Neural Scores: Interim Project Report

Tyler H. McIntosh

University of Greenwich

May 10, 2021

Abstract

The project discussed in this report uses emerging creative technologies to explore the relationship between music and emotions. The aim is to create a new artistic medium in which emotional responses to music are visually depicted. The project also aims to develop and create a new AI interface for the classification of emotions in EEG data, for use in further artistic projects.

Keywords
Music, Emotions, EEG, AI, Machine Learning

1 Introduction

Music is seen to have a great deal of effect on human emotions. In the progression of a classical movement, a composer can raise or lower the emotions of an audience at will. With the musical expression of an idea, an artist can inspire the will of the masses. As displayed by the Baltic singing revolution, the power of a song can unify a country under a common goal (Smidchens, 2014). While, as discussed in *"The auditory culture reader"*, the repetition of a theme can be exploited as a form or torture (Bull and Back, 2020). Whether the source is symbolism or just mere exposure, the emotional effect music has on people is observably powerful.

Despite this, it is arguably not typical or commonplace for people to deeply consider or investigate their emotional responses to a piece of music. This is reflected in the ways music is visually represented; the focus lies on the qualities of the sound or the musical notation, rather than the emotions it invokes in the listener. This lack of illustrative medium leaves a conceptual gap in the way people visualise the subjective qualities of music. The project discussed in this report aims to bridge this gap by generating both an interface and a medium in which emotional responses to music are captured and visually represented – in a similar light to traditional forms such as sheet music, graphical scores, and audio-reactive visualisations.

In order to facilitate a medium in which emotional responses are visualised, an interface that can capture these responses must first be constructed. However, this is a reasonably difficult thing to do; not only can there be a requirement for expensive recording equipment, but an appropriate method of data classification is also required (Mauss and Robinson, 2009). However, recent advancements in the commercial availability of consumer electroencephalography (EEG) equipment has made this problem significantly more accommodating. Therefore, this project suggests that, by using advanced machine learning techniques, it is possible to train and implement artificial intelligence (AI) models that can recognise patterns and features in EEG data that describe emotional responses.

Through the use of these emerging creative technologies, the project hopes to contribute to the discussion of AI as a creative platform – while also exploring new forms and methods of music visualisation and interaction. The research generated by this project is built on the question: "How can we model the relationship between music and emotions, using electroencephalography and machine learning". It is also the hope of the investigator to create a new interface for human-computer interaction, drawing on inspiration from the work and discussions of the *"Creative AI lab"* (Bunz and Jäger, 2021).

2 Emotion Measurement

The task of measuring emotional responses is by no means an easy one; there are many intricate elements to consider when conducting such experiments, and any endeavours should be well-informed. As described in *"Measures of emotion"*, there are many techniques whereby an investigator may measure identifiers of emotional states - each of which entailing their own advantages and disadvantages (Kaplan, Dalal and Lunchman, 2013). While methods such as self-reporting and body-language observation may provide seemingly satisfactory results, they often accompany some degree of unintentional bias (Kaplan, Dalal and Lunchman, 2013). This is where more transparency, quantitative psychophysiological methods become effective (Kaplan, Dalal and Lunchman, 2013).

While emotions are typically described as single or multiple instances of discrete events (i.e., happy and sad or just happy), many studies suggest they are actually a convergence of various central nervous system (CNS) signals (Mauss and Robinson, 2009). It is recommended in *"Measures of emotion: A Review"* that, when approaching emotion measurement, an investigator should base their analysis in dimensions of valence and activation – rather than attempting to capture emotions as discrete states (Mauss and Robinson, 2009). This concept is further explored in *"The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology"*, where it is outlined that valence and activation exist as two discrete systems of the CNS (Posner, Russell, and Peterson, 2005). The article further suggests that the intensity of these two systems can be plotted to create a 2-dimensional model of affective experience (Posner, Russell, and Peterson, 2005). The existence of this valence-activation relationship enables an investigator to ground their measurements in two absolute values, rather than examining arbitrary components and features of individual emotional states.

The recognition and adoption of the valence-activation model is prevalent throughout recent studies and practical examples. There are numerous instances where the valence-activation model has been used to evaluate measurements from various sources. Moreover, the psychophysiological recording method electroencephalography (EEG) is frequently used and referenced as a viable source of data for the retrieval of valence-activation information. In a project report by eNTERFACE, titled *"Emotion Detection in the Loop from Brain Signals and Facial Images"*, the investigators use this technology in their measurement of valence-activation values (Savran et al., 2006). While this study provides affirming results for the use of EEG as a valence-activation measurement device, it rightly acknowledges the problems associated with EEG based measurements – namely, electronic interference or "noise" (Savran et al., 2006). The presence of noise in EEG data is relatively commonplace, as the electrodes used are extremely sensitive to the voltage fluctuations caused by muscle movement. This means actions like jaw clenching, blinking, and smiling can cause abnormalities and spikes in the data (Savran et al., 2006). While there is no attempt to filter the recordings in this example, it is absolutely necessary to clean the incoming data to prevent inaccurate or contaminated results (Jiang, Bian and Tian, 2019). However, in the case of the eNTERFACE project, this was considered in the methodology and steps were put in place to reduce the impact.

