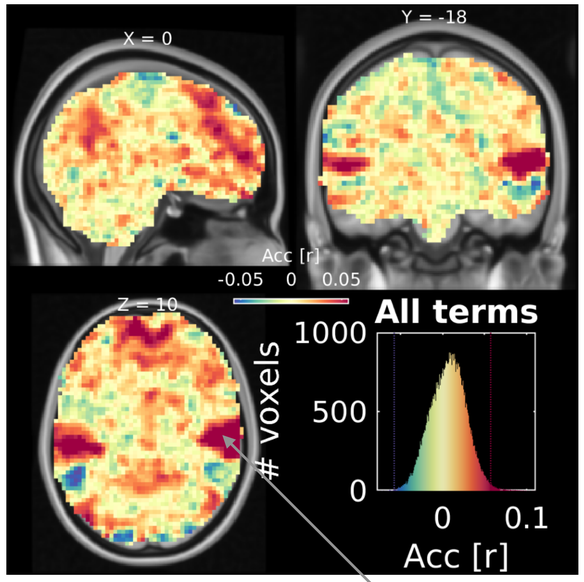


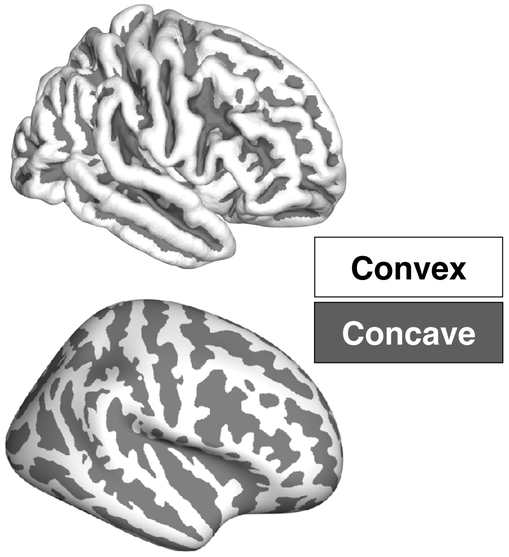
Kim et al., *In prep*.

Changes in CNN embeddings showed higher relevance for changes in emotional ratings than broadband envelope.

Audio-brain correlation 🎶→🧠

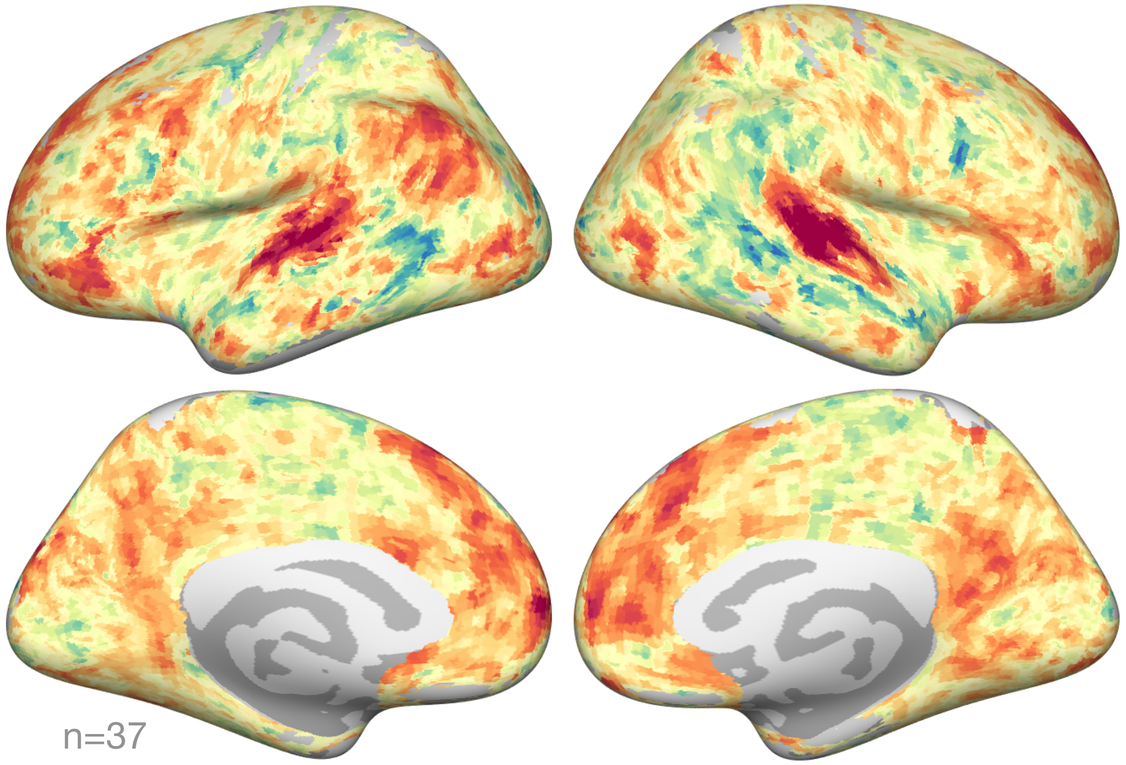
fMRI = Env + |Env'| + |OpenL3'| + |VGGish'| + error



