NeuroView: Generative Visualization of the Diversity of Brain Responses to Jazz

Brandon S. Coventry
Department of Neurological Surgery
Wisconsin Institute for Translational Neuroengineering
University of Wisconsin-Madison
Madison, WI USA 53705
coventry@wisc.edu

Claudia Krogmeier
National Institute for Research in Digital Science and Technology (INRIA)
Université de Bordeaux
Talence, FR
claudia.krogmeier@inria.fr

1 Description of Proposed Project

1.1 Artist Statement

How we perceive the world around us is intrinsically linked to the environments in which we live, the people with whom we interact, and the experiences we've had. This subjective reality has explained in part why we like the music we do, which films make us cry, and how certain smells can so quickly bring us back to key moments in our lives. A monumental discovery in neuroscience is that these subjective experiences we share can in part be measured through electroencelography (EEG). EEG is a non-invasive technique which utilizes electrodes placed on the surface of the head to measure electric fields resulting from activity of collections of neurons acting in concert. These electrodes are positioned across the entire head, allowing for measurement of different neural structures related to diverse activities such as auditory processing, volitional movement, and visual processing, among many more (Fig 1).

EEG has thus allowed for scientific probing of one of humanities most intimate and salient traits: emotion. Specifically, through EEG recordings, neuroscience has found that markers of emotion, specifically the intensity (arousal) and pleasantness (valence) of emotion can be measured in a region of the brain called the prefrontal cortex[1]. Specifically, valence and arousal are correlated to the relative difference in alpha band (signals with frequencies between 8-12Hz) EEG activity of the left and right prefrontal cortex[2]. While these recordings do not capture the entirety of the complex emotional process, artists can use this tool to graphically represent and illustrate personal receptions of media. To this end, Claudia has been exploring the use of EEG activity to create novel experimental films in brain-computer interface settings[3].

We have also been enamoured with the ability of autonomous and semi-autonomous systems to generate art which, by its very nature, is external to the human process of creation. The advent of generative deep learning models has given rise to an exponential leap in creative capabilities of computational systems.

Enticed by the ability to visualize and explore our emotions in new ways mixed with exciting new generative tools, we sought to explore the diversity of our collective experiences to music listening by marrying EEG emotional processing techniques with deep generative networks. As such, we present NeuroView, an exhibit and generative art system with the following aims:

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

- To explore similarities and differences in the reception of music through use of EEG signals and deep generative networks.
- To create surrealist videos generated only using participant brain waves to highlight unique participant listening experiences
- To create an open source framework for others to freely generate EEG-based artistic works.
- To create and present an exhibit showcasing EEG-based generative art which celebrates humanity's collective experience of music.

1.2 Media and Creativity

Generative-adversarial networks (GANs) have proven to be a transformative technology not only for artificial intelligence, but also for the creation of art. GANs have allowed for the formally impossible; creating works which imitate Beethoven and Bach, creation of alien visual landscapes from user text inputs, and completing unfinished works from artists long passed. As such, we present NeuroView, a gallery and software framework which marries generative video art with EEG signal processing to generate novel videos using listener's thought patterns alone.

The heart of NeuroView is a vector-quantized generative adversarial network (VQGAN)[4], a combined transformer and convolutional neural network facilitating the efficient generation of high resolution images in a way that mimics biological image processing. Recently, VQGAN has been modified with contrasting language-image pretraining (CLIP)[5,6] allowing for robust text to image generation. In this work, we modified CLIP-VQGAN to accept inputs from time series EEG data to generate novel videos directly from human neural firing patterns.

As emotional valence and arousal can be directly modulated from sensory stimulation, we decided to visualize the diversity of EEG activity which is elicited when individuals listen to avant-garde jazz music. We chose music from Greg Spero's Spirit Fingers [1], a jazz group pushing the frontiers of avant-garde music. In this work, CLIP-VQGAN was trained on ImageNet with trained models used to generate novel videos with EEG input. Participants in the generation of this exhibit are equipped with a commercial-grade dry EEG setup and are asked to listen Spirit Fingers "Nails"[2]. During EEG acquisition, frontal alpha asymmetries (FAA), the difference of right and left side prefrontal cortex activity, is fed directly into CLIP-VQGAN and a new video is generated.

