

# Neural Encoding of Musical Emotions Evoked by Naturalistic Stimuli

Seung-Goo Kim<sup>1</sup>, Tobias Overath<sup>2</sup>, Daniela Sammler<sup>1</sup>

<sup>1</sup> Research Group for Neurocognition of Music and Language, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany

<sup>2</sup> Department of Psychology and Neuroscience, Duke University, Durham, NC, USA

## Introduction

**Background** Music often evokes strong emotions. Which acoustic features are at the basis of such felt emotions, and how these features and the emotions are encoded in the brain activity, are still debated.

Recent fMRI studies with naturalistic music showed the neural encoding of acoustic features<sup>1,2</sup> as well as structural information<sup>3</sup>. Recent DNN studies revealed that the embedding of audio “semantic” models can predict the *recognition* of musical emotions<sup>4</sup>. Here, we explored the relevance of DNNs to *felt* emotions and neural activity in an open fMRI dataset.

### Research Questions

**Q1** Are the embeddings of pretrained audio “semantic” models predictive of felt emotions and neural activity?

**Q2** How are felt emotions and musical enjoyment associated with neural activity over time?

## Methods

### Data openneuro-ds003085<sup>5</sup>

- Participants:  $n = 39$  ( $W = 21$ , mean age = 24.1 years)
- Imaging: 3-T GE-EPI ( $MB = 8x$ ,  $TR = 1$  s, 3-mm iso, whole brain), Prisma at the University of Southern California
- Musical stimuli: “happy” (2 min 48 s), “sad-short” (4 min 16 s), “sad-long” (8 min 35 s) in styles of movie soundtracks
- Continuous ratings: “Felt Emotionality” and “Enjoyment” after scanning (fader, 40 Hz)

