Decoding of Imagined Speech and Music

Maryam Maghsoudi Shaghaghi[1], Mohsen Rezaiezadeh[1], Rupesh Kumar Chillale[1], Guilhem Marion[2], Ali Ibrahim Ali Mohammed[1], Abhinav Uppal[3], Gert Cauwenbergs[3], Giovanni M. Di Liberto[4,2], Jonathan Simon[1,5], Shihab Shamma[1,2]

[1]Institute for Systems Research, Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, [2]Laboratoire des Systèmes Perceptifs, Département d'Étude Cognitive, École Normale Supérieure, PSL, 75005, Paris, France, [3]Dept. of Bioengineering, University of California San Diego, La Jolla, CA 92093-0412, [4] School of Computer Science and Statistics, Trinity College Dublin, ADAPT centre, Trinity College Institute of Neuroscience, Ireland, [5] Biology, University of Maryland, College Park, MD 20742.

Background

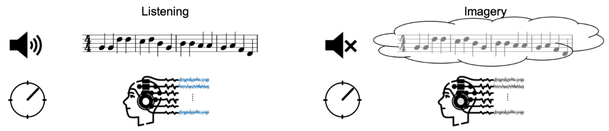
Listening to music evokes spectrotemporally detailed neural responses measured with noninvasive EEG and MEG. Studies by *Di Liberto et al, 2021, Marion et al, 2021* using EEG recordings from professional musicians revealed that listening and imagery responses induced analogous responses, but with roughly the inverse polarity and slightly delayed. This was presumably because imagination is a top-down *predictive* process in contrast to bottom-up listening.

Objectives

- To explore the exact relationship between listened and imagined music responses, and to extend these results to other acoustic signals like speech.
- To predict the listened responses of music and speech from their imagined counterparts, and vice versa, using linear and non-linear mappings.
- Generalizations of these results and algorithms to other recordings from other systems such as MEG, Dry-EEG, and also ECoG.

MEG Experiment Setup

- Recordings of listened and imagined music and speech of 15 musicians (6 females and 9 males, mean age = 33.07) was made in KIT (Kanazawa Institute of Technology) axial gradiometer MEG setup.
- We collected data at 1 kHz sampling rate with an online 500 Hz low pass filter, and a 60 Hz notch filter. Subjects rested in the supine position in a magnetically shielded room (VAC), while the MEG data were recorded.
- Stimuli: Two melodies and two speech snippets – each 27s long were used.
- The melodies were derived from a monophonic MIDI corpus of Bach chorales (BWV 263, BWV 354). Both melodies use similar compositional principals.
- The speech stimuli consisted of two distinct part of a poem ("A Visit from St. Nicholas," Moore or Livingston, 1823), and the two parts were recorded by a professional voice actor.
- Participants were provided with the stimuli a few days before the recording to practice and prepare for the imagery task.
- 10 listening and 10 imagery trials per stimulus presented in a randomized order.
- To ensure that participants performed the mental imagery task with high temporal precision, a visual metronome marking the start of 120 bpm bars was presented on the screen.



Dry-EEG Experiment Setup

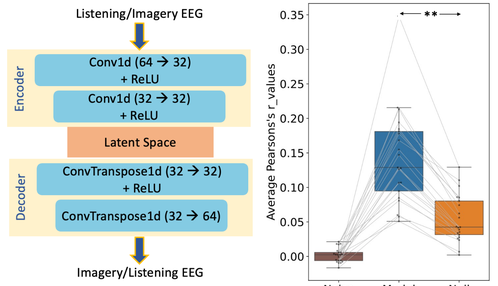
- Recordings of listened and imagined speech were made using CGX Quick-32r Dry-EEG wireless headset.
- The experiment included 3 participants (1 female and 2 males) who listened and imagined six words in five trials.

Non-linear Mapping Between Listening and Imagining Responses

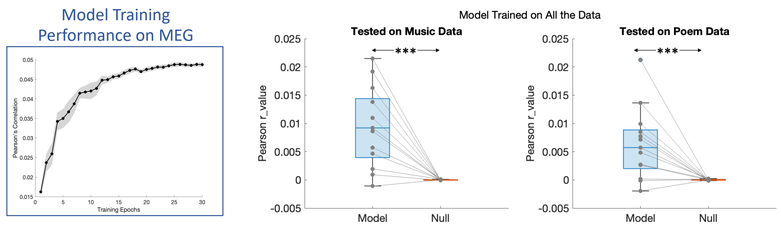
1. The EEG data from Marion et al, 2021 were recorded from 21 musicians while they listened to and imagined four melodies from Bach chorales. We attempted to find a mapping to predict the imagery EEG responses from their counterparts during the listening task. We specifically aimed for a mapping that performs robustly on unseen subject data. A non-linear encoder-decoder architecture was used to map the listening EEG to the reconstructed imagery EEG.