For this exhibit, we will acquire EEG-generated video from a diverse group of participants. The primary exhibit will consist of currated videos organized in a 2x3 or 2x4 grid (Fig 2.) Each tile represents generated video elicited from FAA while listening to jazz from different individuals. As such, the videos will be synched to Spirit Fingers' Nails which will also play as part of the exhibit. We will also bring printed sheets detailing how the videos were formed and information on the history of jazz.

We also do not constrain video generation to scenes which are easily understood, and instead hope for surrealist videos. This design decision was made to help viewers enter into the uncanny valley of thought and neural processes. We aim for these surrealist videos to help viewers think deeply about the nature of thought and emotion, and to participate in the diversity of perception present to humanity.

A secondary component of our exhibit will involve the use of a simple, commercial grade 2 channel EEG device which can be quickly donned, cleaned, and setup so that participants can generate their own EEG-videos which will be displayed on a second screen.

We have already started acquiring participant EEG-video (See video portfolio) and will continue to obtain more videos through the summer of 2023 (see Fig 3).

[1] https://www.gregspero.com/spirit-fingers; https://downbeat.com/news/detail/greg-spero-commands-a-multitude-of-styles

[2] Nails: `https://www.google.com/search?q=Spirit+Fingers+Nails&rlz=1C1GCEA_enUS1023US1023&oq=Spirit+Fingers+Nails&aqs=chrome..69i57j35i39.3119j0j4&sourceid=chrome&ie=UTF-8#fpstate=ive&vld=cid:3a4e0376,vid:87WbDGGzJNg`

1.3 Brief Technical Description

1.3.1 EEG Acquisition and Computational Systems

EEGs were recorded using a CGX Quick32 dry EEG system. Participants are seated during recordings and music played free-field through a speaker. Electrode impedances were checked to ensure good signal quality before recording. Participants are given 30 seconds before onset of music to reach a stable EEG baseline. Participants are instructed to remain as still as possible during the recording to prevent excess motion artifact.

AI models were implemented in Pytorch V2.0. Training of models was performed on a custom deep learning workstation with an Intel Xeon CPU and Nvidia Titan X GPU.

1.3.2 EEG Processing

EEGs are sampled at 500 Hz, allowing for time frequency decomposition down to $< 250Hz$. Signals were low pass filtered to remove high frequency ringing. Changes in mean alpha (8-12Hz) EEG power were calculated by short-time fourier transforms using a hamming window of equally placed time segments. Frontal alpha asymmetry (FAA) was calculated from time-binned mean alpha power above left side (F3) and right side (F4) prefrontal cortex electrodes. FAAs in each time bin were calculated as:

$$FAA(t) = ln(\overline{F4(t)}) - ln(\overline{F3(t)}) \quad (1)$$

where $\overline{F3}, \overline{F4}$ represent the mean alpha power of electrodes F3,F4 respectively. Timeseries FAAs were then fed as inputs to CLIP-VQGAN for image generation. Resulting videos were generated from output image frames of CLIP-VQGAN and combined to a film using OpenCV-Python.

2 NeuroView and the Theme of Diversity

In this exhibit, we aim to illuminate and celebrate the differences that arise in participating in the experience of music. Participants in this project are based in Madison WI and Bourdeux, France and come from a multiplicity diverse backgrounds with varied tastes in music and media. The organization of our exhibited videos (Fig 2) also aims to show that, while our individual experience of music may differ, our collective perceptions create a beautiful tapestry of the human experience.

Jazz also represents a fusion of diverse cultures and experiences while also being reflective of the general human experience. Originally formed in the African American communities of New Orleans, jazz strongly reflects the communities in which it is played, with unique forms arising in New York, Minneapolis, New Orleans, and Los Angeles. What unites all jazz is the drive for the musicians to collectively form a voice through their instrument while also listening and supporting fellow musicians in improvisational jam sessions. As such, jazz takes influence from all people who are open to its form and creates from it something more. It is for these reasons we chose to use jazz as the stimulus for our EEG recordings.

Taken together, NeuroView aims to explore how differences in brain activity can be visualized in ways that highlight this diversity of emotional experience as something beautiful to be celebrated and contribute to our collective humanity.

