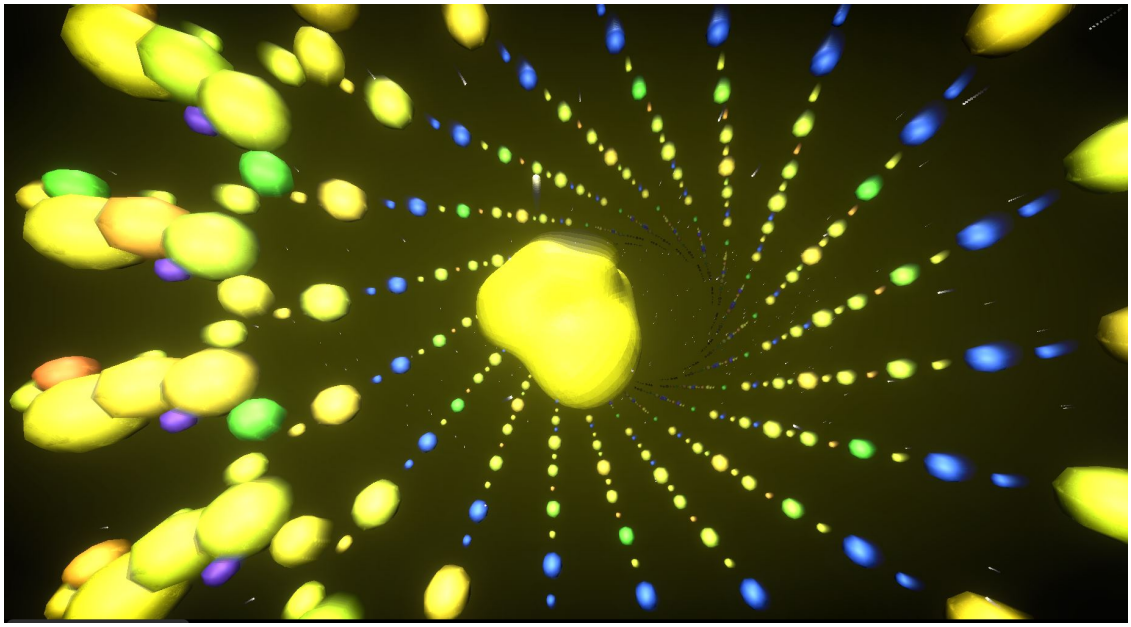

Applying Affect Estimation to 3D Music Visualization

Journeying Through the Development of an Affect-Based Music Visualizer

Master's Thesis in Computer Science and Engineering

Anton Eriksson

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2023

Master's thesis 2023

Applying Affect Estimation to 3D Music Visualization

Journeying Through the Development of an Affect-Based Music Visualizer

ANTON ERIKSSON

UNIVERSITY OF GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2023

Applying Affect Estimation to 3D Music Visualization
Journeying Through the Development of an Affect-Based Music Visualizer
ANTON ERIKSSON

Supervisor: Kivanc Tatar, Department of Computer Science and Engineering

Examiner: Palle Dahlstedt, Department of Computer Science and Engineering

Master's Thesis 2023
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: A screenshot of the visualizer in action while representing a happy song.

Typeset in LaTeX
Gothenburg, Sweden 2023

Applying Affect Estimation to 3D Music Visualization
Journeying Through the Development of an Affect-Based Music Visualizer
ANTON ERIKSSON
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

This project aimed to develop a 3D music visualizer using affect estimation and real-time audio features. Methods used during the project ranged from research through design, focus groups, prototyping, and an experimental survey. The final iteration of the visualizer could generate 3D scenes based on the music and extracted affect values, which participants found to be visually pleasing and fitting to the music. However, the visualizer struggled to accurately communicate the affect values visually, which is attributed to the lacking aesthetic profile and the subsequent implementation. Despite this, the project successfully manifested an affect-based visualizer that offers potential for further development. Future research is needed to investigate the use of abstract means for communicating emotions, and standardization of measures and language is required for this multidisciplinary domain.

Keywords: Music visualizer, affect estimation, audio-visual experience, computer graphics, research through design

Acknowledgements

I would like to express my gratitude to my supervisor, Kıvanç Tatar, for his feedback and guidance throughout this project and my previous work on creating the development environment. Thank you for always being enthusiastic and motivating, and for providing valuable insights that helped me to improve my work. I'm also thankful to my examiner, Palle Dahlstedt, for his diligent efforts in reviewing and assessing my work.

I would like to thank my friends and family for their unwavering support and encouragement throughout this epic journey. Your belief in me and your great ideas have been a constant source of inspiration.

Furthermore, I extend my thanks to all the participants who took part in my studies, without whom I would have lacked the necessary data to create my visualizer. Your contributions have been instrumental in making this project a success.

Lastly, I want to express my appreciation to the open-source community and the creators of the amazing libraries Meyda, Essentia.js, and Three.js. Without your contributions, this project would not have been possible. Thank you for making your work available to others and for inspiring us to push the boundaries of what is possible.

Anton Eriksson, Gothenburg, 2023-06-05

Contents

List of Figures

List of Tables

1

Introduction

Music is an auditory experience. It's an intriguing thought to try to extend the auditory sensation of music into a different sensory domain, such as vision. Visualizing sound is not a new concept, and audio purposefully combined with visual stimuli in a meaningful way can embellish the connection between the audience and the music [1]. Visualizing sound usually takes one of two routes, either as an aid for understanding the audio or as an artistic experience [2].

Artistic interpretations of audio can take many forms and involve multi-modal stimuli. Research on the combination of audio and visual perception is of particular interest and some examples in the audio-visual domain used auditory features to procedural animate the visuals of trees blowing in the wind [2], create 3D landscapes [3] or psychedelic effects [4]. Common with most visualizations, the focus lies on attributes such as pitch, rhythm, frequency, and timbre. Mood and emotional content are seldom represented in audio visualizations [5]. Modern affect estimation models can predict the mood of a song [6][7] and mood can be interpreted as the junction between valence level, happy to sad, and arousal level, calm to energetic [8]. Since the findings of how mood can be artistically visualized are limited [5] additional research on the topic is justified.

Previously I created a web-based audio feature extraction environment that outputs features such as affect (*sadness*, *happiness*, *aggressiveness*, *relaxedness*, *danceability*) as well as conventional real-time audio features such as loudness and amplitude spectrum. These parameters are made available to objects in a 3D scene. By investigating the domain of affect visualization and implementing the findings in the audio-visual environment I aim to create a perceptually and emotionally meaningful audio visualizer.

How can affect estimation be utilized in an abstract music visualization?

The project consists of a planning and literature review phase to get a fundamental insight into the domain audio-visual possibilities. Furthermore, the unexplored area of artistically representing auditory affect using abstract stimuli requires empirical data and due to its complexity I will conduct a focus group discussion to identify crucial features and phenomenological insights. The goal of the focus group is to explore how musical affect can be communicated through abstract stimuli, such as shapes, colors, and animation. Ideas, themes, and conclusions will be extracted from the transcription through thematic analysis and prioritized using Moscow analysis.

Brainstorming, importance/difficulty matrix structuring, and lo-fi prototyping will be utilized to explore and challenge potential solutions. I will conduct developmental work to manifest a prototype based on the insights and requirements in conjunction with the solution determined to be best suited for the project. Finally, the prototype will be evaluated in an experimental survey to determine its effectiveness to communicate affect and representing the audio.

The problem is complex and can be defined as a wicked problem; there are a million ways to represent audio. The area of artistically visualizing music affect has not been thoroughly explored and contributions to this domain are required. The wicked nature of the problem makes it unsolvable, however, the process and artifact development in this project will act as a contribution and aid in the understanding of the audio-visual domain in regard to affect and abstract visualizations. Instead of trying to solve the wicked problem, the project will act as an interpretation of the data and insights gathered, as well as a reference point for future research in the domain.

The goal is to combine empirical insights with theoretical principles and manifest them in an abstract visualization. The multimedia experience could potentially transcend the isolated experience of audio or visual stimuli. Perceptual features and natural mapping is determined to be of importance however particular attention will be put on how affect attributes extracted from audio can be represented in the visualization.

2

Background

2.1 The State of Music Visualization

Musical visualizations have been on an evolutionary trajectory for ages, traversing various forms ranging from traditional music notation to highly intricate digital software [5] [9]. To make significant contributions to this domain, it is crucial to gain a comprehensive understanding of the current state of music visualization. This knowledge helps us position the research in this work while providing an opportunity for gaining familiarity with the existing paradigm, identifying opportunities for innovation, and learning from the pitfalls and successes of past work in the field of music visualization.

Lima et al. conducted a literature survey consisting of 51 papers to summarise the current state of music visualization. In addition to identifying certain trends in data collection and technologies used they encountered three distinct types, or personas, of music visualization. The three major personas of music visualization are the music composer, the music engineer, and the music academic. The music composer is familiar with Common Music Notation (CMN) and MIDI, and visualizations are used as a tool to aid in chord progressions and harmony. The music engineer is concerned with signal processing techniques and visual feedback to complement the audio. The music academic is a blend of the previous personas but has a more exploratory approach and can in addition to investigating structure and chord progression also involve Machine Leaning, stylistic preferences, and artistic expression [5].

Khulusi et al. reviewed 129 works related to musicological data visualization and identified four main types of musicological data: musical works, musical collections, musicians, and instruments. Musical work is the most common type of musicological data visualization, and it can be divided into subcategories of musical scores and musical sounds. Musical scores refer to visualizations that transform traditional music notation structures into other visual formats. Musical sound visualization instead uses a data foundation made up of auditory features, such as loudness and pitch. Musical collections transcend isolated music pieces and visualize macro attributes such as genres, combined listening statistics, and music alignments. Musicians themselves can also be visualized in regard to biography/discography, collaboration networks, and similarities. Instruments can also be visualized also help communicate materialistic and structural properties as well as functional analysis [9].

With these personas and musicological data types in mind, I can classify this project as a musical sound visualization from the perspective of the music academic.

2.2 Generative Art

Generative art has become a powerful tool for music visualizations. The computational power of modern computers affords the possibility to create and produce unique visuals that synchronize to the music in real-time [2]. Not only does this technology make it possible to construct algorithms that explore the audio-visual domain, but it also opens new doors for artistic expression [10]. In the context of my project, generative art will be a cornerstone, and understanding the basics of generative art methods is integral to contributing to the domain of affect-based audio visualization.

Generative art can be defined as the practice where the artist to some degree involves a system that contributes to the creation of the artwork. Generative art has been around for a long time and the system can be as simple as throwing a dice to decide how to structure a musical score or the involvement of specific algorithms to create intricate tile patterns. The act of involving a system in the creation of art is determined to be a generative art process. Whether it's a dice, a set of rules, or a programming language it can be seen as generative art [10].

Generative systems can vary in degree of order/disorder, from orderly to completely random, as well as in degree of complexity. Most of our attention in generative art deals with very complex systems dealing with both order and disorder, such as neural networks, genetic algorithms, and dynamical mechanics [10].

Most of the early experiments of computer-generated art were based on randomly generated numbers [11] and recently sophisticated AI models such DALL-E-2 have been used to generate images [12]. Noise is still very prominent in computer graphics, in particular in regard to textures. Gradient noise algorithms such as simplex noise are used to generate height maps which can be applied to primitives to displace vertices, which can lead to dynamic 3D dimensional landscapes [13]. Instead of using noise to generate landscapes, Ox demonstrated how audio could be used as a way to create 3D environments. The pitch of the audio was used to determine the height of the landscape section and saturation was used as the determinant for hue. Instead of using a predetermined input or pure noise, the instrument audio was used as an input to the generative algorithm [3]. This is similar to the audio-visual building blocks approach mentioned by Brito and Fernandes [2] where the input is not completely random and the algorithm is comprehensible. While neural networks consist of algorithms they are often very high in complexity and describes as black boxes, therefore it's difficult to determine how specific nodes affect the outcome. Constructing an algorithm that takes audio as input and consists of building blocks based on audio-visual relations could help in the exploration of perceptually meaningful audio visualization. This approach would afford the researchers to be actively involved in the artistic aspects of how the visualization is realized.

Many of the visualizers are available as tools for understanding and identifying

characteristics of the music, such as structure or spectral content [5]. Audio visualizations can take on a more artistic approach and psychedelic visualizers with a pure aesthetic goal such as MilkDrop [4] has been around for a long time. Algorithms utilizing audio features have also been applied to generate 3D landscapes [3]. Brito and Fernandes create a procedurally animated tree blowing in the wind with the goal of exploring the expressive potential of audio-visual animation. Instead of solely creating an algorithm they took a more procedural approach to music visualization by creating reactive building blocks that can be configured to represent complex animations. In their research, they found that it's natural for most people to imagine visual metaphors for music and the users were able to add new ideas to the established tree metaphor that they had constructed. The huge processing power of modern computers and the seemingly unexplored concept affords additional research and the authors encourage exploration of the domain of audio-visual artwork and animation [2]. Procedural audio-visual animation in conjunction with affect estimation is an unexplored domain of audio visualization, and therefore worth investigating.

This project aims to create art using the building block approach mentioned by Brito and Fernandes [2]. The system will have an orderly quality in the sense that discreet building blocks determine the outcome of the artwork. The system does not include randomization however the audio signal will be used as an input for the system and to some degree introduce disorder. The complexity required by the system is still unknown and will be determined during the design process. Worth noting is that a complex generative system is not necessarily better than a simplistic one.

2.3 Natural Mapping and Glyphs

A music visualizer requires coupling between audio and visual parameters, and these can be implemented arbitrarily, but it's hypothesized that an obvious connection between audio-visual parameters will result in a more meaningful visualizer [14]. While the focus of this project is in regard to affect and emotions, it can be beneficial to investigate the perceptual relations between audio and visuals. A deeper understanding of how to efficiently couple these parameters can lead to a more immersive experience and enhance the emotional impact of the visualizer on the audience.

Norman famously brought up the importance of mapping controls to outcomes. Natural mapping in particular denotes the pursuit of creating relations between controls and outcomes that are intuitive and immediately understandable [15]. Creating a visualization that represents the audio requires conceptual natural mapping. Without sufficient natural mapping, the scene will just consist of arbitrary shapes and colors. In the case of this project, there is limited user input, but since it's the audio that controls the visuals outputted to the scene natural mapping between audio and the visuals is desirable.

Glyphs are widely used within music visualization [5] and are a staple within the field of information visualization. A glyph is a graphical object and data is mapped to the graphical attributes of the glyph. Some examples of attributes are color, size, spatial position, orientation, surface texture, motion coding, and blink coding.

Natural mapping is important to glyphs to make the interpretation intuitive to comprehend. For example, it's better to map temperature data to color than to size, since I have an understanding of the concept of heat and its connections to colors; blue is cold and red is warm, etc. Mapping city population density data to the size/volume of the glyph would also be considered natural mapping [16].

Glyphs are used to effectively communicate multidimensional data. Artistic visualizations are however not required to communicate exact values and Brito and Fernandes demonstrate how audio features can be mapped to seemingly arbitrary structures such as trees. For example, the size of the leaves was controlled by the frequency spectrum's average [2]. However, it's assessed that natural mapping between audio features and graphical attributes is crucial to be able to capture and represent the audio in a meaningful way.

Attempts to create perceptually meaningful representations of audio have been made. The software AudioScope based the visual aesthetics on attributes of sound with natural mapping in mind. Pure sounds, that consist of a few sine waves are represented as round since sine waves are round. Loud sounds are represented by large objects and quiet sounds are represented by small objects. Percussive and transient sounds are fast and therefore briefly flash on the screen. Brighter sounds consist of many frequency components and are therefore portrayed as spiky [17]. Note that this is just an interpretation of the connections between audio and visuals. In addition, the AudioScope visualizer does not attempt to visualize the emotions of the music.

2.4 Emotional Determinants

The visualizer will utilize affect estimates to determine the emotional impact of music, translating these emotions into abstract visuals that effectively convey the desired affect. In order to achieve this goal, a thorough understanding of the emotional qualities of both music and abstract visuals is essential. Researching how musical features such as rhythm and mode, as well as visual features such as hue and movement, have been employed in the past to convey emotions, will inform the design decisions. This knowledge, in combination with the affect estimates and the aesthetic preferences of the focus groups, will establish the foundational structure for representing emotions in the music visualizer I'm developing.

Synesthesia is a condition where individuals can experience strong connections between stimuli and perceptual features, such as the associations between certain sounds and colors [18]. It can have an impact on the emotional perception of sounds. However, since synesthesia is relatively uncommon, the project does not specifically focus on individuals with synesthesia. Instead, the aim was to identify general determinants of emotions. Nevertheless, the idea of developing a visualizer based on sound-color synesthesia could be a fascinating project worth pursuing.

2.4.1 Emotional Determinants of Music

Conveying emotions in music is very complex and can involve many parameters such as lyrical semantics, tension, melody, mode, tempo, loudness, structure, etc. [19]. The complexity of the problem makes it difficult to couple musical attributes and connect them to specific emotions. Some patterns have however been identified. Tension can be determined by increased loudness and low pitch [20]. Arousal is affected by rhythm and tempo, and a fast song induced psychophysiological reactions related to arousal such as faster breathing and heart rate. Valence is often related to mode and harmonic complexity [21]. Minor and major scales have been shown to affect mood [22] [23] and slow songs in minor mode are often perceived as sad, while fast songs in the major key are often perceived as happy [24]. This notion would be applied to Russell's circumplex model of affect [25] and Thayer's model of the emotional plane [26], given that tempo can be a determinant of arousal and mode can be a determinant of valence.

2.4.2 Emotional Determinants of Abstract Visuals

There are many ways to communicate emotions with visuals, such as text or images of expressive faces. For this project, I'm concerned with how abstract visuals with minimalist semantic value can be utilized to communicate emotions. Grasping the conventions of how specific visual features can be utilized will be valuable in the creation of the visualizer.

2.4.2.1 Color

Traditionally colors have been divided into plus colors and minus colors. Plus colors being warm red-yellow hues were thought to induce positive feelings while minus colors being perceived as cold blue and green hues representing restlessness and anxiousness. Red and yellow have also been associated with outward forceful action while blue and green have been associated with more calm and stable action. Modern empirical research of color psychology problematizes the traditional notions of the meaning of colors. Colors have different contextual connotations and can vary between cultures. It's clear that the use of colors transcends aesthetics and can influence thought and behavior [27].

A cross-cultural study based on orchestral music demonstrated that faster music based in the major mode was associated with colors of high saturation, brightness, and a yellow hue. Slow music in the minor mode was associated with the opposite color pattern, meaning a less saturated, dark blue hue [28]. Arousal and valence have been shown to have connotations to hues and degrees of saturation. Music high in valence and arousal is viewed as saturated hues of yellow, orange, and red, while music low in valence and arousal is seen as sad, calm dark, and blue hues [29]. These findings can be applied to the circumplex model of affect.

2.4.2.2 Movement and Animation

Motion design is crucial to create engaging animations and is wildly used to give characters life in movies and to create captivating user interface designs. The principles of animation and motion design are valuable when trying to convey emotions, even when dealing with abstract means. There are 12 principles of animation famously brought up by animators at Disney [30] which I will account for below. These principles can be utilized in this project to communicate affect, increase engagement and to some extent bring a narrative to an otherwise static illustration.

"Squash and Stretch" refers to giving objects the illusion of weight and volume by altering their shape as it moves through space. "Anticipation" denotes the build up before a movement is set in motion. "Staging" refers to drawing attention to important aspects of the scene to communicate where users should focus. "Straight ahead action" and "pose to pose" refers to two ways to construct animations. "Straight ahead" entails drawing each frame in the animation whereas "pose to pose" entails drawing separate keyframes/poses and adding frames in between. "Follow through" and "overlapping action" refers to the fact that different elements of the scene will have different physics, which becomes clear at abrupt stops. "Slow in and slow out" denotes that objects do not have a constant speed. Often they have a wind up and slow down, which can be depended on their perceived volume. "The arc" principle refers to the effectiveness of curves to simulate natural movement. "Secondary action" refers to the additional sub-animations taking place as the object moves through space. "Timing" refers to the temporal aspects of transforming objects. For example, a heavy object is more difficult to move and would be slower to displace. "Exaggeration" refers to the possibilities of challenging the natural world. While physical accuracy is important exaggeration can make the animation more appealing. Too subtle cues can go unnoticed. "Solid drawing" entails the correct dimensions, shadows, light, and anatomy of objects. Much of this will come for free when using computer graphics software. "Appeal" refers to the goal of striving to create aesthetically pleasing objects and scenes since they generally gather more appeal. This principle is similar to the famous quote said by Don Norman, "Attractive things work better" [31].

In the classical Heider-Simmel illusion experiment participant watch a video of moving shapes. At first glance, it does not represent much but upon further inspection, the shapes can be interpreted as characters experiencing love and abuse. This effect is solely reached through the shapes and their movement. Apart from demonstrating the effects of attribution and anthropomorphization the study also showcases how emotions can be created through very limited means [32]. Lima et al. denote the importance of the relationship between temporal and spatial logic. Temporal and spatial logic should match each other [5]. In the context of a 3D visualization, it entails a causal relationship between when audio events occur in relation to the responsiveness of the objects in the 3D scene. Using the estimated affect values to calibrate the magnitude and speed at which objects transform when triggered by audio cues could be an effective way of representing the song sentiments by movement. For example, one could argue for natural mapping by having a song that are rated high on *aggressiveness* be more responsive and transform objects faster, etc.

2.5 3D Computer Graphics

To bring my vision of an affect-based music visualizer to life, I require a capable medium, and that's where computer graphics come into play. The medium has the ability to manipulate a vast array of parameters that can be utilized to convey emotions. Additional complexity but also utility is introduced when working with a 3D scene, which is expected to enhance the experience and immersion. The algorithmic foundation of computer graphics is a natural fit for my mission to create a music visualizer with generative capabilities, providing the perfect canvas to represent artistic expressions. To fully leverage the potential of this medium, it's crucial to gain a deep understanding of integral parts of 3D graphics. This is particularly important when dealing with delicate subject matter such as affect, where every aspect of the visualizer must be carefully crafted to communicate the intended emotions effectively.

Computer graphics is the area of engineering dealing with generating digital images using technical means and computers. Computer graphic is ubiquitous in our modern life since it's involved in all screen-based rendering. The discipline concerns the creation of digital representations of data and the technologies involved to do so. Computer graphics is a vast area and tasks such as simulating realistic light reflections or optimizing Arithmetic Logic Unit (ALU) computations in the graphics card are worthy domains in their own right [13].

WebGL is a low-level system used for creating and rendering computer graphics on the web. The low-level workflow of WebGL requires a lot of code to create interesting scenes. Wrappers such as Three.js can streamline the processes of creating web-based computer graphics by providing sophisticated tools to create and combine the components of a 3D scene. For example, WebGL requires the developer to specify points and draw lines to create objects whereas in Three.js you can use predefined classes to draw geometries such as a cube. Computer graphics afford many subsystems and manipulable parameters that can be mapped to data to create interesting visualizations. Below, I account for many of the basic components required to construct a 3D scene and how they can be manipulated [33].

2.5.1 Primitive

Primitives are generated shapes made up of polygons. Shapes can be simplistic such as cubes or spheres, or more advanced such as torus knots or polyhedrons. Primitives can also be constructed using parametric curves such as bezier curves or created in separate modeling software and imported into the Three.js scene. The predefined classes in Three.js have attributes that can be manipulated, such as increasing the width of a cube to create a rectangle [33].

2.5.2 Material

Materials can be applied to primitives to control the properties of how the light reflects and how the surface is perceived. Each material is given a base color and a

material type. Materials can have properties such as shininess, roughness, metalness, and emission which determine how the light reflects from the surface [13].

2.5.3 Texture

Textures are generally images that can be applied to objects in the scene. There are however different types of texture maps such as displacement maps, which manipulates the position of specific primitive points, and normal maps which alter how light is calculated and can increase the perceived detail of images without increasing the polygon count [13].

2.5.4 Transformation

Transformation denotes ways to change a primitive's orientation, position, size, and shape. Most parameters can be manipulated and deployed in an animation loop. Of particular interest are the primitive transforms for the position, rotation, and scale which can be used to create moving dynamic objects [13].

2.5.5 Light

Lights are used to brighten the scene as well as create reflections and shadows. There are many different types of light such as ambient light, directional light, point light, and spotlight that can be added to the scene, and light color and intensity can also be manipulated to create different effects [33].

2.5.6 Camera

Cameras are used to determine what in the scene that will be rendered. The view frustum denotes the volume in front of the camera where objects will be rendered. Objects too close to the camera, too far away, or to the sides will be culled and not rendered. The field of view (FOV) as well as the position and angle of the camera can be animated [13].

2.5.7 Renderer

The renderer in itself can also be manipulated. During rendering the vertex and fragment shader are called to determine the light and color of pixels, and even the position of vertices. Post-processing effects can be created by customizing how the shaders manipulate the vertices and pixels [13]. Certain effects come prepackaged with Three.js, such as bloom and glitch pass. Worth noting is that post-processing can be computationally heavy since additional calculations have to be done for each fragment.

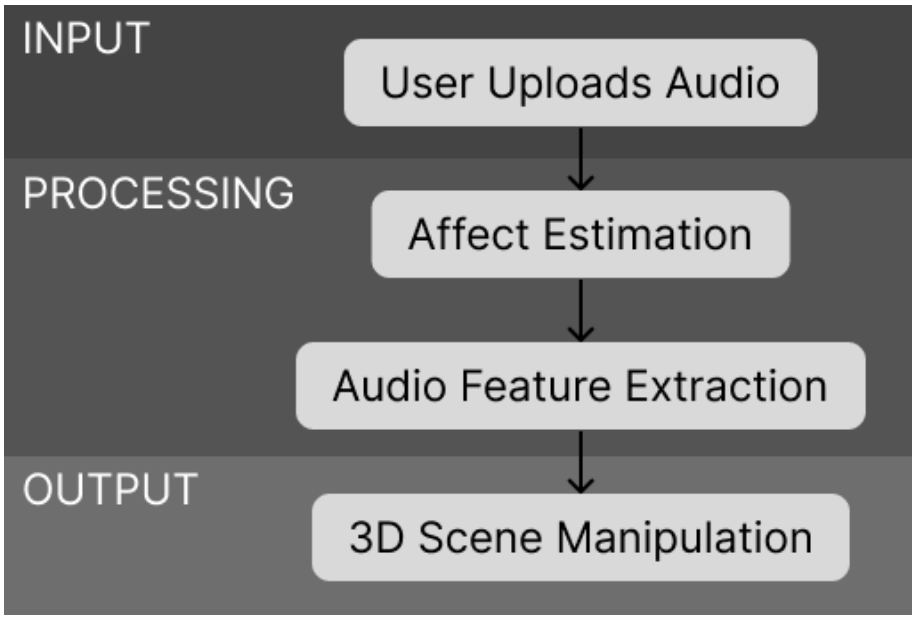

Figure 2.1: The pipeline architecture of the development environment.

2.6 The Development Environment

When designing software, it's essential to have a dedicated space where we can bring our ideas to life and test our manifestations. This allows us to explore different features and settings, iterate on design choices, and refine the prototype. Having a sandbox environment dedicated to iterating on my music visualizer is integral to its design and implementation.

2.6.1 Fundamentals

I have previously created a browser-based developmental environment for creating audio visualizations in 3D. The developmental environment streamlines the process from raw audio to responsive 3D objects and can be utilized by practitioners and developers who do not what to set up their own visualization environments from scratch. The pipeline is built in common JavaScript as an accessible web application. The pipeline makes use of the libraries Meyda[34], Essentia.js [6] [7] and Three.js[33].

The environment requires a user to upload an audio file. The audio is analyzed and affect attributes are extracted in a pre-visualization step. The visualization and music then start and real-time audio features are extracted and visualized using responsive objects in a 3D scene. Both the labeled and real time features are stored in a global object which is mapped to parameters in the scene.