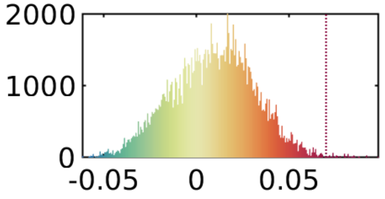


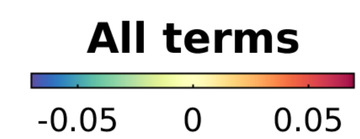
Convex

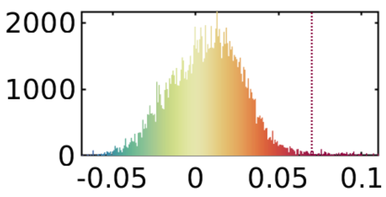
Concave

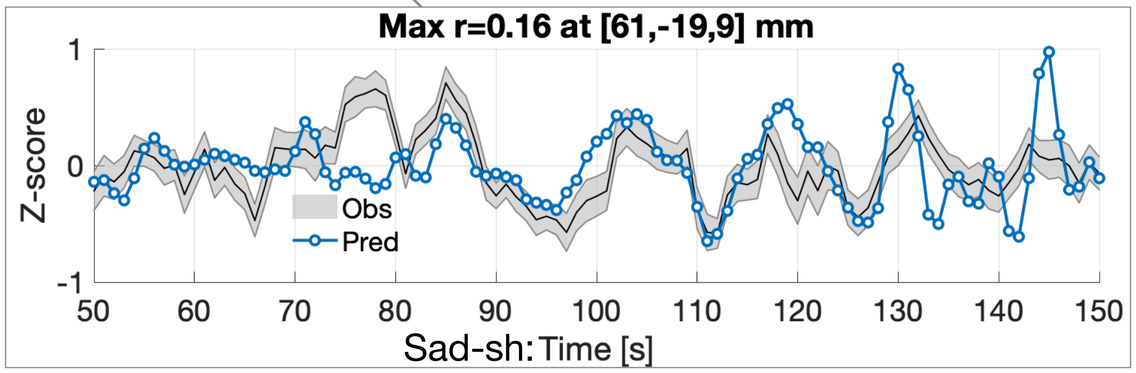


n=37





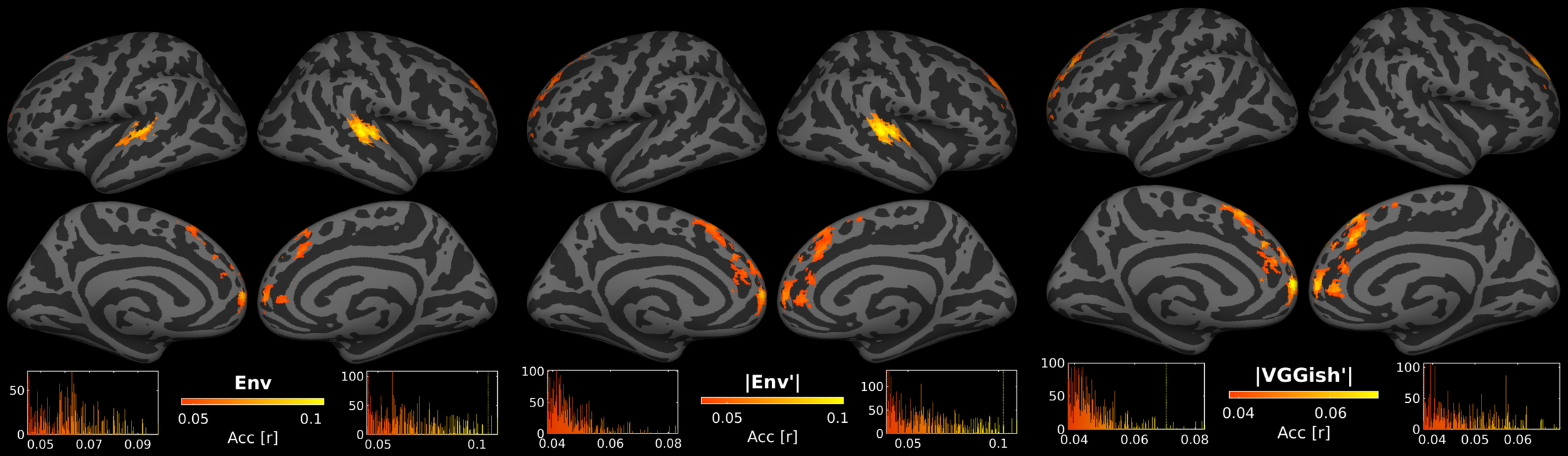




Kim et al., *In prep*.

Audio-brain correlation 🎶→🧠

fMRI = **Env** + **|Env'|** + **|OpenL3'|** + **|VGGish'|** + error (clus-P < 0.05)