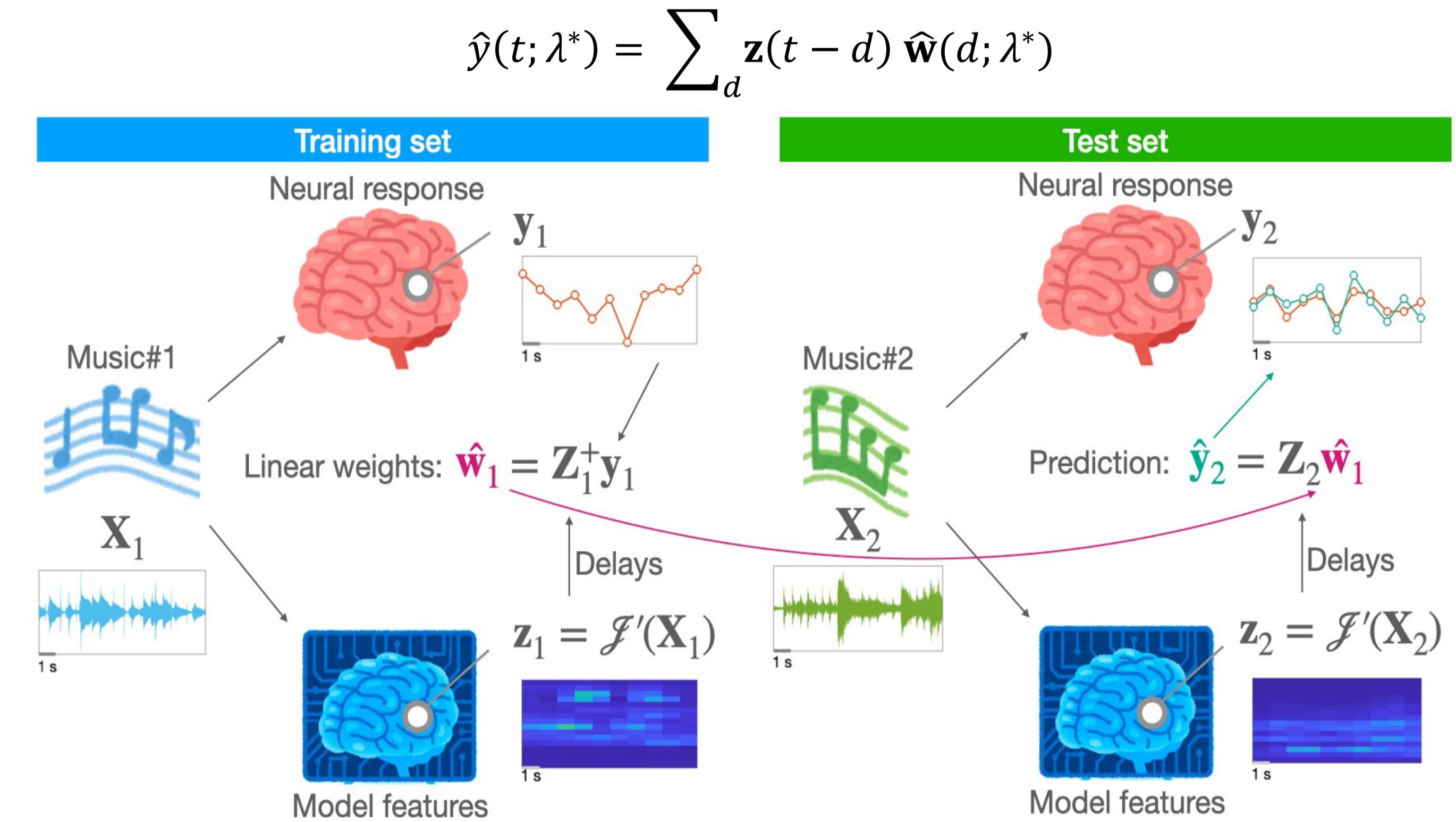
### Image processing

- SPM12 (v7771) + ANTs (v2.3.5), 3-mm-iso, MNI152, FWHM = 6 mm, ICA-AROMA (v0.4.4-beta), detrending (rel. pow. ( $<1/128$  Hz)  $< 5\%$ ), excluding  $n = 2$  for small FOV