2.6.2 Essentia.js: Estimating Music Affect

Essentia.js is a JavaScript audio analysis library with a C++ back end. The library affords many audio processing and analysis tools but affect estimation is the most important feature for this project. Essentia.js has trained machine-learning models

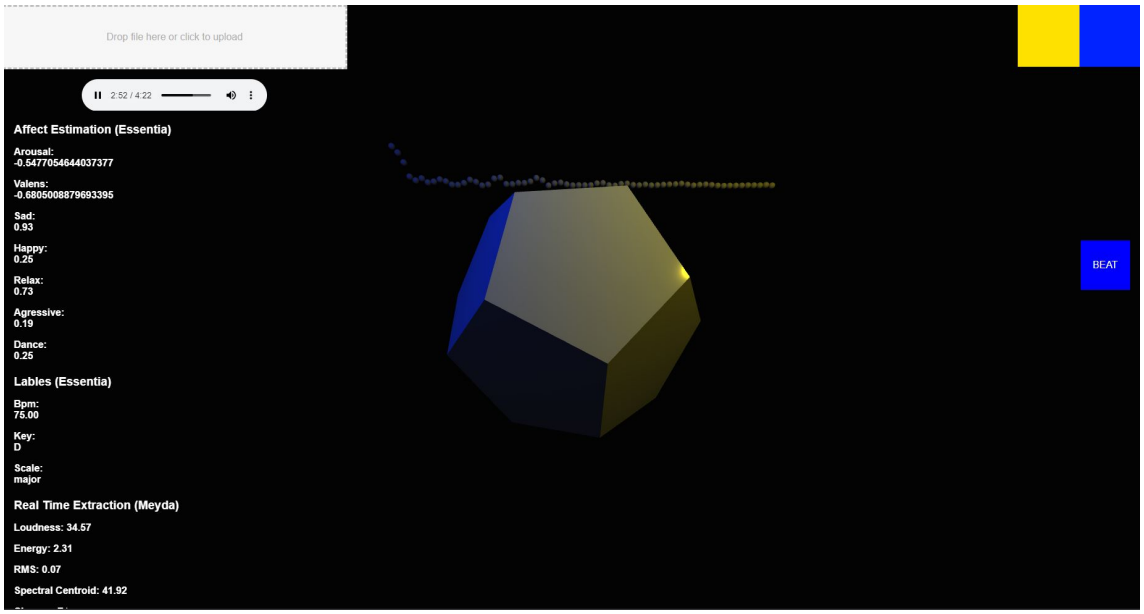

Figure 2.2: The initial state of the development environment.

that can be utilized in browser applications using WebAssembly and TensorFlow [6][7]. The Essentia.js model implemented in the development environment takes a raw audio file as input and extracts the following attributes:

- ***Sadness***: an estimation of how sad the song is.
- ***Happiness***: an estimation of how happy the song is.
- ***Relaxedness***: an estimation of how relaxed the song is.
- ***Aggressiveness***: an estimation of how aggressive the song is.
- ***Danceability***: an estimation of how danceable the song is.
- **Beats Per Minute (BPM)**: an estimation of the number of beats per minute.
- **Key**: the estimated main key.
- **Scale/Mode**: the estimated main mode.

Sadness and *Happiness* are related to valence whereas *Relaxedness* and *Aggressiveness* are related to arousal. Abstract features such as *Danceability* demonstrate the power of classification models. The song is also classified in musical terms such as BPM, key, and scale.

2.6.3 Meyda: Extracting Real Time Audio Features

Meyda is a real-time audio feature extraction library written in JavaScript. The library uses WebAudio API which is a tool for manipulating audio in the browser [34]. WebAudio API is based on modular routing where audio nodes are linked together inside an audio context [35] and Meyda can connect to the audio context and extract audio features as the audio is playing in real time. Meyda can extract a

range of attributes, but the following real-time features are anticipated to be most important for the visualizer:

- **Root mean square (RMS)**: rough representation of loudness.
- **Loudness**: perception of sound related to amplitude.
- **Amplitude Spectrum**: distribution of energy across frequencies.
- **Spectral Centroid**: center of mass of frequency distribution.
- **Spectral Rolloff**: frequency at which energy rolls off to zero.
- **Chroma**: representation of pitch content, independent of the octave.
- **Perceptual Spread**: perceived width of sound in stereo field.
- **Perceptual Sharpnes**s: perceived "sharpness" or "hardness" of sound.
- **Perceptual Spread**: perceived width of sound in stereo field.
- **Mel-Frequency Cepstral Coefficients (MFCC)**: spectral envelope representation based on mel scale.

The developmental environment consists of two Meyda analyzers, one that extracts features from the raw audio and one with a lowpass filter at 200hz. The lowpass analyzer is detecting the energy level of low frequencies and triggers when the value reaches a set threshold to simulate a simplistic beat tracker.

2.6.4 Three.js: Manipulating a 3D Scene

Three.js is a versatile computer graphics library that utilizes WebGL to create 3D scenes directly in the browser. With a range of tools and features, it allows for the extensive manipulation of computer graphics parameters such as primitives, cameras, and lights, giving designers and developers great creative control [33]. Three.js affords the ability to create a wide variety of projects, from interactive games [36] to hyperrealistic models with advanced light calculations [37]. It has even been used before to create a music player that has dynamic graphics which react to the real-time audio features of the music [38]. This adaptability and flexibility make Three.js an ideal choice for developing a browser-based 3D music visualizer. By mapping and manipulating the scene parameters based on audio features and affect estimates, I aim to create an emotionally engaging experience for the audience.

3

Theory

3.1 Research Through Design

There are three types of research related to design. Research for design describes the preparation process for design. Research into the design is a meta-commentary of how is design conducted. Research through design denotes learning and development when being practically involved in the design process. Research through design includes developmental work, meaning-making, or customizing technology to do something that no one has done or considered before. Treading in this unexplored domain leads to playful tinkering as a means of learning [39].

Theory and technology can come together through design. Zimmerman et al. [40] suggested four criteria for evaluating design research within human-computer interaction (HCI). The process of how the choice of methods is motivated is important, as well as whether the process can be reproduced. The research must produce an invention that combined subject matters and address a specific case. Validity is difficult to use as a measure in the domain of design and therefore design research should instead be evaluated on relevance. Finally, the design research should be extensible meaning it should be possible to expand upon and/or derive knowledge from the body of work.

Developmental work is a subcategory of research through design particularly focused on the processes involved when making and constructing new artifacts [39]. Developmental work inherently includes testing, tinkering, and discovery and is closely related to prototyping and the practical involvement of design.

3.2 Phenomenology and First Person Design

Phenomenology is the study of experiences, in particular, how we experience. The structures of the conscious first-person experiences are of interest and the concept of intentionality is central. Intentionality denotes how experiences are directed towards or about an object. A sentence like "I imagine a fearsome creature like that in my nightmare." captures the fundamental attributes of a first-person experience and demonstrates intentionally towards an imagined nightmare creature [41]. This project will tread into the domain of how music is imagined visually, which is a question that resides in the phenomenological domain. The question "What does it

entail to imagine music visually?" captures some of the characteristics of the problem at hand.

The involvement of first-person perspectives is a long time practice within HCI and the personal experience can intuitively fill in the gap which can not be sought after by objectivity [42]. HCI practitioners use a first-person perspective in several aspects of their design profession [43].

In this project, phenomenology and first-person experience will be involved in both the focus group discussion as well as when prototyping. The focus group will revolve around getting insight into the participant's experience of audio-visual perception and exploring how intentionality is directed in relation to affect, music, and visuals. When prototyping commences insights drawn from the literature review, empirical data, as well as the first-person experience of the developer, will be combined in a melting pot to shape the final manifestation.

3.3 Wicked Problem

Design is a complicated area and the challenges can be classified as wicked problems. Rittel [44] famously outlined ten characteristics of wicked problems and this is how they relate to the problem of visualizing music sentiment. First of all, there is no definite formula for this problem. There are no axioms that can be utilized to deduce a perfect visualizer. The problem doesn't have a stopping rule, meaning that there is no way of knowing when the solution is complete. In the case of this project, the solution is simply determined by time and the completion of one design cycle. This however doesn't entail that the solution is optimal or finished. The solution to the problem at hand will not have a true-or-false answer. The solution will instead be evaluated, by looking into how well the visualization communicates emotions and contribute to the experience of music. There is no ultimate test for proposed solutions to wicked problems. Designs and artifacts can be evaluated on specific measures but designs are unique, materialized by the designer, and can be viewed as universal in scope, which makes design a practice of unrivaled wickedness [45]. The solution is a one-shot operation, meaning that the solution is determined by the time and place in which it was conceived. The data collected from the focus group, and the derived insights, will be unique and it's impossible to replicate the exact same conditions in which it was conducted. It's a fleeting operation and the design artifact should be viewed as a product of its circumstances. There is no set number of solutions to a wicked problem. There are unlimited ways to visualize music and interpret emotions. Wicked problems are essentially unique in nature. While similar problems can have overlapping characteristics there is no set way of finding a solution to a wicked problem. Wicked problems also consist of symptoms of other problems. It's challenging to create emotional visuals based on music since it's dependent on subjective qualities of emotions which is in itself a complex problem. This also entails that there is more than one way to explain wicked problems and explanations might even vary depending on individual perspective. This is yet again true for how music should be portrayed. Wicked problems deal with solutions to actual real-world challenges and the planners are responsible for their actions and

approach [44].

In short, this means that there are no all-around waterproof solutions to wicked problems. There is no optimal solution to visualizing the emotional contents of music since it deals with many determinants, fleeting empirical evidence, and subjectivity. Wicked problems go against much of the scientific paradigm of rules and laws [45]. Rigorous scientific methodologies and approaches can be utilized in the practice of design, but the wicked nature of the design is faulty from the viewpoint of scientific scrutiny.

In this project, Zimmerman's criterion for design research [40] will be adopted to contribute to the academic domain of design with the insights gained from the creation of the artifact and the attempt to solve a wicked problem.

3.4 Perceived and Felt Emotion

Studies dealing with emotions in music have been rather ambiguous with respect to the distinction between perceived emotion and felt emotion. In the case of music, perceived emotion would be the emotional content expressed by the musical piece, while felt emotion would be the emotional content induced in the listener. These qualities are indeed distinct. One can understand what kind of emotional content a stimulus communicates without experiencing the emotion [23].

There are multiple ways in which perceived emotions can affect felt emotions. The far most common notion is a positive relation, meaning that a stimulus perceived as happy will be felt as happy. A negative relation would entail that stimuli perceived as happy are received as sad. A perceived emotion can also have a neutral effect on felt emotion [46]. In regards to music, perceived emotion doesn't always coincide with felt emotion. In particular, musically trained individuals appraised the perceived emotions significantly differently from felt emotions in regards to certain musical characteristics, such as note density [23]. A multitude of factors such as context, psychological state, and individual variation makes it difficult to ensure a positive relationship between perceived and felt emotion, however, the convention in practice is to assume a positive relation [46].

In the scope of this project, I focus on perceived emotion. The complexity of constructing a bulletproof stimulus that is perceived and felt in the same manner is out of the scope of this research.

3.5 Arousal And Valens

Russell's circumplex model of affect [25] and Thayer's model of the emotional plane [26] describe human emotions on two axis. The first axis represents arousal, which accounts for the intensity of the emotion and goes from low arousal to high arousal. The second axis is valence which determines how pleasant the emotion is and goes from negative valence to positive valence. The axis of valence and arousal have been

applied in Music Information Retrieval (MIR) to estimate the affect of the music [8] [47] [26]. Machine learning models can be used to classify the track sentiments.

Affect estimation tools, such as Essentia.js, have broken down arousal into two distinct values, denoted as the degree of *relaxedness* and the degree of *aggressiveness*. The same is true for valence which is broken down into the degree of *sadness* and the degree of *happiness*. Sophisticated models can estimate mood with high accuracy, up to 90% accuracy [8], but the accuracy for arousal is often higher than the accuracy for valence [8][47]. High levels of cross-cultural agreement have been demonstrated when evaluating experimental music in terms of arousal and valence [48].

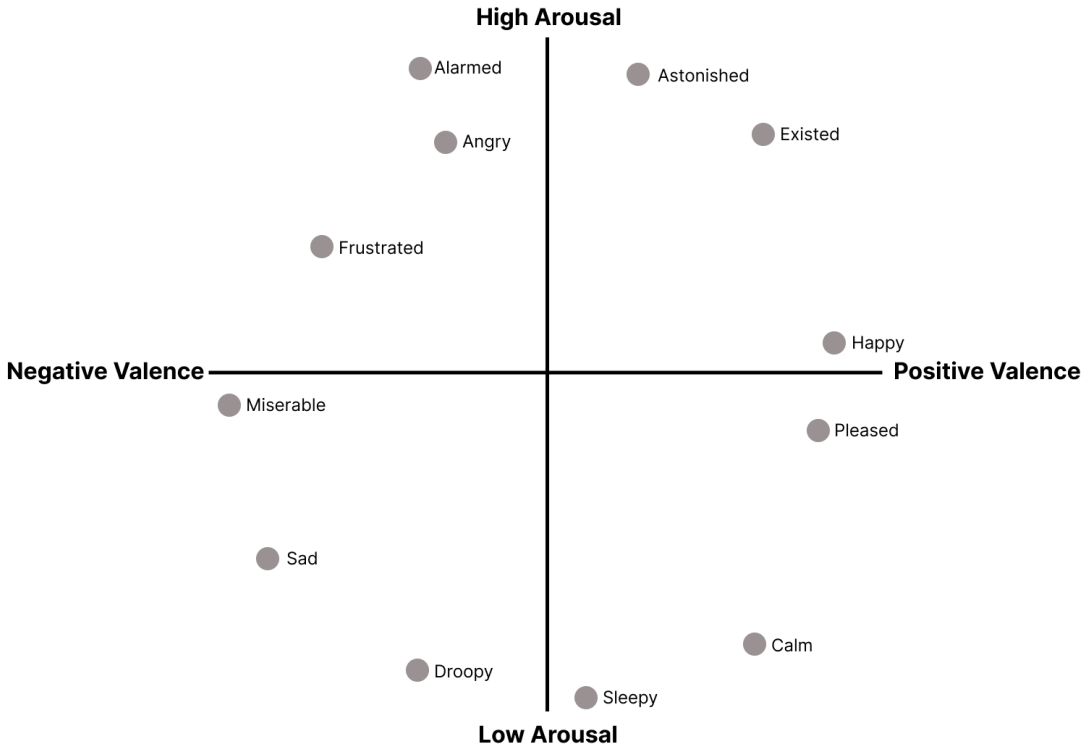

Figure 3.1: Russell's Circumplex Model of Affect. Note that the marked emotions are estimations.

The traditional models of emotion have an antagonistic view of emotion, meaning that sadness is the direct opposite of happiness etc. [25]. This view has been challenged and rich experiences describe complex emotions with both pleasant and unpleasant qualities. Complex emotions such as longing and sympathy can have both qualities of sadness and happiness in them [49]. The affect estimation model used in this project can output values for both *sadness*, *happiness*, *aggressiveness*, and *relaxedness* at the same time, which means for example that a song can be classified to be both sad and happy at the same time. There is however no claim that this should be interpreted as a complex emotion. In addition, the traditional models of emotion suggest that emotions can be derived based on the level of arousal and level of valence, such that high arousal and high valence become exited and low arousal and high valence become serene [25]. While the attributes extracted by

the affect estimation models are closely related to arousal and valence they do not claim to estimate these kinds of relational emotions. Classification models such as Essentia.js have been trained to classify complex attributes such as *danceability* [6] [7] and given the right data set the model could be calibrated to estimate complex emotions.

In the scope of this project, the estimated affect values will simply be taken at face value and regarded as accurate predictions for the music sentiment. Worth noting is that the Essetntia.js model used for this project does not take lyrical semantics into consideration when determining emotional content. To make this project implementation feasible and pragmatic I follow these assumptions when dealing with the Essentia.js affect estimates:

- High *Aggressiveness* $\approx$ High Arousal
- High *Relaxedness* $\approx$ Low Arousal
- High *Happiness* $\approx$ High Valence
- High *Sadness* $\approx$ Low Valence

It is important to note that when discussing the impact of media content on emotions, the content itself does not possess inherent emotions, but rather it influences and induces emotions. Alternative terminology, such as pleasantness for valence and eventfulness for arousal, have been suggested to denote the emotional qualities of media content [50]. However, for the purpose of this report, I have chosen to use the established terminology of valence and arousal to describe all emotional responses. It should be noted that eventfulness and pleasantness could have also been used interchangeably to describe the perceived emotions of the visualizer.

4

Methods

4.1 Double Diamond Model

There are many ways to describe the design processes, and for this project, we focus on the double diamond model because of the following reasons. The double diamond model is a way to represent the creative design process, it structures the process into divergent and convergent sections and consists of four different phases. The discreet division of converging and diverging sections creates separate sections for peripheral exploration and decision-making. The first is the divergent discovery-phase where the nature of the problem is explored in a broad sense. Research is required at this stage. The second phase is the convergent define-phase where the problem is narrowed down and a deeper understanding of the important aspects and dynamics of the problem are laid out. The third phase is the divergent develop-phase where different solutions to the established problem get explored. Prototypes and concepts can be explored. Finally, the fourth phase is the convergent deliver-phase where the solution is tested and finalized [51]. The area of affect-based music visualization is rather unexplored which makes the double diamond model well-suited for our project since it makes room for both exploration and structuring of the problem and the solution.

Below I account for how these sections and phases will be utilized in the project.

4.1.1 Discover: Problem Diverging

The problem at hand is unexplored and requires a broad discovery-phase. The phase will yield an understanding and contribute to a pool of information from which I can extract insights. It's in this face I conduct a literature review and explore the topic through focus groups.

4.1.2 Defining: Problem Converging

In the context of this project, the problem converging phase entails identifying important aspects of the problem. This would entail extracting insights from the focus group and the literature to get a comprehensive understanding of why it's difficult to visualize mood, what aspects that are important, and suggestions on how to do it. The task of developing a prototype visualizer requires certain guidelines and at the end of the problem-converging phase a structured priority list of important aspects,

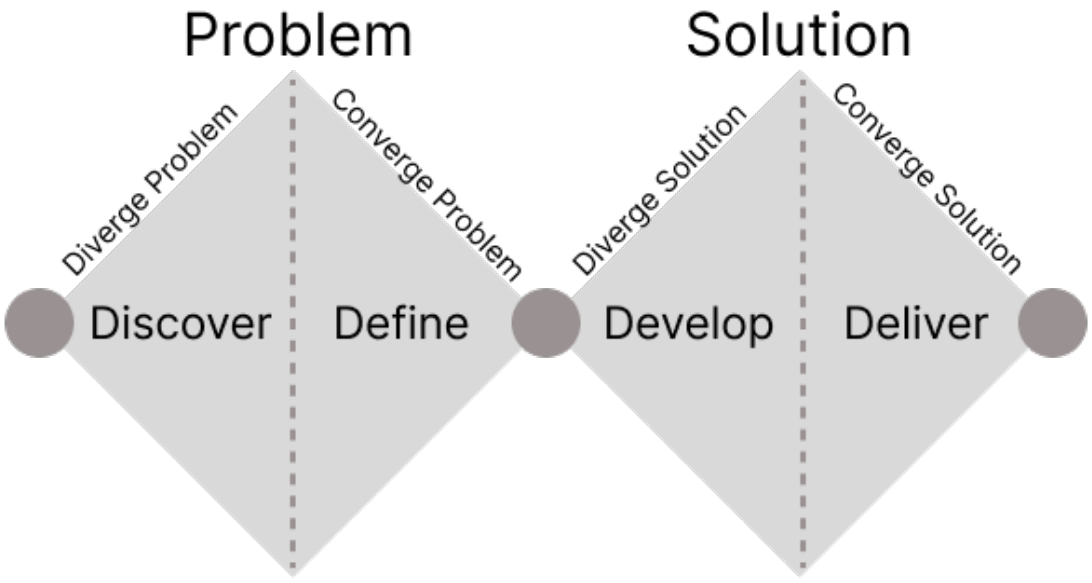

Figure 4.1: The Double-Diamond Model

such as ideas or an aesthetic profile, will materialize. The insights and ideas can be used in the development phase to explore potential solutions.

4.1.3 Develop: Solution Diverging

At this point I'm expected to have a prioritized list of requirements, with statements such as "High arousal songs should be portrayed as fast and aggressive" and "visual appearance should be responsive to the audio". These requirements can however be realized in many different ways depending on which metaphor, approach, and solution to proceed with. Exploring different solutions and ways to fulfill the characteristics of the problem is important, and the focus can be shifted from the problem to the solution. With that said, due to the limited time span and resources of this project, all solutions can not be exhausted, but exploration through ideation and prototyping is required. I will require structuring and prioritization so that the next phase can elapse without pitfalls.

4.1.4 Deliver: Solution Converging

The ideation and exploration of the previous phase will yield several low-fi prototypes and the best of them will be refined. Due to time restrictions, I anticipate only being able to conduct one design cycle meaning that the development will result in a single prototype that can be tested but the feedback will not lead to additional iterations. Visualizations are rarely evaluated on tests with proper measures [5] and I want to make sure that I take my time to test the creation to contribute to the Zimmerman criteria for extensibility [40].

With that said, this evaluation step of the project is highly intriguing but optional in the scope of this project. If problems occur along the way that delays the timeline

this phase will be culled.

4.2 Focus Group

A focus group is a qualitative study where themes and questions are discussed in a social context. The goal is to get an insight into new perspectives and generate ideas. Focus group conduction is widely used as a research methodology but not standardized [52]. A focus group requires a clear definition of the aim of the study and a list of questions that create a base for the session. While groups consisting of participants with similar characteristics can contribute to more engaging discussion this notion has been challenged. Breaking the homogeneous groups can help to overcome pre-existing patterns and yield more honest views. Focus groups usually have between six to eight participants and a minimum of three focus groups are recommended per research topic. Large research areas have been suggested to strive for theoretical saturation, meaning continuously conducting more focus groups until they do not produce any new data [53]. Three focus group sessions are approximated to reach 80% theoretical saturation, whilst four to five sessions will reach about 90% saturation [54]. The session should not take longer than two hours to not exhaust the participants [53]. Social context and group dynamics are important aspects of the method and contribute to the data collection. A skilled focus group leader should facilitate the group and encourage varied perspectives by asking follow-up questions and promoting discussion between the participants [52]. The data from a focus group session should be transcribed, analyzed, and presented [53].

4.3 Thematic Analysis

Thematic analysis is a method that can be used to extract themes from qualitative methods such as interviews or focus groups. The method can be seen as a bottom-up approach where insights are extracted and emerge from the data. The method can be broken down into five steps, compiling, disassembling, reassembling, interpreting, and concluding.

Compiling entails the transcription of qualitative data as well as getting a sense of the entirety of the captured data.

In the disassembling step, the transcription gets broken down into meaningful groupings. Through the process of coding meaningful segments represent ideas and thematic keywords are extracted. Coding can be decided a priory or generated as the data is proceeded [55]. Different disciplines and paradigms utilize different terms [56] and coding should be domain specific [55].

In the reassembling step, the initial codes get combined into focused codes which represent larger ideas and coherent themes. To get an overview of the thematic landscape reassembling can be done through a matrices approach where participants, concepts, and themes are structured in rows and columns, or in a thematic hierarchy where themes get sub-themes, etc.

While the data gets interpreted from the start of the method a discreet interpretation step can help with the understanding of the relational and structural contents of the themes and ideas. Visualizing the thematic clusters or creating a thematic map can help in understanding the relationship between themes and codes.

The concluding step simply states that some insight should have been derived from the initial data. Worth noting is that the findings from a thematic analysis are difficult to replicate and seldom generalizable due to the context of the qualitative data collection. The conclusion can still be valuable to the research and problem at hand but generalizing the findings should be done with care [55].

4.4 The MoSCoW Analysis

The MoSCoW analysis is a prioritization technique that structures deliverables in tiers of importance. The technique aids development and creates an order in which tasks should be explored or implemented. The first tier is the "Must have"-tier where features crucial to the project are listed. The second tier is the "Should have"-tier where high-priority items are listed which would make great contributions to the project. The third tier is the "Could have"-tier where desirable features are listed that could be implemented given that time and resources are available. The final tier is the antagonistic "Won't have"-tier which lists requirements that should be avoided [57].

4.5 Crazy 8 Brainstorming

This method is a simple brainstorming technique used to rapidly generate ideas. Take 1 minute to come up with an idea for the problem, draw it on a piece of paper, repeat this eight times, and discuss the generated ideas. The purpose of the method is to generate a large number of ideas, which can help start the design process and can help to spark innovative creative thinking [58]. The technique can be an effective creative tool for designers and problem solvers when exploring potential solutions.

4.6 Importance/Difficulty Matrix

Ideas can be structured on a grid with two axis, one axis for ease of implementation and one for impact significance. This creates four discrete fields which each symbolize different types of ideas. The grid is used as an aid for understanding the return on investment [59], where ideally you want an impactful idea that's easy to implement.

4.7 Prototyping

Prototyping is a fundamental element of the design process and affords designers to explore and refine their ideas, test assumptions, and gather feedback. Both

basic Low-fidelity (Lo-Fi) and complex High-fidelity (Hi-Fi) prototyping become important tools for the iterative development process of this project. The usage of prototypes in this project is best described as mediums used to manifest the ideas, as fidelity increases incrementally.

4.7.1 Lo-Fi Prototyping

Lo-Fi prototyping employs using rudimentary means and techniques, such as sketching, to explore and iterate design concepts. It affords the designer to put an emphasis on conceptual qualities and temporarily ignore specific details. Lo-Fi prototyping is highly modifiable and can aid in the discovery of pitfalls and problems at an early stage in the design process [60].

4.7.2 Hi-Fi Prototyping

Hi-Fi prototypes are typically employed in the later stages of the design process after the project has been thoroughly outlined. They frequently employ the same techniques and technologies as the final product, rendering them more expensive to produce. Hi-Fi prototypes offer superior interaction capabilities and convey a wider range of design possibilities. However, due to their narrower focus, they are more resource-intensive should the design space require further exploration [60].

4.7.3 Prototyping as Filters and Manifestation

Conventionally prototyping has been divided into low and high-fidelity, but utilizing prototypes as filters or as manifestations has been suggested to support design exploration. Prototypes can be viewed as filters of ideas by breaking down the concept into specific sections. For example, prototyping functionality can aid in idea generation and can be viewed as a separate process from prototyping appearance or means of interactivity. By viewing prototyping as a filter, discreet subareas of an idea can be explored in isolation [61].

In addition, the process of prototyping can be viewed as a manifestation of the idea. Manifestation can vary in aspects such as material, fidelity, and scope, and over iterations, the prototype is refined and altered. The idea is manifested in the prototype and as the prototype evolves in regards to material, fidelity, and scope, it becomes a better representation of the idea [61].

4.8 Evaluation

4.8.1 Summative Evaluation

Summative evaluation traditionally refers to the quality assessment of a completed product, and it is typically carried out using formal measures [62]. This evaluation is usually performed at the end of the design process to determine the effectiveness of the product in meeting its goals and expectations [63].

4.8.2 Affect Rating

The Positive and Negative Affect Schedule (PANAS) is a widely used instrument in psychology research for measuring affective states and traits, and it has been shown to have good reliability and validity. Participants assess stimuli using a self-report questionnaire based on an array of emotions. The point scale ratings are then used to derive the values of arousal, valence, positive affect (PA), and negative affect (NA) [64].

Although established evaluation techniques and measures are rare for music visualization [5], PANAS has been used to assess the felt and perceived emotions of music [46]. Conveniently, PANAS uses a 5-point Likert scale, but research has shown that altering the scale to a 7-point scale can reduce certain unwanted answering behaviors [65]. This alteration does not affect the formula used to extract the PANAS ratings; rather, it increases the range from -8 to 8 to -12 to 12.

4.8.3 Experimental Survey

Experimental surveys can be used to determine distinctions between conditions. Both between-subjects and within-subjects designs have their strengths and weaknesses. Within-subjects design increases statistical power since it subjects each participant to each condition, but it can be subject to ordering effects and require counterbalancing [66]. To minimize potential order effects, a Latin square balancing method can be employed to make a feasible effort to minimize ordering effects [67] [66].

5

Process

5.1 Pre Project Development

As a starting point for this project, we utilized a developmental environment that we had constructed in a previous project course. The environment consisted of a pipeline for uploading a file, estimating affect values, extracting audio features, and making the data available in a 3D scene. In addition to having a solid base in terms of technology and code, we could also account for a significant chunk of learning and experience with the intricate technologies needed to create the visualizer. In particular with the real-time feature extraction library Meyda [34] and the 3D rendering library Three.js [33]. This made it possible to accelerate development and focus on the vital parts of the project, such as exploring music's affect and how it could be represented.

Overall, having the development environment as a starting point was a huge advantage in the development process of the visualizer. With that said, the development environment went through many changes during the project's runtime and required a lot of continuous learning to adapt to the new requirements and features of the visualizer.

5.2 Discover Phase

5.2.1 Focus Group Study

The project required some form of empirical data as a foundation due to the uncharted nature of the domain. A focus group was chosen as it would yield qualitative data that could accurately describe emotions and visuals. Additionally, techniques such as a sketching session could be integrated to facilitate communication of the complex domain of emotion. The focus group resulted in an insight into how music can be visually imagined, which would serve as the cornerstone of the visualizer's aesthetic profile. A detailed plan of the focus group can be found in the appendix (Appendix A.1.4).

5.2.1.1 Material

The focus group was conducted in a quiet 8-person group room at the Chalmers University of Technology and the audio was recorded. Each participant where given a pencil for writing and sketching as well as nine pens with the colors red, orange, yellow, green, light blue, blue, purple, pink, brown, and black.

5.2.1.2 Participants

12 participants were recruited through convenience sampling on social media platforms such as Facebook and Instagram. Seven participants reported that they had *less than 1 year* of visual expertise and five participants reported that they had *between 1 and 5 years* of experience. Regarding musical expertise four participants reported having *less than 1 year* of experience, three participants reported having *between 1 - 5 years* of experience, and five participants reported having *more than 10 years* of experience.

The focus groups were divided into three separate sessions and each session took 1 hour to conduct. Initially, the focus group was to be conducted in English since the participant invitation pool was international, but since all participants were fluent in Swedish it was instead conducted in Swedish, which was thought to capture the experiences of the participants in higher detail.

5.2.1.3 Procedure

The participants signed a consent form (Appendix A.1.1) that informed them about the study procedure and ethical concerns. Participants also filled in an expertise form (Appendix A.1.2) by reporting their years of experience within the audio and musical domains respectively. Participants where given chocolate bars. The focus group session was initiated with a simplistic ice-breaker question and then the topic of the session was introduced. The moderator showed and introduced the participants to the base state of the music visualizer, which was meant to aid in an understanding of the scope and the possibilities of the project. The first part of the session consisted of a brainstorming-sketching session.

