EEG2Audio

2024 UCSD Cognitive Science Summer Research Fellowship

Jacob Lamadrid

Project Outline

Audio Reconstruction	Seeking to extract/reconstruct auditory stimuli via deep learning models with some input neural signal, possible via time-frequency domain “images” (Spectrograms)
EEG2Image Models [1]	Inspired by existing work in image reconstruction from EEG/MRI, including current work performed at the lab
fMRI Audio Reconstruction	Majority of existing work in audio reconstruction is in the context of fMRI which is partially relevant to EEG but not entirely replicable
Benefits of EEG in Context	EEG provides an easy, non-invasive method for capturing neural signals over fMRI as well as a finer temporal resolution in recording
Challenges	Code from previous works in EEG to audio modeling is not publicly available and datasets for this task are extremely limited

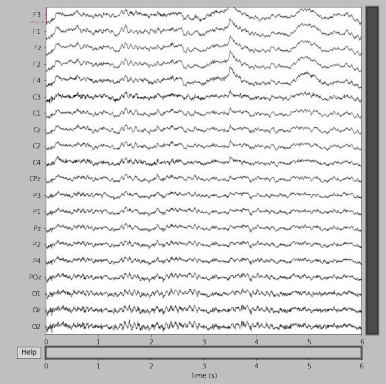
Data

Naturalistic Music EEG Dataset - Tempo (NMED-T) [6]
- 125 channel recording
- *Full length songs*

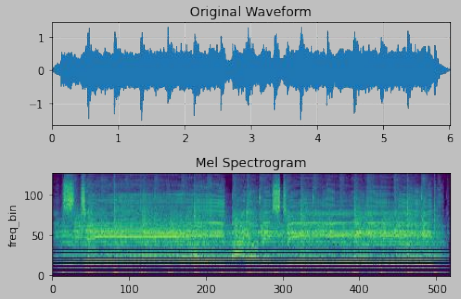
MAD-EEG: an EEG dataset for decoding auditory attention to a target instrument in polyphonic music [7]
- 20 channel recording
- Repeating music segments

Note:
- Current publicly available datasets for EEG audio reconstruction restricted to music stimuli

Audio processed into MelSpectrograms*



Aggregated & Filtered EEG Data



Example MelSpectrogram

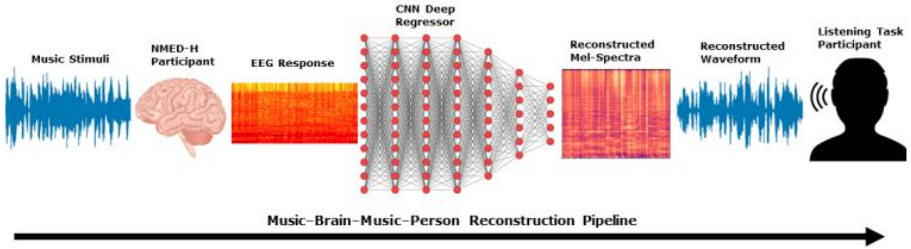
* "An object of type MelSpectrogram represents an acoustic time-frequency representation of a sound: the power spectral density P(f, t). It is sampled into a number of points around equally spaced times and frequencies (on a Mel frequency scale)" [5]

Overview of Methods

CNN (EEG2Mel) [2]

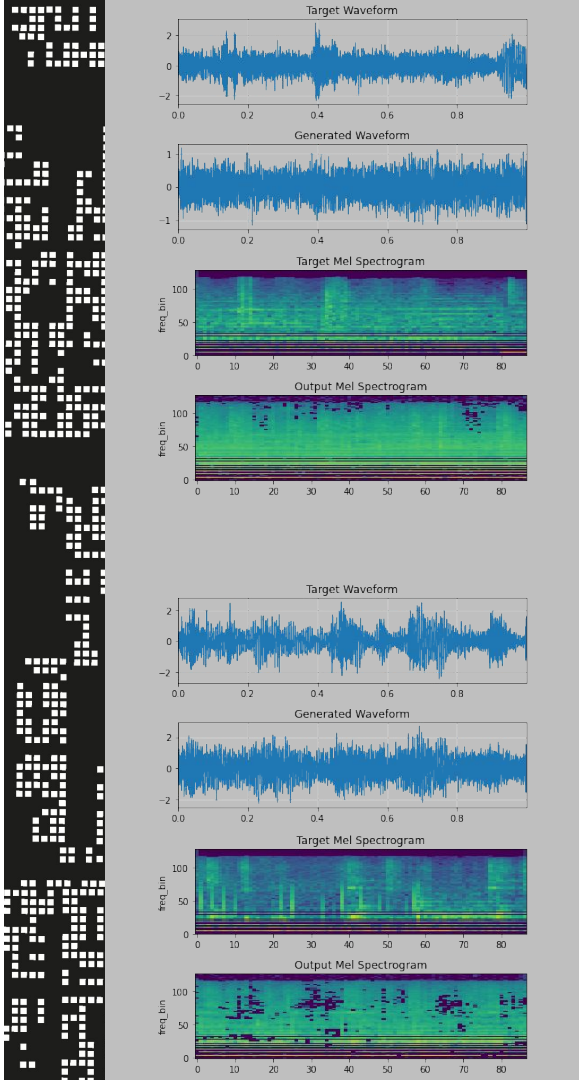
Convolutional Neural Network directly mapping EEG to audio