3 Data Classification

Arguably, the most significant challenge when attempting to measure emotion state identifiers is finding a way to extract the relevant information from the raw/filtered data. There are many theories and suggestions on how to do this, in regard to EEG data, but the emerging consensus appear to focus on machine learning – specifically, neural networks. This is because, as explained in *"Neural Networks and Deep Learning"*, neural networks possess an immense ability to intelligently learn the rules of a training dataset – especially when the training dataset is given identifying labels (Aggarwal, 2018). This "black-box" learning process makes them ideal for situations where manually implementing all the rules of a relationship is not feasible (Aggarwal, 2018) – such as in stock market predictions and self-driving cars. Furthermore, the node-based structural nature of neural networks makes them extremely versatile and robust in their trained application (Aggarwal, 2018). These factors make neural networks a perfect candidate for a task such as the valence-activation retrieval of EEG data.

While standard neural networks are excellent at understanding the rules and relationships of data in linear instances, they lack the ability to analyse temporal and spatial patterns. This is where more complex structures such as convolutional neural networks (CNN) and long short-term memory (LSTM) networks become necessary (Aggarwal, 2018). This is a relevant issue in the retrieval of valence-activation values from EEG data as the network would be required to assess signal patterns over time, rather than just the values reported every time the headset updated. This problem was addressed in *"A Study on Mental State Classification using EEG-based Brain-Machine Interface"* in which the investigators applied a short-time windowing technique to the data for time-series statistical feature extraction (Bird et al., 2018). This method gives the network a sense of temporal context and allows it to make more general and reasonable predictions of mental state over a given time. While this method produces highly accurate results, a similar outcome can be achieved by converting the data to an image, such as a spectrogram, and classifying it with a CNN.

4 Related Work

As stated in the introduction, this project aims to create a new medium in which the emotional responses to music are visualised. There are many different ways and formats in which this could be achieved – for instance, real-time reactive visualisations, generative paintings, and procedural drawings. Each of these carries their own significance to the artistic impact of the project. A source of particular inspiration has been the work of Javier Casadidio. Using tools such as TouchDesigner, Casadidio produces generative artworks and posts them online (Casadidio, 2021). A number of these artworks are dynamic posters that retain a still image until interacted with, then becoming videos (Casadidio, 2021). This is an interesting way of displaying art because it allows the artists to produce a still image while also enabling them to show the inner workings of the generative process.

As a secondary goal, this project aims to create a new interface for human-computer interaction – one that can capture human emotions, using machine learning, and output them to any other creative applications. This goal was developed in response to a panel discussion titled *"Aesthetics of New AI Interfaces"* (Bunz et al., 2021). In this discussion, speakers commented on how developers could use AI to create new modes of human-computer interaction and the forms in which such interfaces could manifest (Bunz et al., 2021). It was concluded that while one might assume artists aren't interested in breaking open the "black-box" of machine learning, they regularly rise to the challenge and bend existing programs to create new forms of art (Bunz et al., 2021). This is an important concept to internalise when designing an artistic interface; the more data you make available to the user, the more artistic freedom they have.

In a commissioned work funded by the Saatchi Saatchi Wellness agency, the studio *"Random Quark"* lead a project named *"The Art of Feeling"* (random quark, 2017). In this project, the creators used EEG to capture the emotions of participants as they recalled a particularly emotional moment in their life (random quark, 2017). These emotions were painted to a canvas, using a particle system, and displayed in a gallery (random quark, 2017). While this project was highly successful, the method used to calculate emotions is arguably somewhat misleading. According to the source code, the data collector calculated a valence value by comparing the amplification lateralization of the alpha bandwidth, and an activation value from the average amplitude of the alpha and gamma bandwidths (random quark, 2017). While these calculations have some basis in scientific reasoning, the implementation is likely too rudimentary and rigid to accurately reflect the emotion state of the participant (Altenmüller, 2002). However, given the artistic aims of the project, this was likely of minor concern, as it did not have a negative effect on the visual or conceptual output.

5 Methodology

The methodology proposed by this project is primarily inferred from the related literature. However, due to the emerging nature of the research field explored by this project, there are methods for data classification used which are unprecedented and experimental. A brief explanation on why these methods were chosen and why they are effective will be given.