3 Project Schedule and Readiness

We are confident that our proposed exhibit will be complete for NeurIPS 2023. To illustrate work which has been completed, work to be done, and stretch goals that are not strictly necessary for project success but would add to the exhibit are presented in a Gantt chart (Fig 3). Critical drains on project time largely lie in training time for CLIP-VQGAN. The model is fully trained and accepting input EEGs and is able to process data as it is obtained. We have actively begun recruiting participants to contribute videos to the exhibit, and hope to have between 4-10 videos in the exhibit. Participant recordings require only 20-30 minutes of time and will be completed during summer of 2023. Videos will be generated during participant recordings and thus will not consume a considerable amount of time. A large recruitment window was allocated so that we could recruit a diverse set of participants. After generation of all videos for the experiment, videos will be edited to exhibit format (See Media and Creativity) which can be easily displayed using a a screen and personal laptop.

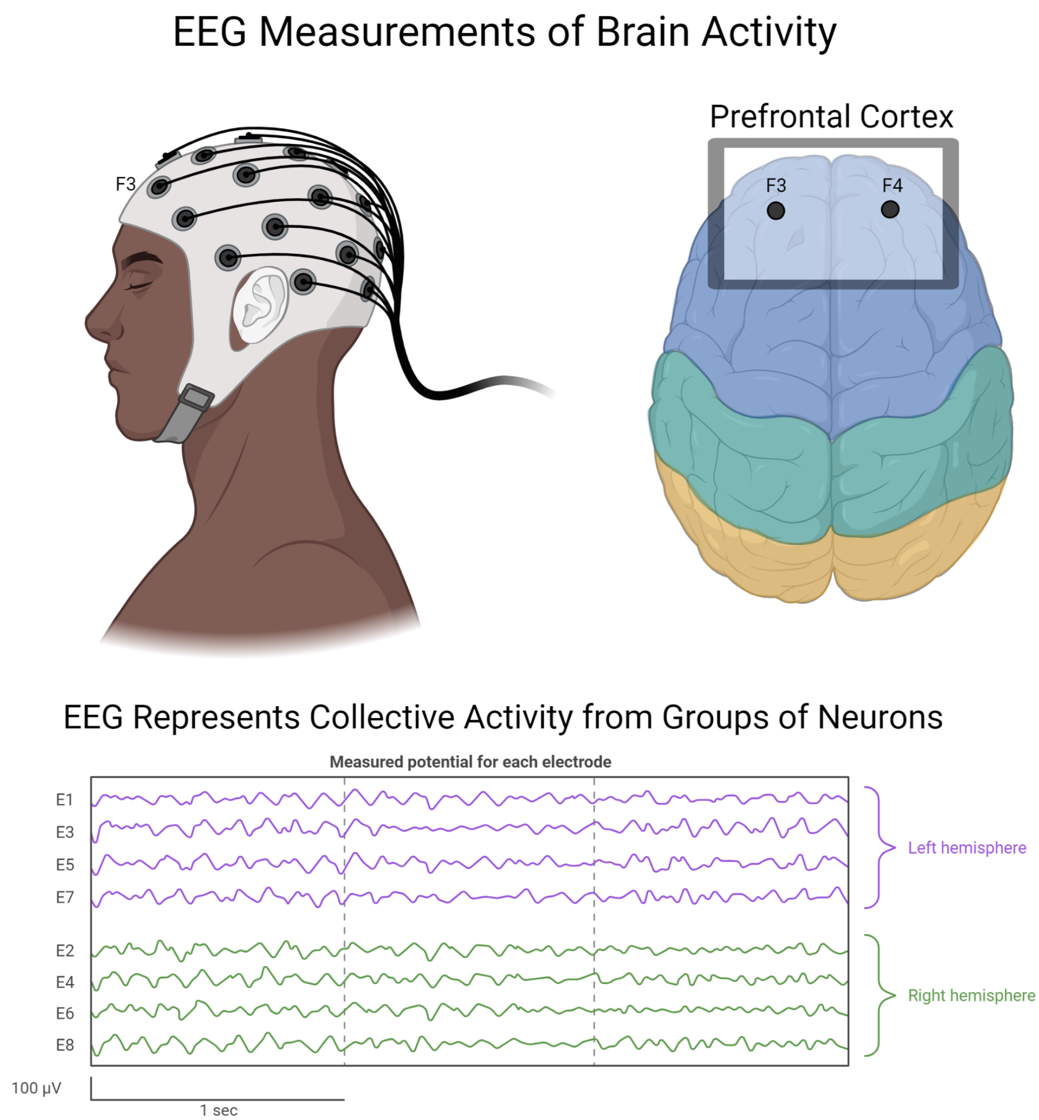