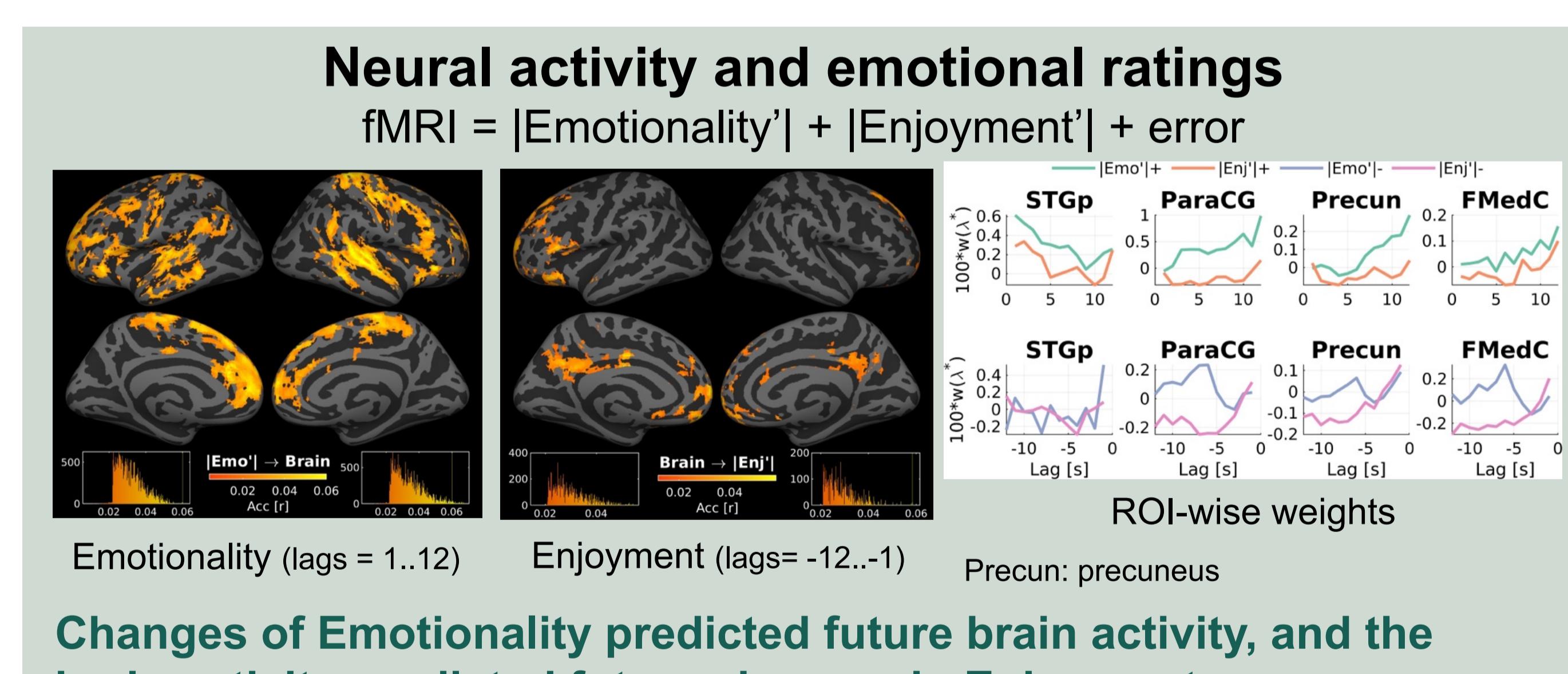
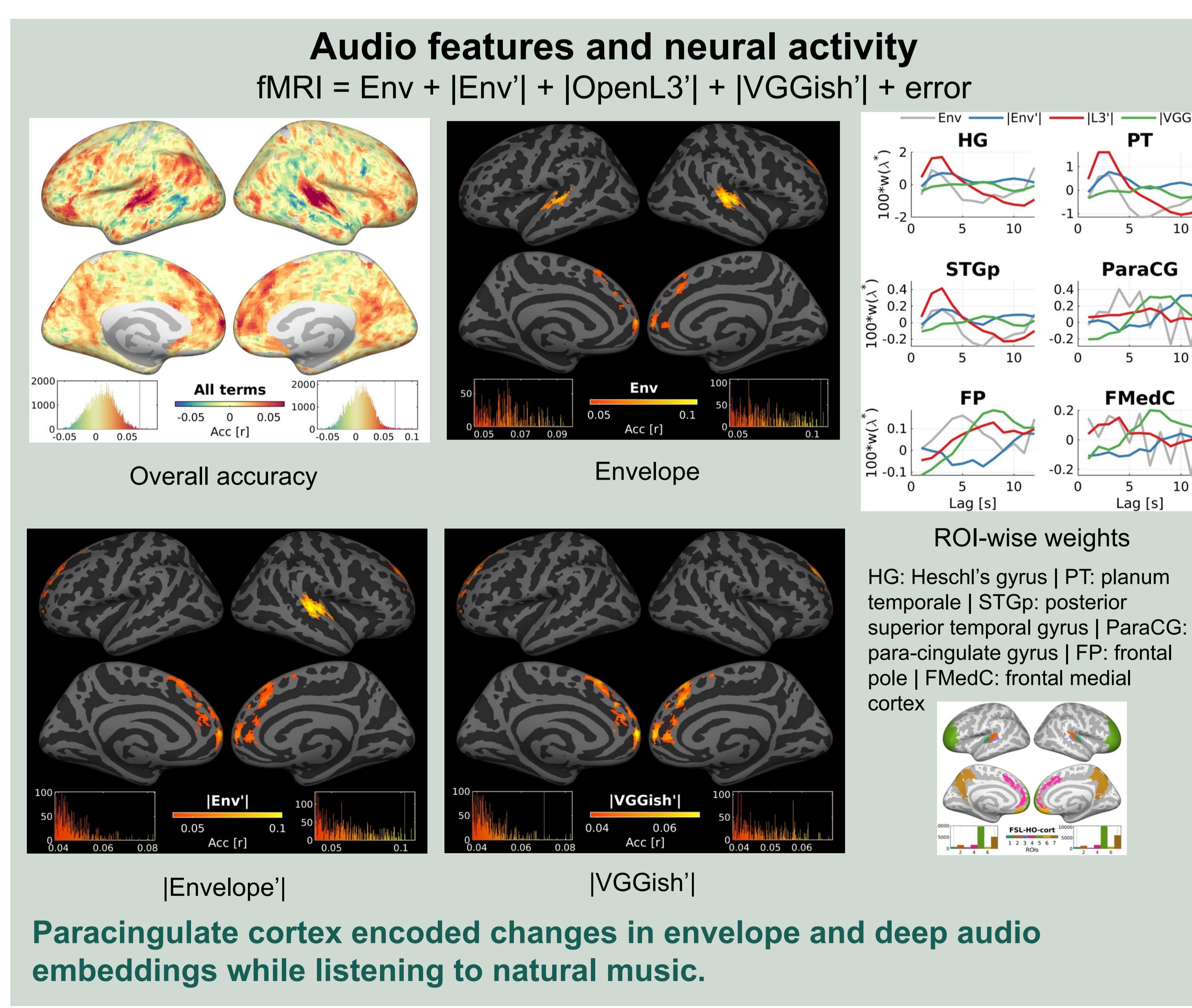