5.1 Participants & Materials

As this project relies on people as a source of primary data, participants make up an important part of the planned procedure. While there are no explicit requirements regarding the selection of participants, there are other constraints that must be noted. Namely, due to the current national coronavirus lockdown restrictions, any participant interaction will be limited to an opportunity sample of members of the investigator's household. These are the only candidates that it is safe and legal to involve. Furthermore, according to the NHS website, EEG procedures carry a very small risk of causing seizures in people with epilepsy (NHS, 2018). Therefore, the selection process will not involve anyone with a history of seizures.

The project uses EEG technology to record brainwave data from participants, for the purpose of training and classification. As demonstrated in the related literature, brainwave data contains the encoded values of emotional valence and activation; it is the job of the investigator to implement a way to decode these values. In this case, the Muse: Brain Sensing Headband is used. The Muse headband is a consumer grade EEG device designed for meditation activities. However, it comes with a suite of developer tools for recording and playback purposes. This device was chosen for its low cost and high resolution –

supplying 4 independent channels with the electrode positions: TP9, AF7, AF8, TP10 (Bird et al., 2018). The placement of these electrodes provides a spatial context of brainwave signals – as well as raw values.

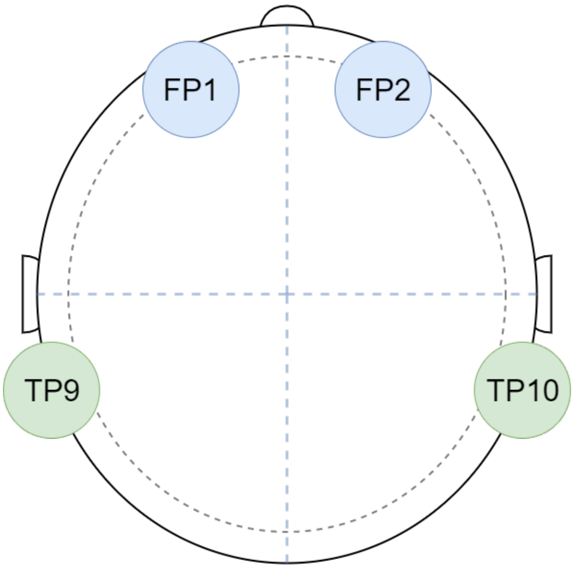

Figure 1: Muse headband electrode placement

To interface with the Muse EEG headband, the *"Mind Monitor"* mobile app is used. The Mind Monitor app acts as a middle-man between the headset and the end user and can interpret the raw EEG data transmitted by the headset. The app also analyses the EEG data to create a profile for the current user, which is used to adaptively normalize the signals (Clutterbuck, 2021). A Fast Fourier Transform (FFT) is then applied, to convert the signals from the time domain to the frequency domain – showing how much activity lies in a frequency bandwidth, rather than the frequency of the oscillations themselves (Clutterbuck, 2021). This information is then broadcasted to the host machine, over the Open Sound Control (OSC) protocol.

5.2 Data Collection

As the project is a data driven one, the process of data collection is an important part to get right. Therefore, reasonable attention has gone into the procedure for preparing a participant for recording and preparing the equipment for use. Before recording takes place, it is important to make sure the participant knows to not move their head or face and to keep their eyes closed – any one of these can potentially cause abnormalities and biases in the data (Boytsova and Danko, 2010; Bird et al., 2018). It is also important to slightly dampen the electrodes before placing the headset on the participant, as this increases the conductivity around the contact points, resulting in a cleaner signal. Once all preparation steps are complete, the headset is placed on the participant, and the session can begin. Recordings are made on the host machine, using the *"MusePlayer"* developer tool.

5.2.1 Training Data

For the purpose of accuracy and precision, it is important to record training data that is specific to each participant. The training data consists of four 3 minute recordings that each represent a different extreme of the valence-activation scale. This data is used to train the internal tooling of the emotion measurement system by building a profile of the patterns specific to the participant – much like a fingerprint. While recording, the participant is played music that the investigator has chosen to represent happiness, sadness, excitement, and calmness. A library is prepared, but the participant is welcome to suggest alternatives if they believe it would provoke a stronger response. There is a 20 second priming period before the recording starts, to allow the participant to enter the desired mood. The participant is also encouraged to recall thoughts and memories that make them feel the emotions that are currently being captured (in attempt to cross-validate the emotion state). There are a few minutes rest between the recording sessions, to allow the participant to return to a typical emotion state.

This method was chosen as it accounts for interpersonal variations of standard mood and predisposition (that is, a typically angry or sad person) – as raised in *"Measurement of emotions"* (Kaplan, Dalal and Lunchman, 2013). By building individual profiles, the system can make more accurate predictions of mood that reflect a variation from the typical state.