Participants were prompted with an emotion and were tasked to draw a representation of how the visualizer could represent a song with the given emotion. Describing the experience of emotions, visuals and music in words can be an expert skill, and manifesting it as a Lo-Fi prototype was meant to aid in creativity and the sketches worked as a tool of communication. The prompted emotions/states were the features extracted by Essentia.js, that being *happiness*, *sadness*, *aggressiveness*, *relaxedness*, and *danceability*. Participants were given two minutes of sketching per emotion. The participants had access to the same types of pens and color variations. A paper naming different types of visual features was displayed to incite inspiration during sketching (Appendix A.1.3).

After sketching the participants were asked to motivate how they had used visual features to represent the different emotions. Participants were also queried about the relations between visuals and audio as well as musical metaphors. The moderator

asked follow-up questions and at times asked participants to motivate certain aspects of their drawings.

From the first group, it became clear that more emphasis should be put on the three-dimensional aspects and subsequent groups underlined this to a greater extent in the instructions.

5.3 Define Phase

5.3.1 Thematic Analysis

The thematic analysis was conducted to extract tangible meaning and patterns from the data gathered during the focus group sessions.

During the compiling phase, the qualitative data captured from the focus group was transcribed using the auto-transcription feature of Microsoft Word [68] to get a rough parse of the audio recording, in addition, the rough transcription went through two manual passes by the moderator to accurately represent what was included in the context of the focus group.

Through a bottom-up disassembling approach, the transcription was explored and inspected to find quotes that represented meaningful ideas or labels. The coding was done using a rows-and-columns approach in a spreadsheet. 93 unique codes were found during this stage.

During the reassembling phase, the codes were grouped into seven overarching themes, that being "emotions", "color", "shape", "behavior", "visual elements", "musical elements" and "metaphors".

Colors denoted the hues, range, brightness, saturation, and amount of colors used to represent the emotion. Shape denotes attributes such as size, length, density, symmetry, and roughness. Behavior includes ways to describe how the objects could act in the scene, which includes speed, responsiveness, vibrations, and predictability. Visual elements denote specific mentions of objects available in the 3D scene, such as particles, fog, and lights. Musical elements include mentions of musical attributes or instruments, such as bass rhythm and tempo. Metaphors include words that were used as parables to emotions or to music as a whole.

In addition, the sketches were clustered into their respective emotional categories to aid in pattern recognition. The four most representative sketches for each category were selected to be ambassadors for their respective emotions. The sketches were selected based on how well they captured the colors and shapes of the generally expressed attitudes, as well as how well they represented specific concepts or ideas. Representative quotes were translated from Swedish to English.

5.3.1.1 Interpreting

In the interpreting step of the thematic analysis, I identified patterns in the collected data, images, and metaphors.

Figure 5.1: The sketches from the sessions was set up on a wall. The rows represent emotions, and the columns represent participants. The orange post-it notes denote representative sketches.

Happiness: The color of happiness was described to be bright and saturated. Most prominent was the yellow hue, but green and blue were also mentioned. Specific importance was put on the brightness of the color. *Person 11 "Yes same, I tried to pick bright happy colors like yellow, blue, and bright green."*. The shapes associated with happiness were described as soft, sprawling, and wavy. *Person 3: "Outwards, I wanted to move outwards on the paper. It's not really something contained, it's big and spread out."* The higher intensity of happiness the more exuberant, rapid, and spread out the visuals were imagined. Happiness was described using fire and weather metaphors, in particular sunlight. *Person 2: "Like a gust of wind, like a spring wind, it comes and gives hope about sunlight. "*

Sadness: Sadness was described by using dark hues of colors, in particular black and blue. *Person 6: "Yes, I think only in hues of blue, different dark blue hues."* Participants expressed that if they were given a wider range of colors with different brightness they would have selected the darker ones. Sad colors were imagined to be less sharp and more dialed down in regard to saturation. Sadness was described to be round, compact, wavy, contained, and flat. *Person 4: "Yes, sorrow never gives, it always lays down, It is like, sand in the ocean or something. It sinks and then it lays there and so on."* The movements of sadness were described as slow careful movements which take time to respond to the musical features. The visualization performing sad music was imagined to be less reactive and not instantly answer to events, but mold slowly between different states. *Person 12: "I would say rather slow and soft, but which at times can act out. But not particularly pointy."* Rain, drops, and fog were the most frequent metaphors describing sadness. Rain could also be described as a liquid with high viscosity. *Person 3: "It's like, it's not yes, more like a sharp fog that shouldn't look too good... However, here in the middle, there is a lump, it has very high density, a lump that spread like the plague."*

Figure 5.2: The four most representative sketches of the aesthetics of happiness.

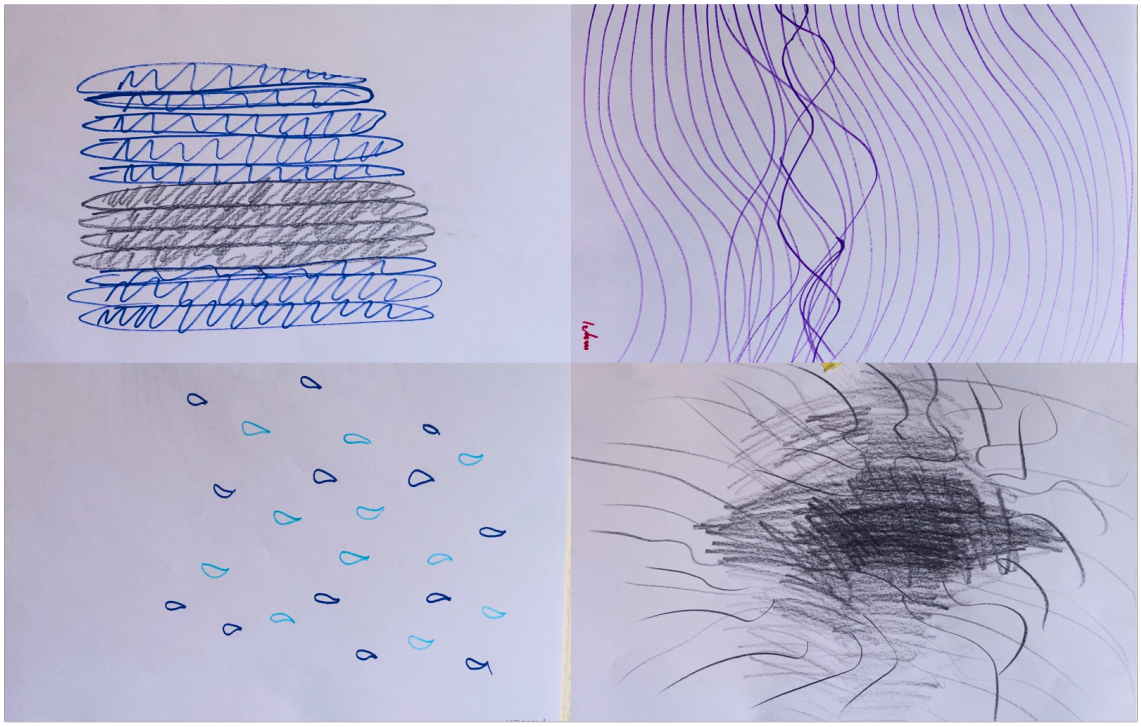

Figure 5.3: The four most representative sketches of the aesthetics of sadness.

Figure 5.4: The four most representative sketches of the aesthetics of aggressiveness.

Aggressiveness: Red, orange, and black were the most prominent colors used to describe aggressiveness. Yellow was used on occasion. The colors were described as sharp and dark. *Person 1: "It should be dark and harsh colors, I think it should be sharp colors, red is very good."* The shape of aggressiveness was described to be pointy and solid with sharp angles. *Person 4: "I also went with that, like triangles. It's the most pointy shape we got."* The behavior of an aggressive visualization was described to be of high intensity, including sudden and fast jerky changes. Aggressiveness was also described as actionable with a decisive momentum towards a direction. *Person 1: "I really agree about direction, it feels like there's a course toward a specific direction, at least for me."* Metaphors used for aggressiveness were fire, explosions, and chaos. *Person 6: "I think about chaos. Like, it doesn't feel like there's any logic to how things move around."*

Relaxedness: Colors used to represent relaxedness were, in general, dialed down, and bright, green, and blue were most frequently used. *Person 6: "I went for nature, so green and bright blue, quite naturalistic."* The shapes used were big, soft, light, round, and wavy. *Person 11: "Rather soft large waves, not at all hard, but softy soft".* Most notably relaxedness was denoted as a slow emotion and movements were sparse and careful. The visualization was described to showcase changes over time, meaning it should not react to every instance. *Person 11: "Maybe not every bass drum, like, gives results, more like changes over time than every beat."*Nature elements such as cloud, forest, and the ocean was used as metaphors to describe relaxedness.

Danceability: Danceability was described as a colorful display of a wide range of

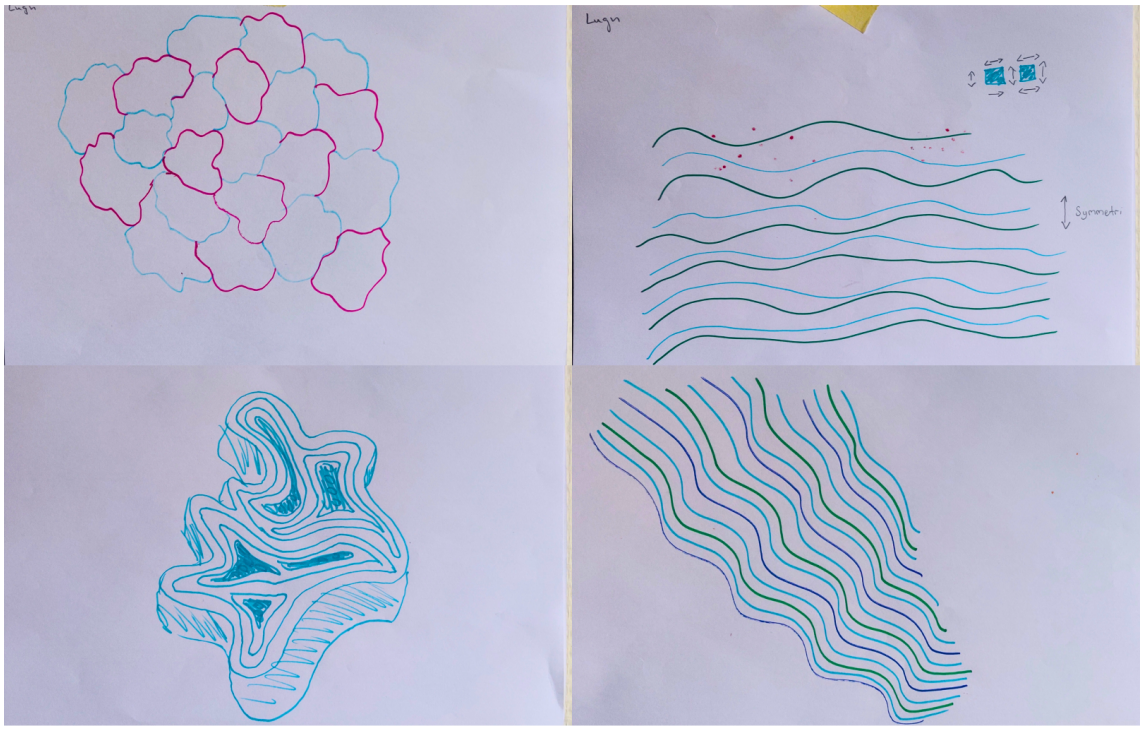

Figure 5.5: The four most representative sketches of the aesthetics of relaxedness.

saturated colors. Neon colors were also brought up. No particular shapes could be identified to be closely related to danceability. Danceability denoted the importance of showcasing the rhythm and bass. *Person 4: "Dance music is often very rhythmical. That's the point with dance music, that it's recurrent and often mathematically rhythmical. That's why I drew everything at an even distance."* . *Person 9: "I think someone deaf should be able to look at it and be like, without feeling the vibrations, know exactly how to dance."*. Bass as an audio feature was described as large and wide in scale. Metaphors used for describing danceability were stars, the 70's, and rainbows.

Interactions: In general, participants expressed no particular interest in actively interacting with the visualizer and engaging in the visualization was seen as a passive activity. The discussion, however, arouse about whether active involvement would advance the experience of the audio-visual experience.

Importance: The most important features were the colors and movement of the objects in the scene. Closely followed by their shape. *Person 11: "Color and movement are the most important. Everything can be represented by a ball, in some way."*. Colors were said to be closely linked to emotions and the behavior and movements of objects in the scene were denoted as crucial since without them it would just be a still image.

Strucutral Metaphores for Music: Participants thought that weather was a good metaphor for music and that it has a lot of potentials to convey different emotions. Participants brought up sunshine, rain, and storms as variations of weather that could represent emotions. The usage of weather metaphors to describe emo-

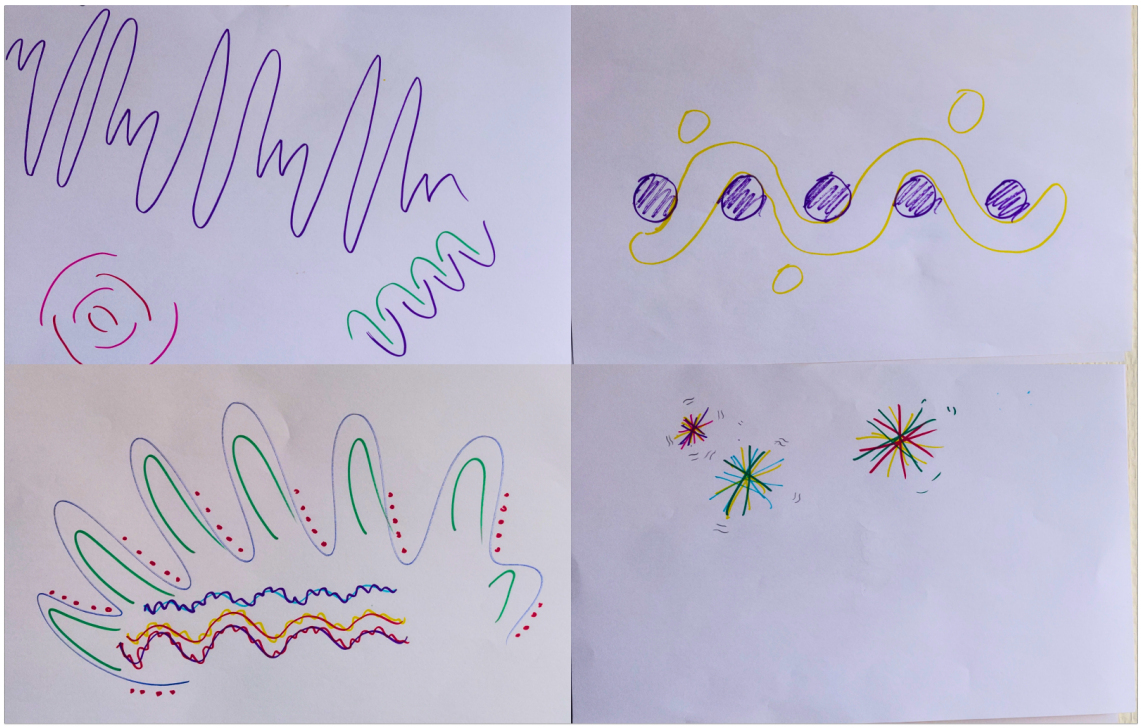

Figure 5.6: The four most representative sketches of the aesthetics of danceability.

tions were frequent across all groups. Participants also described music as a trip. *Person 5: "I like the idea of a song being some type of trip, like it constantly moves forward in a way. It moves, somewhere, towards something or away from something...It doesn't matter if it's Vivaldi or Taylor Swift. I feel like every song is some type of trip."*Music could also be represented as a liquid. *Person 11: "My initial thought is, like water. It can take many shapes, in a way. It can be aggressive or calm."* Music represented as aerodynamics affecting objects also came up during the discussion. *Person 9: "It's something that blows and depending on the mood the object's shape changes and is affected."*

5.3.1.2 Concluding

The results of the thematic analysis revealed multiple aesthetic properties uniquely related to specific emotions. The visualizer was designed with these patterns in mind. Many of the patterns were antagonistic, such as happiness being bright and sadness being dark, or aggressiveness being fast and relaxedness being slow.

No music was played during the study to avoid priming the participants which might have caused the focus to shift slightly from the experience of emotions in music to just mere the experience of emotions.

The emotions used for the study were based on the terms used in Essentia.js, however, translated into Swedish. It is possible that the Swedish translations had different connotations, and the same goes is true for the translation of the quotes. It became clear that the participants had metaphors in mind when describing the visual elements of emotions.

Type	SubType	Happiness	Sadness	Agressivness	Relaxedness	Danceability
Color	Brightness	Bright	Dark	Dark	Bright	
Color	Saturation	High	Low	High	Low	High
Color	Hue	Yellow, Lightblue, Green, Pink	Blue, Purple, Black	Red, Orange, Black	Green, Blue	Colorful, Neon
Shape	Solidness	Soft	Solid	Solid	Soft	
Shape	Angles	Round	Round	Pointy	Round	
Shape	Size	Big	Small	Big	Big	
Shape	Spread	Spread	Contained	Spread	Spread	
Shape	Wavyness	Wavy	Wavy	Straight	Wavy	
Shape	Compactness	Wide	Compact	Wide	Wide	
Behaviour	Speed	Slow/Fast	Slow	Fast	Slow	
Behaviour	Reactivness	Fast	Slow	Fast	Slow	High
Behaviour	Carfullness	No	Yes	No	Yes	
Behaviour	Jerkyness	Yes/No	No	Yes	No	
Behaviour	Intensity	Low/High	Low	High	Low	
Behaviour	Momentum	Yes/No	No	Yes	No	
Behaviour	Amount of Movment	Low/High	Low	High	Low	

Figure 5.7: Summary thematic analysis results.

Participants also commented on the lack of colors. Ideally, more resources should ahve be allocated to the range of colors, in particular in regard to brightness and saturation. With that said, participants verbally announced their color preferences in most cases where the available color range was insufficient,

5.4 Develope Phase

5.4.1 Crazy 8 Brainstorming

The ideation method was used to kick off the next phase of the project with a spurt of creativity and playfulness. I engaged in a rapid Crazy 8 brainstorming session. The goal was to generate eight unique ideas on how to realize the visualizer in regard to the overarching structural metaphor for music or to highlight important aspects that could bring value to the visualizer.

The first sketch showcases how the visualizer could be based on a black hole radiating Hawking radiation. The music could affect how the center pulsates, its density, and when particles are shot out from the center. A user could possibly interact with the camera or rotate around the center.

The second sketch portrays a landscape made up of audio waves with a glowing orb at the end of the horizon. The landscape would change dynamically to the audio. The orb in the background would symbolize the essence of the song and could radiate color.

The third sketch is inspired by the game Guitar Hero. Shapes glide towards the camera in different lanes depending on frequency.

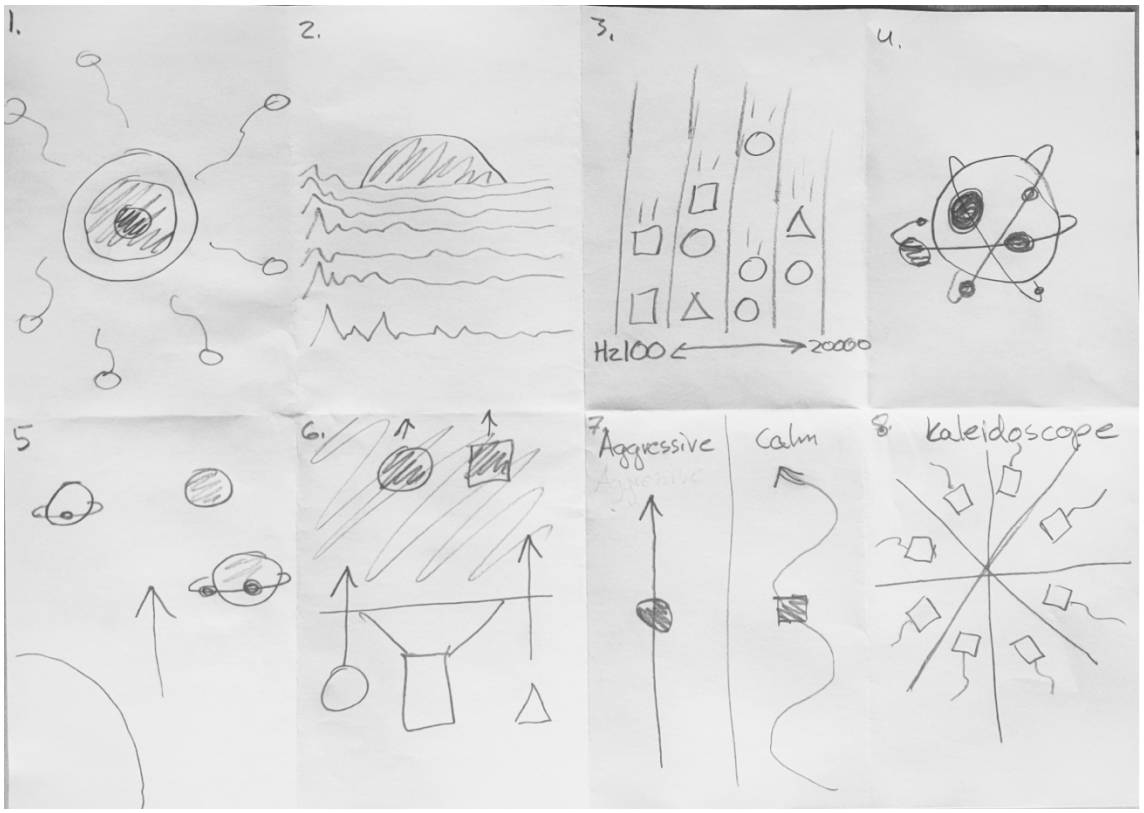

Figure 5.8: The sketch from the Crazy 8 brainstorming session.

The fourth sketch is using a solar system or an atom as a metaphor for music. The orbs could represent and react to different aspects of the music, and their speed and trajectory could be determined by the tempo of the music.

The fifth sketch represents music as a trip through galaxies. The journey passes the galaxy which represents a segment of audio features.

The sixth sketch is similar to the fifth, with the slight difference that instead of traveling into a galaxy the trip is traveling away from objects. The objects could spawn off the screen and be moved into the scene to create the effect of falling or traveling away. This would have temporal advantages since musical events can be made into objects and instantly be put into the scene, close to the camera which would create better mapping than spawning objects in the distance.

The seventh sketch simply states that the manner in which objects travel through space could be determined by emotions. Songs that are mainly aggressive should have a straight and clear course while calm songs could have objects travel in a sine wave through space.

The eight sketch showcases the music as a symmetrical kaleidoscope.

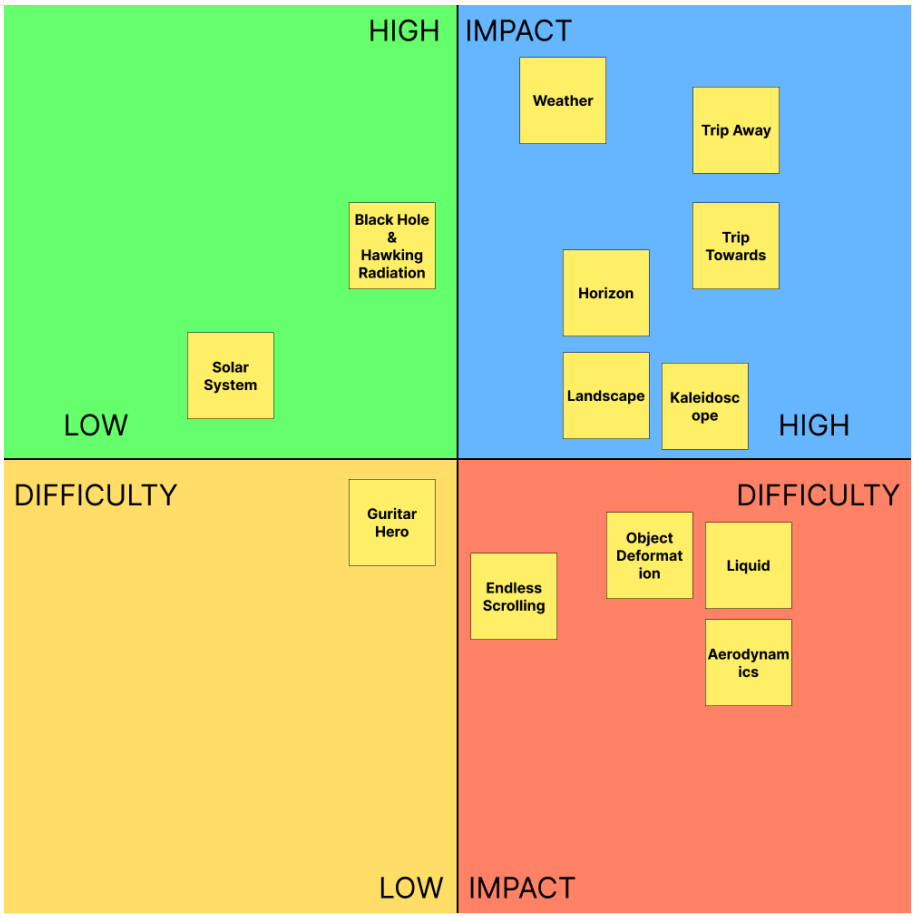

Figure 5.9: The ideated potential music metaphors displayed in an importance difficulty matrix.

5.4.2 Importance/Difficulty Matrix

The method was used to get a structured overview of the project's possible directions and ideas. The ideas generated through the focus group and the ideas generated by the brainstorming session were written down on post-it notes. The ideas were evaluated on their potential to communicate musical affect and features as well as how difficult it would be to program them.

Difficulty	Impact	Idea
Low	Low	"The Guitar Hero" idea was determined to be rather unoriginal and there were no clear-cut ways to represent the music's emotion.
High	Low	"Endless Scrolling" as a metaphor for music would not utilize the 3D scene to its full potential.

Continued on next page

Table 5.1 – continued from previous page

Difficulty	Impact	Idea
High	Low	"Object Deformation" would be interesting to utilize in a visualizer but would be challenging to code since it would require mesh manipulation. In addition, the shape of an object was deemed by the focus group results to be less important than other aspects such as movement and color.
High	Low	"Liquid or Aerodynamics" as metaphors for music would be very difficult to code since it would require an involved physics engine.
High	Low	"A Solar System" concept would be easy to create and the orbiting "planets" could represent different audio features. Communicating emotions would however be rather vague.
High	Low	"Black Hole and Hawking Radiation" could represent the essence and density of the song and the behavior of the radiating particles could aid in the communication of emotions and auditory features.
High	High	"Weather" is an exceptional metaphor for emotions and has connotations to music.
High	High	"Horizon and Landscape" could be constructed in a way to represent audio features and emotions, it is however determined to be rather uninspired.
High	High	"A Kaledioscope" as a means to represent music would be aesthetically pleasing but not as unique as some of the other concepts.
High	High	"A Trip Towards or A Trip Away" as a metaphor for music was highly regarded by the focus group participants. A distinction was made in regard to what direction the trip would travel. A trip traveling away from objects seemed like the best option since the music can trigger events to spawn objects and bring them into the scene close to the camera. This entails that objects close to the camera are temporarily close, and vice versa, objects in the distance represent music events that occurred in the past.

Table 5.1: Short descriptions of the considered ideas.

5.4.2.1 The Selected Metaphore

The best concepts were determined to be the weather, the trip away, and the black hole. The ideas and metaphors were combined into one executable concept.

The fundamental structure was imagined to be a trip moving through space. This

Figure 5.10: The image to the left represents a song with neutral arousal and high valence while the song to the left represents a song with high arousal and slightly negative valence.

illusion would be realized by having particles move at a constant rate away from the camera. The trip represented the duration of the song. Based on musical events objects would spawn into the scene. In the middle of the scene, there would be an object which represented the essence of the song, ideally, this would be a highly moldable object. The object would pulsate, bounce, and radiate particles, which would afford many visual parameters that could be used to represent the song, such as the trajectory and behavior of the radiating particles. The illustration would not portray weather, however, it would be used as inspiration from the manners in which weather successfully communicates emotions.

5.4.3 Lo-Fi Protityping

A brief Lo-Fi prototyping session was conducted using simple means to explore and get confident with the parts and visuals involved in the project without expending significant time and resources. Lo-Fi sketching was conducted to get an idea of the general aesthetic profile of the visualizer. Two songs with different affect values were imagined, the first with a generally positive mood and neutral intensity, and the second one with a slightly negative mood and high intensity. The sketches aimed to capture the color choices, shapes, and directions of the music visualizations. The Lo-Fi sketches were recreated in Figma [69] to increase the fidelity slightly and to be able to add a suitable background gradient color.

5.4.4 The MoSCcoW Analysis

The goal of the method was to divide and conquer the complexity of the problem. Breaking down the scope of programming into smaller pieces and ranking them based on importance would aid in the pursuit of a minimum viable product. Given time and resources the program could be expanded upon.

