Shared generative representation of auditory concepts and EEG to reconstruct perceived and imagined music

André Ofner, Sebastian Stober {ofner, sstober}@uni-potsdam.de
University of Potsdam, Germany - Research Focus Cognitive Sciences

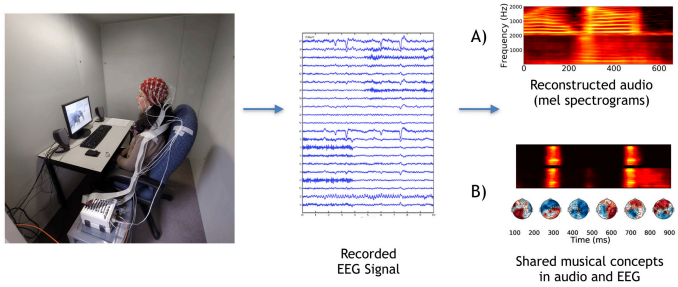
Can we reconstruct perceived and imagined music from Electroencephalography (EEG) data?

Can we learn and describe mental concepts?

Motivation and hypotheses

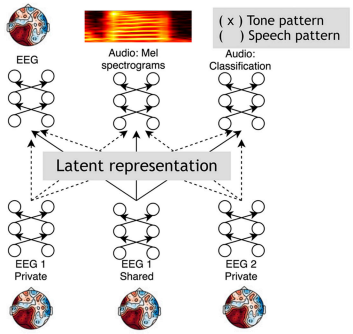
- music imagination is grounded in perceptual processes
- tight coupling between auditory and conceptual processing [1]
- EEG is modulated in correlation to the perceived auditory stimuli, its internal representation and cognitive processing
- deep neural networks for non-linear analysis of noisy EEG [2]

Experiment setup



Shared representation learning of audio and EEG

- based on Variational Canonical Correlation Analysis (VCCA) [3]
- adapted to multi-task learning from multiple modalities for simultaneous decoding of audio and EEG signal.
- multi-view Variational Autoencoder (VAE) with multiple encoder and decoder subnetworks connected by a single latent layer
- trained on music perception EEG data from two datasets
- deep convolutional neural networks for non-linear observation



OpenMIIR speech dataset

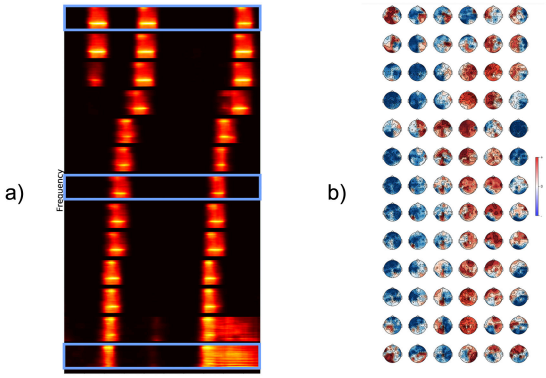
- 4 looped speech and corresponding sine wave patterns
- cued perception directly followed by cued imagination
- 64 EEG channels at 512 Hz within 7 subjects

NMED-T dataset

- perception condition in 10 naturalistic full length songs
- cued perception and imagination
- 125 EEG channels at 125 Hz within 20 subjects

Learned auditory concepts

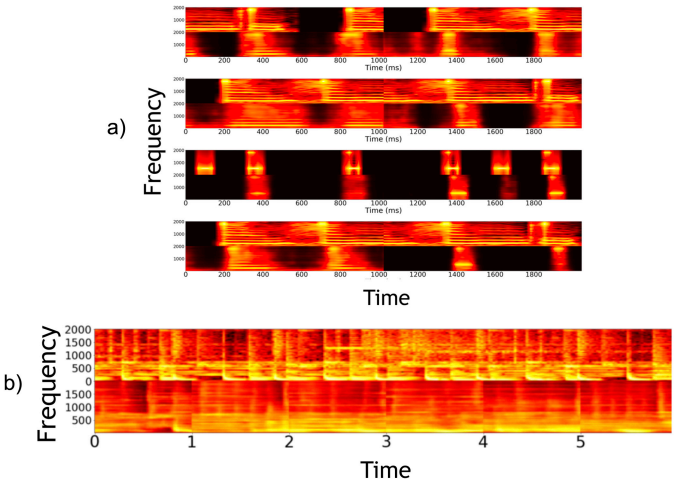
- simultaneous representation and retrieval of audio and EEG is possible
- latent space allows linear interpolation between various real EEG inputs
- OpenMIIR speech: visually understandable latent representation of brain activity and audio for simple motifs across different instrumentations
- NMED-T: more complex, less intuitively understandable representation
- hinting at disentangled representation of rhythm, tempo and timbre



a) Interpolation in the shared latent space. Reconstructions for three real inputs are indicated in blue. Remaining reconstructions are interpolated.
b) Temporal sequences of EEG brain states decoded simultaneously with the audio reconstructions.

Audio spectrogram reconstruction

- representation allows reconstruction of perception and imagination
- stimulus complexity and the amount of training data show strong effect on the reconstruction quality



a) Reconstructed Mel spectrograms of cued perceived rhythmical trials for a single subject of the OpenMIIR speech dataset.
b) Reconstructed Mel spectrogram from the NMED-T dataset after training on all subjects. Target stimuli are presented above their reconstructions.

VCCA framework: Multi-modal extensions

- The proposed model can be extended to incorporate multiple inputs
- inputs could stem from different subjects other neuro-imaging modalities
- additional decoders could decode other cognitive processes from the shared representation

Machine Learning in Cognitive Science

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

OpenMIIR
public domain dataset of EEG recordings for music imagery information retrieval

[1] M. Kiefer, E.-J. Sim, B. Herrnberger, J. Grothe, and K. Hoenig. The sound of concepts: four markers for a link between auditory and conceptual brain systems. Journal of Neuroscience, 28(47):12224–12230, 2008.
[2] S. Stober, D. J. Cameron, and J. A. Grahn. Using convolutional neural networks to recognize rhythm stimuli from electroencephalography recordings. In Advances in neural information processing systems, pages 1449– 1457, 2014.
[3] W. Wang, X. Yan, H. Lee, and K. Livescu. Deep variational canonical correlation analysis. arXiv preprint arXiv:1610.03454, 2016.