5.2.2 Experiment Data

In this section, data that will be used by the experiment to generate a visual output is recorded. The experiment data consists of four recordings that are taken as the participant is listening to a piece of music. The recording starts when the music starts and ends when the music ends. The resulting file is a recording that can be replayed in-sync with the source music, which is important for the mapping of the visual output. There is a priming period of 20 seconds, where music of the same style and volume is played, to avoid any spikes in activation as the music begins. The music is selected from a library, or the participant can request their own music. There is no encouragement for particular thoughts or memories, as the aim is to capture the emotions the music makes the participant experience and not any specific emotion in particular.

5.3 Emotion Detection

As discussed earlier, standard neural networks lack the ability to analyse temporal and spatial patterns in data, due to their instance-based linear nature. Therefore, for the purposes of this project, a convolutional neural network (CNN) is used for the classification of valence and activation values. In short, CNNs are able to detect patterns in multidimensional data by grouping areas of close proximity; instead of viewing the data as one line, they examine it square by square. This makes them very effective for tasks that involve recognising features and patterns in a spatial context (Aggarwal, 2018). This is especially useful in the case of EEG analysis, as the network is able to interpret psychophysiological patterns, such as hemispherical activation and lateralization, natively.

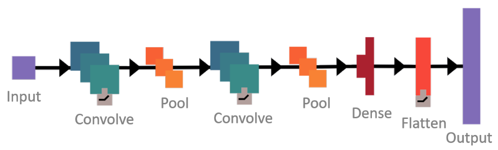

Figure 2: Proposed Neural Network Structure

5.3.1 Data Formatting

Before any data can be interpreted by the network, it must first be formatted in a way that the network can understand. This is done in real-time as the recordings are played back, whether it be in the training or the classification stage. The investigator starts playing back the recordings, using the MusePlayer developer tool. This sends the data over OSC to the Neural Scores application – a program made in the development of this project. The Neural Scores application stores a 3 second short-term memory of the EEG data it receives, which it then renders as a spectrogram. In the spectrogram, the channels are ordered by frequency bandwidth and assigned a colour (that is, delta becomes red). This provokes the network to differentiate between the bandwidths. Inside each colour band are the electrode positions (TP9, FP1, FP2, TP10). The colour displayed in each individual pixel is the amplitude of that frequency bandwidth, at that electrode, at that point in time.

Not only does this method provide a visual representation of the EEG data, but it also allows the system to perform image classification – a technique whereby a convolutional neural network analyses an image by inspecting the red, green, blue, and alpha values of each pixel. Furthermore, the sliding window nature of the spectrogram inherently contextualises the data in the time domain by presenting a running history of the values.

Figure 3: EEG spectrogram

5.3.2 Training & Classification

Before the network can be expected to make any classifications of the experiment data, it must first be taught how to examine the spectrograms and what to look for. Due to the black-box nature of machine learning, the network only requires a dataset that contains labelled examples of the parameters of its task – it will work out the rules by itself. This means the user can simply broadcast the training data to the Neural Scores application and give it a label for what it is receiving. This will convert all the streamed data into a single dataset, which can then be passed to the network for training. The training dataset can be created with live EEG data, but this is not recommended as there would be no opportunity for quality control. For the purpose of this experiment, as there are two independent variables, two CNN models are created for the exclusive classification of each valence and activation. Once all training is complete and the loss/accuracy functions are nominal, the trained model can be saved and exported.

The classification process operates in a similar way to the training setup, but, instead of storing the generated spectrogram images in a single dataset for training, they are instantly passed through the models for classification. Due to the low resolution of the spectrogram images and the nature of trained models, this process is very lightweight and does not require a lot of processing power. The system outputs a classification for positive valence or negative valence, and positive activation or negative activation. The system also produces a confidence score for each classification. The confidence score is typically a ranking of how sure the network is of its classification. However, in this situation, as the only other possibility is the direct opposite of the classification, the confidence ranking can also be seen as a linear intensity of the classification. The two confidence scores of each classification are processed to produce one value between -1 and 1 for each valence and activation. The values produced by the valence-activation classification system can be mapped directly to the circumplex model of affect proposed by Posner, Russel, and Peterson (2005). The values produced are also transported out of the Neural Scores application via OSC.

5.4 Visualisation

As the Neural Scores application outputs the valence and activation values over OSC, the data can be used in most other creative applications. As mentioned in the introduction, this project aims to create a new medium in which the emotional responses to music are depicted – the same way notation is depicted in sheet music or timbre is depicted in graphical scores. It will do this by using the classified valence and activation values in the production of procedural generative art. While the possible styles and methods of this are still being investigated, it is likely that the project will make use of applications such as TouchDesigner for the creation of generative art. It is also likely that the artworks will exist as still images, with a video counterpart – drawing on inspiration from the dynamic posters by Casadidio (2021). The images/videos will be displayed in a public gallery, or online, depending on the lockdown restrictions at the time.