Based on the concept in mind potential features were written down on post-it notes and prioritized using MoSCoW analysis. The features were evaluated based on how easy they would be to implement as well as how efficient they would be to commu-

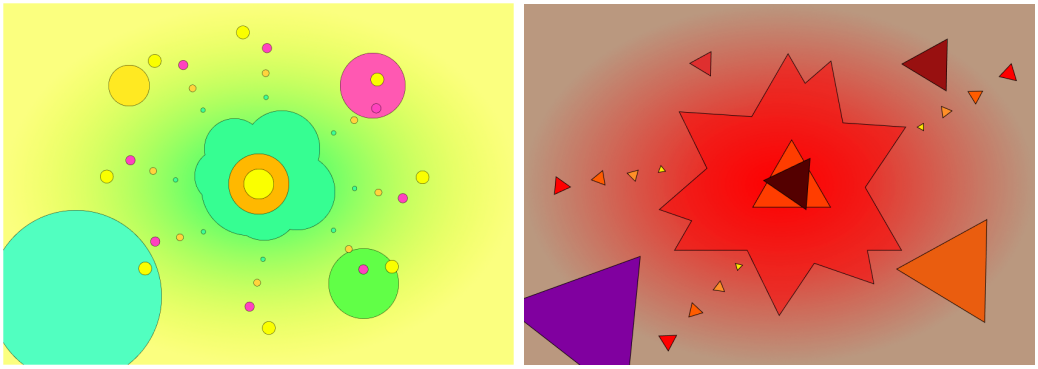

Figure 5.11: The image to the left represents a song with neutral arousal and high valence while the song to the left represents a song with high arousal and slightly negative valence.

nicate emotions, audio features, or the selected concept. The aesthetic contribution of the features was also considered during the prioritization.

5.4.4.1 Must Have Features

5.4.4.1.1 Travel Illusion Particles To sell the illusion of movement particles need to spawn behind the view frustum and move into the scene at a constant pace. The pace could be determined by the song's tempo which would utilize natural mapping since faster songs would also move faster and vice versa. The particles should spawn at random locations at semi-random intervals and move on the z-axis away from the camera. Which would create the illusion of movement. The particles could be created from planes to minimize the number of polygons in the scene.

5.4.4.1.2 Determine Color The results from the focus group underline the importance of colors and the affect estimations will determine the range of colors used in the scene. An algorithm based on the affect values was required which selects, hue, brightness, saturation, and variations of colors. Ideally, the algorithm would select a range of suitable colors.

5.4.4.1.3 Determine Shape Affect values could determine what kind of shapes are used in the scene.

5.4.4.1.4 Essence Shape The essence shape was imagined to be a central point of the visualization. It could be an object with multiple layers varying in opacity which could communicate the density of the object. Sad songs would be presented as more compressed while aggressive and happy songs would be seen as more spread out and airy. The object's size could be mapped to the loudness of the song.

5.4.4.1.5 Radiation Behaviour The essence shape could radiate smaller objects and the way in which these radiating particles move through space could be dependent on affect values. Emotions described as soft and wavy could have particles move in a sine wave while emotions described as jerky could have a triangular

movement trajectory. Radiation particles could also have their own emission. Visualizations with high "aggressiveness" should be more unpredictable which will affect how the direction of the particles is determined,

5.4.4.1.6 Spawn and Despawn Objects A function for spawning and despawning objects was required. Failing to despawn objects would severely affect performance.

5.4.4.2 Should Have Features

5.4.4.2.1 Basic Materials Basic material controls how light would be reflected on the surface of the object and could be used to create the illusion of hard metallic objects. Affect values would control the type of material that is used for the objects in the scene. Materials should be created before rendering to increase runtime performance.

5.4.4.2.2 Fog Affects values would determine the amount and color of the fog in the scene.

5.4.4.2.3 Responsiveness Speed Affect values would determine how reactive objects are and how fast they respond to audio events. This could be done by varying how often the transformations are updated.

5.4.4.2.4 Essence Shape Hover Movement A hover effect of the essential shape could make the visualization feel less static which would aid the illusion of moving through space.

5.4.4.2.5 Distortion Shape Three.js is not a modeling software, however, it would be possible to utilize mesh manipulation to create interesting shapes. Deforming objects dynamically in tune with the auditory features of the scene could help communicate the "pointiness" of objects and contribute to the aesthetics of the scene.

5.4.4.2.6 Derivable Changes By deriving the mean of a sequence of real-time audio values it would be possible to track how the song changes over time. For example, a song might have quiet and loud segments, and utilizing a derivative could aid in tracking the dynamics of the song.

5.4.4.2.7 Trail effects Trail effects behind moving objects could be added for aesthetic purposes.

5.4.4.2.8 Firework/Explosion Creating a function that could spawn objects in a fireworks/explosion-like display could contribute to the aesthetics of the visualizer.

5.4.4.3 Could Have Features

5.4.4.3.1 Post Processing Post-processing effects such as bloom would surely contribute to the aesthetics of visualization, however, they were assessed to be computationally heavy and would only be added if time was available and the effect on performance was negligible.

5.4.4.3.2 Optimization Optimization of the code and rendering could be conducted if time was available.

5.4.4.3.3 Loading Bar When uploading a song the affect estimation takes a few moments to classify the song, and a loading bar could inform the user that the information is being retrieved.

5.4.4.3.4 Interface Creating an interface in which the user can toggle the display on or off would benefit the immersion and usability of the application.

5.4.4.3.5 Usability The general usability of the application would not be a priority but would make the application more enjoyable, such as being able to reupload a song without refreshing the page.

5.4.4.4 Wont Have Features

5.4.4.4.1 Interactability Due to the low amount showed interest in regard to interactivity with the visualizer it was assessed that the visualizer should solely be a passive experience.

5.5 Deliver Phase

5.5.1 Hi-Fi Prototyping

The Hi-Fi prototyping consisted of the realization of the MoSCoW requirements and translating the mere concepts into a working prototype. A scrum-like methodology was applied, where short sprints related to specific issues or requirements were processed in a methodological fashion. The tickets were managed in order of importance with a few exceptions, which I will comment on below. With that said, the Hi-Fi prototyping was highly iterative, and many implemented features were tweaked continuously, refined, and altered until the end of the development phase. The aesthetic profile as well as performance optimization was to some extent discovered during the development when iterative testing and problem-solving commenced. Due to the fact the previous Lo-Fi prototypes didn't manage to capture all the essential features of the visualizer, such as audio and movements, which also led to completely new never before thought of features. For clarity, each discrete block of features will be presented in isolation even though overlap and iterative backtracking were ingrained in the process. The features below are presented in roughly the order in which they were developed.

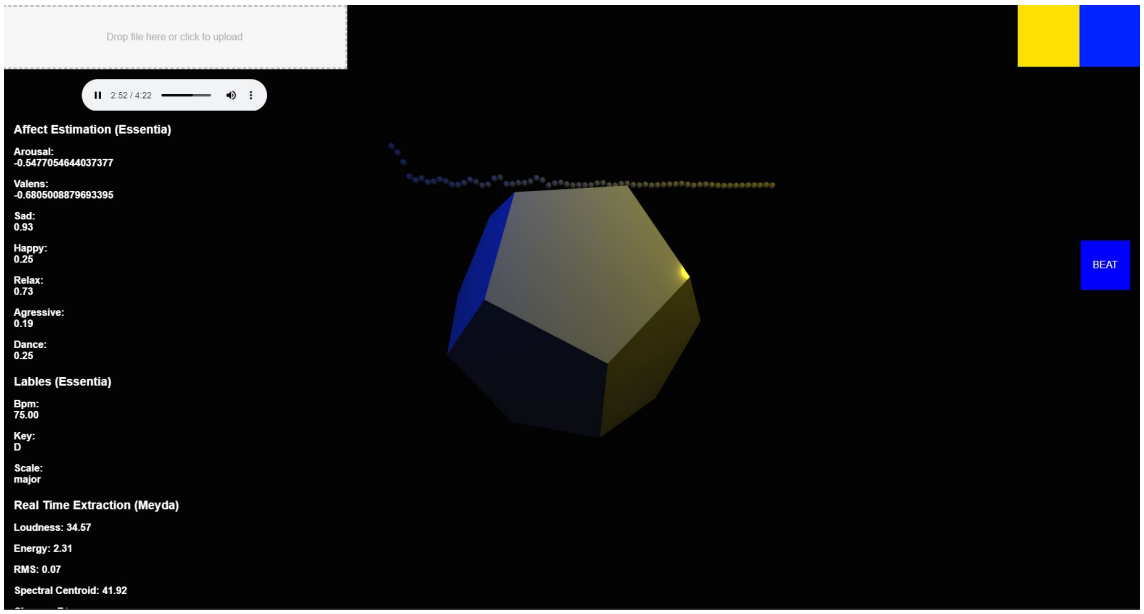

Figure 5.12: The initial environment.

5.5.1.1 The Starting Point

While this environment looks simplistic it remained the fundamental backbone of the project. Major changes were made to the 3D components of the scene and slight changes were made to the interface. The affect estimation and real-time processing were for the most part kept completely constant during the project. Changes to the affect estimation parameters would alter the components built on top of it. Slight calibration of the real-time audio extraction was made in regards to the low pass-filtering Meyda instance where the low pass filter was lowered to 100Hz, from 200Hz.

When development commenced the 3D scene was wiped clean so that new components based on the gathered insights could be explored.

5.5.1.2 Travel Illusion Particles

Spawning particles would be crucial to create the illusion of traveling through space. Simplistic plane geometry was selected for the particles to keep the polygon count low. Early in the development, the particles changed colors as they traveled through space but as more elements were added to the scene the particle color was changed to an emissive white color. This reduced the visual noise of the scene. The particle's x and y position as well as rotation was randomized. The z position was set to always be slightly behind the camera so that the particles never spawned directly in the view frustum.

The particle had a default movement speed of 0.01, however, as a song was loaded the movement speed was updated based on the tempo of the song. This was meant to create the illusion that faster songs travel through space at a rapid pace, and slow songs move leisurely. When a particle had traveled far enough into the distance to barely be visible it was removed from the scene to reduce the number of unnecessary

computations.

5.5.1.3 Determine Color

The results from the focus group denoted the importance of getting the colors right for the scene. While the focus group exhibited patterns in how colors and emotions could be associated it became clear that using mere one color would be insufficient, both in terms of emotional communication but also in regard to aesthetic engagement. To capture the emotional range, an algorithm was developed that generates twelve parts color palettes. Color pallets with more than twelve colors were explored but twelve parts were beneficiary since they could be mapped to the chroma array which captures the twelve different pitch classes.

There are multiple ways to represent colors in code, for example, HSL, RGB, and hex code. HSL stands for "hue, saturation, and light" and affords convenient manipulation of each of these attributes. Since each of these attributes had been mentioned in the focus groups as well as in previous research it would be necessary to have an effective way to alter these features independently, and therefore HSL as a format was valuable.

For each song, I selected a main hue and a complementary hue. To determine the main hue, I calculated the atan angle of a point based on the highest values of *aggressiveness/relaxedness* and *happiness/sadness*. I used the atan angle of the lowest values to determine the complementary hue. These atan angle values, together with a hardcoded bias value, were used to determine the hue circle's rotation, which in turn influenced the hue's input value in the HSL color. The rotation of the hue circle in relation to arousal and valence was decided based on the insights I gathered from the focus group.

Each color pallet sampled eight additional hues around the main hue as well as two additional hues around the complementary color. The range of the sampling, meaning the color variation, was determined by *danceability*, which slightly increased the sampling range, as well as *sadness*, which slightly decreased the sampling range.

At one point during development, the main colors were used to denote the pitch classes within the scale of the song and respectably the complementary colors denoted the pitch classes outside of the scale. This however resulted in five complementary colors instead of the previous three, which aesthetically made the visuals look noisier, and to some extent, the main colors representing the song were perceived to be less important. In addition, this system was based on the notion that Essentia.js was able to classify key and mode accurately, however, this seemed faulty at times. Therefore I reverted back to the previous system of nine main hues, and three complementary hues.

Multiple ways were tested how to calculate saturation and brightness. Exploring different calculations initially resulted in either too bright or too dark values for colors. The most efficient and stable way to decide saturation and brightness was to determine a range in which the values could reside, then use the minimum value as a bias and the remaining range as a modifier. The saturation modifier was determined

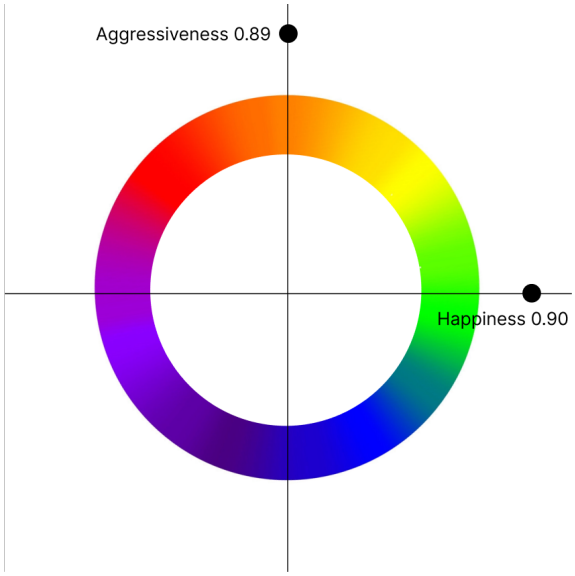

Figure 5.13: The x-axis denoted valence and the y-axis denotes arousal. In this example, *aggressiveness* is higher than *relaxedness* and will therefore be used to calculate the main hues, and the same goes for *happiness* since it's higher than *sadness*.

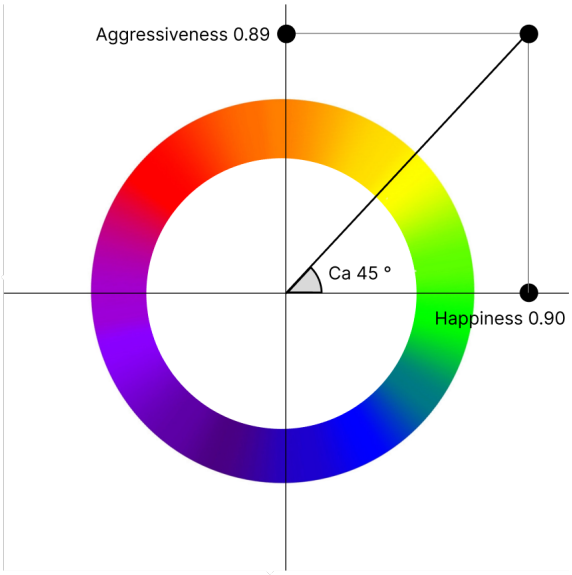

Figure 5.14: Calculating the atan angle given the point denoted by the *aggressiveness* and *happiness*. 45 will be used as the hue value for the main central HSL color function.

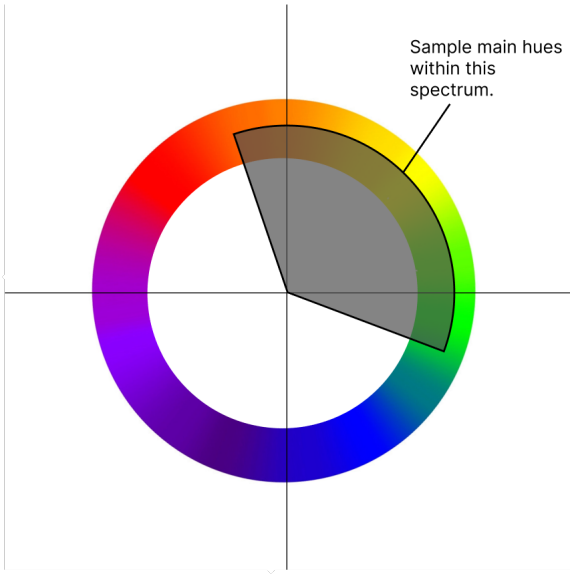

Figure 5.15: Approximately within this range eight additional main hues will be sampled. A high *danceability* value would extend this range.

Figure 5.16: An example of a color pallet where the song is in the key of G.

by "happiness" and the brightness modifier was determined by *aggressiveness*.

Each of the twelve colors in the pallet represents one chroma, starting with the main central hue and the tonic in the track scale.

The background color was initially set to the same color as the main central hue but was over time changed to a highly shaded version of the main hue to better represent a trip through the darkness of space.

Over all other features and aspects of the application, colors were by far the one aspect that was tweaked the most.

5.5.1.4 Essence Shape

The central shape, "Essence Shape", was meant to represent the core of the particular song. Initially, the essence shape was supposed to be represented by a static variation of one of the basic Three.js shapes but vertex manipulation was investigated early on which open up more possibilities. An icosahedron was used to represent the default essence shape, and a basic formula utilizing all affect estimates was used to determine the number of vertices. For example, aggressive sad songs would yield rougher shapes, while calm happy songs would create a more detailed round icosahedron.

By using gradient noise to control the vertices of the essence shape more intricate shapes could be created. In addition, fluid motions were created by using 4D simplex noise, where the fourth degree controls how the noise changes over time. By using the real-time audio feature RMS as the input of the fourth dimension the traversal through time got controlled by the energy level of the song. This resulted in a shape that responded to changes in the music. The pace of RMS traversal was also multiplied by the mean of *aggressiveness* and *danceability*, which entailed that highly aggressive and danceable songs have made an essence shape that was fast-moving and reactive. RMS was also used to alter the size of the essence shape.

The *danceability* was used as a seed for the noise algorithm which entailed that the song will always have the same unique starting position.

Slight rotation was added to the essence shape to make it look more like a comet moving through space.

5.5.1.5 Post Processing

Post-processing was meant to be implemented at the late stages of development if the time plan afforded it. However, post-processing was anticipated to be computationally heavy, so it moved to the forefront of development. If performance dropped too much the effects could simply be removed, but if they did not affect performance they were determined, to more, accurately depict a final visualization which would aid development. Three.js afforded a great way to implement basic post-processing effects an "effectVignette", "bloomPass" and "afterImagePass" was added to the scene without limiting performance. The effectVignette pass added darkened borders along the outline of the scene, which contributed to the immersion. The bloomPass brightened certain already bright areas of the scene to make them bleed light beyond their natural borders. This effect had to be tweaked each time lights or emissions were altered, however, it made a great contribution to the aesthetics of the scene. The afterImagePass pass created fading traces of moving objects which also made a huge contribution to the final product, in particular in regards to conveying the illusion of traveling through space.

A brief dark fog was also added to the scene to make objects disappear in the distance. The fog density was altered by the *happiness*, so happier songs had a slightly less dense fog.

5.5.1.6 Radiation Behaviour

In addition to the main essence shape, the scene required a secondary prominent visual feature, and having shapes radiate from the center was determined to be a good fit, and it would also mimic some of the characteristics of more well-known audio visualizers. The focus group repeatedly brought up waves so a sine wave seemed like a perfect match to control the radiation movements. Overtime triangle waves were added for sad and aggressive songs to incorporate sharper angles and a measure of unpredictability.

Initially, the radiation only moved on the x-and y-axis but to utilize the 3D nature

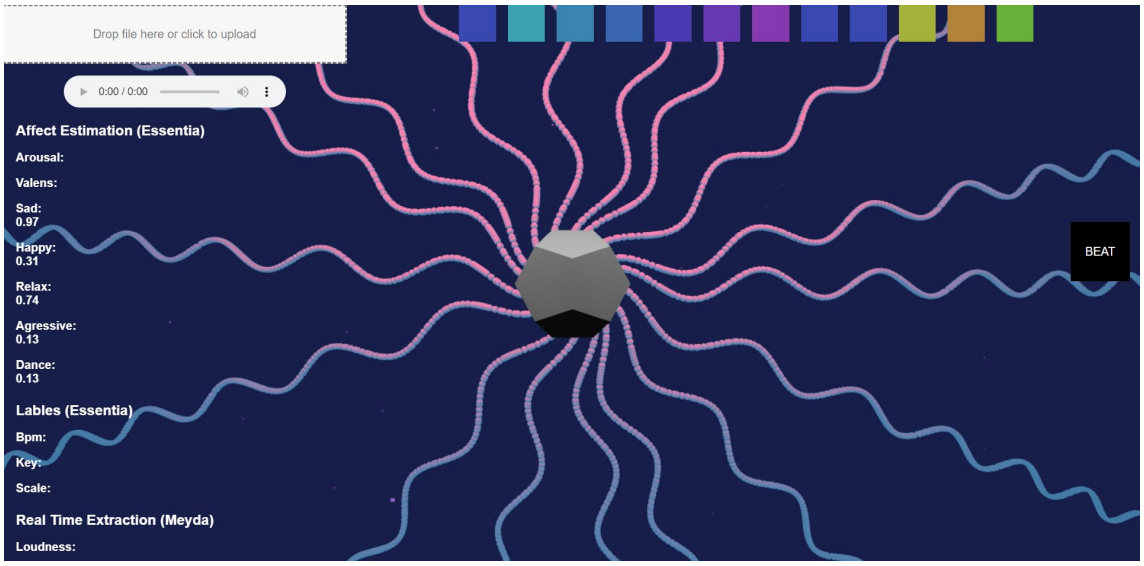

Figure 5.17: Early sine wave radiation tinkering.

of the scene the z-axis was added. Having objects move towards the camera was engaging but quickly managed to feel noisy. Having objects travel away from the camera would, on the other hand, utilize the 3D aspect and could, in addition, communicate the temporal history of the song in an intriguing way.

The shape of the radiation was determined in a similar manner to the essence shape, whereas aggressive songs had rough pointer shapes and calm songs had rounded shapes. Radiation circles spawned when a beat was detected, meaning when the energy in the low-frequency band reached a set threshold. The threshold was determined by the mean of the detected peak RMS multiplied by 0.9 and a mean of the latest 1000 RMS values. This ensured that the threshold would be responsive and suited for each unique song. To control performance the radiation function was passed through a throttle function, which controlled the rate at which the function could be called, and a value of 500ms was set. When the function fired the radius of the spawned objects was determined by the RMS, meaning loud sounds spawn larger shapes. For each chroma value that surpassed 0.95 at the time of the function call additional colors were added based on the color pallet and the chroma indices. This made it possible to differentiate the pitch classes in each beat.

In the early drafts and sketches, the essence shape was imagined to travel passed planets which in themself represented certain events of the music. This concept was only briefly explored but deemed to infer with the radiation shapes and the aesthetics of the visualizer.

5.5.1.7 Data Set

To be able to tweak parameters and customize the visualizer to match sentiments I needed a corpus of songs and data. It would be impossible to test all combinations of affect estimates so instead I created a data set that could be sorted to find songs with prominent attributes, which could be used as guidelines in the design.

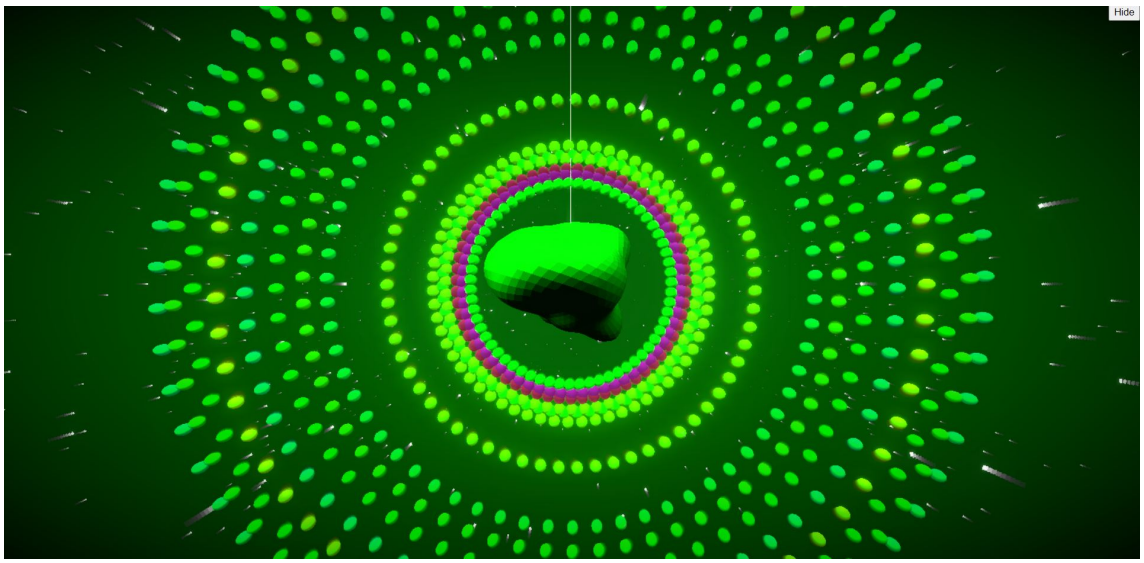

Figure 5.18: A more developed version of the radiation aesthetics.

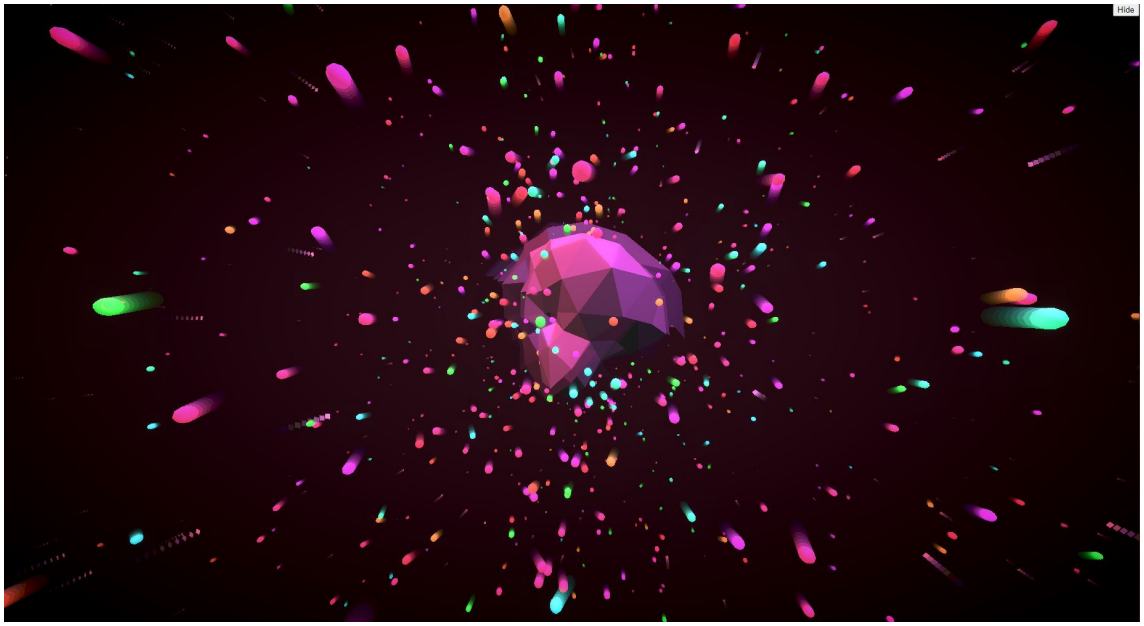

Figure 5.19: This type of radiation behavior was determined to be disorganizing.

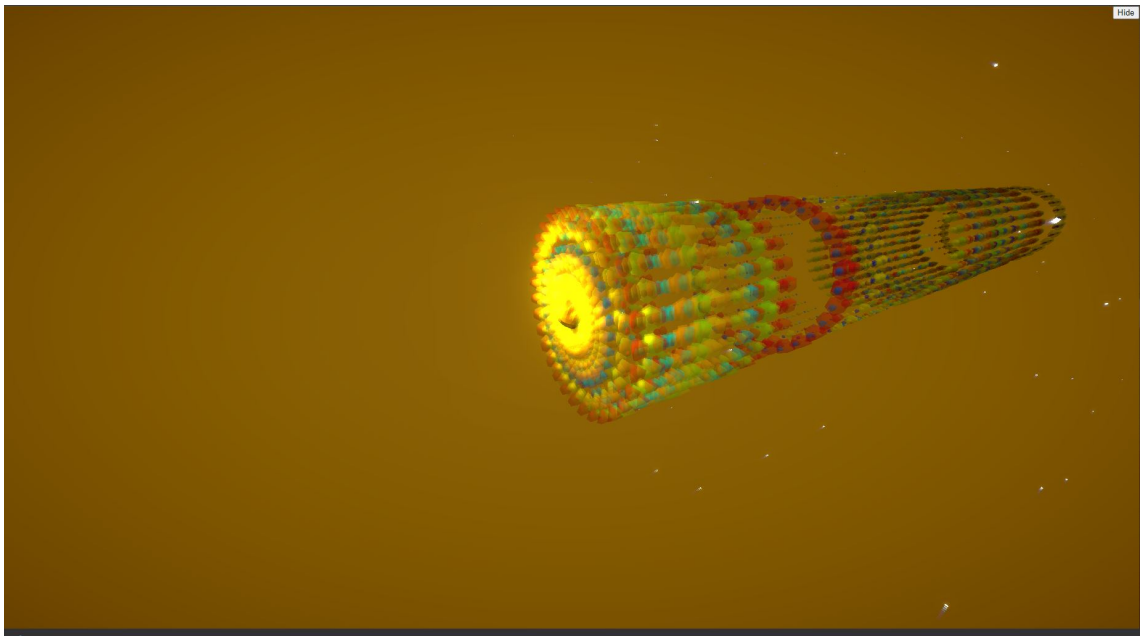

Figure 5.20: A side view of the visualization. Note that the z-axis showcases the temporal aspect of auditory events that took place in the past.