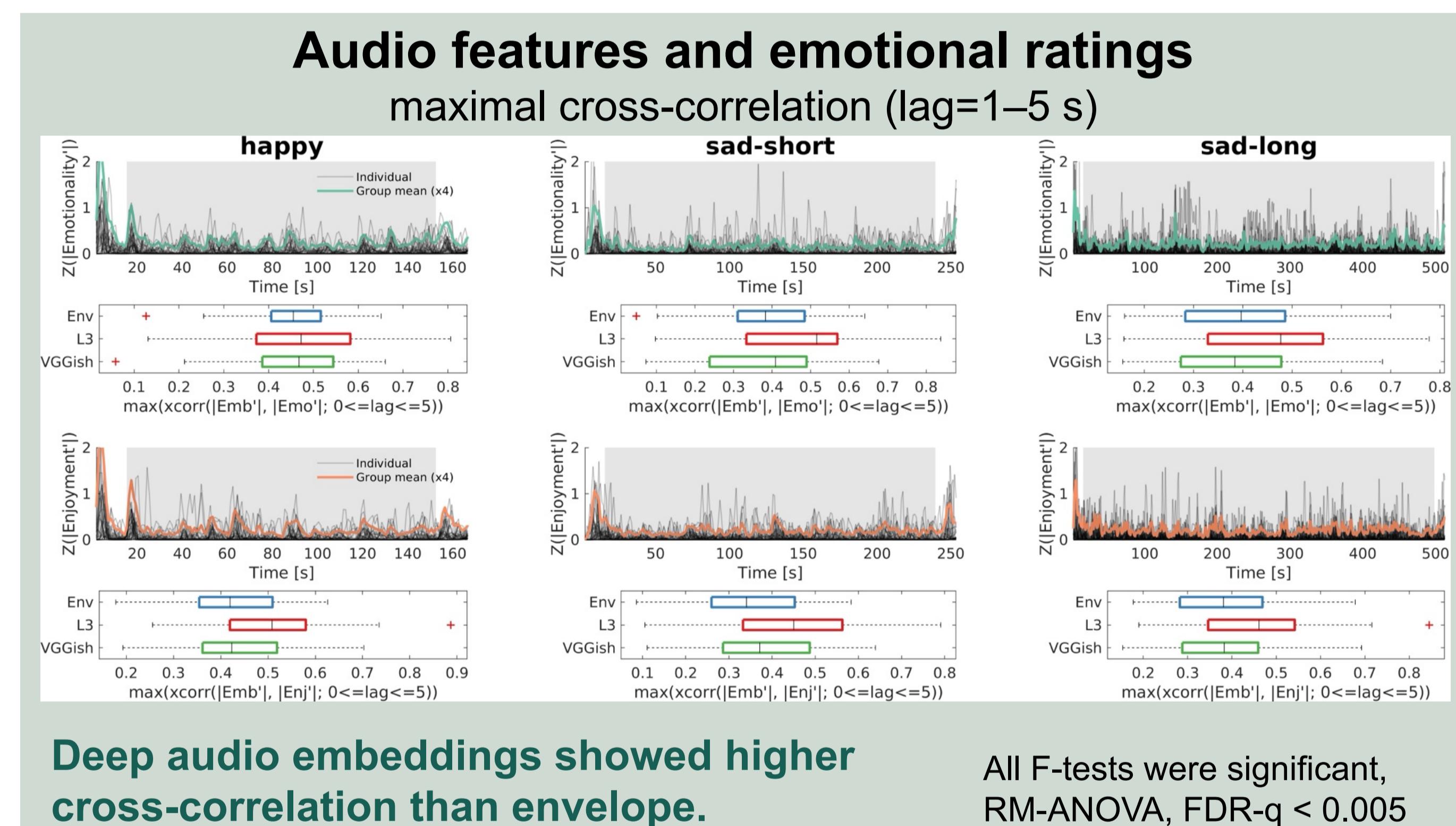