6 Progress

In the development of this project, as mentioned in the methodology, an application was created by the investigator (see Appendix A). Not only does this program house all the neural network and data handling functions of the experiment, but it is also standalone; it doesn't require any other software to function. The investigator has also built and trained all neural network functions of the experiment and has deployed models in preparation for visualisation experiments. Furthermore, preliminary proof-of-concept visualisations have been created using the TouchDesigner software and the p5.js library for JavaScript (see Appendix B). The first prototypes were generated using a similar particle system to the Random Quark project and were created at the start of the project. The second prototypes were generated using TouchDesigner and were created after the Neural Scores application was deployed. Finally, an ethics application for the involvement of participants has been submitted and is pending review.

7 Evaluation

In the development of this project, many milestones were met, and significant challenges have been overcome. Despite starting with no knowledge of programming, not only has a neural network based emotion detection system been created, but it has been packaged into a lightweight standalone application that anyone with the same hardware can use. However, this achievement symbolises a significant pitfall the project has had so far; the artistic identity. While some consideration was given to the visual output at the start of the project, this aspect was neglected throughout the development of the emotion detection system. It is only recently that the artistic goal has been established; previously, the research question was "What form does music take in the mind". This was ambiguous and did not convey the purpose of the project. Since, the question has been refined to "How can we model the relationship between music and emotion, using electroencephalography and machine learning". This establishes both the purpose of the project and the interests it regards. Furthermore, it contains a description of the ideal artistic output of the project, with the word: "model". This announces the aim of the project; to create visual models that display the emotional responses people have to music.

While the system for valence-activation classification proposed by this project typically produces nominal loss/accuracy values in training, it is important to note its limitations. While it is statistically likely the results from the classification are accurate, they can only be as accurate as the training data. If the training data does not capture enough of the desired state, it would not be feasible to expect the system to learn how to correctly identify that state. While it may be possible to observe psychophysiological identifiers of emotional state, emotions are still a subjective experience. Therefore, any system that uses people in its calibration will inherit this limitation. A possible solution to this is to reduce the calibration window, at the cost of training data volume, in the hopes of capturing a higher concentration of the desired state.

The next steps of this project will see the development of the artistic output. While some example images have been generated, as visible in appendix B, these renders lack the artistic depth desired by the project. While they represent the immediate emotions the participant is feeling, they do not display the emotional changes over time, or the progressing journey through emotions that music can cause. The next goal of the project is to contextualise this journey in the generated artworks; to show the emotions felt in their entirety, rather than just the momentary fluctuations. This is where the heart of the project lies; in the creation of a graphical score of sensation, a musical score of passion – a neural score of emotions.

References

Aggarwal, C., 2018. Neural Networks and Deep Learning. Springer International Publishing.

Altenmüller, E., 2002. Hits to the left, flops to the right: different emotions during listening to music are reflected in cortical lateralisation patterns. Neuropsychologia, 40(13), pp.2242-2256.

Boytsova, Y. and Danko, S., 2010. EEG differences between resting states with eyes open and closed in darkness. Human Physiology, 36(3), pp.367-369.

Bird, J., Manso, L., Ribeiro, E., Ekart, A. and Faria, D., 2018. A Study on Mental State Classification using EEG-based Brain-Machine Interface. 2018 International Conference on Intelligent Systems (IS), pp.795-800.

Bull, M. and Back, L., 2020. The auditory culture reader. Taylor & Francis.

Bunz, M. and Jäger, E. 2021. Creative AI Lab. [online] Creative-ai.org. Available at: <https://creative-ai.org/> [Accessed 28 February 2021].

Bunz, M., Jäger, E., Fiebrink, R., Andersen, C. and Cameron, A., 2021. Aesthetics of New AI Interfaces. Serpentine Gallery. [Panel Discussion]. Available at: <https://www.serpentinegalleries.org/whats-on/aesthetics-of-new-ai-interfaces-panel-discussion/> [Accessed 12 Feb 2021].

Casadidio, J., 2021. yop3rro. [Instagram]. Available at: <https://www.instagram.com/yop3rro/> [Accessed 3 March 2021].

Clutterbuck, J., 2021. Technical Manual. [online] Mind Monitor. Available at: <https://mind-monitor.com/Technical_Manual.php> [Accessed 3 March 2021].

eNTERFACE. Available at: <http://www.enterface.net/enterface06/docs/results/reports/project7.pdf>.