The mapping performance shown below is evaluated using the Pearson correlation between reconstructed and true imagery. The model significantly outperformed the null model generated by shuffling the stimulus labels across trials (t-test p-value << 0.05).

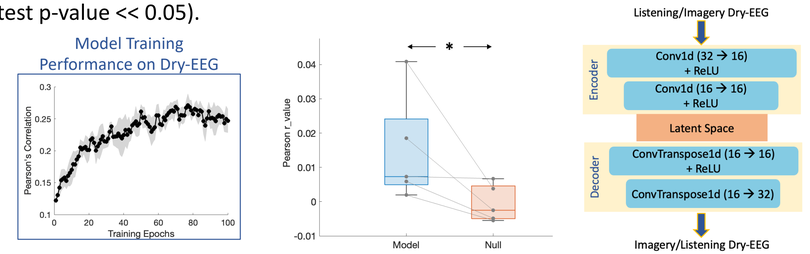
2. Similar mappings were then computed for music with MEG recordings and shown subsequently to perform well also on speech recordings. We trained the networks here using all music and speech data from all subjects except for one left-out to examine the network performance with unseen data.



The mapping below significantly outperformed the null model generated by shuffling the stimulus labels across trials (t-test p-value << 0.005).

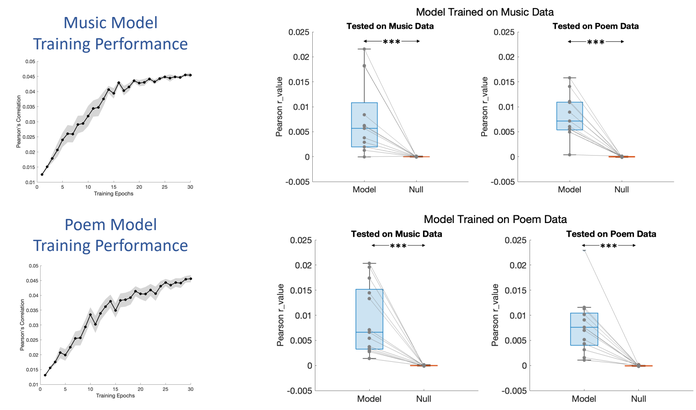


3. A two-layer encoder decoder convolutional neural network was then trained in a similar way on the dry-EEG recordings to map the listening signals to their corresponding imagery for speech in the form of words. Compared to the previous recordings, this experiment had far fewer subject and number of trials. The network is evaluated based on the correlation between the true imagery of each word with its corresponding reconstructed imagery. For 5 words (out of 6) the Pearson correlation value of the model was significantly higher than the null (Permutation test p-value << 0.05).



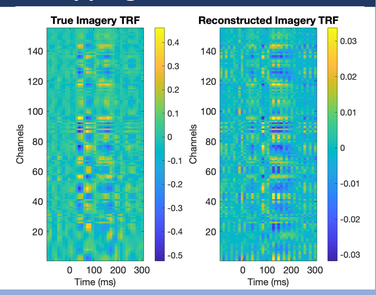
Transfer of MEG Mappings from Music to Poem and Vice Versa

To evaluate the generalizability of the mappings, performance of the model trained on MEG music was tested on unseen MEG poem responses and vice versa. These models were also trained on all subjects except one left-out and were tested on the left-out subject's responses. The models significantly outperformed the null model based on Pearson's correlation values (p-value << 0.005).



TRFs of the Non-linear Mappings

To visualize the non-linear mappings, we calculated the TRFs of the reconstructed imagery signals. The correspondence between the two sets of TRFs (here shown for the MEG music data) is remarkable demonstrating the similarity between the true imagination recordings and their predictions from the listening recordings.



Conclusion

We were able to successfully train small encored-decoder convolutional neural networks to map the listening and imagery responses to their counterparts. These mappings were successful in three different settings (EEG, MEG, and Dry-EEG), as well as in both music and speech stimuli. They were also generalizable to recordings from all subjects. Another significant finding was that the mapping computed with music or speech responses in MEG were equally effective for both, indicating that the nature of the measured responses was similar and likely reflected an auditory representation in early auditory cortical stages.

References

1. Marion, Guilhem, Giovanni M. Di Liberto, and Shihab A. Shamma. "The music of silence: part I: responses to musical imagery encode melodic expectations and acoustics." *Journal of Neuroscience* 41, no. 35 (2021): 7435-7448.
2. Rezaiezadeh, M., Decoding the brain in complex auditory environments Doctoral Dissertation, University of Maryland (2022)