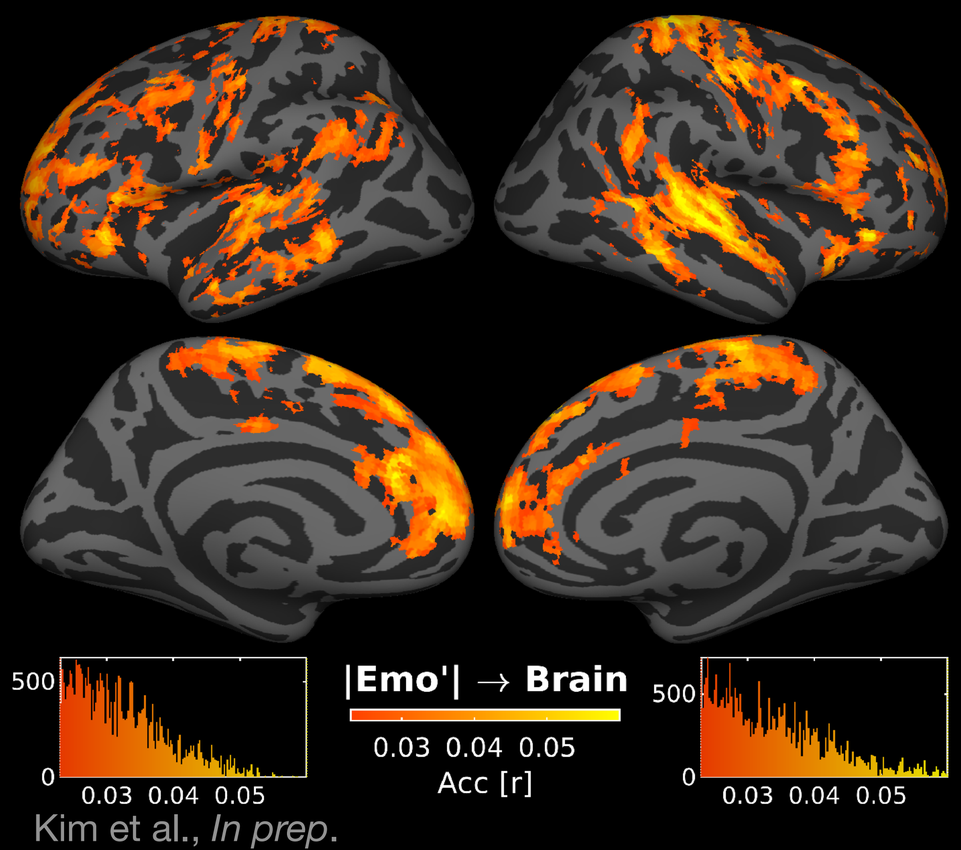


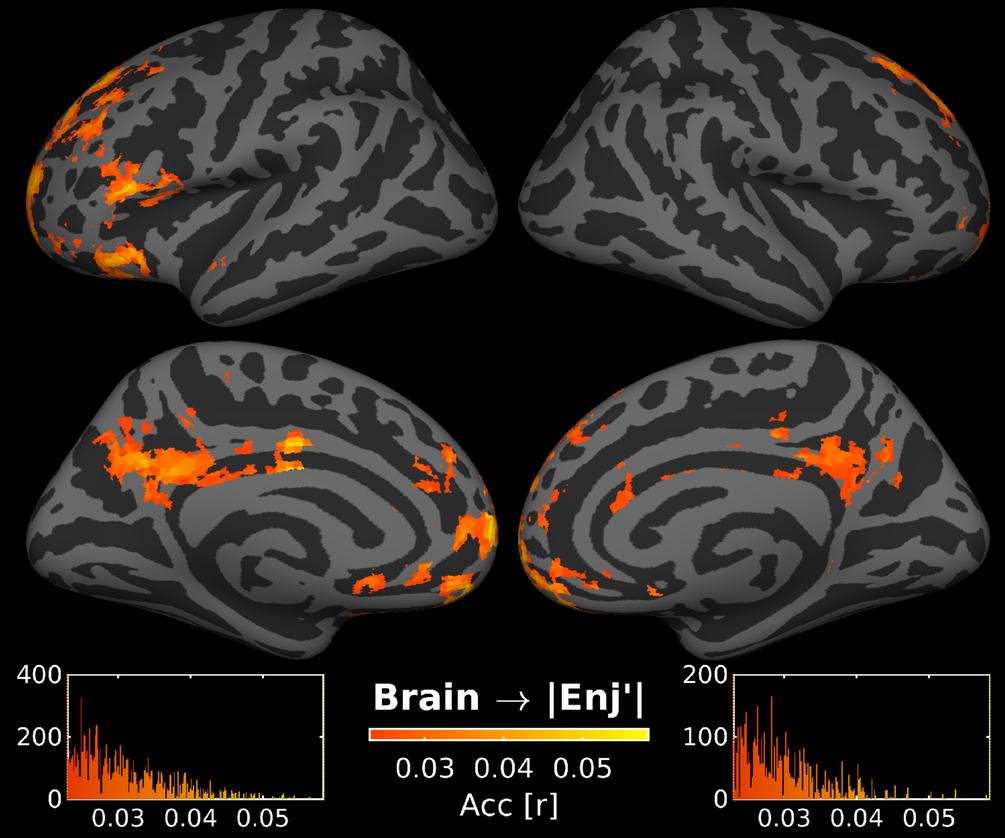
Kim et al., *In prep*.

CNN embedding was only encoded in the mPFC,
but not in the auditory cortex, which reflects its abstractness.