Jiang, X., Bian, G. and Tian, Z., 2019. Removal of Artifacts from EEG Signals: A Review. Sensors, 19(5), p.987.

Kaplan, S., Dalal, R. and Lunchman, J., 2013. Measurement of emotions. In: L. Tetrick, M. Wang and R. Sinclair, ed., Research Methods in Occupational Health Psychology, 2nd ed. Routledge, pp.61-75.

Mauss, I. and Robinson, M., 2009. Measures of emotion: A review. Cognition Emotion, 23(2), pp.209-237.

NHS. 2018. Electroencephalogram. [online] Available at: <https://www.nhs.uk/conditions/electroencephalogram/> [Accessed 3 March 2021].

Posner, J., Russell, J. and Peterson, B., 2005. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Development and Psychopathology, 17(03).

Randomquark.com. 2017. random quark. [online] Available at: <https://randomquark.com/case-studies/mindswarms.html> [Accessed 2 March 2021].

Savran, A., Ciftci, K., Chanel, G., Mota, J., Viet, L., Sankur, B., Akarun, L., Caplier, A. and Rombaut, M., 2006. Emotion Detection in the Loop from Brain Signals and Facial Images. [online] Dubrovnik.

Smidchens, G., 2014. The Power of Song: Nonviolent National Culture in the Baltic Singing Revolution. University of Washington Press.

