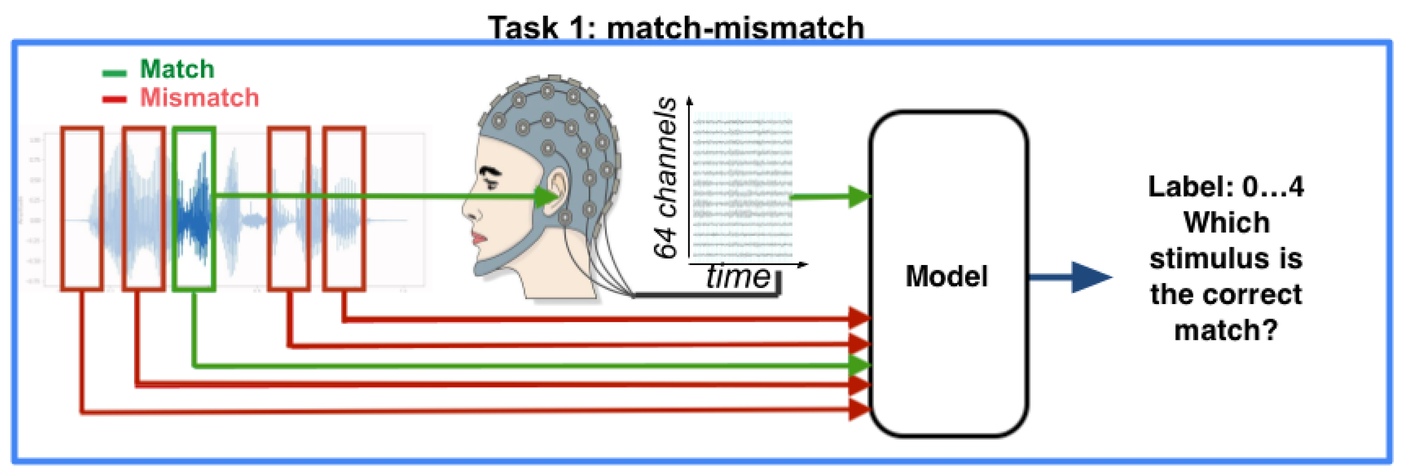
Improving Auditory EEG Decoding with Time-Frequency Analysis

Alexander Karbowski [1], Sophia Xiao [2], Corinne Orton [3], Dylan Marchlinski [4]

[1] Harvard University, [2] Emory University, [3] University of Utah, [4] University of Pennsylvania

Introduction

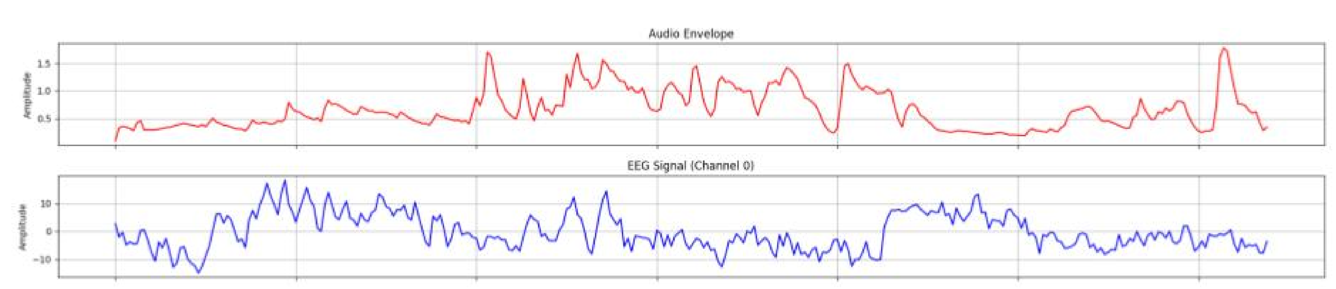
ICASSP 2024 Grand Challenge:
We address the matching task of the 2024 ICASSP Auditory EEG Grand Challenge: given a 5-second segment of EEG and five 5-second audio stimuli, determine which audio is the correct match.