25 royalty-free songs were gathered from the site Pixabay [70]. Each of the five affect estimate values was used as a search tag on the site and for each tag five songs were selected and downloaded at random. The songs were then passed through the affect estimation model and their respective affect values were recorded in a spreadsheet. This made it possible to sort songs based on particular affect values and by that isolate features or attributes. While some songs might have similar *relaxedness* and *sadness* values it does not mean they will behave in the same mannerism since additional parameters such as RMS have a grand effect on the outcome. This data set was mostly used to calibrate colors.

A handful of popular songs were bought on iTunes [71] and passed through the visualizer to estimate the effectiveness of the visualizer. The preconceived emotional engagement to these well knows songs aided in determining how far off the mark the visualizer was.

5.5.1.8 Developer Mode

The Essentia.js estimation could at times, take a couple of moments to load, especially when classifying longer songs. Two ways to run the program were set up. The first was the unaltered pipeline where Essentia.js computed the affect values and iniated Meyda and the 3D scene when loading was finished. The second way to run the application simply preloaded affect estimates for predetermined songs, which made it possible to surpass the loading of the Essentia.js. This became a huge time saver and made it possible to iterate and tweak minor details at a much faster rate. The preloaded data in itself could also be manipulated to investigate the affect values effect on the visualizer.

Num	Song	Sadness	Happiness	Relaxedness	Aggressiveness	Danceability	Bpm	Key	Scale
1	agg1	0,21	0,3	0,44	0,98	0,99	52,5	C#	minor
2	agg2	0,19	0,86	0,41	0,82	0,98	121,95	G	minor
3	agg3	0,95	0,05	0,92	0,3	0,7	100	D	major
4	agg4	0,37	0,25	0,33	1	0,98	131,58	F	minor
5	agg5	0,62	0,18	0,5	0,81	0,82	94,94	G	minor
6	dance1	0,93	0,16	0,73	0,14	0,8	100,67	C#	minor
7	dance2	0,66	0,4	0,6	0,1	0,81	100	G	major
8	dance3	0,72	0,07	0,63	0,07	1	110,29	C	minor
9	dance4	0,43	0,58	0,55	0,59	1	115,38	F#	minor
10	dance5	0,37	0,35	0,46	0,5	1	120	G	minor
12	happy1	0,74	0,56	0,5	0,24	0,94	94,94	F#	major
13	happy2	0,67	0,5	0,64	0,39	1	90,36	C	minor
14	happy3	0,87	0,51	0,88	0,2	0,88	120	D	major
15	happy4	0,94	0,35	0,72	0,07	0,68	85,23	F	minor
16	happy5	0,77	0,64	0,54	0,07	0,8	85,23	Bb	major
17	relax1	0,98	0,31	0,73	0,04	1	75	E	minor
18	relax2	0,97	0,13	0,93	0,05	0,16	119,05	E	minor
19	relax3	0,99	0,17	0,95	0,05	0,05	107,91	E	major
20	relax4	1	0,06	0,94	0,05	0,06	79,79	F	major
21	relax5	1	0,03	0,97	0,08	0,58	83,33	E	major
22	sad1	0,98	0,09	0,96	0,06	0,13	93,75	A	minor
23	sad2	1	0,07	0,89	0,05	0,14	100	A	minor
24	sad3	0,99	0,05	0,97	0,05	0,1	92,59	A	minor
25	sad4	0,98	0,06	0,9	0,08	0,16	75,76	B	minor
26	sad5	0,94	0,1	0,9	0,1	0,25	120	Bb	minor

Figure 5.21: The dataset with the royalty-free Pixabay music. The song selection is available in the Appendix (Appendix A.2).

5.5.1.9 Basic Materials

The focus group denoted the density variation among emotions. While colors are of great importance it was also crucial to control how light reflects off the objects, to convey how solid and dense the objects were to be perceived. Initially, not much thought was put into this but as the development proceeded more modern aesthetic was brought up as feedback when showcasing the progress to other designers. The roughness and metalness properties of Three.js materials were investigated but were insufficient to convey the density of the objects in the scene. Therefore the addition of textures got involved. Nine textures, with normal and height maps, simulating materials with a range of roughness were trialed, from a rough corroded surface to a smooth paper surface.

The texture heightmaps can in isolation contribute extensible to create a modern aesthetic. However, heightmaps require shapes with an extensive amount of polygons to work properly. In the case of this visualizer, the polygon count of each radiant shape was kept to a minimum, which in turn rendered height maps unusable. Without heightmaps, some of the tested textures simply didn't look pleasing.

Utilizing the normal map of the paper texture in conjunction with affect responsive metalness and roughness values proved to be the best all-around option. Some textures looked strange when put on spheres, with clearly visible streeks, and distinct color changes however the sleek nature of the paper texture added a subtle but welcomed dimension to the scene. By varying the material roughness and metalness values the objects could mimic the reflective properties of a range of values such as crystal, steel, and rubber.

5.5.1.10 Reactivness

The focus group participants denoted reactiveness as an important characteristic of emotions and the visualizer. Emotions such as *aggressiveness* were described to be jerky and highly responsive to the changes in the music and *sadness* was described as a more careful evolving display. Much of the visualizer was controlled by the real-time RMS values, however, RMS can jump rapidly between high and low values which could result in jerky animations. To counteract this a mean of the n latest RMS values was calculated which would ensure a continuous smooth animation. The amount of RMS values used to calculate the RMSmean was determined by *relaxedness* so that calmer songs became slightly less responsive. Responsiveness was still kept high to create a connection between the music and spatial components. When the RMSmean value was developed it replaced most other instances of RMS used in the application, such as the RMS values used to control the essence shape.

5.5.1.11 Lights

Several different versions of lights were tested during the design iterations. From the inception, the lights used for the scene had different colors but as the color template got more advanced it became difficult to balance the hues. Therefore the default white light was the most sensible. Four different light sources were used, ambient, directional, and two point lights. Ambient light simply increases the brightness of everything in the scene and was set to a minimal value. The directional light was positioned high on the y-axsís to mimic the atmospheric light from above. One point light was placed slightly above and behind the view frustum to add a slightly angled light source. The second point light was placed at the center of the essence shape to make it radiate light to the surrounding objects. The light intensity of the point lights was controlled by *happiness* so the happier songs had brighter lights.

Shadows were only briefly explored and not really fleshed out in this visualization. The involvement of shadows could have contributed to the visuals.

5.5.1.12 Camera Movement

Initially, the visualization only consisted of the essential shape and naturally the camera was placed in close proximity to it, but as more objects were added to the scene the camera was moved back. Two camera options were then determined, one very upfront to the center of the scene and one more encapsulating far-off view. Eventually, movement based on RMSmean was added to the close-up camera which was meant to create a cinematic experience. The closeup camera in conjunction with its animations made use of the 3D capabilities of the scene while the far-off view had similarities to a 2D visualization.

5.5.1.13 Interactability

Interactability was deemed to be a very low-priority feature and while no conscious interactable capabilities were developed for the users, some interactable features were implemented. Buttons for hiding the hud, changing the camera position, and

disabling the frequency wave were implemented. These options made it possible to slightly customize the visualizer and were seen as quality-of-life improvements. To create a more immersive experience the buttons were also programmed to disappear if the user idle for five seconds.

In addition, orbit control was added to the scene, which affords the user to zoom as well as move the camera freely with the mouse. This was initially used as a developer tool when customizing the 3D scene but after consideration, it was left in since it can be compelling to explore the scene from different angles. The user can also press the camera button to revert back to one of the two main views.

5.5.1.14 Moon Object

The visualization managed to showcase the temporal history, energy, beats, and chroma but struggled to showcase the behavior of different bands/frequencies. A tiny sphere orbiting the essence shape like a moon was created to convey some frequency information. The idea was to divide the frequency spectrum into five bands and derive which of the bands changed the most in close temporal proximity, and based on that make the moon object behave in different ways. Some effort was put into this but the main setback was that it was difficult to make the tiny object matter in the context of the visualizer. Both the object's size, movement pattern, and light emission were tested but one of these parameters was deemed significant enough to be pursued. The moon object was left in but its orbit is simply mapped to RMSmean. Given more time exploration into how to utilize the frequency spectrum would be a main priority.

5.5.1.15 Frequency Wave

As a reaction to the troublesome quest to implement a subtle way to showcase frequency, I went with a more direct approach and visualized the whole audio buffer. The buffer is an array of 512 values that represent a snapshot of the audio signal, and by mapping these values to the y-value of objects I create a representation of the audio signal. To dress up the visualizer the array was divided into segments where the even indexes were put on the bottom of the screen and the uneven indexes were put on the top of the screen. At first, the buffer wave was put in the middle of the screen but this made the scene look very crowded so filling up the already empty space of the scene was a cleaner procedure. The elements were at first created by low polygon spheres but were later changed to simple plane geometry to optimize performance. A button option to disable the signal wave was also implemented.

5.5.1.16 Interface and Usability

During the runtime of the development, minimal effort was placed on the Head-Up Display (HUD) elements and the styling of the associated interface. The HUD was used as a convenient way to present the values of importance during development. Since the option to hide the HUD was implemented the display of affect and audio features were left untouched in the HUD. It's hypothesized that users might even

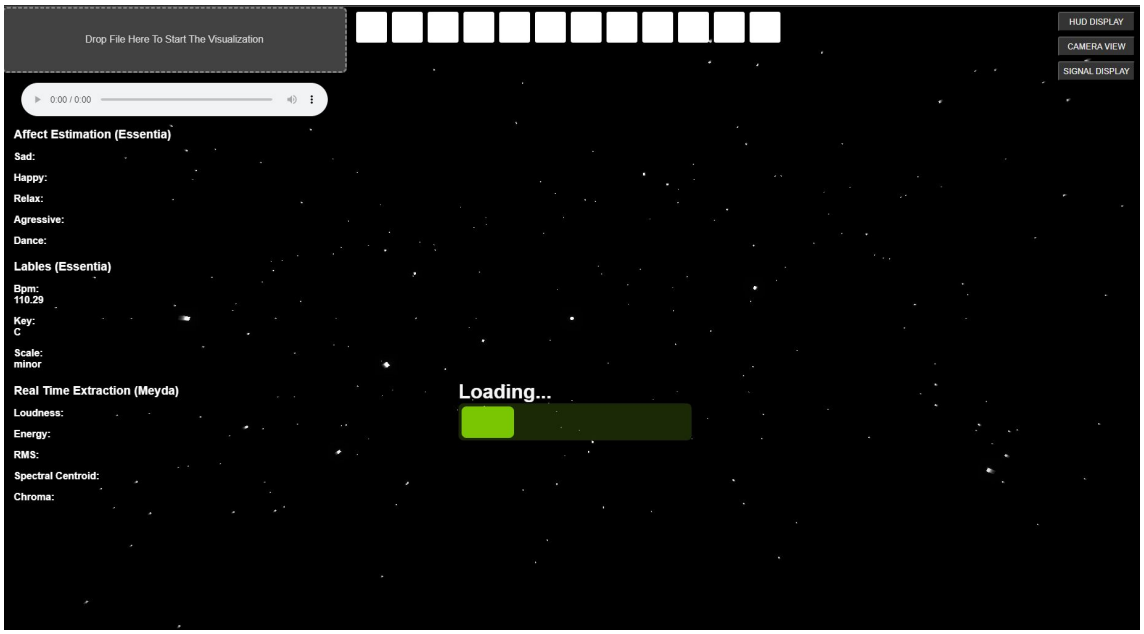

Figure 5.22: The view of the visualizer when an audio file has been uploaded and affect estimates are being extracted.

find it interesting to get access to the affected values of the songs they upload to the application.

The usability of the application is rather barebones. The user can upload a song, control a player and manipulate some basic HUD elements. More attention to the interface design would aid usability. A user also needs to reload the page to upload another song which is far from an optimal solution. In addition, when uploading a song, it takes a few moments to extract the affect values of the song. To counteract this and communicate that the application is running and computing behind the scenes I developed a loading bar. The loading bar has twelve segments and refreshes when certain milestones in the Essentia.js extraction are reached.

The development of this visualizer focused on the 3D scene and the means to communicate affect, however, more attention to detail in regard to the interface and usability would be of high importance if the fidelity would increase further.

5.5.1.17 Optimzation

The application was built with low requirements of computational resources to be able to run on a variety of computers. To minimize frame rate drops some optimizations in regard to performance were kept in mind during the development. The following measure was taken to increase the performance of the application:

- **Despawning objects** that moved too far away from the camera.
- **Keeping the polygon count low** by using low-detail objects, as well as planes when 3D objects are not necessary.
- **Implementing a throttle function** to control the frequency of radiation

objects spawning.

- **Model the design to keep the number of objects low.** Some design choices, such as the number of waves radiating from the center, were solely dependent on the performance of the application. The final amount of radiating waves where about a fourth of the value that was initially set.
- **Utilizing the tools provided by Three.js** when available, such as translating groups of objects instead of moving each object separately and utilizing the built-in Three.js shaders for post-processing.
- **Preloading textures and materials** before starting the visualization to reduce load times.
- **Avoiding shadow calculations** which can be computationally intensive and negatively impact performance.
- **Removing console operations** during runtime to minimize unnecessary overhead and improve performance.

While these measures were to some degree successful in keeping a balanced framerate, I'm not claiming this application to be optimized or that it utilized best practices of 3D rendering.

5.5.1.18 Aesthetics

Throughout the project, the insights gleaned from the literature review and focus group analysis provided a solid foundation for the aesthetic profile of the visualizer. However, the design and development process relied heavily on the creativity and intuition of the designer's first-person perspective. As the project evolved, the initial concept was refined through numerous iterations, with the team exploring multiple paths and solutions. Although feedback from colleagues and associates was occasionally sought for valuable input, external assistance was not always feasible for each design decision. As a result, I relied heavily on my own intuition and expertise. While some solutions were discarded simply because they subjectively didn't look aesthetically pleasing, others were pursued due to biases and inclinations towards a particular design aesthetic.

It also became apparent that compromising in regard to aesthetics was inevitable. For example, songs with different affect values were used as references when modeling the color pallet algorithm. The algorithm was tweaked until it matched the reference songs, the empirical insights, and the designer's preferences. At a later moment, another song was trialed and didn't seem to fit the color pallet as well. Compromising was necessary and the algorithm was then recalibrated to take this into account. It is truly a wicked problem and a monumental task to try to manually calibrate a color pallet algorithm that fits every song in every genre.

As an additional caveat, the uniqueness of each visualization became a factor. While the real-time audio features would be unique for each song, the building blocks and affect estimates that constituted the majority of the scene's main attributes could at times be rather similar. A truly one-of-a-kind aesthetic for each visualizer would be

preferable. I took some liberty in pursuing uniqueness, for example, by implementing opacity under specific affect conditions; however, this was not fully explored.

5.5.1.19 Clean Up, Comments and Publishing

As the project commenced the code was constantly being committed and pushed to the project repository on GitHub. Towards the end of the development, more attention was put on code readability so that other developers can utilize the repository. Some code blocks, such as the creation of the main materials array, were rewritten for the sole purpose of making the code less bloated. A pass to make sure the code followed the established naming conventions and had sufficient comments were also conducted. To finalize the publishing of the code a README-file was created which gave credit to the libraries involved and gave brief instructions on how to run the application.

5.5.1.20 What I Didn't Do

For the most part, I managed to implement the tickets that were created during the MoSCoW analysis. As iterations commenced and became more ingrained some of the previously highly anticipated features became obsolete, such as the concept of involving passerby planets, and in exchange, some new features were discovered along the way.

A number of features and improvements were discovered but due to lack of time, they did not make the cut. The following features or changes are estimated to improve the visualizer:

- **Sophisticated shadows and graphics** can add a touch of modernity to the render, elevating its visual appeal.
- **Varying textures and materials** can convey the density of objects more effectively while adding an element of visual interest to the scene.
- **Including timbre and frequency bands** can introduce highly prominent audio features that are currently missing from the visualization, enhancing its overall impact.
- **Computing song dynamics**, such as novelty, can track how the song evolves over time, resulting in a more progressive and compelling visualizer.
- **Reducing load times** while uploading a song can improve the application's usability, making it more accessible to users.
- **Revamping general usability**, including the interface, can bring tremendous value to the application, making it more user-friendly and intuitive.
- **Optimizing performance** can significantly enhance the application's capabilities, enabling the inclusion of features like height maps and higher polygon counts.

- **More variation** would prolong the novelty effect and create a more substantial engagement over time.

6

Results

6.1 The Final Design

The project resulted in an application that can process audio files uploaded by the user, extract affect estimates, and output an abstract 3D visualization that is calibrated to represent the emotional content of the audio. The affect values and real-time features which are unique to a particular audio file determine a range of parameters. The manipulable parameters include colors, shapes, movements, materials, post-processing effects, cameras, and lights.

This scene is designed to simulate a journey through the galaxy, with a morphable shape at its center transmitting radiation in a vivid and multicolored display. The central shape transforms dynamically, responding to changes in the energy level of the accompanying music. When a beat is detected, radiation waves are generated, with their hues determined by the active pitch classes. This radiation moves in response to the song's energy, initially traveling towards the sides before receding away from the camera to create a sense of depth and facilitate the differentiation of previous song events.

In the following sections, I will account for the final design of the visualizer and how the affect values and real-time audio features influenced the visualizer. Note that I report the implementation in an abstract manner and do not extensively elaborate on how the affect estimates and audio features were coded. Figure 6.1 showcases an example of how the code implementation was realized, however, for further elaboration, I refer to the code repository [72]. Also, note that the static images used in this report may not fully encapsulate the dynamic and interactive nature of the visualizer; review the repository [72] to attain a holistic and immersive representation of the visualizer's design.

6.1.1 The Mapping of Affect Values

6.1.1.1 Happiness

The *happiness* affect value has a significant impact on the visualization, imbuing scenes with a bright, soft, and vibrant aesthetic. Figure 6.4 showcases an example of how the visualization represents a song with a high *happiness*-value. *Happiness* made the following contributions to the aesthetic profile:

```
// Material values
let metalness = 0;
let roughness = 0;
let reflectivity = 0;
let clearcoat = 0;
let clearcoatRoughness = 0;
let emissiveIntensity = 0.0;
let opacity = 1;

// Update material values based on mood predictions.
function updateMaterial() {
  metalness = 0.6 - audioFeatures.predictions.mood_happy;
  roughness = 0.3 + 0.7 * audioFeatures.predictions.mood_happy;
  reflectivity = audioFeatures.predictions.mood_happy;
  clearcoat = 1;
  clearcoatRoughness = audioFeatures.predictions.mood_sad;
  emissiveIntensity = audioFeatures.predictions.mood_happy / 10;

  if (
    audioFeatures.predictions.mood_relaxed > 0.6 &&
    audioFeatures.predictions.mood_happy > 0.4
  ) {
    opacity = 0.5;
  }
  setMaterial();
}
```

Figure 6.1: A code example where mood predictions are used to manipulate material properties.

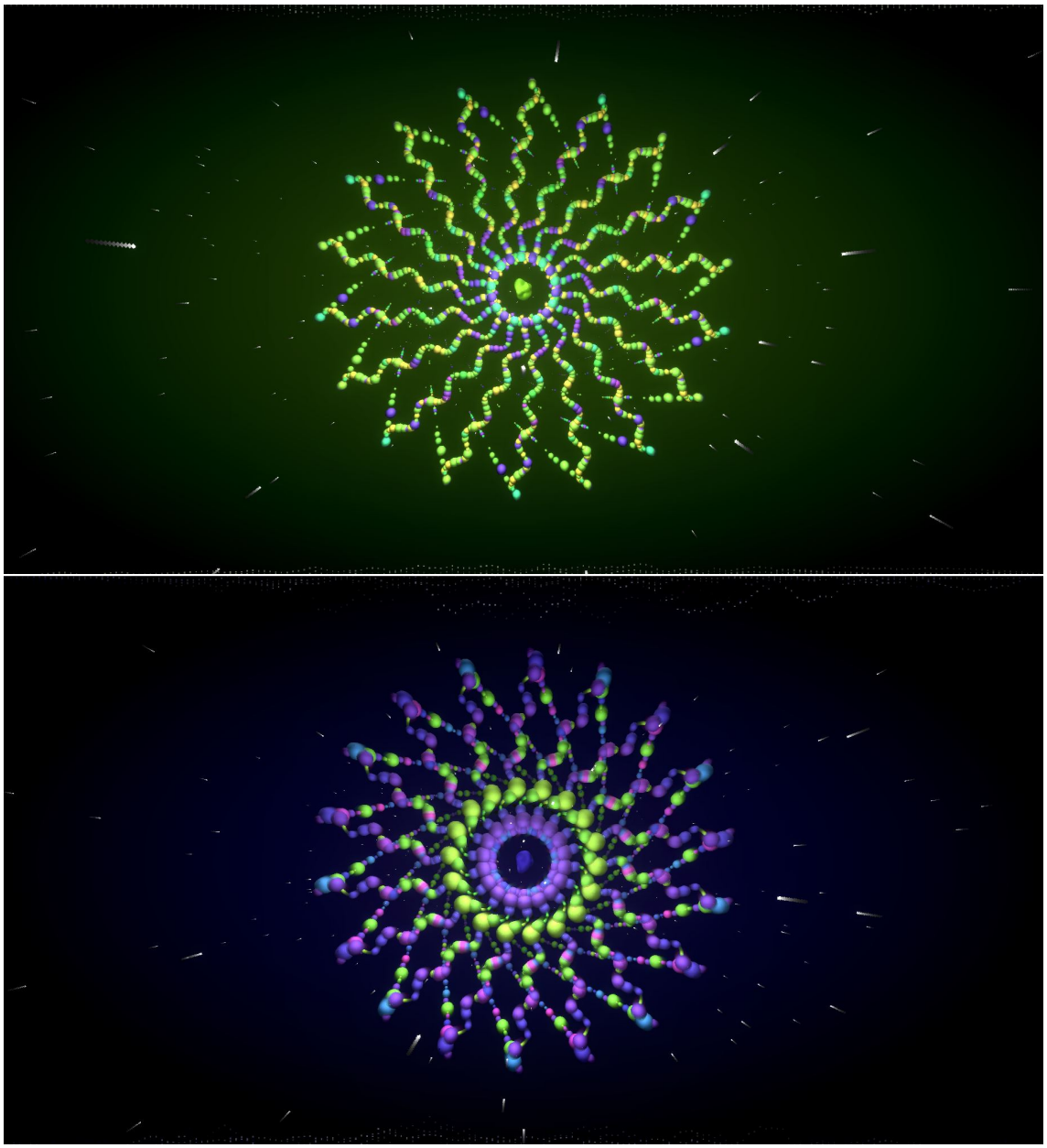

Figure 6.2: Two zoomed-out examples of the final visualizer.

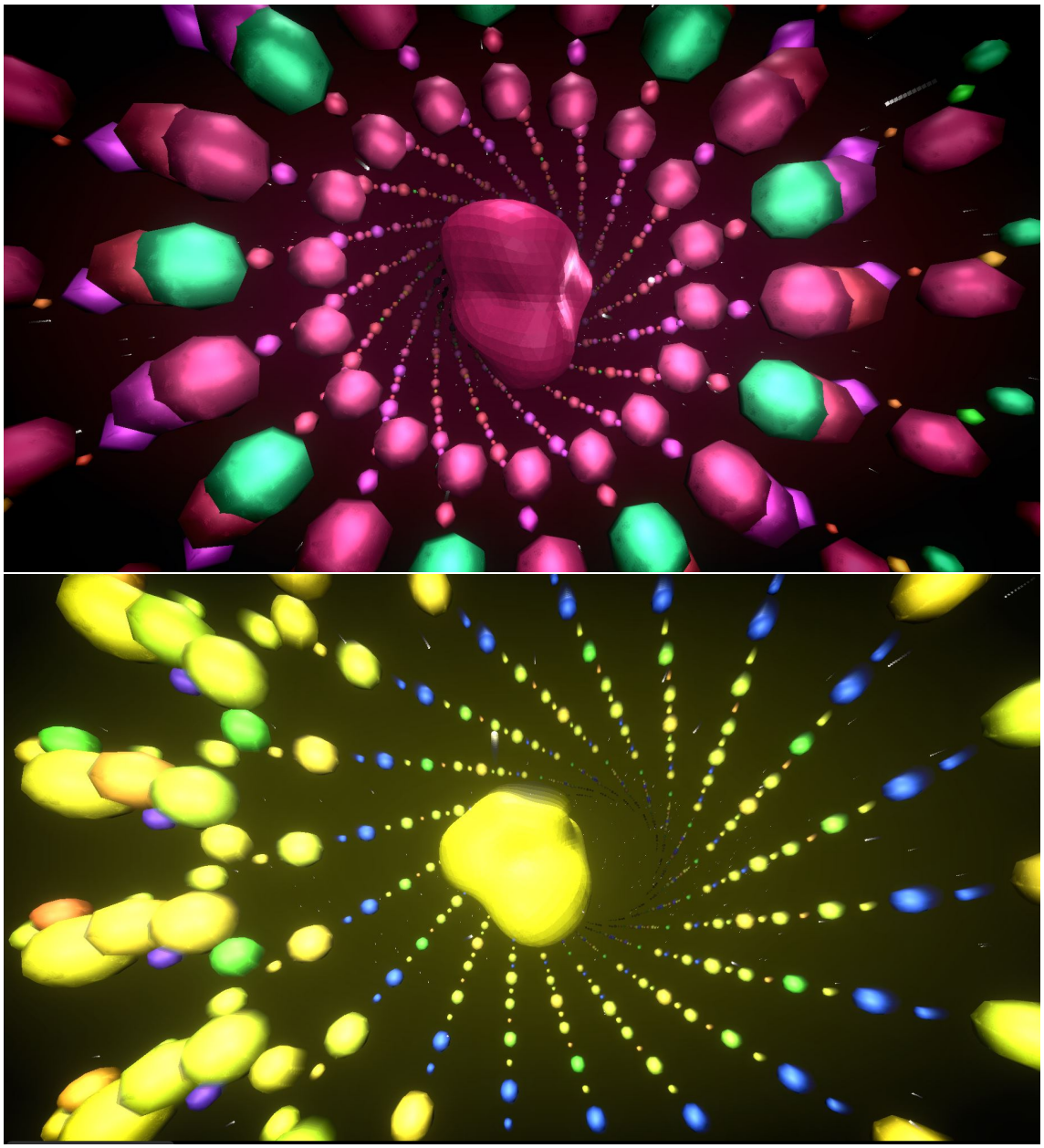

Figure 6.3: Two zoomed-in examples of the final visualizer.

- Introduced bias towards the usage of orange, yellow, and green hues.
- Made colors more vibrant by increasing saturation.
- Increased the number of vertices used for the essence shape and radiation geometry.
- Favoured the usage of a sine waveform.
- Determined the amount of fog in the scene.
- Determined the intensity of light sources.
- Decreased the vignette post-processing effect, making the scene brighter.
- Determined material parameters, such as reflectivity properties metalness, and roughness, as well as material emission.

6.1.1.2 Sadness

The affect value of *sadness* can have a profound impact by infusing scenes with a solid, slow, and dark aesthetic. Figure 6.5 showcases an example of how the visualization represents a song with a high *sadness*-value. *Sadness* made the following effects on the aesthetic profile:

- Introduced bias towards the usage of blue and purple hues.
- Decreased the range of which colors could be sampled.
- Increased the number of vertices used for the essence shape and radiation geometry.
- Favoured the usage of a sine waveform.
- Increased the vignette post-processing effect, making the scene darker.
- Determined the clearcoat material property.

6.1.1.3 Aggressivness

The *aggressivness* affect value can transform a visualization by making it more responsive, faster, sharper, and less predictable. Figure 6.6 showcases an example of how the visualization represents a song with a high *aggressivness*-value. *Aggressivness* made the following contributions to the aesthetic profile:

- Introduced bias towards using purple, red, and orange hues.
- Increased the color brightness.
- Increased the morphing speed of the essence shape.
- Decrees the number of vertices used for the essence shape and radiation geometry.
- Favoured the usage of a triangle waveform.