Appendix A: Neural Scores Application

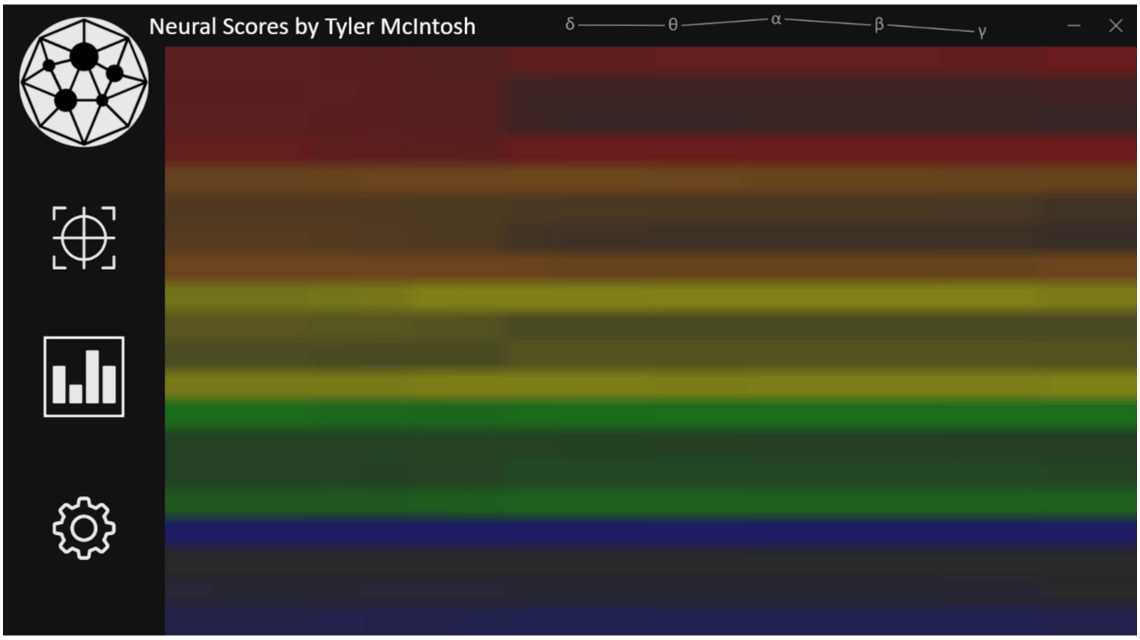

Figure 4: Spectogram View

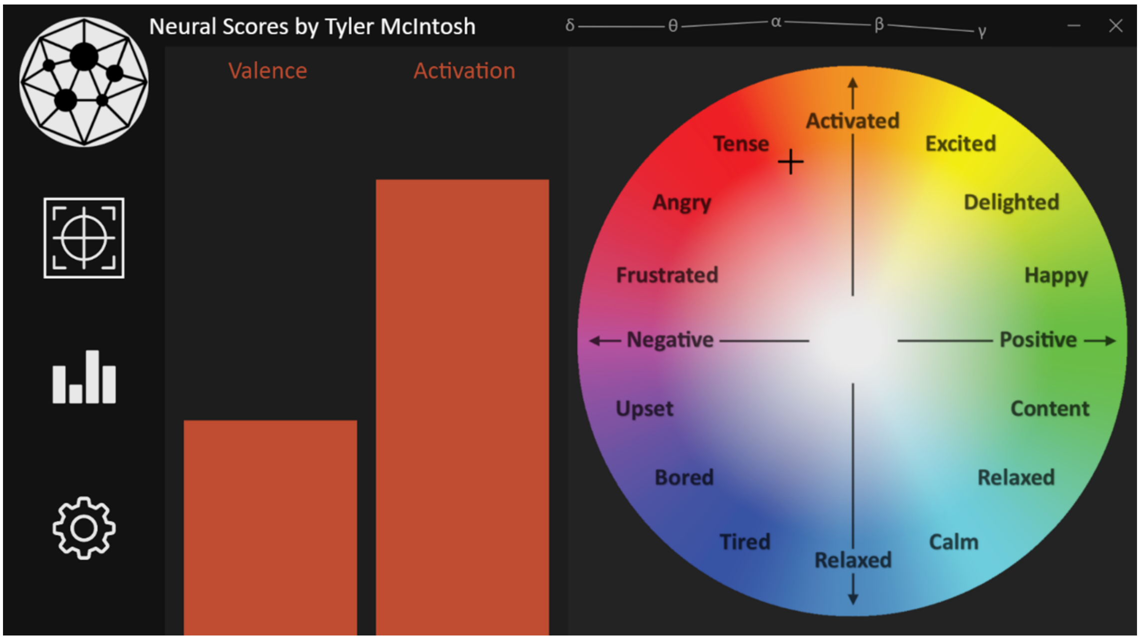

Figure 5: Classification View

Appendix B: Artwork Images