Proposed Approach
Employ time-frequency and time-scale analysis for feature extraction to train shallow and lightweight neural networks to solve the challenge.

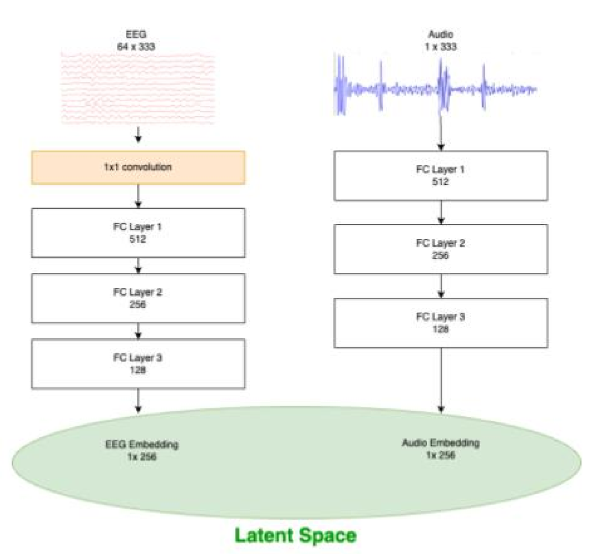
Dataset Details

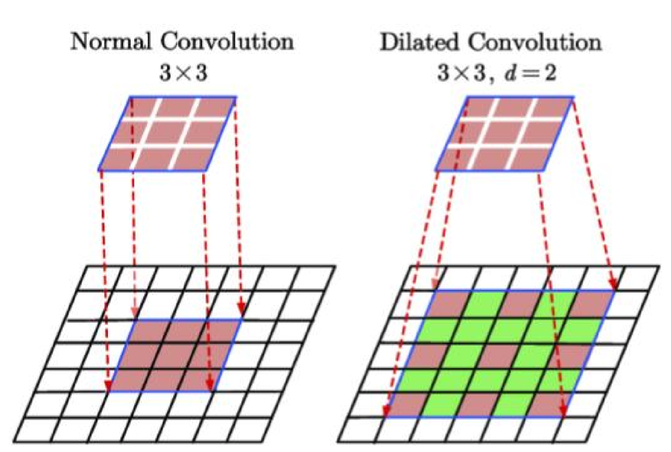
- 64-channel EEG recordings from 85 young individuals with normal hearing, each of whom listened to 90-150 minutes of natural speech.
- EEG collected at 8192 Hz, preprocessed with high-pass filter, downsampled to 1024 Hz, processed with a Wiener filter, re-referenced to a common average, and downsampled to 64 Hz.
- Audio collected at frequency 48kHz. The envelope was computed, then downsampled to 64 Hz.



Neural Networks

Two Branch Encoder: fully connected layers are applied to the EEG signals and audio separately, and outputs are embedded in latent space.
Dilated Convolutional Neural Network (DCNN): fully connected layers replaced by 2D dilated convolutional layers.