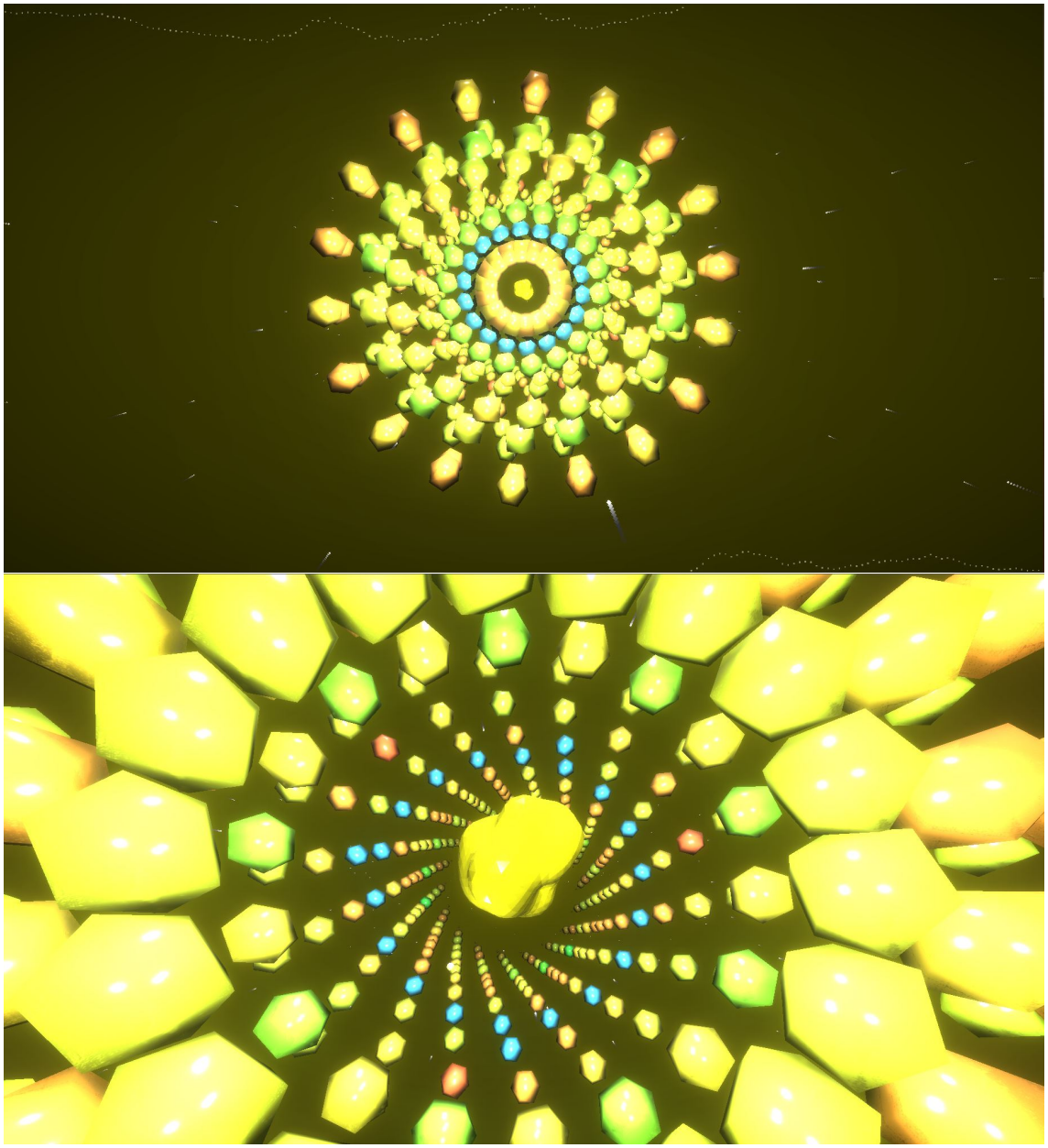

Figure 6.4: Ordering the data set by *happiness* this song (agg2) was the the highest rated.

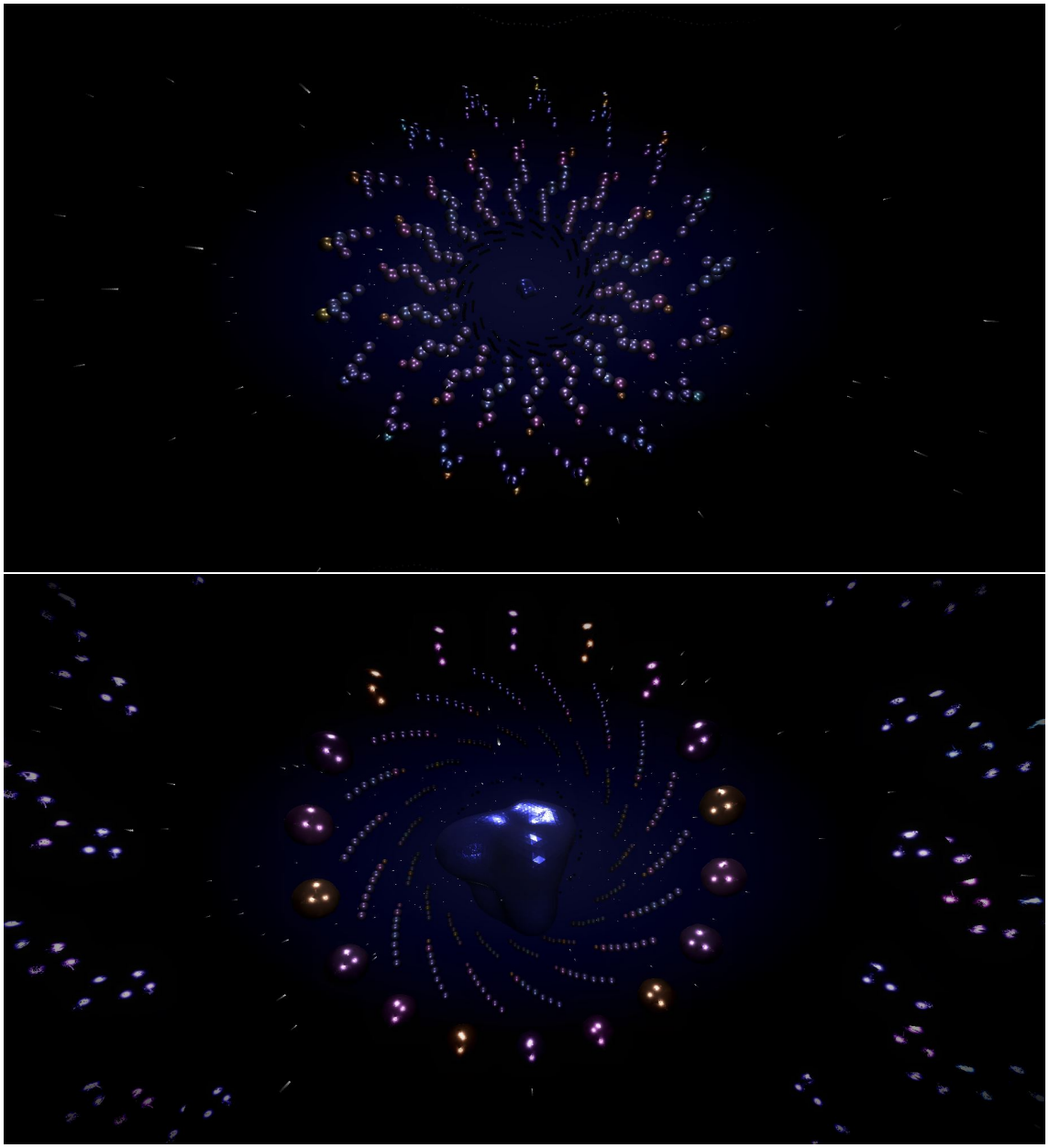

Figure 6.5: Ordering the data set by *sadness* this song (relax5) was the the highest rated.

- Determined the clearcoat material property.

6.1.1.4 Relaxedness

The *relaxedness* value made the scene softer and less responsive. Figure 6.7 showcases an example of how the visualization represents a song with a high *relaxedness*-value. *Relaxedness* made the following contributions to the aesthetic profile:

- Introduced bias towards using blue and green hues.
- Increased the number of vertices used for the essence shape and radiation geometry.
- Determined the material opacity.
- Control responsiveness by varying determining the number of values needed to determine rmsMean.

6.1.1.5 Danceability

The *danceability* affect value affected the visualizer by inciting movement and increasing color variation. Figure 6.8 showcases an example of how the visualization represents a song with a high *danceability*-value. *Danceability* made the following contributions to the aesthetic profile:

- The *danceability*-value was used for the gradient noise seed, which controlled the essence shapes initial shape.
- Controlled the speed of radiation movements.
- Increased the morphing speed of the essence shape.
- Increased the range in which the color pallet was sampled.

6.1.2 The Mapping of Beats Per Minute and Musical Key

6.1.2.1 Beats Per Minute

BPM controlled several parameters related to speed, to connect the pace of the music with the pace of the music. Among these were:

- Controlled the speed and rotation of the radiation objects.
- Controlled the speed of the particles.

6.1.2.2 Musical Mode

The *musical mode* was solely used as a binary value that was to select specific secret options. Among these were:

- Determined the direction of the spiral/the radiation when it travels into the distance.
- Determined the variation of sine wave/triangle wave used.

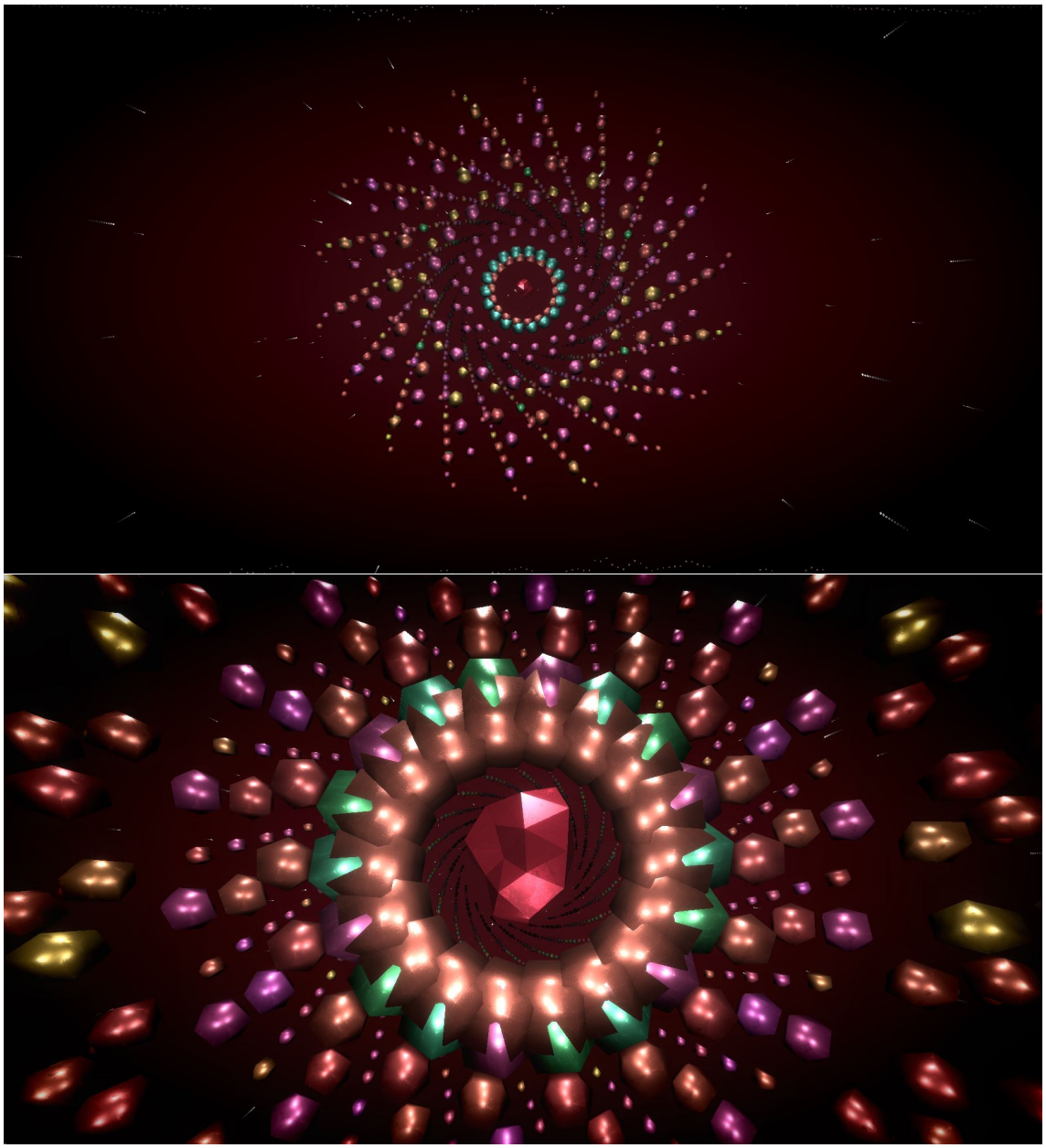

Figure 6.6: Ordering the data set by *aggressiveness* this song (agg5) was the the highest rated.

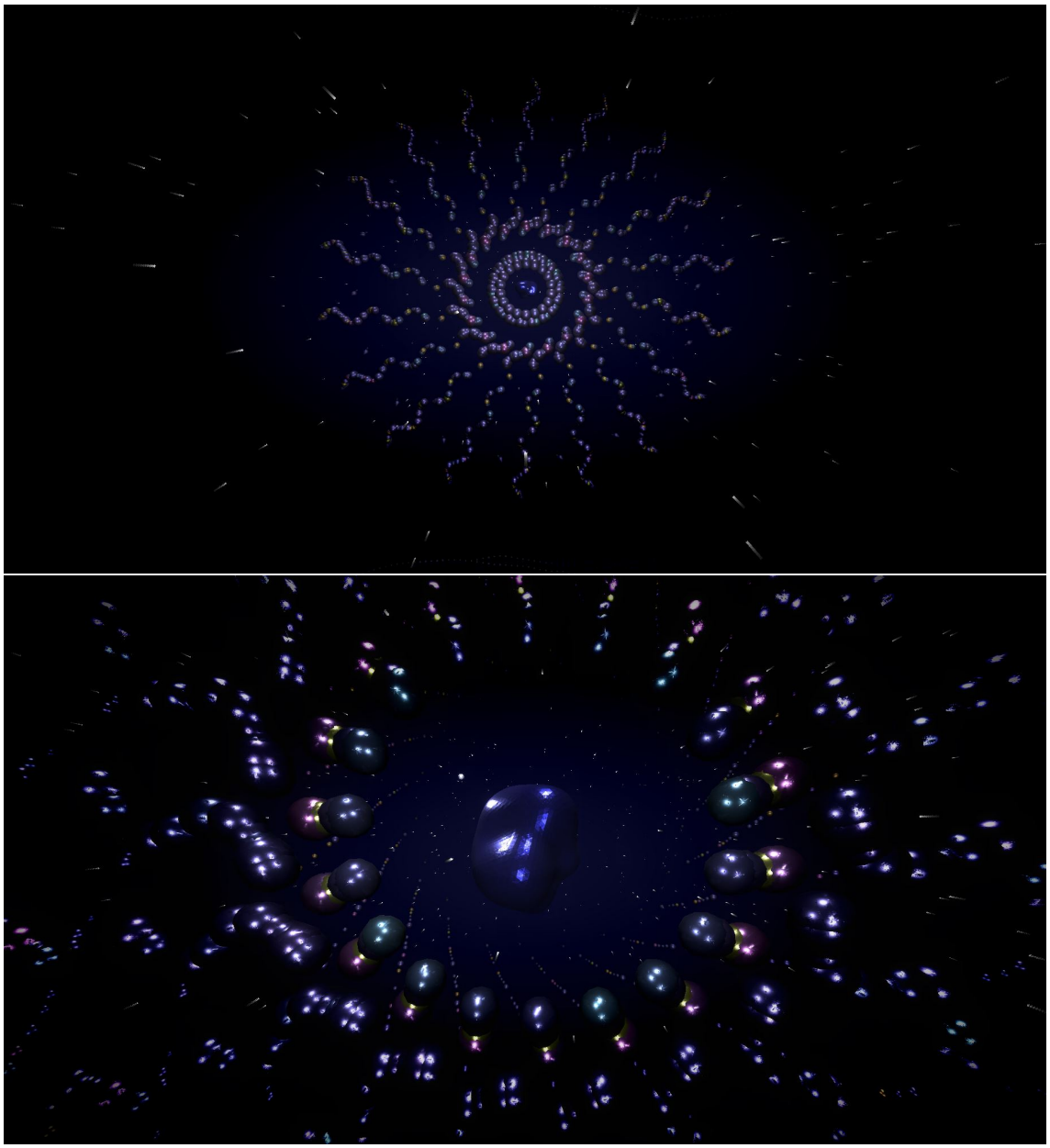

Figure 6.7: Ordering the data set by *relaxedness* this song (sad3) was the the highest rated.

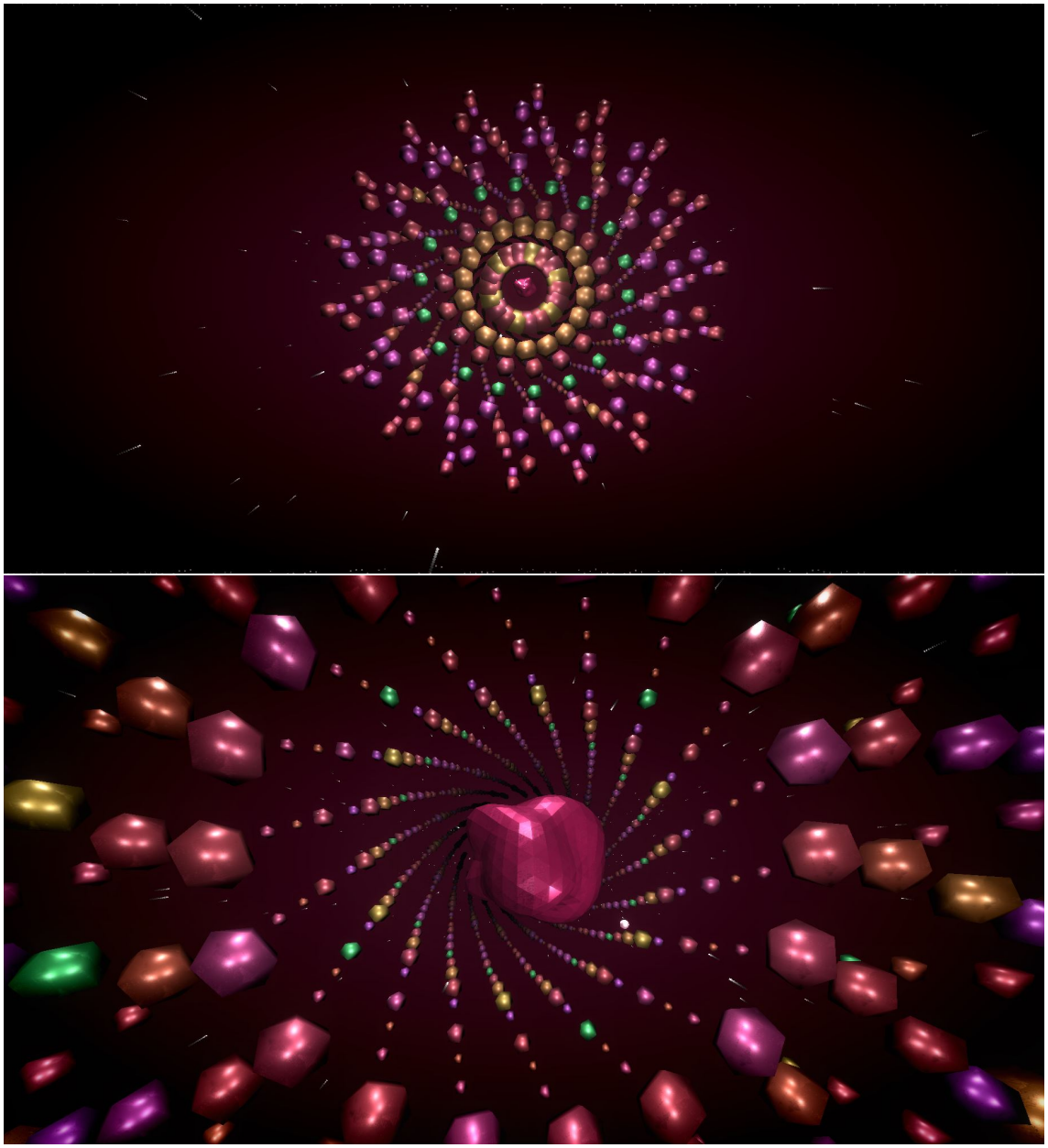

Figure 6.8: Ordering the data set by *danceability* this song (dance5) was the highest rated.

6.1.3 The Impacts of Real-Time Values

6.1.3.1 RMS

Since *RMS*, and the derived value *rmsMean*, represents the energy of the music it was used as the driving force of the visualizer and was involved in a range of features. The following is a list of how *RMS* was involved in the visualizer:

- Determined the size of the radiation and essence shape.
- Controlled the morphing speed and responsiveness of the essence shape.
- Controlled the movement and rotation speed of the radiation objects.
- Controlled the movement speed of the zoomed-in camera.
- Controlled the moon object's trajectory and movement speed.
- Used as a value for responsiveness.
- Determined the beat threshold.

6.1.3.2 Chroma

Chroma was simply used to determine which colors were used when the radiation spawned.

6.1.3.3 Audio Buffer

The 512 discrete samples of the *audio buffer* were mapped to simple plane geometries to create a real-time representation of the audio wave.

6.1.4 The Interface

The interface affords the user to drag and drop an audio file to start the visualizer. The user is treated with a loading bar while the audio is being processed. The user can control the audio player and the visualization responds accordingly. For example, if the music is paused the visualizer stops. The interfaces showcase the unique color pallet, affect estimates, and real-time features. The user can hide the buffer wave and interface display as well as jump between preset camera views. In addition, the user can control the camera with the mouse to explore the visualization through different angles. The buttons will disappear if the user idles for over five seconds.

6.2 Summative Evaluation

6.2.1 Objectives and Aims

The creation of the music visualizer was successful in visualizing music using abstract means, which was the primary goal of the project. Nonetheless, the true

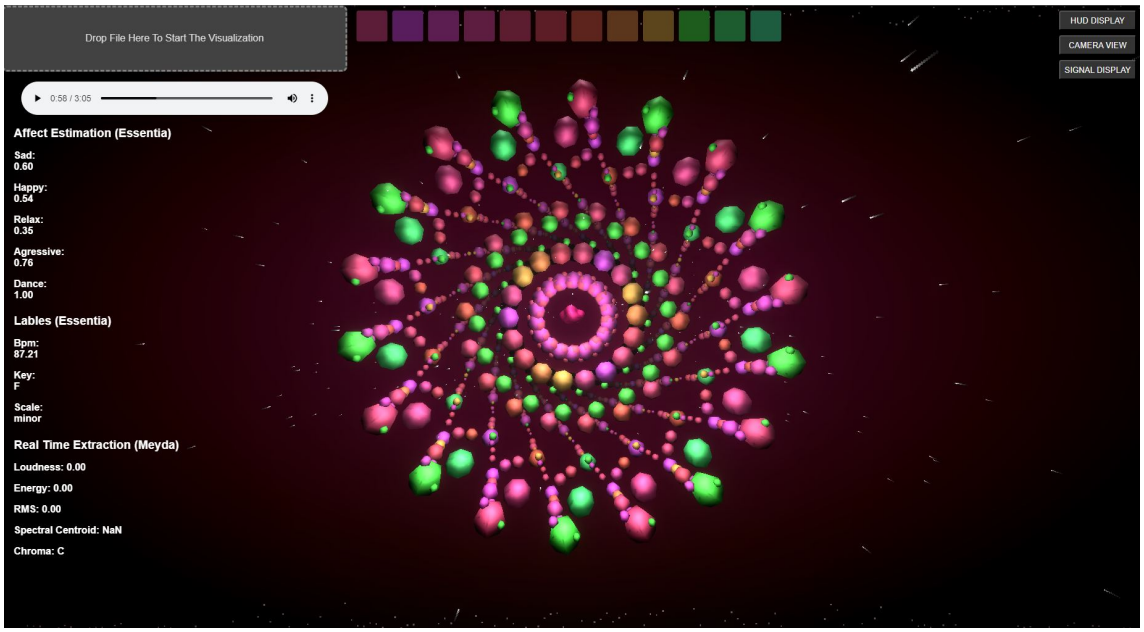

Figure 6.9: The visualizer with the interface displayed.

potential and effectiveness of the visualizer remained unknown. Therefore, a summative evaluation was conducted to assess its performance. The prototype underwent a summative evaluation to determine its ability to convey affect, and valuable feedback was gathered from stakeholders. The insights gained from this evaluation could prove crucial in refining future iterations of the visualizer.

6.2.2 Design

The summative evaluation took the shape of an experimental survey. The survey was a quantitative-qualitative hybrid since it involved both qualitative measures and free text inputs. The experiment had a within-subjects design since all participants were subject to each condition. The Latin square counter-balancing method was applied to minimize order effects. The experiment had one independent variable (the variation of visualization), with four levels. The visualizations were selected by ordering the data set based on the Essentia.js affect values (*aggressiveness*, *sadness*, *relaxedness*, *happiness*) and choosing a visualization that exhibited high values in the specific affect of interest, while also being visually distinct from other conditions.

Condition	Song	Sadness	Happiness	Relaxedness	Aggressivness
1	agg2	0,99	0,05	0,97	0,04
2	sad3	0,98	0,31	0,73	0,04
3	relax1	0,43	0,58	0,55	0,59
4	dance4	0,19	0,86	0,41	0,82

Table 6.1: Affect estimates for the songs used in the summative evaluation.

The experiment had six dependent variables, Positive Activation (PA), Negative Activation (NA), Valence, Arousal, Fitness, and Approval. The (a) Positive Affect

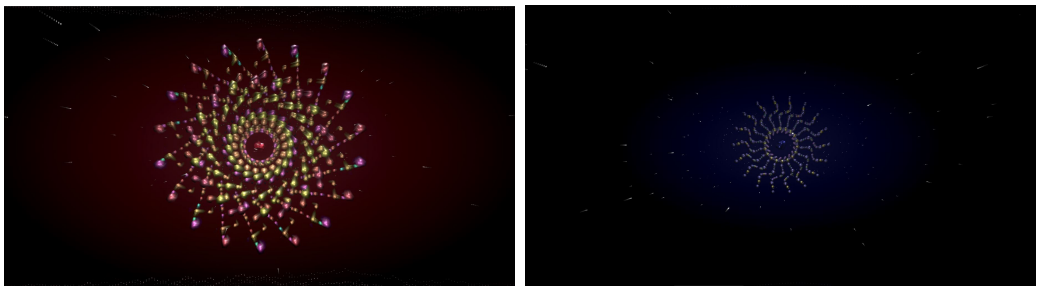

Figure 6.10: Left: Condition 1 showcased the visualization for the song "agg4". Right: Condition 2 showcased the visualization for the song "sad3".

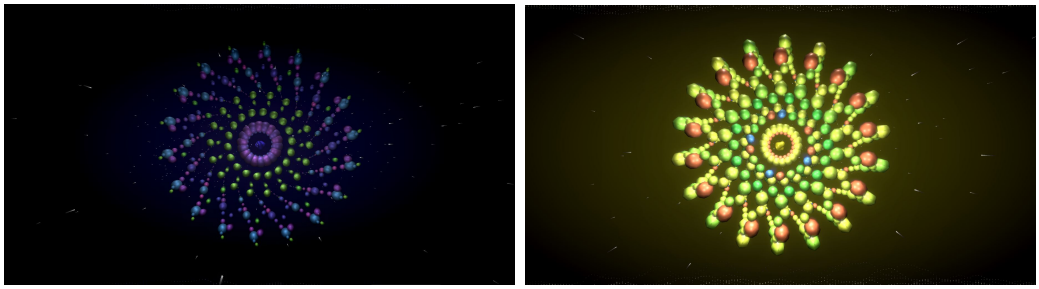

Figure 6.11: Left: Condition 3 showcased the visualization for the song "relax1". Right: Condition 4 showcased the visualization for the song "dance4".

(PA) scores are obtained by subtracting the sum of Low PA items (bored, vegetated) from the sum of High PA items (energetic, peppy), (b) Negative Affect (NA) scores are obtained by subtracting the sum of Low NA items (relaxed, calm) from the sum of High NA items (fearful, angry), (c) Valence scores are obtained by subtracting the sum of Negative items (sad, hopeless) from the sum of Positive items (happy, satisfied), and (d) Arousal scores are obtained by subtracting the sum of Low Activation items (inactive, sleepy) from the sum of High Activation items (alert, surprise). Notably, a 7-point Likert scale was used in this study, resulting in a range of -12 to 12 for the maximum and minimum values on these dimensions. Fitness and Approval were simply determined by explicitly nominal questions. Quantitative data were obtained by querying participants on potential improvements to the music visualizer in terms of its ability to communicate affect and enhance its aesthetic profile.

Four hypotheses were developed based on a combination of factors. The visualizations were created using affect estimates from Essentia.js, which closely aligned with the constructs of arousal and valence. The focus group data was utilized to make informed decisions about how to construct the visualization. These two factors together formed the basis for the formulation of the four hypotheses. The reasoning was that the affect estimates would be conveyed through the visualizations so the PANAS ratings would be predictable. Hypotheses 1, 2 are based on the notion that the visualizations with a particularly high affect value will result in the complementary high PANAS rating, such as:

- High *Aggressiveness* $\approx$ High Arousal
- High *Relaxedness* $\approx$ Low Arousal

- High *Happiness* ≈ High Valence
- High *Sadness* ≈ Low Valence

Hypotheses 3 and 4 propose that the powerful combination of unique affect estimates and the real-time capabilities of the visualizer makes a strong connection between the visuals and audio as well as a likable impression.