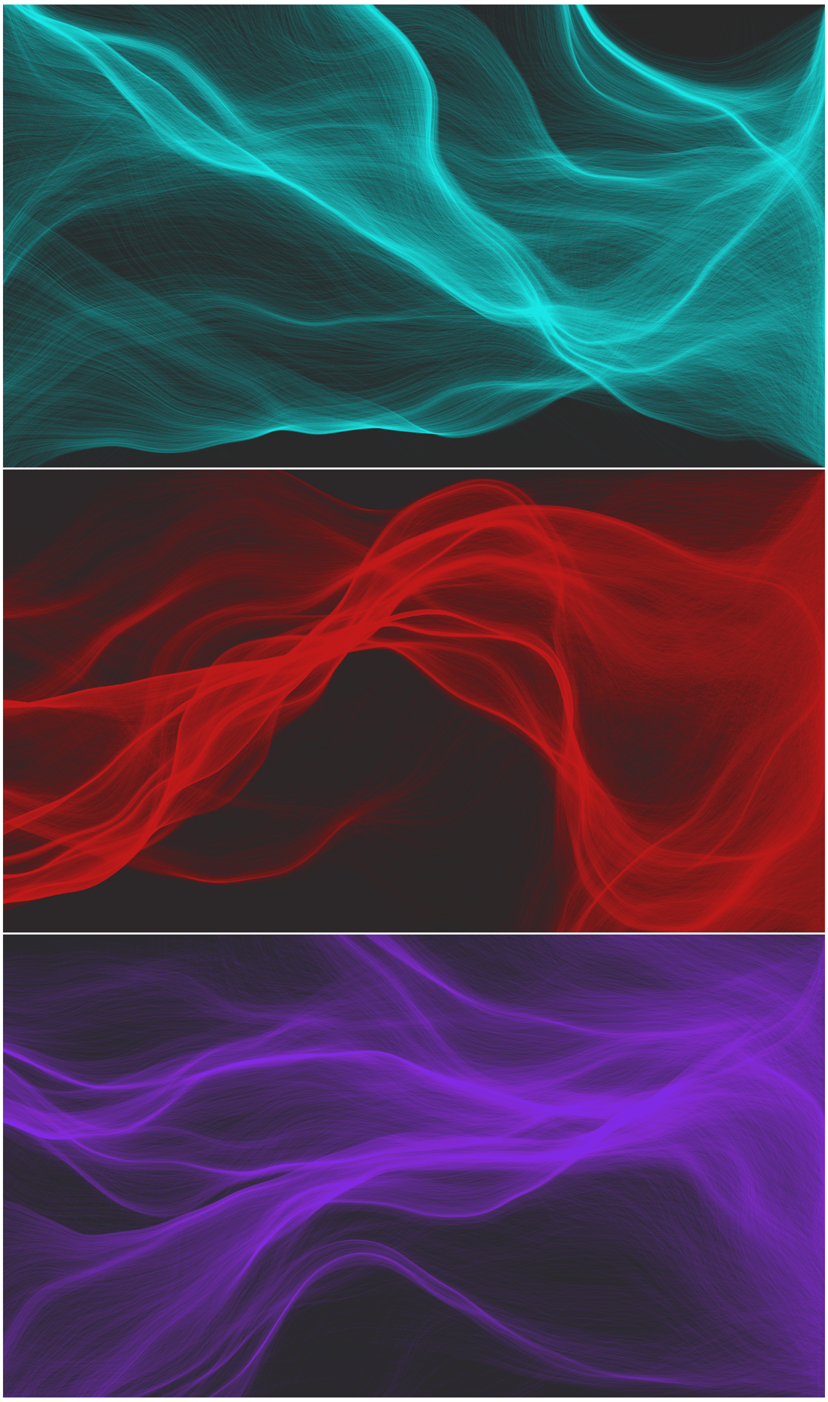

Figure 6: First Prototypes (Sadness, Anger, Tranquility)

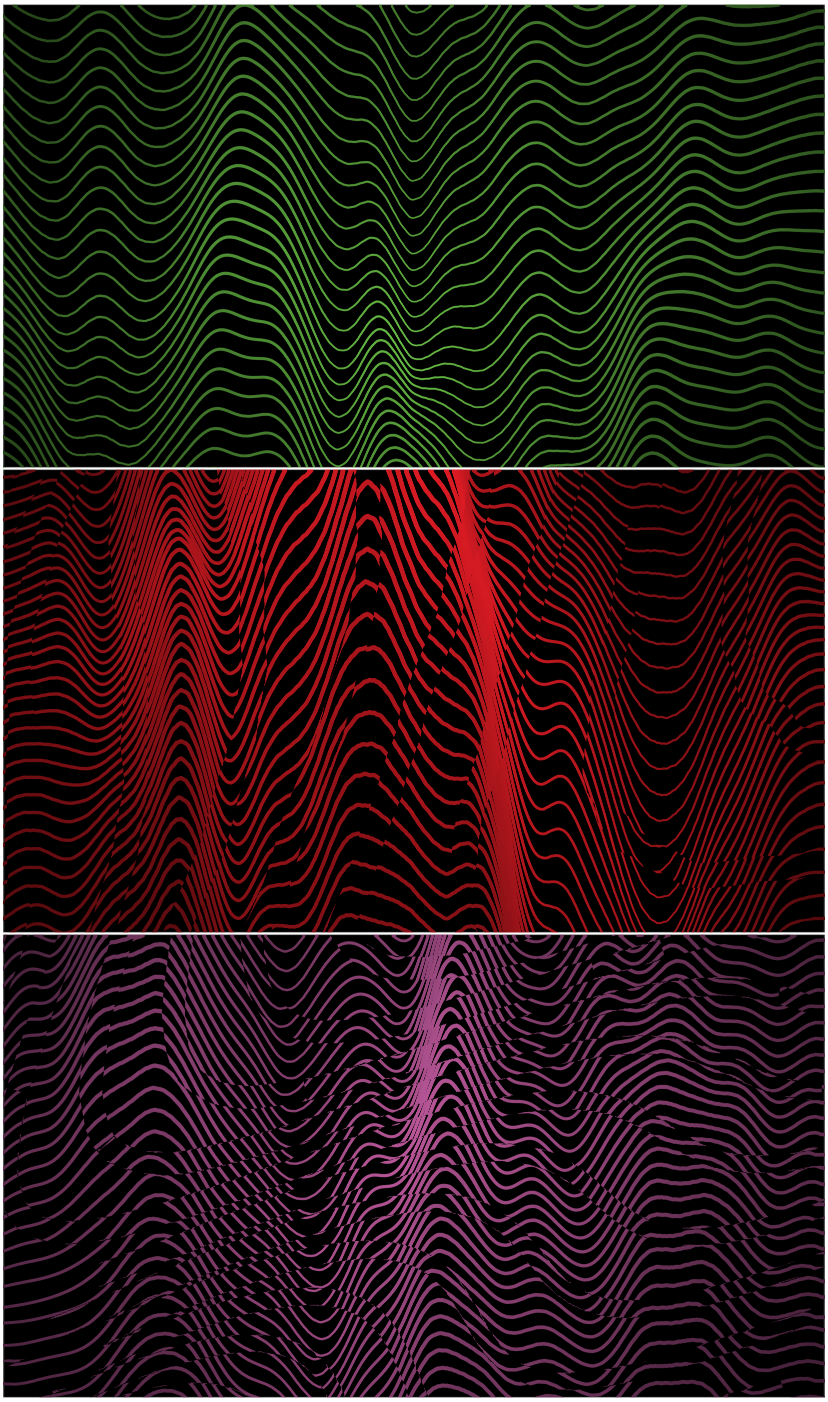

Figure 7: Second Prototypes (Happiness, Anger, Fear)