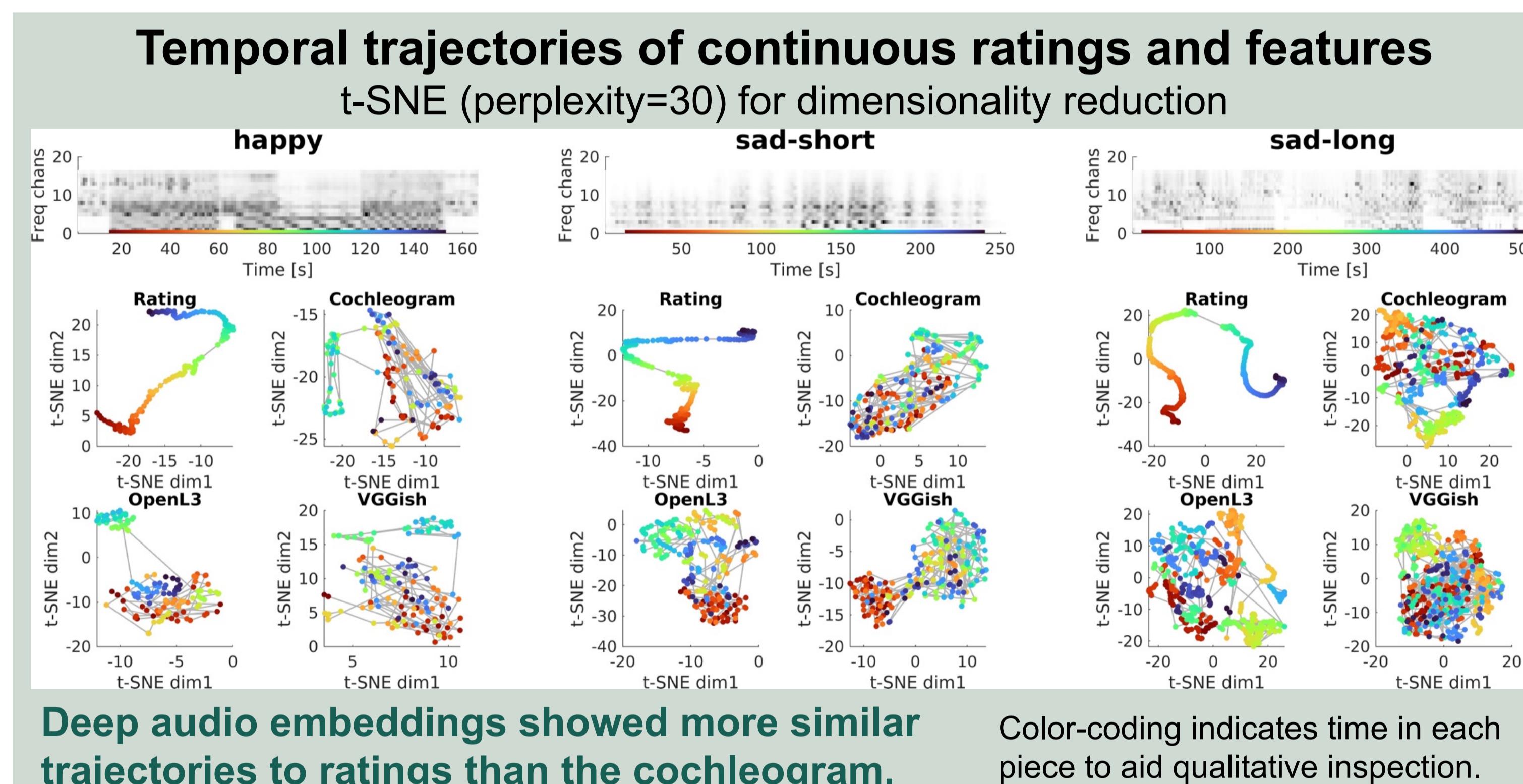
### Feature extraction