Brain-emotion correlation 😀😢↔🧠

fMRI = |Emotionality'| + |Enjoyment'| + error (clus-P < 0.05)



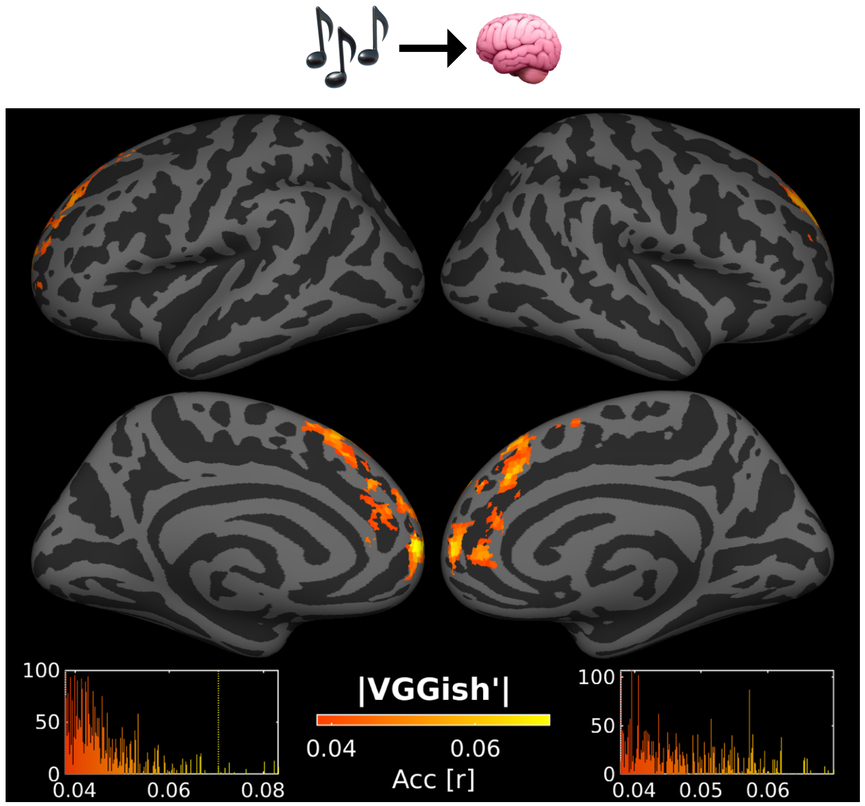


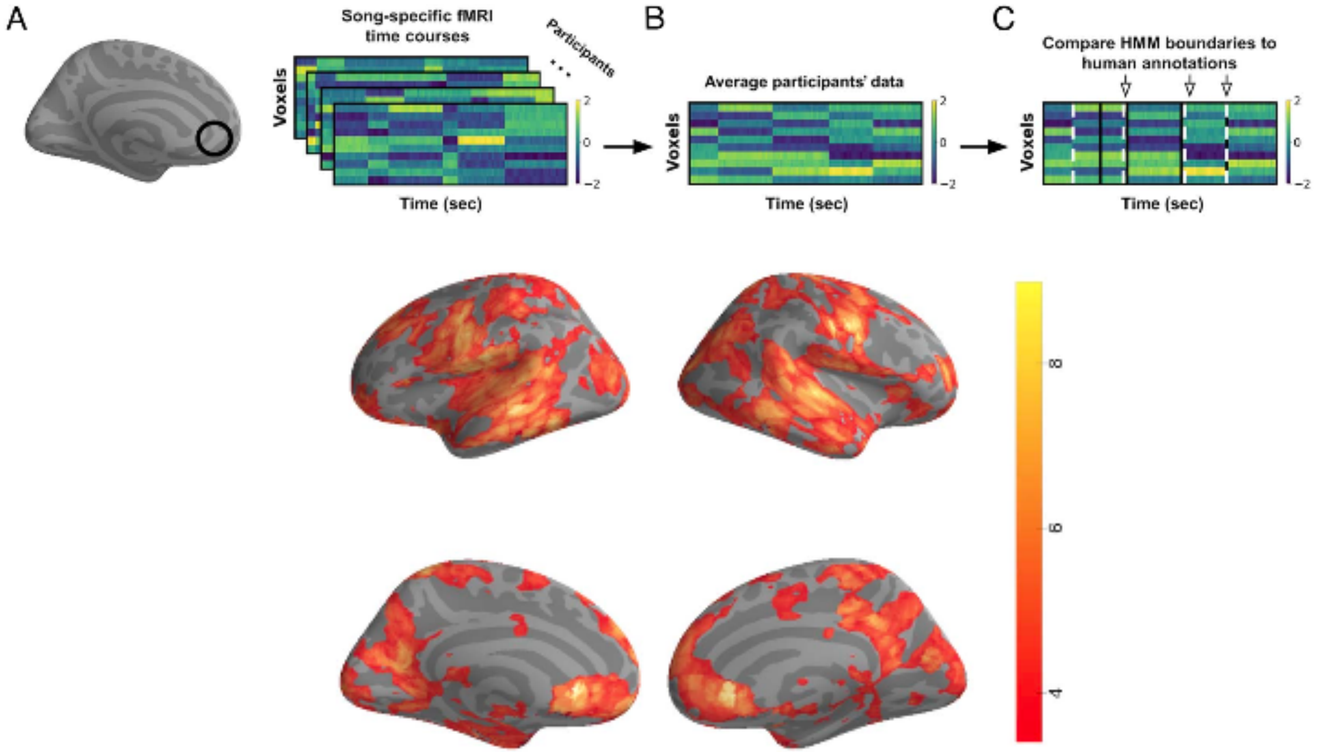
Kim et al., *In prep*.