- Input: Power Spectral Density array of EEG recording
- Output: MelSpectrogram*



Findings

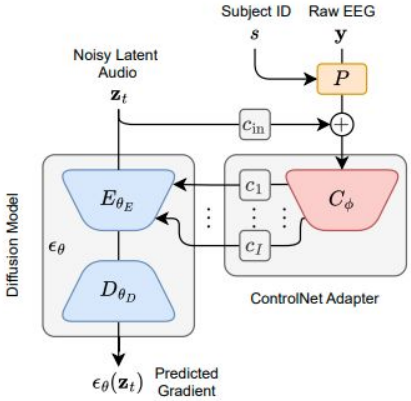
+ Least computationally intense
+ Easy model/procedure interpretability
– Poor with high variance
– Poor in fine temporal properties



Overview of Methods

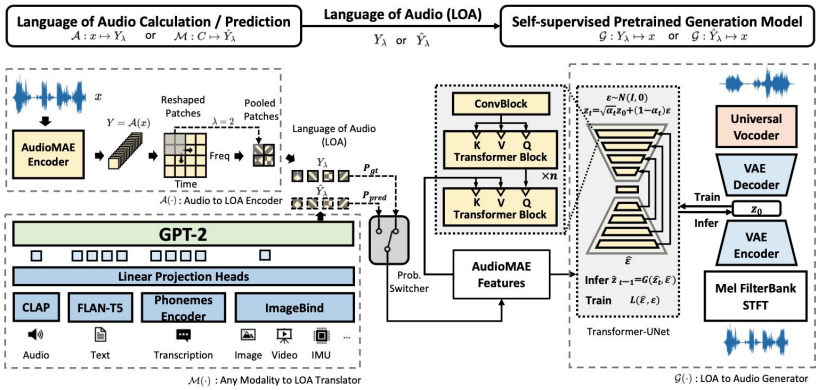
Diffusion (AudioLDM) [3]
Latent Diffusion Model utilizing a prompt generative model for conditioning outputs based on EEG

- Input: Latent noise image + Projected EEG (Conditioning)
- Output: Latent MelSpectrogram* (Ready for VAE decoding)



Findings

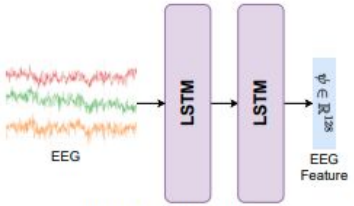
+ Pretrained model
+ Standard EEG to image method
– No publicly available implementations
– Data dimensionality issues



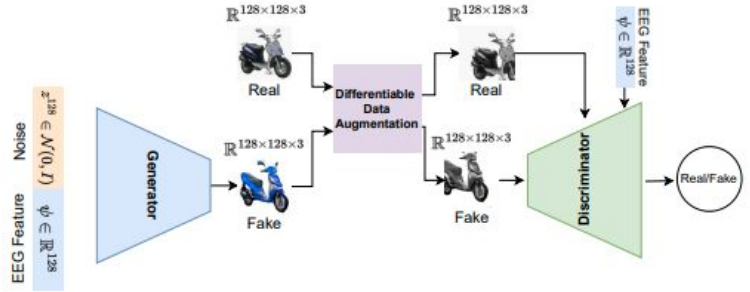
Overview of Methods

DCGAN (Future Work) [4]
Generative Adversarial Network generating audio based on EEG features then classifying real/fake audio

- Input: EEG Features + Noise vector
- Output: MelSpectrogram*



(a) EEG feature extractor



(b) EEG2Image

Findings
+ Fast inference
+ Enhanced feature extraction
+ Applications outside of music
+ Improved interpretability
– No existing work in audio

Outcomes & Future Work

Data and Model Preparation	CNN model successfully set up for training on new data and 2 other models drafted
Comprehensive EEG Audio Training Data	Current datasets are primarily utilizing fMRI or are limited to applications in music response
Working GAN Implementation	DCGAN model with a more comprehensive input
Thank you!	Special thanks to Simon, Ian, and Prof. de Sa

References

[1] Yunpeng Bai, Xintao Wang, Yan-pei Cao, Yixiao Ge, Chun Yuan, Ying Shan: “DreamDiffusion: Generating High-Quality Images from Brain EEG Signals”, 2023; arXiv:2306.16934.

[2] Adolfo G. Ramirez-Aristizabal, Chris Kello: “EEG2Mel: Reconstructing Sound from Brain Responses to Music”, 2022; arXiv:2207.13845.

[3] Emilian Postolache, Natalia Polouliakh, Hiroaki Kitano, Akima Connelly, Emanuele Rodolà, Luca Cosmo, Taketo Akama: “Naturalistic Music Decoding from EEG Data via Latent Diffusion Models”, 2024; arXiv:2405.09062.

[4] Prajwal Singh, Pankaj Pandey, Krishna Miyapuram, Shanmuganathan Raman: “EEG2IMAGE: Image Reconstruction from EEG Brain Signals”, 2023; arXiv:2302.10121.

[5] Jansson, A., Humphrey, E., Montecchio, N., Bittner, R. M., Kumar, A., & Weyde, T. (2018). Wave-U-Net: A Multi-Scale Neural Network for End-to-End Audio Source Separation. arXiv preprint arXiv:1806.03185.

[6] Steven Losorelli, Duc T. Nguyen, Jacek P. Dmochowski, and Blair Kaneshiro (to appear). NMED-T: A Tempo-Focused Dataset of Cortical and Behavioral Responses to Naturalistic Music. In Proceedings of the 18th International Society for Music Information Retrieval Conference, Suzhou, China.

[7] Giorgia Cantisani, Gabriel Trégoat, Slim Essid, Gael Richard. MAD-EEG: an EEG dataset for decoding auditory attention to a target instrument in polyphonic music. Speech, Music and Mind (SMM), Satellite Workshop of Interspeech 2019, Sep 2019, Vienna, Austria. ⟨hal-02291882v1⟩