Figure 1: Measuring brain activity using EEG. EEG represents the collective activity of groups of neurons recorded on scalp electrodes (left). Prefrontal cortex is a portion of the frontal lobe involved with emotional processing which can be measured using EEG electrodes F3 and F4 (Right). We can visualize the electrical potentials from surface electrodes (bottom). Image generated using BioRender.

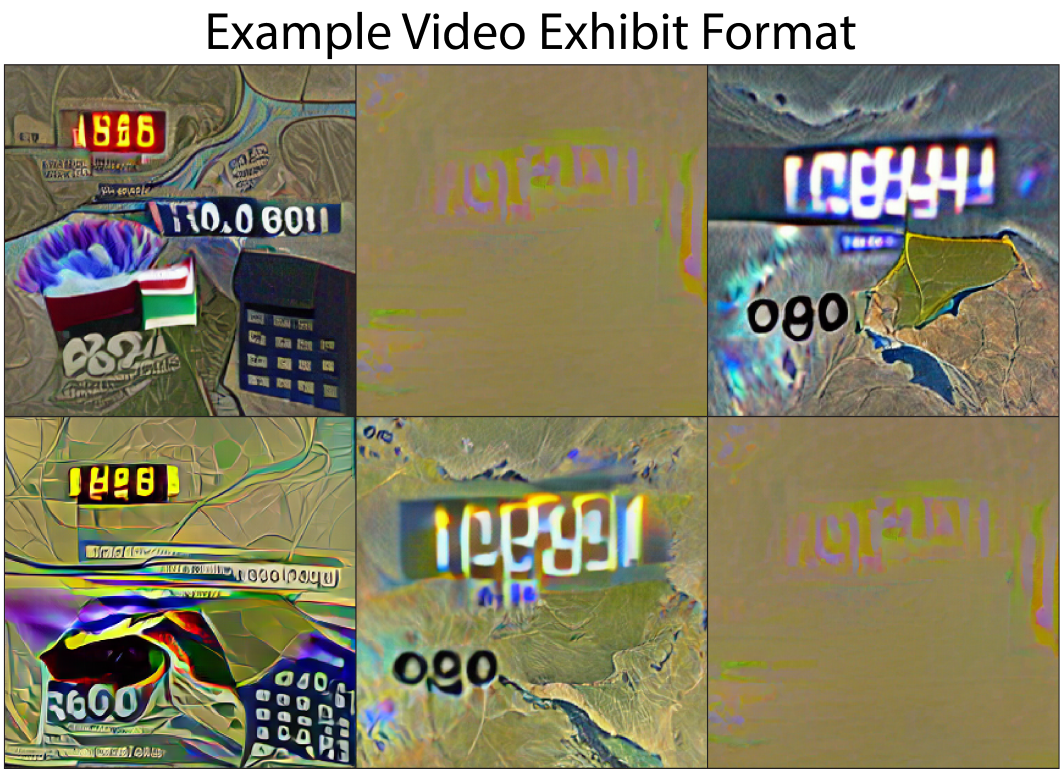

Figure 2: Example video exhibit format. Our video exhibit will display 4-10 concurrent videos generated in response to EEG activity elicit from avant-garde jazz. Videos and song will be synced and played concurrently. Each video tile will be derived from EEGs of different participants.

We have also added in several stretch goals that would add to the project, but are not contingent on project success. The first of these is to acquire and label our own image sets to facilitate a more unique visual experience. As most AI-art based models (including ours in its current state) use similar training sets, the output images and videos can sometimes have a similar artistic style. We hope to create more unique experiences by applying transfer learning on our CLIP-VQGAN model with our own labeled image sets. As such, we have allocated more training and EEG processing time if needed.