Medial prefrontal cortex encoded both past emotionality and future enjoyment.

Discussion

Deep audio embedding and musical emotions

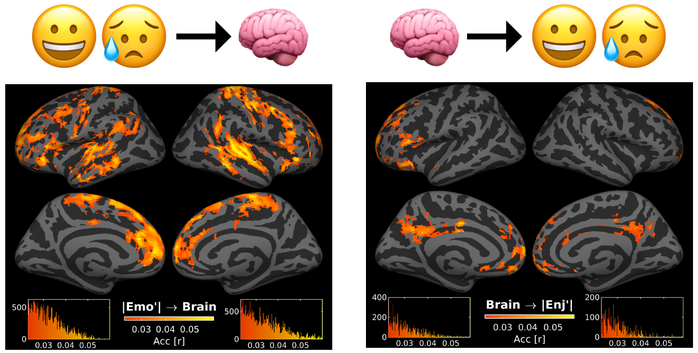


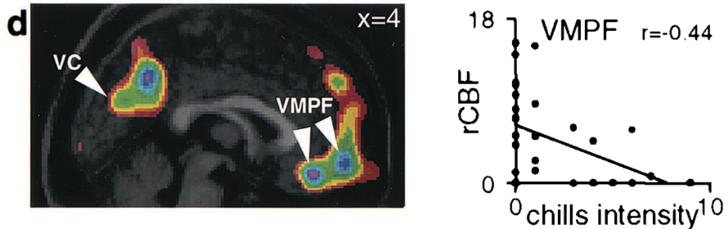


Williams et al., 2022, *J Cog Neurosci*.

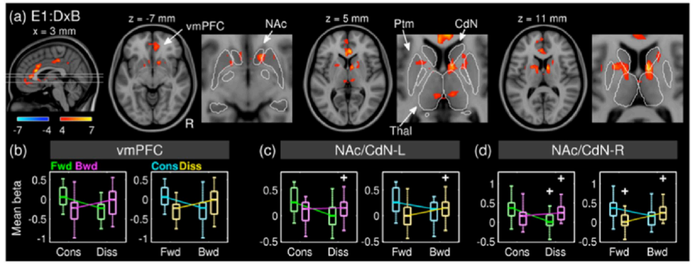
Audio semantic model changes were encoded in the mPFC, which showed a sensitivity to musical structures ("boundaries").