- Auditory features: broadband envelope (185–6840 Hz; dim = 1) and narrowband cochleogram (dim = 16; only for t-SNE analysis) using NSL tools<sup>6</sup>
- Deep audio embedding (final layer): OpenL3<sup>7</sup> (dim = 512) and VGGish<sup>8</sup> (dim = 128)

**Linearized encoding analysis** For stationarized (absolute temporal derivatives) features and standardized (Z-scoring) neural responses, voxel-wise linearized encoding models with ridge regression<sup>9</sup> were optimized via nested cross-validation and fit (lags = 1, 2, ..., 12 s or -12, -11, ..., -1 s):



**Statistical inference** Prediction accuracy (Pearson’s  $r$ ) of observed model vs. phase-randomized models for each term, 1,000 randomizations,  $\min(p)^{10}$  cluster-level inference, one-sided,  $\alpha = 0.05$



Changes of Emotionality predicted future brain activity, and the brain activity predicted future changes in Enjoyment.

## Conclusions

**C1** Deep audio embedding seems sensitive to information that is relevant for emotional responses, beyond low-level acoustics. In particular, the encoding of the DNN embeddings in the paracingulate gyrus may link to the encoding of musical boundaries in the DMN<sup>10</sup>.

**C2** Changes of two continuous ratings (Emotionality and Enjoyment) were differentially encoded in the brain, potentially suggesting distinct mechanisms of the felt emotions and aesthetic judgements<sup>11</sup>.

### Future directions

- Two-step (feature to brain, and brain to behavior) encoding analysis
- Modeling the global dissimilarity (e.g., t-SNE) in addition to the local dissimilarity (e.g., derivatives) over time

### Correspondence

Seung-Goo Kim  
seung-goo.kim@ae.mpg.de



### Acknowledgment

We thank Matthew Sachs for the openneuro data and continuous rating data and irasutoya.com for image sources.

- References**
- [1] Santoro et al., 2014, doi:10.1371/journal.pcbi.1003412
  - [2] Nakai et al., 2021, doi:10.1002/brb3.1936
  - [3] Williams et al., 2022, doi:10.1162/jocn\_a\_01815
  - [4] Koh & Dubnov, 2021, doi:10.48550/arXiv.2104.06517
  - [5] Sachs et al., 2020, doi:10.1016/j.neuroimage.2019.116512
  - [6] Chi et al., 2005, doi:10.1121/1.1945807
  - [7] Cramer et al., 2019, doi:10.1109/ICASSP.2019.8682475
  - [8] Hershey et al., 2016, doi:10.48550/arXiv.1609.09430
  - [9] Huth et al., 2016, doi:10.1038/nature17637
  - [10] Geerligs & Maris, 2021, doi:10.1002/hbm.25399
  - [11] Juslin, 2016, doi:10.1016/j.jplrev.2013.05.008