Hypothesis:

- ***H0:****The variation of music visualization has no effect on the PANAS Ratings.*
- ***H1:*** *Visualization 1 and Visualization 4 will, respectively, elicit a significantly higher degree of perceived arousal than Visualization 2 and Visualization 3*
- ***H2:*** *Visualization 2 and Visualization 3 will, respectively, elicit a significantly lower degree of perceived valence than Visualization 1 and Visualization 4*
- ***H3:*** *The majority of participants will report positive fitness ratings for all visualizations.*
- ***H4:*** *The majority of participants will report positive likability ratings for all visualizations.*

6.2.3 Material

The chosen songs were subjected to the visualizer, and 20-second clips from the beginning of each song were screen recorded using Windows Game Bar [73]screen recording. The video included sound and was recorded in 30 fps, however, due to the lack of computational power the frame rate drops occurred in the visualizations. Additionally, muted versions of these clips were also created using the free video editing software VideoProc Vlogger [74] and uploaded to YouTube. The survey platform that was utilized for the study was Survio [75], due to its European Union (Czech Republic) location and its ability to support full-screen video embedding. Four variations of the survey were created with varying condition orders based on the Latin square method. Using the redirection service Nimblelinks [76] a link was created which distributed the participants evenly to each survey variation.

6.2.4 Participants

Participants were recruited through convenience sampling on Facebook [77] and LinkedIn [78], and in total 16 participants participated in the survey.

6.2.5 Procedure

Upon entering the survey, participants were given an introduction to the study and provided with information about confidentiality, data handling, risks, and participation agreement. Participants had to acknowledge their participation consent. The study was divided into four visualizations, each of which had two sections: the first section focused on perceived emotion, while the second section focused on fitness, approval, and qualitative feedback. In the first section, participants watched a muted

version of the visualizer and rated 16 statements based on the PANAS measures [46], using a seven-point Likert scale. They rated terms such as alert, surprise, energetic, peppy, happy, satisfied, relaxed, calm, inactive, sleepy, bored, vegetated, sad, hopeless, fearful, and angry. In the second section, participants watched the same video as before but this time with the audio active. They were then asked if the visualizer suited the emotional content of the music and whether they liked the visualizer. To collect qualitative data, they were asked for suggestions on how to improve the emotional communication of the visualizer, as well as how to make it more appealing. Once all four visualizers were evaluated, the participants were thanked for their time and engagement.

The results gathered were processed in a spreadsheet and the PANAS scores were calculated; a statistical analysis using SPSS [79] was conducted to identify potential effects. Fitness and Approval were determined by the distribution of "yes" and "no" responses. The qualitative data went through a brief thematic analysis, to extract the valuable insights of the feedback.

6.3 Results of the Summative Evalaution

A multivariate analysis of variance (MANOVA) was conducted to examine the effects of visualization variation (IV) on the PANAS scores (DVs). The analysis revealed a significant multivariate effect for Wilks' lambda ($p < 0.001$) which led to further analysis.

6.3.1 Arousal Results

The visualization variation had a significant effect on Arousal (DV1), F(3, 12) = 20.136, $p < .001$, 2 = 0.50. Further post hoc pairwise comparisons using Tukey's Honestly Significant Difference (HSD) test indicated that there were significant differences between Condition 2 Arousal (M=-6.19, SD=3.99) and all other conditions respectively; Condition 1 Arousal (M=3.31, SD=4.35), (MD = -9.50, SE = 1.534, $p < 0.001$, 95% CI [-13.55, -5.45]); Condition 3 Arousal (M=3.25, SD=5.36), (MD = -9.44, SE = 1.534, $p < 0.001$, 95% CI [-13.49, -5.39]); Condition 4 Arousal (M=4.00, SD=3.43), (MD = -10.19, SE = 1.534, $p < 0.001$, 95% CI [-14.24,-6.14]).

6.3.2 Valence Results

The visualization variation had a significant effect on Valence (DV2), F(3, 12) = 15,380, $p < .001$, 2 = 0.44. The pairwise comparisons indicated that there were significant differences between Condition 2 Valence (M=-4.75, SD=4.98) and all other conditions respectively; Condition 1 Valence (M=3.88, SD=4.15), (MD = -8.63, SE = 1.66, $p < 0.001$, 95% CI [-13.01, -4.24]); Condition 3 Valence (M=3.75, SD=4.30), (MD = -8.50, SE = 1.66, $p < 0.001$, 95% CI [-12.88, -4.12]); Condition 4 Valence (M=5.37, SD=5.25), (MD = -10.12, SE = 1.66, $p < 0.001$, 95% CI [-14.51, -5.74]).

6.3.3 PA Results

The visualization variation had a significant effect on PA (DV3), F(3, 12) = 16.384, p < .001, 2 = 0.45. The pairwise comparisons indicated that there were significant differences between Condition 1 PA (*M*=4.44, *SD*=4.91) and Condition 2 PA (*M*=-3.35, *SD*=4.24), (MD = 8.19, SE = 1.48, p < 0.001, 95% CI [5.24, 11.14]), as well as between Condition 2 PA and Condition 4 PA (*M*=5.75, *SD*=3.79), (MD = -9.50, SE = 1.48, p < 0.001, 95% CI [-12.45, -6.59]). No significant effects were found for Condition 3 PA (*M*=1.31, *SD*=3.65).

6.3.4 NA Results

The visualization variation, Condition 1 NA (*M*=-0.56, *SD*=6.00), Condition 2 NA (*M*=-4.63, *SD*=5.71), Condition 3 NA (*M*=-4.44, *SD*=4.18) and Condition 4 NA (*M*=-2.50, *SD*=5.49),dit not have a significant effect on NA (DV4).

6.3.5 Fitness Results

The survey results indicated that the visualizer was well-suited to the emotional content of the music. Out of 64 total responses, 75% (n=48) reported that the visualization matched the music.

Condition 3 had the highest rate at 100% (16 participants responded "yes", and 0 participants responded "no"), followed by Condition 2 with 81.25% (13 participants responded "yes", and 3 participants responded "no"), and Condition 4 with 68.75% (11 participants responded "yes", and 5 participants responded "no"). Condition 1 received the lowest approval rate with only 50% (8 participants responded "yes", and 8 participants responded "no") of the participants reporting that the visualization matched the music.

6.3.6 Approval Results

Based on the survey results, the approval rate of the visualization varied across the different conditions. Out of 64 total responses, 82.81% (n=48) reported that they liked the particular visualization.

Condition 3 received the highest approval rate and 100% (16 participants responded "yes", and 0 participants responded "no") of the participants reported that they liked the visualization. Conditions 1 and 4 had similar approval rates of 81.25% (13 participants responded "yes", and 3 participants responded "no"), while Condition 2 had a slightly lower approval rate of 68.75% (11 participants responded "yes", and 35participants responded "no").

6.3.7 Improvements Results

The participants in the study provided valuable qualitative feedback on how the visualizations could be improved to better convey the emotional content of the music and enhance their aesthetic appeal.

6.3.7.1 Feedback on Visualization 1

Participants recommended using darker and harsher colors to better match the music's emotions, as well as introducing more chaos and randomness to the visualization. They also suggested using sharper and stronger colors and asymmetrical visuals to make it more visually appealing.

6.3.7.2 Feedback on Visualization 2

Participants highlighted the importance of representing chord changes, tones, and rhythm to capture the music's emotions. They recommended adding more interactions between particles and objects, as well as introducing rain to enhance the visualization's impact. To make it more aesthetically pleasing, they suggested zooming in on the visualization and creating better representations of different tones and rhythms.

6.3.7.3 Feedback on Visualization 3

Participants suggested slowing down the movement of objects to better match the music's tempo and representing frequencies better, especially the bass. They also recommended using engaging shader effects and letting objects move in from the sides to enhance the visualization's engagement and attractiveness.

6.3.7.4 Feedback on Visualization 4

Once again the visualization was missing representation of melody and specific instruments. The respondents wanted the visualization to move faster and use brighter colors to better portray the music's emotions. Both accounts of the usage of less and more color were reported to make the visualization more aesthetically pleasing. The visualization was reported to sync poorly with the music and have bleak colors, which were aspects that brought down the visualization's appearance.

6.3.8 Hypothesis

H1: *Visualization 1 and Visualization 4 will elicit a significantly higher degree of perceived arousal than Visualization 2 and Visualization 3*

While Visualization 1 had a significantly higher arousal than Visualization 2 the survey did not demonstrate any other predicted effects, and therefore the hypothesis was rejected. The visualizer was not able to translate the affect estimates into a visual representation of arousal.

H2: *Visualization 2 and Visualization 3 will elicit a significantly lower degree of perceived valence than Visualization 1 and Visualization 4*

The results do not support the hypothesis, since only Visualization 2 demonstrated a significantly lower degree of perceived valence. The visualizer was not able to translate the affect estimates into a visual representation of valence.

H3: *The majority of participants will report positive fitness ratings for all visualizations.*

The results support the hypothesis since 75% of the participants thought the visualizer suited the emotional content of the music.

H4: *The majority of participants will report positive likability ratings for all visualizations.*

The results support the hypothesis since 82.8% of the participants reported that they liked the visualizers.

7

Discussion

7.1 Research Question

How can affect estimation be utilized in an abstract music visualization?

In this project, I successfully created an abstract music visualizer prototype using a unique combination of affect estimates and real-time data. I have documented the process and final design and hope that it can serve as a reference for future development in this field. As I progressed with the project, I came to a realization that creating the visualizer was just one part of the equation. It was of equal importance to determine whether the visualizer would prove to be effective in its intended purpose.

While it is possible to extract affect estimates and make informed design decisions based on theory and empirical data, this does not guarantee that the visualizer accurately represents the emotional content of a song. Emotions are extremely complex, and visualizing them is a challenging task, given the subjectivity and interpretation involved. Furthermore, music is also subjective, and participants' genre and style preferences could influence how they interpret the visualizer.

In attempting to create a generalizable visualizer, I encountered the challenge that emotions take different shapes in different genres. A classical piece like Prokofiev's "Dance of the Knights" can be viewed as quite aggressive in terms of classical music, but it does not compare to any modern "hardstyle" track. Comparing affect values across such different domains creates additional challenges. Although narrowing the scope to a specific genre might have led to more subtle nuance, the project's ambitious breadth was what made it captivating.

My project highlighted visual characteristics of musical affect that merit further investigation, such as "waviness," "responsiveness," and "density," and their relationship to different genres and musical dynamics. Although it was too ambitious to extract rigorous guidelines for the aesthetics of musical affect, I believe that doing so is necessary to create an effective visualizer.

It is worth noting that my project focused solely on perceived emotion, and I based my design on the assumption that it would incite felt emotion. However, I cannot make claims that the visualizer makes users feel a particular way. Nonetheless, the fact that a majority of the evaluation participants mentioned that they liked the

visualizer suggests that it can incite some form of enjoyment. More research is needed to determine if the visualizer can enhance the experience of music.

The cross-modality of affect estimation and the integration of audio-visual representation is a tricky topic since it is unclear how it is perceived and which sensory stimuli, if any, dominate. Vision accounts for a significant part of our perception, and it is possible that the visual elements of the visualizer affect how the emotional content of music, as well as the overall experience of the representation, is interpreted. Conversely, it could also be the other way around, meaning that the music influences the interpretation of the visuals. The goal was to create a multimodal representation that combines different sensory domains into one entity that communicates the associated emotions through a merged experience. However, further research is required to explore the influence of the sensory elements involved and the causal relationship.

I want to address the project's dimension of wickedness. My approach was to create an affect-based visualizer, which was primarily based on the findings of the focus group. I acknowledge that there are numerous ways to build a visualizer, and if the project were to be reproduced, a different kind of visualizer could be developed. I believe that this design is just one manifestation of the ideas gathered along the way and that other researchers can use my lapses, findings, and knowledge for better iterations. The manifestation is just one initial interpretation of the aesthetic profile extracted from the focus group, but delving deeper into the discipline of visualizing emotions could yield a better aesthetic profile which in turn would yield a better visualizer.

7.2 Methodology Discussion

7.2.1 Focus Group Discussion

The focus group resulted in several important insights which became a fundamental part of the subsequent design. However, there are several aspects of the conduction of the study that should be addressed.

Due to limited resources, no note-taker was present during the study, which mitigated the possibility to analyze subtle interpersonal cues. By exploring the phenomena of the visual appearance of emotions, I aimed to discover patterns. The nature of the domain of emotions and how they are perceived and felt are riddled with subjectivity, and a large variety of stances and characteristics of emotions were declared. People can experience emotions in different ways. Due to the complexity of the topic, it would have been beneficial to conduct additional focus groups to reach theoretical saturation. In line with the essence of a wicked problem, the results of the focus group and the associated sketches are estimated to be difficult to replicate.

During the focus group, the attention shifted from the topic of emotions in music to the experience of emotions. I did not play music during the study to avoid priming effects, but an alternative focus group design could have incorporated music better to more efficiently immerse the participants in the topic. The participants reported

subjective accounts and used metaphors to describe their experiences and intentionality toward the visuals of emotions. While these metaphors and descriptions were bundled together in the thematic analysis to extract patterns, I did not dig deep into the underlying meaning of the metaphors that were used to describe their experiences. It is possible that the metaphors, such as sunshine and rain, carry additional connotations in the context of emotions and music, which I might have neglected.

The usage of the prompting terms for the sketching segment was based on the Essentia.js values, but in hindsight, these words were deemed to carry too many semantic connotations. The focus group was also carried out in Swedish instead of English, which could have led to subtle semantical differentials being lost in translation. While it is true that first-person reports are susceptible to excessive scrutiny, I found that the study's environment allowed participants to provide detailed explanations of their thought processes and justify their sketch designs, thus enabling the extraction of valid information.

7.2.2 Development Discussion

The Lo-Fi prototyping segments were limited in their ability to represent the envisioned ideas and concepts for the project due to their 2D static nature. As a result, I had to explore movements, dynamics, and 3D aspects in the coding segment, which was a labor-intensive process. While this exploration could have been done in an additional 3D modeling step to more accurately represent the scene before programming, I skipped this step due to time constraints and the necessity to learn the Three.js library and its potential in conjunction with affect and real-time data.

I began the Hi-Fi development with a predetermined set of features, which I for the most part managed to implement. However, I learned that due to the nature of the problem, there is an endless number of parameters to control and numerous possibilities for tinkering. To simplify the development process, I adopted a building block approach [2], but I believe that the design could be more modular, making it easier to test and refine specific components of the visualizer. Parts of the visualizer have modular aspects, such as the color selection, but ideally, I could have created a more modular structure, which could be further developed as we gain more knowledge on how to visually represent music affect.

Throughout the development process, I encountered numerous pivots and ideas. I initially intended to create a "trip-through-the-galaxy" metaphor, but I also incorporated other concepts, such as the central "distortion/essence shape" and aspects of other ideas such as "the kaleidoscope". It became clear that development is an iterative process, and ideas come and go as we tweak, tinker and explore the design process.

A larger and more varied data set, including affect, dynamic range, and genre, would have greatly aided development. With an almost endless combination of affect parameters, having a larger range of songs would have made it easier to single out different parameters and improve our understanding of how they influence the visualizer.

my focus was on affect estimates, which led us, for the most part, to ignore bands and frequencies. However, feedback from the summative evaluation suggested that representing instruments and frequencies are essential elements of an audio-visualizer and should be realized. I could have given more attention to this aspect of the visualizer. Since music can switch dynamically between different sentiments, deriving affect in real-time and controlling the dynamics of the visualizer are also critical factors worth further exploration.

Some visual features in the scene, such as the radiation color and size, directly map to auditory features. However, this is not explicitly stated and apart from the attempts to create a suitable mapping between the visuals and audio, the user is not given hints on how to distinguish these features. I believe that introducing the underlying structure of the visualizer to the user could enhance their understanding and allow them to encode the visuals into auditory features, similar to common music notation.

Some of my design choices deviated from the established aesthetic profile due to implementation difficulties or concerns about aesthetics. For example, I included edge cases for features like opacity to create a more unique aesthetic variation, even though opacity was not mentioned during the literature review or focus group. I made these types of decisions to create a more unique and engaging visualizer or to make the development process more feasible. Throughout the developmental work and research through the design process, first-person decision-making and the designers' intuition played a significant role. While the first-person methodology in academics has been under fire it is actively being used in the practice of HCI [80][42]. The reason why I incorporated a focus group and summative evaluation was to reduce tunnel vision and gain overlooked insights and metrics.

7.2.3 Evaluation Discussion

In this summative evaluation, I utilized a 7-point scale for PANAS ratings, which provides greater precision, than the 5-point scale, in the rating process. However, it should be noted that this alteration may make it challenging to compare the results to other studies using PANAS scores.

During the screen recording of the visualizations, there was a drop in resolution and framerate, which likely affected how participants experienced the conditions in terms of responsiveness and aesthetics. Moreover, the visualization only showcased a specific set of features due to the zoomed-out camera view. Using the zoomed-in view could have better portrayed the visualizer as an immersive 3D scene.

The experiment utilized a within-subjects design to reduce the number of participants required. However, only 16 participants were included, and a sample size of 24 would have been necessary to achieve an 80% power level if large effects were present. This entails that the results of the summative evaluation are likely not generalizable to the population. Additionally, participant bias is suspected as most participants were recruited through convenience sampling, which may have led to higher scores for Fitness (DV5) and Approval (DV6). A more diverse and larger

sample of participants would have been preferable to address this issue.

To some extent, I attribute the significant results of the evaluation, meaning approval and likability, to the potential bias introduced by the method of recruitment. Due to the use of convenience sampling in the recruitment process, there is a concern that participants may have provided inaccurate reports. To address this, I suggest a follow-up study with a larger and more diverse pool of participants recruited through varied channels. This would help mitigate any bias and provide a better representation of the performance of the visualizations. A larger study with a wider range of visualizations and music could also yield more informative results. Narrowing the scope down and focusing on a specific artist or genre for song selection could also be of interest.

The methodology used in this study was chosen for its reproducibility and use of established measures. However, it should be noted that the results are difficult to compare to the affect estimates obtained from Essential.js. If Essentia.js had utilized PANAS scores instead it would be more straightforward to determine if it was possible to convey the effect estimates through the visualizations.

7.2.3.1 Hypothesises Discussion

Both H1 and H2 hypotheses were rejected, indicating that representing music visually using extracted affect estimates is a complex task. Visualization 2 and Visualization 3 had high values for *sadness* and *relaxedness*, leading to the assumption that they would elicit similar visualizations and emotional representations. However, they had slightly different *happiness* values, which could explain the varied PANAS scores. Visualization 3 had a higher *happiness* and *danceability* value, affecting movements and colors, which might be the reason it scored higher than Visualization 2 on both arousal and valence. The original premise for H1 and H2 was that the visualizer would represent the affect estimates and convey corresponding PANAS scores, but this proved to be more complicated. Although the hypotheses were simplistic, the experiment suggests that there might be complex interaction effects between the parameters of the visualizer. Slight changes in one or two affect values can result in distinctly different visuals, evoking different emotions. Further research is needed to accurately represent emotions through aesthetics, including calibration of the visualizer and exploration of different genres and styles of music.

Interestingly, the results were consistent with both H3 and H4 since the participants reported that they liked the visualizer and that the visuals suited the music. This suggests that the visualization managed to represent the song to some extent, and this could be attributed to the responsive real-time features. For example, RMS was used extensively to manipulate the scene and make it responsive which is suspected to have made the music and visuals feel connected and enjoyable to watch. Visualization 4, which had the most frame drops, was also the least liked visualization, emphasizing the importance of a temporal and spatial relationship between audio and visuals.

7.3 Ethics

Ethical consideration was of crucial importance during the project's run time, in particular in regard to reporting and conducting studies.

Participants were asked to sign consent forms for the focus group studies and summative evaluation, which outlined the study's purpose and how their data would be used. By obtaining informed consent, participants were made aware of the study's purpose and were given the option to withdraw at any point. The collected data was stored offline to ensure that the data wasn't accessed by unauthorized individuals. In line with the General Data Protection Regulation (GDPR) regulations pseudonymization was performed to limit that data can be attributed to specific participants. The survey platform used in the evaluation study was Survio [75], a platform based in the European Union (EU). This decision was made to ensure that the study complied with the EU GDPR requirements [81].

The results and procedure of the project have strived to be transparent, fair, honest, and as unbiased as possible. The procedure and methodology of the project have been described in detail to make it possible to reproduce. This includes being transparent about the pitfalls and undesirable outcomes I encountered, such as insignificant results and methodological lapses. I have aimed to accurately report my findings and made the final prototype open source on GitHub so that anyone interested can inspect my work [72].

In the course of development, the calibration of the visualizer's aesthetic was carried out using royalty-free music licensed from Pixabay [70] and purchased tracks from iTunes [71]. To prevent debates related to fair use, only royalty-free music was employed in the dataset and summative evaluation. This approach ensured compliance with intellectual property rights and minimized ethical concerns related to the use of copyrighted material.

Proper attribution is crucial when using third-party libraries in a project. In line with this principle, I have made sure to properly credit the libraries used to accomplish this project. The three primary libraries utilized were Meyda [34], Essentia.js [6], and Three.js [33]. These libraries were credited in both the GitHub repository and the project report.

7.4 Zimmerman's Criterion of Design

According to Zimmerman [40], design research should be evaluated on the choice of methods and description of the process, aim to combine subject matters to address a specific case, be evaluated on relevance rather than validity, and be extensible for future knowledge derivation. The following sections denote how this project related to these criteria.

7.4.1 Method Selection and Process Description

The methodology for this project was rooted in research through design, first-person design, and the double diamond model. This approach provided a clear structure for the various phases required to create the music visualization tool. Since the domain of visualizing music effects through empirical means is relatively unexplored, the project employed a focus group to establish a foundation for development. The subsequent methodologies were then built on the focus group findings and a thorough literature review, allowing for a comprehensive and well-informed approach to the project. Discrete features were identified, described, and ranked before starting to code. These measures were employed to keep a solid structure and transparency throughout the project, which would in turn make the project reproducible.

7.4.2 Case Specificity

The project dealt with the creation of a prototype for a music visualization that was based on the estimated effects of the audio. The purpose of the project was to explore this domain and come up with an idea that could serve as inspiration for complementary research and design. This is a distinct area of design that has not been thoroughly explored before. In addition, the developmental environment also prompts the design and research to address specific conditions.

7.4.3 Relevance

The current state of technology goes against the hands-on creation of a visualizer and instead adapts a big data approach. While the affect estimates of the visualizer are in fact based on the Essentia.js classification model the aesthetic properties of the visualizer are completely designed by human artistic craftsmanship. AI-generated art can win art competitions and the discussion of artificial images can be considered art more relevant than ever [82]. The visualization includes generative art features and utilizes machine learning-derived features, but its fundamental design is still based on the human touch. This was not intended to be an artistic statement against the AI art community, but rather it raised further questions about hybrid versions of human/computer art generation. With that said, whether we realize affect-based music visualization by human or machine-based decision-making, exploring the usage of affect in the audio-visual domain is unexplored and intriguing.

Additional applications for an affect-based visualizer could be as a compliment to music applications such as Spotify[83] or as an addition to music performance software such as Resolume [84] and TouchDesigner[85].

A visualizer has the possibility to tap into the domain of slow technology. Subtle anticipation, meaning await and actively ponder upon how an event will play out [86] [87], could be utilized as a meaningful feature of artistic audio visualizations. The audience might be very familiar with the song they want to analyze and can take an active role in speculating how the music will be performed by the software. This could spark reflection and create a more meaningful way to consume music.

7.4.4 Extensibility

First of all, the project is open source which means that anyone can take a hands-on approach and add to the repository. The evaluation was classified as a summative evaluation, but it can also be considered a formative evaluation if development continues through additional iterations. The feedback received from the evaluation and the insights gathered allows for both refinements of the current design and extraction of valuable information. For instance, the aesthetic profile of affect, which was derived from the focus group, can be applied in similar projects.

7.5 Issues with Affect Estimation

The usage of the terms "arousal" and "valence" is well established in emotion estimations. The affect estimation model, Essentia.js, however, used its own terms which have been part of an ongoing discussion during the project's runtime. For example, to describe the intensity, Essentia.js uses two values, *aggressiveness* and *relaxedness*, instead of the sole value of arousal. Spotify, on the other hand, lacks a value for intensity but has an explicit valence value [83]. The affect estimation community seems to be riddled with unique company practices. Standardized values for affect estimation would not only make it possible to compare different ML models to each other but could also make it easier to tie the extracted values to established theories of emotion.

During the project's focus group, the participants were prompted with the terms used in the Essentia.js model. However, this intrudes semantic connotations, which a standardized terminology would avoid. For example, the usage of the word "aggressive" can be interpreted to have negative connotations, which do not conform with the neutral grounds of arousal. Using the PANAS system, *aggressiveness* could be determined as high negative activation while *relaxedness* could be rated as low negative activation. While the Essentia.js values of *aggressiveness* and *relaxedness*, in practice, seemed to represent the dimension of arousal, other terms for them should have been used in the focus group to avoid semantic confusion. For example, alternatives to the usage of *aggressiveness* could be "high intensity"/"high eventfulness" or using the PANAS terms such as "alert" and "energetic."

Once again, I want to denote that this confusion could have been avoided by introducing standardized terminology between the domain of emotional theory and affect estimation technologies. Tying these domains together could also make the estimation of complex emotions more theoretically sound.

The aesthetics of the visualizer are customized by my subjective decisions, and to a great extent, the audio-visual fitness is determined by the design. With that said, the visualizer is created based on the notion that the affect estimates are, in fact, accurate and representative of the music. Many times this assumption was valid, but other times, lapses in the affect estimation were discovered, such as classical music never being rated high on *aggressivness* and a strong correlation between high values of *sadness* and *relaxedness*. While these findings might be unique to the tiny data set

used for the project, it's worth noting that the visualization is based on the ground truth that the affect estimates are accurate, and if not, the visualization certainly would not fit the song. The visualization is only as good as its affect measures.

Initially, I derived a measure for arousal and valence, such as *arousal* = *aggressiveness* - *relaxedness* and *valence* = *happiness* - *relaxedness*. However, this reduced the number of unique parameters that could be utilized for the visualization, and often the values ended up as a value close to zero. With the Essentia.js model, songs could be estimated to have high *sadness* and high *happiness*, which could independently be hooked up to parameters, and by that, create a larger variety of aesthetics, which was a sought-after feature.

While arousal and valence are well-established measures in emotion theory, ML models offer the potential to extract unique measures such as *danceability*, which are only derivable using AI. I believe that the priority of affect estimation should utilize the established measures. However, discovering patterns in more complex emotions and attributes using ML-extractable values like "liveness," "danceability", and "acousticness" can be of great value. These parameters could serve as unique features in a music visualizer, offering a more comprehensive and holistic view of the music being analyzed. By leveraging the power of ML, we can enhance our understanding of emotions and create novel and engaging ways of experiencing music.

8

Conclusion

This project aimed to create a music visualizer that visually represented the emotions of a song. To do so, I extracted insights about visual aesthetics from literature and focus groups, which were used in conjunction with affect estimates and real-time extracted features.

Creating an audio-visual experience in conjunction with a synchronized emotional output could create a unique and multimodal way to consume music. The visualizer afforded the processing of any audio file and derived its emotions and musical features. The project is open-source and acts as a contribution to the community and can be adopted by any researchers or enthusiasts. The visualizer created was both liked, and the visuals in response to the music were deemed to be fitting, suggesting that it's possible to intertwine these mediums in a meaningful way.

Emotions, music, and visuals are infinitely complex domains riddled with subjectivity, and it's difficult to develop a visualizer that fits every genre, style, and sentiment. While the evaluation reported that the visuals suited the music, it lacked evidence for effective emotional communication. This suggests that the aesthetic profile derived was not accurate enough or that it was not implemented effectively in the visualizer. However, a more comprehensive evaluation would be beneficial. With these shortcomings in mind, I hope that my prototype serves as inspiration for further exploration in affect-based musical visualization. My research has also shed light on several areas of affect-based music visualization that required further investigation.

Further exploration of the emotional connotations of abstract means, and in particular, establishing a thorough aesthetic profile would be of importance. In extension, exploring how different types of visualizers can be derived from the same aesthetic profile could also be of interest. Different types of aesthetic profiles could also be derived, such as culture-specific, genre-specific, or instrument-specific. This would benefit the music visualization community immensely. Refining the aesthetic profile and improving the accuracy of affect estimation could lead to more innovative and engaging ways to visualize music. While there are claims that visuals can enrich the experience of music, this can be further explored. In particular, to what extent can an affect-based music visualizer affect the audience's emotions, in terms of felt or perceived emotions? Furthermore, can a poorly calibrated visualizer have a negative effect on the experience? Further investigation is also needed on the interaction effects of the combination of musical affect and real-time audio feature extraction. This could also include the possibilities of the natural integration of real-time affect

estimation in conjunction with real-time audio feature extraction. Finally, the domains of software development and classical emotional theory use widely different language, and a consistent language that connects the paradigms would make findings more broadly applicable. Standardizing measures of emotion, such as valence and arousal, in the domain of affect extraction is recommended.

I took on an incredibly complex wicked problem and succeeded in bringing to life a functioning manifestation of my initial idea and revealed areas of importance for the future of affect-based musical visualization. The ambitious effort of a combination of research, design, and development has, in addition to shedding light on my lapses, also tested my skill, ingenuity, and tenacity. I hope that my design process and insights can inspire more research in the quest for visualizing the emotions of music.

Bibliography

[1] L. Yunli and F. Revina Nur, "Interactive music visualization for music player using processing," in *2016 22nd International Conference on Virtual System & Multimedia (VSMM)*, 2016, pp. 1–4. DOI: 10.1109/VSMM.2016.7863205. [Online]. Available: https://ieeexplore.ieee.org/document/7863205.