Different encoding of emotionality & enjoyment

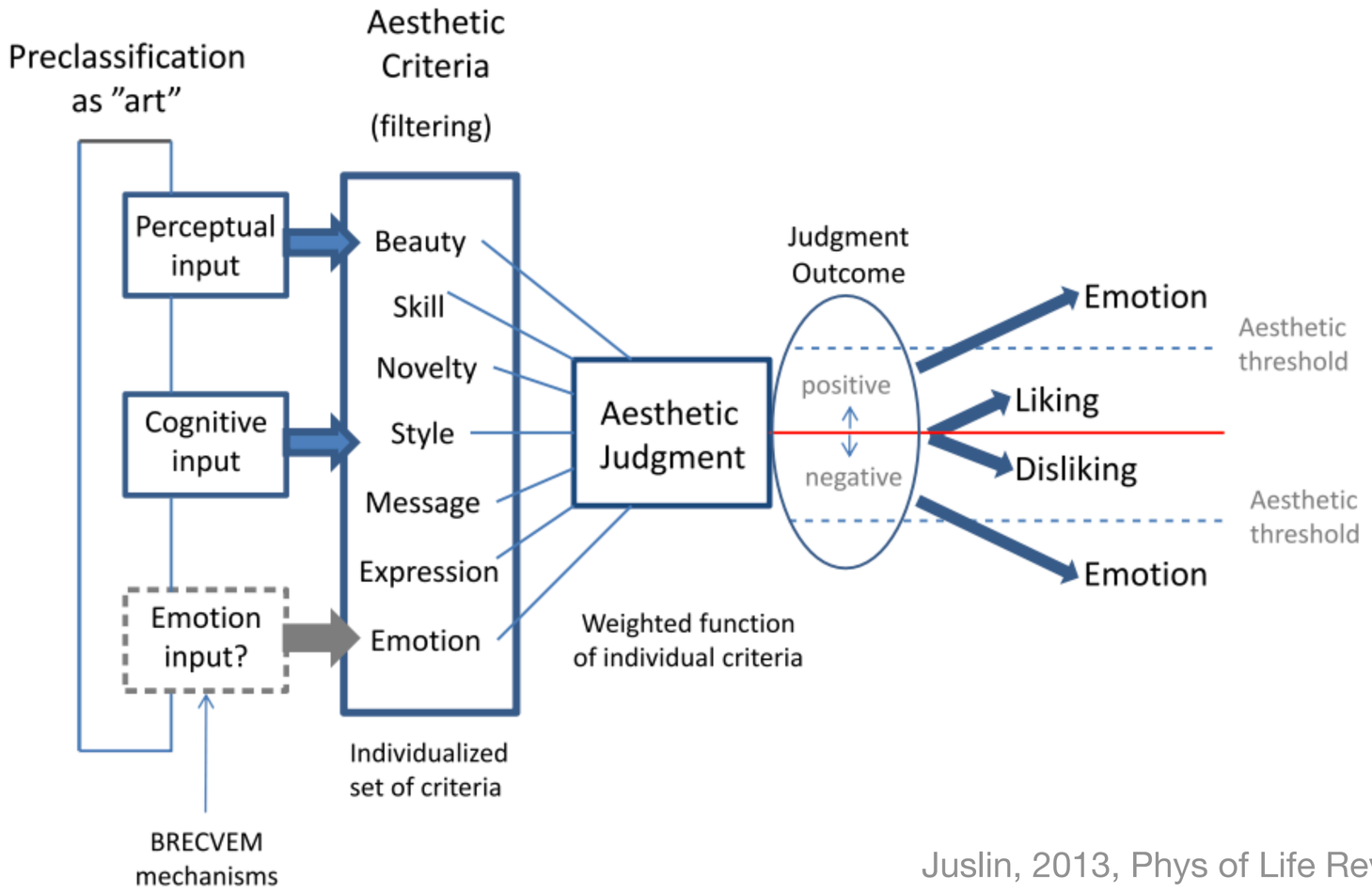




Blood & Zatorre, 2001, PNAS



Kim et al., 2019, Sci Rep.