Based on our schedule, our exhibit will be completed by October of 2023 which allows for extra preparation and processing time as needed.

4 Copyright and Licenses

In the spirit of open source culture and to facilitate further artistic creativity, we will release all AI models and EEG-video generation code at the conclusion of NeurIPS 2023 at the following Github repository: `https://github.com/bscoventry/NeurIPS_NeuroView`. All models and video generation code will be released under an MIT License.

Conference attendees who participate in the exhibit will be granted co-copyright with Coventry and Krogmeier and will be allowed to disseminate their videos as they see fit.

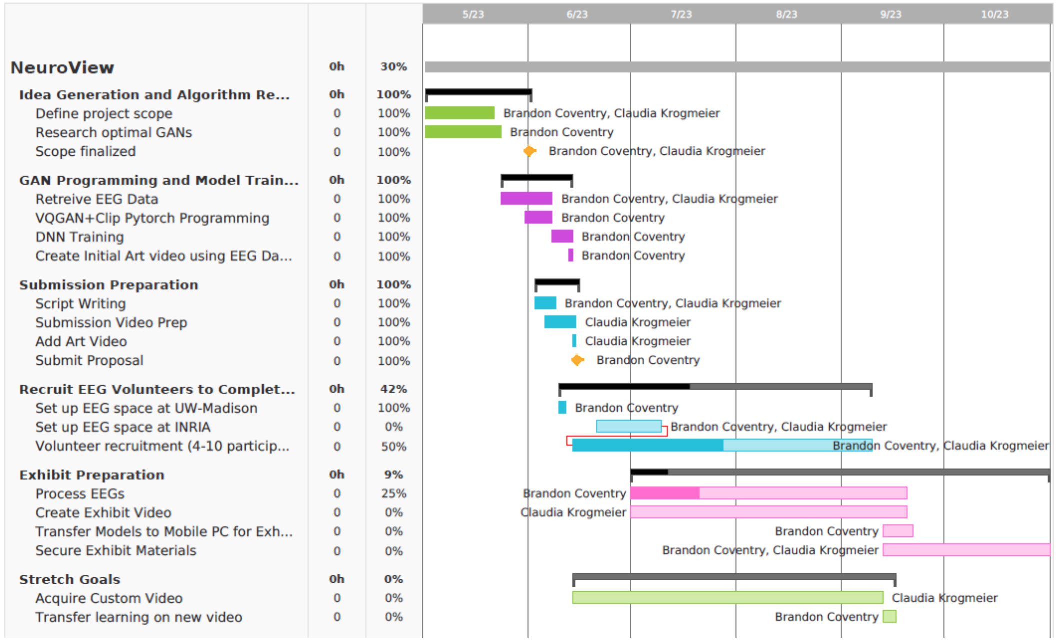

Figure 3: NeuroView Gantt chart detailing completed and future work.

5 Safety, Security, and Ethics

5.1 Security of EEG Signals

While the use of biosignals can present potential issues related to personal identification and revelation of medical diagnoses, this project was designed to ensure that the use of EEG presents no potential of disclosure of medical or identifiable features.

First, devices used to acquire EEGs used in our exhibit are non-medical grade, lacking sensitivity necessary to discern the minutia of relevant medical information. Furthermore, diagnostic EEG relies on patients in hospital environments in well controlled conditions which require the aid of other medical diagnostic tests to identify disease. Second, data is obtained only from 1-2 surface electrodes. Along with inherent variability and noise in EEG signals, it is not possible to use these signals to specifically identify a participant or reveal sensitive information such as sex, ethnicity, medical status, etc. Furthermore, as visualized images are generated from but do not contain EEG data, participants are at no risk of identification through viewing of output video.

Even though our hardware prevents malicious use of biosignals, we will take extra steps to absolutely protect participant identities and biosignals. First of all, EEG signals which are acquired in real time are processed and passed directly to VQGAN for image generation. Raw EEG signals and processed frontal alpha asymmetries are immediately discarded from memory and the protected python class and are not kept for future retrieval. Second, conference participants who wish to keep their generated videos will have their videos de-identified by a randomized naming scheme and sent to participants via secure data transfer links through the University of Wisconsin-Madison. In this way, we aim to go above and beyond to protect all participant identities.