[2] C. Brito and A. Ramires Fernandes, "Towards music-driven procedural animation," in *2019 International Conference on Graphics and Interaction (ICGI)*, 2019, pp. 40–47. [Online]. Available: https://ieeexplore.ieee.org/document/8955060.

[3] J. Ox, "2 performances in the 21st century virtual color organ: GridJam and im januar am nil," in *Proceedings of the 7th International Conference on Virtual Systems and Multimedia*, 2001. DOI: https://doi.org/10.1109/VSMM.2001.969716.

[4] R. Geiss. "MilkDrop," MilkDrop. (2012), [Online]. Available: http://www.geisswerks.com/milkdrop/.

[5] H. Lima, C. Dos Santos, and B. Meigunis, "A survey of music visualization techniques," *Association for Computing Machinery*, vol. 54, no. 7, pp. 1–29, 2022. DOI: 10.1145/3461835.

[6] A. Correya, D. Bogdanov, L. Joglar-Ongay, and X. Serra, "Essentia.js: A JavaScript library for music and audio analysis on the web," presented at the 21st International Society for Music Information Retrieval Conference, 2020, pp. 605–612. [Online]. Available: https://repositori.upf.edu/bitstream/handle/10230/45451/bogdanov_ismir_essent.pdf?sequence=1&isAllowed=y.

[7] A. Correya, J. Marcos-Fernández, L. Joglar-Ongay, P. Alonso-Jiménez Xavier Serra, and D. Bogdanov, "Audio and music analysis on the web using essentia.js," *Transactions of the International Society for Music Information Retrieval (TISMIR)*, vol. 4, no. 1, pp. 167–181, 2021. [Online]. Available: https://transactions.ismir.net/articles/10.5334/tismir.111/.

[8] M. Nuzzolo, *Music mood classification*. [Online]. Available: https://sites.tufts.edu/eeseniordesignhandbook/2015/music-mood-classification/.

[9] R. Khulusi, J. Kusnick, C. Meinecke, C. Gillmann, J. Focht, and S. Jänicke, "A survey on visualizations for musical data," *Computer Graphics Forum*, vol. 39, no. 6, pp. 82–110, Sep. 2020, ISSN: 0167-7055, 1467-8659. DOI: 10.1111/cgf.13905. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/cgf.13905 (visited on 02/14/2023).

[10] P. Galanter, "What is generative art? complexity theory as a context for art theory," 2003. [Online]. Available: https://citeseerx.ist.psu.edu/doc/10.1.1.90.2634.

[11] F. Nake, "Computer art: A personal recollection," in *Proceedings of the 5th conference on Creativity & cognition - C&C '05*, London, United Kingdom: ACM Press, 2005, p. 54, ISBN: 978-1-59593-025-5. DOI: 10.1145/1056224.1056234. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1056224.1056234 (visited on 02/09/2023).

[12] OpenAI. "DALLůe 2." (2023), [Online]. Available: https://openai.com/dall-e-2/.

[13] T. Möller, E. Haines, and N. Hoffman, *Real-time rendering*, 4th ed. Milton: Chapman and Hall/CRC, 2018, OCLC: 1051139762, ISBN: 978-1-351-81614-4.

[14] S. Lipscomb D. and E. Kim M., "PERCEIVED MATCH BETWEEN VISUAL PARAMETERS AND AUDITORY CORRELATES: AN EXPERIMENTAL MULTIMEDIA INVESTIGATION," [Online]. Available: http://www.lipscomb.umn.edu/docs/LipscombKim_ICMPC8_proceedings.pdf.

[15] D. A. Norman, *The Design of Everyday Things.* 2013.

[16] C. Ware, *Information visualization: perception for design*, Fourth edition. Cambridge, MA: Morgan Kaufmann, Inc, 2021, 538 pp., ISBN: 978-0-12-812875-6.

[17] D. Lu. "A perceptually meaningful audio visualizer." (2016), [Online]. Available: https://delu.medium.com/a-perceptually-meaningful-audio-visualizer-ee72051781bc.

[18] M. I. Mylopoulos and T. Ro, "Synesthesia: A colorful word with a touching sound?" *Frontiers in Psychology*, vol. 4, 2013, ISSN: 1664-1078. DOI: 10.3389/fpsyg.2013.00763. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00763/abstract (visited on 05/31/2023).

[19] M. Susino and E. Schubert, "Musical emotions in the absence of music: A cross-cultural investigation of emotion communication in music by extra-musical cues," *PLOS ONE*, vol. 15, no. 11, S. R. Livingstone, Ed., e0241196, Nov. 18, 2020, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0241196. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0241196 (visited on 01/31/2023).

[20] R. Y. Granot and Z. Eitan, "Musical tension and the interaction of dynamic auditory parameters," *Music Perception*, vol. 28, no. 3, pp. 219–246, Feb. 1, 2011, ISSN: 0730-7829, 1533-8312. DOI: 10.1525/mp.2011.28.3.219. [Online]. Available: https://online.ucpress.edu/mp/article/28/3/219/62480/Musical-Tension-and-the-Interaction-of-Dynamic (visited on 02/01/2023).

[21] P. Gomez and B. Danuser, "Relationships between musical structure and psychophysiological measures of emotion.," *Emotion*, vol. 7, no. 2, pp. 377–387, May 2007, ISSN: 1931-1516, 1528-3542. DOI: 10.1037/1528-3542.7.2.377. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/1528-3542.7.2.377 (visited on 01/31/2023).

[22] S. Brave and C. Nass, "Emotion in humancomputer interaction," in *Human-Computer Interaction Fundamentals*, A. Sears and J. Jacko, Eds., vol. 20094635, Series Title: Human Factors and Ergonomics, CRC Press, Mar. 2, 2009, pp. 53–

68, ISBN: 978-1-4200-8881-6 978-1-4200-8882-3. DOI: 10.1201/b10368-6. [Online]. Available: http://www.crcnetbase.com/doi/abs/10.1201/b10368-6 (visited on 02/06/2023).

[23] A. Kawakami, K. Furukawa, K. Katahira, K. Kamiyama, and K. Okanoya, "Relations between musical structures and perceived and felt emotions," *Music Perception*, vol. 30, no. 4, pp. 407–417, Apr. 1, 2013, ISSN: 0730-7829, 1533-8312. DOI: 10.1525/mp.2013.30.4.407. [Online]. Available: https://online.ucpress.edu/mp/article/30/4/407/62565/Relations-Between-Musical-Structures-and-Perceived (visited on 02/13/2023).

[24] P. G. Hunter, E. G. Schellenberg, and U. Schimmack, "Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions.," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 4, no. 1, pp. 47–56, Feb. 2010, ISSN: 1931-390X, 1931-3896. DOI: 10.1037/a0016873. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0016873 (visited on 01/31/2023).

[25] J. A. Russell, "A circumplex model of affect.," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, Dec. 1980, ISSN: 1939-1315, 0022-3514. DOI: 10.1037/h0077714. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/h0077714 (visited on 01/23/2023).

[26] V. L. Nguyen, D. Kim, P. V. Ho, and Y. Lim, "A new recognition method for visualizing music emotion," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, no. 3, p. 1246, Jun. 1, 2017, ISSN: 2088-8708, 2088-8708. DOI: 10.11591/ijece.v7i3.pp1246-1254. [Online]. Available: http://ijece.iaescore.com/index.php/IJECE/article/view/7483 (visited on 01/23/2023).

[27] A. J. Elliot and M. A. Maier, "Color psychology: Effects of perceiving color on psychological functioning in humans," *Annual Review of Psychology*, vol. 65, no. 1, pp. 95–120, Jan. 3, 2014, ISSN: 0066-4308, 1545-2085. DOI: 10.1146/annurev-psych-010213-115035. [Online]. Available: https://www.annualreviews.org/doi/10.1146/annurev-psych-010213-115035 (visited on 02/01/2023).

[28] S. E. Palmer, K. B. Schloss, Z. Xu, and L. R. Prado-León, "Musiccolor associations are mediated by emotion," *Proceedings of the National Academy of Sciences*, vol. 110, no. 22, pp. 8836–8841, May 28, 2013, ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1212562110. [Online]. Available: https://pnas.org/doi/full/10.1073/pnas.1212562110 (visited on 02/01/2023).

[29] K. L. Whiteford, K. B. Schloss, N. E. Helwig, and S. E. Palmer, "Color, music, and emotion: Bach to the blues," *i-Perception*, vol. 9, no. 6, p. 204 166 951 880 853, Nov. 2018, ISSN: 2041-6695, 2041-6695. DOI: 10.1177/2041669518808535. [Online]. Available: http://journals.sagepub.com/doi/10.1177/2041669518808535 (visited on 02/01/2023).

[30] Nataha. "Animation toolworks' library - 12 principles." (Jun. 9, 2016), [Online]. Available: https://web.archive.org/web/20160609091550/http://www.animationtoolworks.com/library/article9.html (visited on 02/14/2023).

[31] D. Norman, "Emotion & design: Attractive things work better," *Interactions*, vol. 9, no. 4, pp. 36–42, Jul. 2002, ISSN: 1072-5520, 1558-3449. DOI: 10.1145/

543434.543435. [Online]. Available: https://dl.acm.org/doi/10.1145/543434.543435 (visited on 02/14/2023).

[32] F. Heider and M. Simmel, "An experimental study of apparent behavior," *The American Journal of Psychology*, vol. 57, no. 2, p. 243, Apr. 1944, ISSN: 00029556. DOI: 10.2307/1416950. [Online]. Available: https://www.jstor.org/stable/1416950?origin=crossref (visited on 02/03/2023).

[33] three.js. "Fundamentals." (), [Online]. Available: https://threejs.org/manual/en/fundamentals.html (visited on 02/03/2023).

[34] H. Rawlinson, N. Segal, and J. Fiala, "Meyda: An audio feature extraction library for the web audio API," presented at the WAC - 1st Web Audio Conference. France, 2015. [Online]. Available: https://wac.ircam.fr/pdf/wac15_submission_17.pdf.

[35] "Web audio API," mdm. (Jan. 4, 2023), [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/API/Web_Audio_API.

[36] "9 three JS games," Free Frontend. (Nov. 19, 2021), [Online]. Available: https://freefrontend.com/three-js-games/.

[37] H. Egloff. "Henryegloff," Awesome Examples of Three.js. (Sep. 3, 2022), [Online]. Available: https://henryegloff.com/awesome-examples-of-three-js/.

[38] D. Fojcik. "Music player with three.js," Webflow. (), [Online]. Available: https://webflow.com/made-in-webflow/website/threejs-music-player.

[39] C. Frayling, "Research in art and design," *Royal College of Art Research Papers*, vol. 1, no. 1, pp. 1–5, 1993.

[40] J. Zimmerman, J. Forlizzi, and S. Evenson, "Research through design as a method for interaction design research in HCI," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose California USA: ACM, Apr. 29, 2007, pp. 493–502, ISBN: 978-1-59593-593-9. DOI: 10.1145/1240624.1240704. [Online]. Available: https://dl.acm.org/doi/10.1145/1240624.1240704 (visited on 01/30/2023).

[41] D. Woodruff Smith, *Phenomenology*, in *Stanford Encyclopedia of Philosophy*, 2013. [Online]. Available: https://plato.stanford.edu/entries/phenomenology/.

[42] A. Desjardins, O. Tomico, A. Lucero, M. E. Cecchinato, and C. Neustaedter, "Introduction to the special issue on first-person methods in HCI," *ACM Transactions on Computer-Human Interaction*, vol. 28, no. 6, pp. 1–12, Dec. 31, 2021, ISSN: 1073-0516, 1557-7325. DOI: 10.1145/3492342. [Online]. Available: https://dl.acm.org/doi/10.1145/3492342 (visited on 02/14/2023).

[43] X. Zhang and R. Wakkary, "Understanding the role of designers' personal experiences in interaction design," in *DIS '14: Proceedings of the 2014 conference on Designing interactive systems*, Vancouver BC Canada: Association for Computing Machinery, 2014, ISBN: 978-1-4503-2902-6. DOI: https://dl.acm.org/doi/10.1145/2598510.2598556.

[44] H. W. J. Rittel and M. M. Webber, "Dilemmas in a general theory of planning," *Policy Sciences*, vol. 4, no. 2, pp. 155–169, Jun. 1973, ISSN: 0032-2687, 1573-0891. DOI: 10.1007/BF01405730. [Online]. Available: http://link.springer.com/10.1007/BF01405730 (visited on 02/01/2023).

[45] R. Buchanan, "Wicked problems in design thinking," *The MIT Press*, vol. 8, no. 2, pp. 5–21, 1992.

[46] A. Gabrielsson, "Emotion perceived and emotion felt: Same or different?" *Musicae Scientiae*, vol. 5, no. 1, pp. 123–147, Sep. 2001, ISSN: 1029-8649, 2045-4147. DOI: 10.1177/10298649020050S105. [Online]. Available: http://journals.sagepub.com/doi/10.1177/10298649020050S105 (visited on 02/13/2023).

[47] C. Laurier and P. Herrera, "Mood cloud : A real-time music mood visualization tool," 2008. [Online]. Available: https://www.researchgate.net/publication/237814955_Mood_Cloud_A_Real-Time_Music_Mood_Visualization_Tool.

[48] J. Fan, K. Tatar, M. Thorogood, and P. Pasquier, "Ranking-based emotion recognition for experimental music," in *International Society for Music Information Retrieval Conference*, 2017.

[49] S. F. Fokkinga and P. M. A. Desmet, "Ten ways to design for disgust, sadness, and other enjoyments: A design approach to enrich product experiences with negative emotions," *International Journal of Design*, vol. 7, no. 1,

[50] J. Fan, M. Thorogood, and P. Pasquier, "Emo-soundscapes: A dataset for soundscape emotion recognition," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, San Antonio, TX: IEEE, Oct. 2017, pp. 196–201, ISBN: 978-1-5386-0563-9. DOI: 10.1109/ACII.2017.8273600. [Online]. Available: http://ieeexplore.ieee.org/document/8273600/ (visited on 05/10/2023).

[51] B. H. Banathy, *Designing social systems in a changing world* (Contemporary systems thinking). New York: Plenum Press, 1996, 372 pp., ISBN: 978-0-306-45251-2.

[52] L. Melissa and K. M. Goodrick, "Focus group research: An intentional strategy for applied group research?" *The Journal for Specialists in Group Work*, vol. 44, no. 2, pp. 77–81, 2019. DOI: https://doi.org/10.1080/01933922.2019.1603741.

[53] T. O. Nyumba, K. Wilson, C. J. Derrick, and N. Mukherjee, "The use of focus group discussion methodology: Insights from two decades of application in conservation," *Special Feature: Qualitative methods for eliciting judgements for decision making*, vol. 9, no. 1, pp. 20–32, DOI: https://doi.org/10.1111/2041-210X.12860.

[54] M. M. Hennink, B. N. Kaiser, and M. B. Weber, "What influences saturation? estimating sample sizes in focus group research," *Qualitative Health Research*, vol. 29, no. 10, pp. 1483–1496, Aug. 2019, ISSN: 1049-7323, 1552-7557. DOI: 10.1177/1049732318821692. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1049732318821692 (visited on 02/15/2023).

[55] A. Castleberry and A. Nolen, "Thematic analysis of qualitative research data: Is it as easy as it sounds?" *Currents in Pharmacy Teaching and Learning*, vol. 10, no. 6, pp. 807–815, Jun. 2018, ISSN: 18771297. DOI: 10.1016/j.cptl.2018.03.019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1877129717300606 (visited on 01/26/2023).

[56] A. Kuper, S. Reeves, and W. Levinson, "An introduction to reading and appraising qualitative research," *BMJ*, vol. 337, a288–a288, aug07 3 Aug. 7, 2008, ISSN: 0959-8138, 1468-5833. DOI: 10.1136/bmj.a288. [Online]. Available: https://www.bmj.com/lookup/doi/10.1136/bmj.a288 (visited on 01/26/2023).

[57] K. Brennan, *A guide to the Business analysis body of knowledge (BABOK guide)*, Version 2.0. Toronto: International Institute of Business Analysis, 2009, OCLC: 426221913, ISBN: 978-0-9811292-1-1.

[58] Google Ventures, *Crazy 8s*, https://designsprintkit.withgoogle.com/methodology/phase3-sketch/crazy-8s, n.d.

[59] H. L. McQuaid and D. Bishop, "An integrated method for evaluating interfaces," in *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, Seattle Washington: ACM, Mar. 31, 2001, pp. 287–288, ISBN: 978-1-58113-340-0. DOI: 10.1145/634067.634237. [Online]. Available: https://dl.acm.org/doi/10.1145/634067.634237 (visited on 01/30/2023).

[60] M. Walker, L. Takayama, and J. A. Landay, "High-fidelity or low-fidelity, paper or computer? choosing attributes when testing web prototypes," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 46, no. 5, pp. 661–665, Sep. 2002, ISSN: 2169-5067, 1071-1813. DOI: 10.1177/154193120204600513. [Online]. Available: http://journals.sagepub.com/doi/10.1177/154193120204600513 (visited on 04/20/2023).

[61] Y.-K. Lim, E. Stolterman, and J. Tenenberg, "The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas," *ACM Transactions on Computer-Human Interaction*, vol. 15, no. 2, pp. 1–27, Jul. 2008, ISSN: 1073-0516, 1557-7325. DOI: 10.1145/1375761.1375762. [Online]. Available: https://dl.acm.org/doi/10.1145/1375761.1375762 (visited on 02/07/2023).

[62] M. Khayat, M. Karimzadeh, D. S. Ebert, and A. Ghafoor, "The validity, generalizability and feasibility of summative evaluation methods in visual analytics," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2019, ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2019.2934264. [Online]. Available: https://ieeexplore.ieee.org/document/8805439/ (visited on 04/19/2023).

[63] A. Cooper, R. Reinmann, D. Cronin, and C. Noessel, *About Face : The Essentials of Interaction Design*, 4th ed. John Wiley & Sons, Incorporated, 2014, 723 pp., ISBN: 978-1-118-76640-8. [Online]. Available: https://ebookcentral.proquest.com/lib/chalmers/detail.action?docID=1762072.

[64] L. A. Clark and D. Watson, *The PANAS-x: Manual for the positive and negative affect schedule - expanded form*, Institution: University of Iowa, 1994. DOI: 10.17077/48vt-m4t2. [Online]. Available: https://iro.uiowa.edu/esploro/outputs/other/9983557488402771 (visited on 04/26/2023).

[65] A.-W. Harzing, J. Baldueza, W. Barner-Rasmussen, *et al.*, "Rating versus ranking: What is the best way to reduce response and language bias in cross-national research?" *International Business Review*, vol. 18, no. 4, pp. 417–432, Aug. 2009, ISSN: 09695931. DOI: 10.1016/j.ibusrev.2009.03.001.

[Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0969593109000353 (visited on 04/26/2023).

[66] G. Charness, U. Gneezy, and M. A. Kuhn, "Experimental methods: Between-subject and within-subject design," *Journal of Economic Behavior & Organization*, vol. 81, no. 1, pp. 1–8, Jan. 2012, ISSN: 01672681. DOI: 10.1016/j.jebo.2011.08.009. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167268111002289 (visited on 04/26/2023).

[67] M. Allen, *The SAGE Encyclopedia of Communication Research Methods*. 2455 Teller Road,ăThousand OaksăCaliforniaă91320: SAGE Publications, Inc, 2017, ISBN: 978-1-4833-8143-5 978-1-4833-8141-1. DOI: 10.4135/9781483381411. [Online]. Available: https://methods.sagepub.com/reference/the-sage-encyclopedia-of-communication-research-methods (visited on 04/21/2023).

[68] Microsoft Corporation, *Microsoft word*, Redmond, WA: Microsoft Corporation, 1983-2023. [Online]. Available: https://www.microsoft.com/en-us/microsoft-365/word.

[69] Figma, Inc., *Figma*, version latest, 2016–present. [Online]. Available: https://www.figma.com/.

[70] "Pixabay." (), [Online]. Available: https://pixabay.com/ (visited on 05/05/2023).

[71] Apple. "iTunes." (), [Online]. Available: https://www.apple.com/se/itunes/.

[72] A. Eriksson. VibifyEnv. (2022), [Online]. Available: https://github.com/AntonErikssonCode/VibifyEnv.

[73] *Windows game bar*, Microsoft Corporation, 2023. [Online]. Available: https://support.microsoft.com/windows/windows-game-bar-what-it-is-and-how-to-use-it-82b1309d-9973-c1c4-2e0e-a49da8246e2f.

[74] Digiarty Software, Inc., *VideoProc Vlogger*, Computer software, 2023. [Online]. Available: https://www.videoproc.com/.

[75] *Survio*, https://www.survio.com, Accessed: May 5, 2023.

[76] *Nimblelinks*, Accessed: May 5, 2023. [Online]. Available: https://nimblelinks.com/.

[77] *Facebook*, https://www.facebook.com/, Accessed: May 5, 2023.

[78] *Linkedin*, https://www.linkedin.com/, Accessed: May 5, 2023.

[79] IBM Corporation, *Spss statistics*, https://www.ibm.com/products/spss-statistics, Accessed: May 5, 2023, 2021.

[80] A. Ball and A. Desjardins, "Revealing tensions in autobiographical design in HCI.," presented at the In Proceedings of the 2018 Designing Interactive Systems Conference, New York, 2018, pp. 753–764. DOI: https://doi.org/10.1145/3196709.3196781.

[81] E. Commission., *Regulation (EU) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (general data protection regulation) (text with EEA relevance)*, 2016. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[82] K. Roose, "An a.i.-generated picture won an art prize. artists arent happy.," *New York Times*, Sep. 2, 2022.

[83] "Discover spotifys features," Spotify. (Jan. 4, 2023), [Online]. Available: https://developer.spotify.com/discover/.

[84] "Introducing resolume wire," Resolume. (), [Online]. Available: https://resolume.com/.

[85] "TouchDesigner features," TouchDesigner. (), [Online]. Available: https://derivative.ca/feature/application-building.

[86] W. Odom, M. Yoo, H. Lin, T. Duel, T. Amram, and C. Amy Yo Sue, "Exploring the reflective potentialities of personal data with different temporal modalities: A field study of olo radio," *DIS '20: Proceedings of the 2020 ACM Designing Interactive Systems Conference*, Pages 283–295, Jul. 2020. DOI: https://doi.org/10.1145/3357236.3395438.

[87] L. Hallnäs and J. Redström, "Slow technology - designing for reflection," *Personal and Ubiquitous Computing*, vol. 5, no. 3, pp. 201–212, 2001.

A

Appendix A

A.1 Focus Group

A.1.1 Focus Group Consent Form

Informed consent regarding the participation in a focus group about music visualization

The purpose of this study is to explore how music can be experienced visually and how music affect can be performed in an abstract music visualizer. The study includes a focus group discussion and a sketching session. The study will take approximately 1 hour.

The participation will benefit Anton Eriksson's master thesis. The study involves minimal risk and the chance of causing discomfort is not greater than in everyday life. This consent form will be stored at a safe location at Chalmers University of Technology.

I understand that I participate voluntarily and can terminate and/or recall my participation at any time without consequences.

I understand that my identity will be kept anonymous.

I understand that I should avoid disclosing personal information about myself or any other participant during the focus group session.

I approve that my participation is audio recorded and transcribed, as well as that the results and direct quotes can be used in the thesis.

I approve that the sketches I draw during the study are collected anonymously by the researchers and can be used in the thesis as a creative commons (CC-0) image.

I understand that the recording, transcription, and sketches from the focus group can be stored for up to two weeks after the completion of the master thesis.

I understand the results generated by my participation only will be used for research purposes.

I understand that I, on request, can get access to the data I contributed as well as get information about how the data is utilized in the thesis.

I understand that I can contact Assistant Professor Kivanc Tatar at tatar@chalmers.se if I have additional questions or need clarification.

Name Clarification: _____

Signature: _____

Date: _____

A.1.2 Focus Group Expertise Form

Expertise

Visual expertise:
How many years of experience do you have with subjects such as animation, computer graphics, illustrations, drawing or information visualization?

- ☐ Less than 1 year
- ☐ Between 1 and 5 years
- ☐ Between 5 and 10 years
- ☐ More than 10 years

Musical expertise:
How many years of experience do you have with subjects such as playing an instrument, composing music, sound engineering or studying music theory?

- ☐ Less than 1 year
- ☐ Between 1 and 5 years
- ☐ Between 5 and 10 years
- ☐ More than 10 years

A.1.3 Focus Group Inspiration

Colors

Shapes

Movement

Interactions

Effects

Textures

Words

Arrows

A.1.4 Focus Group Detailed Plan

Focus Group Detailed Plan

Time: 6-8 March, at 17:00 (location available between 15:00 - 19:00)
Location: Kuggen, Trepunksbältet Grouproom
Participants: Recruited through convenience sampling on social media

Objectives:

- Extract the aesthetic profile of the mood-based music visualizer.
- Find out the important visual attributes of the different emotions.
- Be specific and investigate how the visualization prototype can be developed.
- Find ideas and metaphors for how to realize the visualizer.

Important:

- Minimize bias:
    - Only show the visualizer briefly to not prime the participants to a specific aesthetic.
    - Music will not be played during the focus group since it can influence what song/elements the participant thinks of when sketching. For example, playing a sad song with acoustic elements might influence the participants to sketch their interpretation of acoustic visuals.

Materials:

- Mobile charger
- Mobile with memory
- Zoom Mic
- Mic cable
- Flash Drive
- Computer with memory
- Computer charger
- Penn
- Paper
- Consent forms
- Demographics forms
- Sets of colored pens * participants
- Sketch papers * 5 * participants

Detailed Plan:

1. **Greet Participants**

2. **Sign Consent Form**

3. **Sign Demographics Form**

4. **Start Focus Group**

5. **Ice Breaker:** *My guilt pleasure song is….. Do you have a guilty pleasure song?*

6. **Introduce The Topic:**
   "I am currently writing my master's thesis about music visualization, in particular about how music emotions can be represented in an artistic manner. I am building an abstract music visualization program that performs music and tries to represent the musical sentiment in a 3D environment. I need your help to identify properties in music and visuals that I can utilize to efficiently communicate the song's emotions."

   Show the basic black-and-white visualization environment.

   "This is the program, you can upload a song and it will extract emotions and my question to you is how musical emotions can be visualized. "

   "We will start this session by drawing some sketches based on some emotions, and after that, we will discuss the relations between music and visuals. "

7. **Main Section**
   a. **Sketching Session**
      The sketching session will take 10 min (5 * 2min).

      "I would like you to sketch how this visualizer could look like when performing an emotional song. Think in terms of abstract elements such as color, shapes, movement, interactions, and effects. You can also use arrows or use words to explain your thought process. There is no right or wrong answer so just sketch what comes to mind. "

      What would the visualizer look like when performing a happy song?

      What would the visualizer look like when performing a sad song?

What would the visualizer look like when performing an aggressive song?

What would the visualizer look like when performing a calm song?

What would the visualizer look like when performing a danceable song?

b. Motivating Sketching

For each emotion sketch the following questions can be discussed:

How can colors be used to represent X music emotions in an abstract music visualization?

How can shapes be used to represent X music emotions in an abstract music visualization?

How can movement and animation be used to represent X music emotions in an abstract music visualization?

How can interactions be used to represent X music emotions in an abstract music visualization?

Did you use any additional elements to represent the X music emotions in the abstract music visualization?

Was there anything you would like to add that you couldn't draw?

Did you represent some specific musical feature in your drawing? Which? (I can prompt musical features if needed)

c. Visual Feature Discussing

Are there any relations between musical features and visual features? Which? (I can prompt musical features if needed)

What is the most important thing to get right when representing music emotions in a visualizer?

d. Ideas and Metaphors

Are there any good visual metaphors for music?

e. **Closing remarks**
   Do you have something that you would like to add?

8. **Thank Participants**
   a. They get a candy bar :D

9. **End Session**

A.2 Pixabay Data Set

Num	Song		
1	agg1	Metal (Dark Matter)	AlexGrohl
2	agg2	Blast	AlexiAction
3	agg3	Stylish Rock Beat Trailer	ComaStudio
4	agg4	Crag - Hard rock	AlexGrohl
5	agg5	Epic Dramatic Action Trailer	QubeSounds
6	dance1	Town	BeCorbal
7	dance2	Tropical Sumer Music	Music_Unlimited
8	dance3	Ethnic background music for short video 1 minute. Dance hip hop beat	DMD Production
9	dance4	Funk It	ComaStudio
10	dance5	Disco Groove	QubeSounds
12	happy1	Best Time	FASSounds
13	happy2	Catch It	Coma-Media
14	happy3	Happy Day	Stockaudios
15	happy4	Weeknds	DayFox
16	happy5	Jazz Happy	Music_For_Videos
17	relax1	Lofi Study	FASSounds
18	relax2	Relaxing	Music For Videos
19	relax3	The Beat of Nature	Olexy
20	relax4	In the Garden -final	23624974
21	relax5	Ambient Classical Guitar	William_King
22	sad1	Emotional Piano Sad Background Music For videos	Lesfm
23	sad2	Dramatic Atmosphere with Piano and Violin	Universfield
24	sad3	Sad moment/ Sad and melancholy piano background music	SoulProdMusic
25	sad4	Wander	Monument_Music
26	sad5	Cienmatic Cello	Lexin_Music