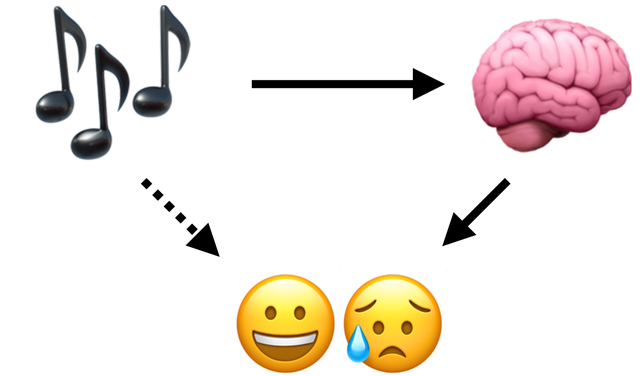


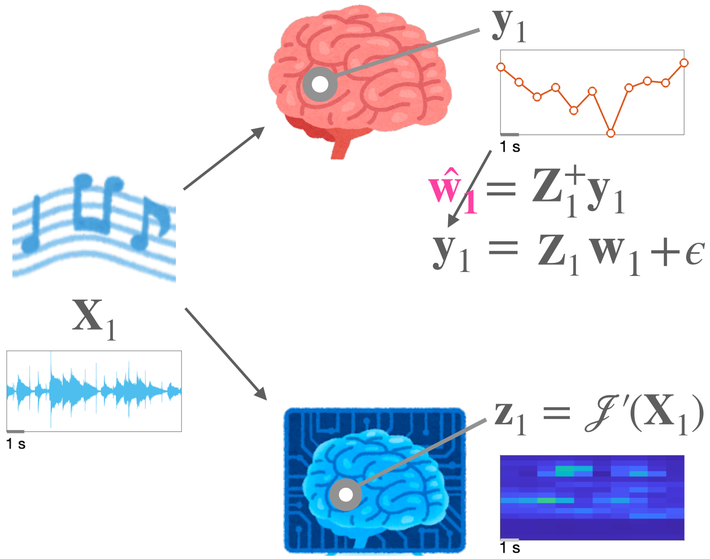
Juslin, 2013, Phys of Life Rev.

vmPFC activity was followed by Enjoyment rating changes.

Future directions

- **Analyses ("how to link 🎶/🧠/😀😢?")**
  - Two-step encoding analysis (audio→brain; brain→emotion)
  - Global dissimilarity (e.g., time-by-time RSA)
- **Model features ("how to represent 🎶?")**
  - Different levels of abstraction (CNN layers)
  - Audio-domain predictive models (i.e., OpenAI-Jukebox, MetaAI-MusicGen, Google-MusicLM)
  - Comparison with symbol-domain predictive models (e.g., IDyOM, RNN)





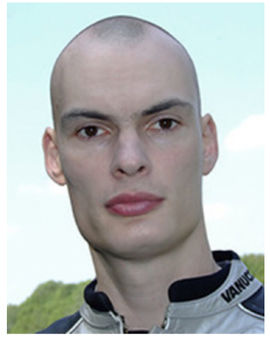
Kim, 2022. *Front. Neurosci.*

Conclusions

- CNN embeddings were sensitive to information that was relevant for emotional responses, **beyond low-level audio features**.
- In particular, the CNN encoding in the mPFC but not in the auditory cortex suggests that **its representation is rather abstract than sensory percepts**.
- Changes of two continuous ratings (*Emotionality* and *Enjoyment*) were differentially encoded in the brain, potentially reflecting **distinct mechanisms** of ***felt* emotions** and **aesthetic judgements**.

Thank you for your attention!

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



Dr. Karsten Müller
MPI-CBS, Leipzig, Germany

Dr. Tobias Overath
Duke University, NC, USA

seung-goo.kim@ae.mpg.de