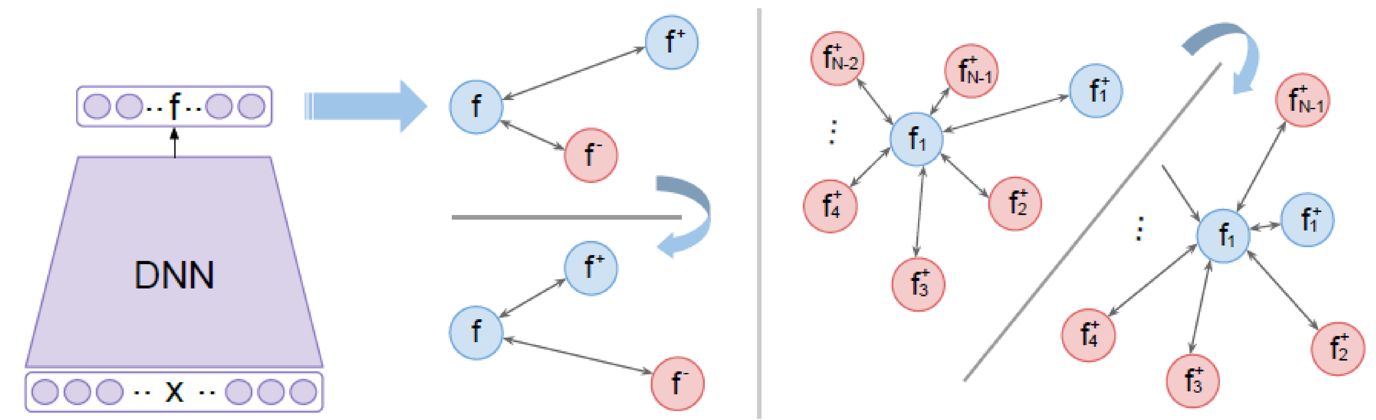




InfoNCE Loss Function

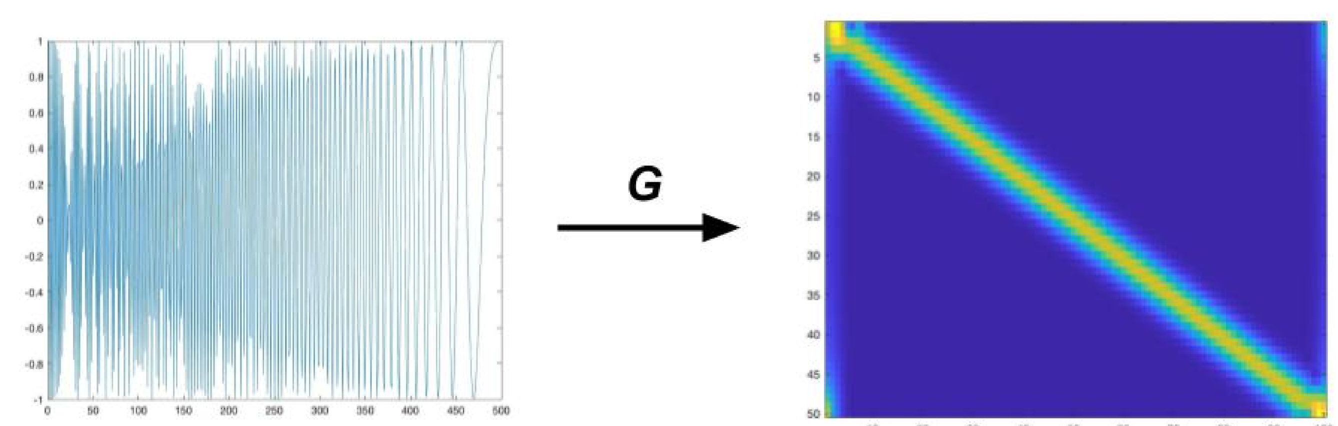
The networks are trained by minimizing the **InfoNCE Loss Function**, which ensures that matched data are close together and mismatched data are far apart in the latent space. Let x_i denote the normalized output of the EEG encoder, y_i^+ the matching normalized output of the audio encoder and y_j^- mismatched normalized outputs of the audio encoder. The InfoNCE loss is defined as

$$\mathcal{L}_{\text{InfoNCE}} = -\log \frac{\exp\left(x_i \cdot y_i^+\right)}{\sum_{j=1}^{5} \exp\left(x_i \cdot y_j^-\right)}$$