5.2 Ensuring Safe Image Generation

Generative video has potential to generate image sequences which might be considered obscene or offensive. To ensure that offensive video sequences do not emerge, our implementation of VQGAN was trained on newer ImageNet databases which have been filtered of illicit content[3]. Furthermore, any video sequences that we augement VQGAN with will be free of people and illicit content. Videos

[3]Yang et al, Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy. https://image-net.org/update-sep-17-2019

which will be used in our primary exhibit will also be screened for illicit image generation as a safety precaution.

6 Resources and Sponsors

At present, NeuroView has no internal or external sponsors. If accepted, Brandon will be applying for a small equipment grant through UW-Madison to facilitate purchase of consumer grade EEG systems for use at NeurIPS.

7 Biographies

Brandon S. Coventry received a bachelors degree in Electrical Engineering from Saint Louis University and a masters degree in Electrical and Computer Engineering from Purdue University West Lafayette. Afterwards, Brandon completed his PhD in Biomedical Engineering from Purdue University West Lafayette where he specialized in neural engineering and artificial intelligence. He is now a postdoctoral research associate at the University of Wisconsin-Madison. Brandon's research interests largely involve understanding the mechanisms of electrical and optical deep brain stimulation and its clinical use for patients with Parkinson's disease and obsessive compulsive disorder. Brandon also works to create intelligent deep brain stimulation systems which utilize deep reinforcement learning to derive statistical models of patient brain activity in response to clinical stimulation in real time. Before starting on his engineering path, Brandon was a professional musician touring the United States with a multiplicity of bands and musical acts. During this time, he also studied jazz performance (guitar and double bass) at the University of Illinois Urbana-Champaign. In his spare time, Brandon still plays live as much as humanly possible and also composes film scores for a variety of short films and art videos.

Claudia Krogmeier is a video artist and creative technology researcher currently based in Bordeaux, France. She received bachelors degrees in Telecommunications and French from Indiana University Bloomington and a masters degree in Computer Graphics Technology from Purdue University West Lafayette. She then received a PhD in Technology from Purdue University where she specialized in creative technology and emotionally-responsive brain computer interfaces. She is currently a postdoctoral researcher at INRIA in Bordeaux, France. Her work is designed to trigger curiosity and propose unusual perspectives. Organic "special" effects which arise naturally during filming or editing are often highlighted rather than dismissed in her final work. She is inspired by ideas of surrealism, nonlinear narratives, the brain, and experiences of emotion. Her next creative work will explore the discontinuity of relationships within virtual social networks, and ideas of narrative identity.

Brandon and Claudia have collaborated on a multiplicity of creative and technological projects which have been seen around the world and resulted in several publications.

References

[1] Bestelmeyer PEG, Kotz SA, & Belin P (2017). Effects of emotional valence and arousal on the voice perception network. *Soc Cogn Affect Neurosci.*, 12(8):1351-1358. doi: 10.1093/scan/nsx059.

[2] Krogmeier C, Coventry BS, & Mousas C (2022). Frontal Alpha Asymmetry Interaction with an Experimental Story EEG Brain-Computer Interface. *Frontiers in Human Neuroscience*. Vol 16. DOI: 10.3389/fnhum.2022.883467

[3] Krogmeier C, Garcia Bravo E, & Mousas C (2023). Using Experimental Filmmaking to Create an Engaging Brain-Computer Interface. *Proceedings of the ACM on Computer Graphics and Interactive Techniques (Proc. of ACM SIGGRAPH 2023 Art Papers)*. Vol. 6, No. 3.

[4] Esser P, Rombach R, & Ommer, B (2021) Taming Transformers for High-Resolution Image Synthesis. *CVPR* pp. 12873-12883

[5] Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, & Sutskever I (2021). Learning Transferable Visual Models From Natural Language Supervision. *arXiv*. DOI: 10.48550/arXiv.2103.00020

[6] Crowson K, Biderman S, Kornis D, Stander D, Hallahan E, Castricato L, & Raff E (2022). VQGAN-CLIP: Open Domain Image Generation and Editing with Natural Language Guidance. *arXiv*. DOI: 10.48550/arXiv.2204.08583