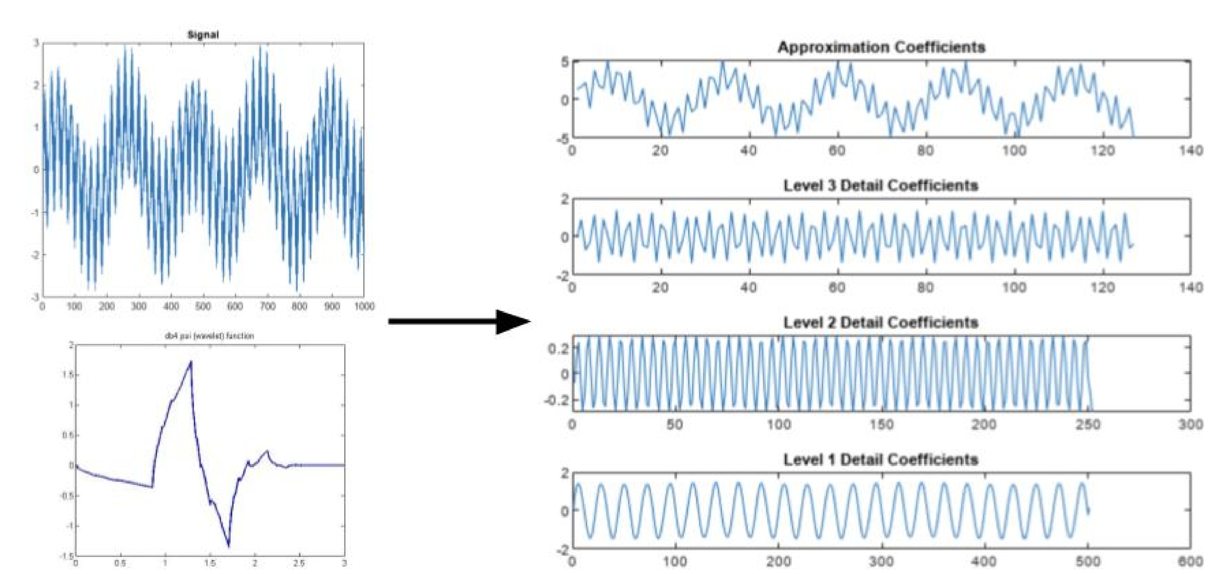


Time - Frequency Analysis

Discrete Gabor Transform: transforms a signal from the time domain to time-frequency domain, $G : \mathbb{C}^L \to \mathbb{C}^{M \times N}$. Captures change in frequency over time. Main parameters are the frequency (M) and time (N) resolutions.



Discrete Wavelet Transform: measures the *scale* of a wave over time, by projecting signals onto wavelets. Main parameter is type of wavelet used.



Results

Matching Accuracy:

	Raw	Gabor	Wavelet
2-Branch	33.8	32.4	**34.19**
DCNN	34.16	**39.15**	30

- The 2-branch model's accuracy was highest with the wavelet data.
- The baseline DCNN's accuracy was highest with the Gabor data, and this was the highest accuracy attainedd.

Overlap of Correct Guesses:

		DCNN Gabor	
		Correct	Incorrect
2-Branch Wavelet	Correct	0.19	0.15
	Incorrect	0.20	0.46

- Of the two top-performing configurations, at least one is correct 54% of the time.

Discussion

- Using transformed data to train networks provides an advantage over raw data.
  - Using the Gabor transform, we reach nearly 40% accuracy with a small DCNN, compared to around 34% accuracy on the raw data.
  - Gabor data may yield higher accuracy because it is higher-dimension (redundant), giving the model more information to work with.
- Fourier features achieve better performance than wavelet features
- Mathematically transforming data before feeding it into a neural network is a computationally cheap way to improve matching accuracy and interpretability overall.
  - Feature extraction can be used in conjunction with DNNs to reach higher accuracy, improve transparency, and lower computational cost.

Acknowledgements

This work is sponsored by the National Science Foundation grant DMS 2149913. We thank Prof. Czaja and Gabriel Vilarrubi for their mentorship during the REU program.

References

[1] Bernard Accou et al. "SparrKULee: A Speech-evoked Auditory Response Repository of the KU Leuven, containing EEG of 85 participants". In: *bioRxiv* (2023). Now published in *Data*, doi: 10.3390/data9080094. DOI: 10.1101/2023.07.24.550310. URL: https://www.biorxiv.org/content/10.1101/2023.07.24.550310v1.

[2] Bernd Accou et al. "Modeling the relationship between acoustic stimulus and EEG with a dilated convolutional neural network". In: *2020 28th European Signal Processing Conference (EUSIPCO)*. 2021, pp. 1175–1179. DOI: 10.23919/Eusipco47968.2020.9287417.