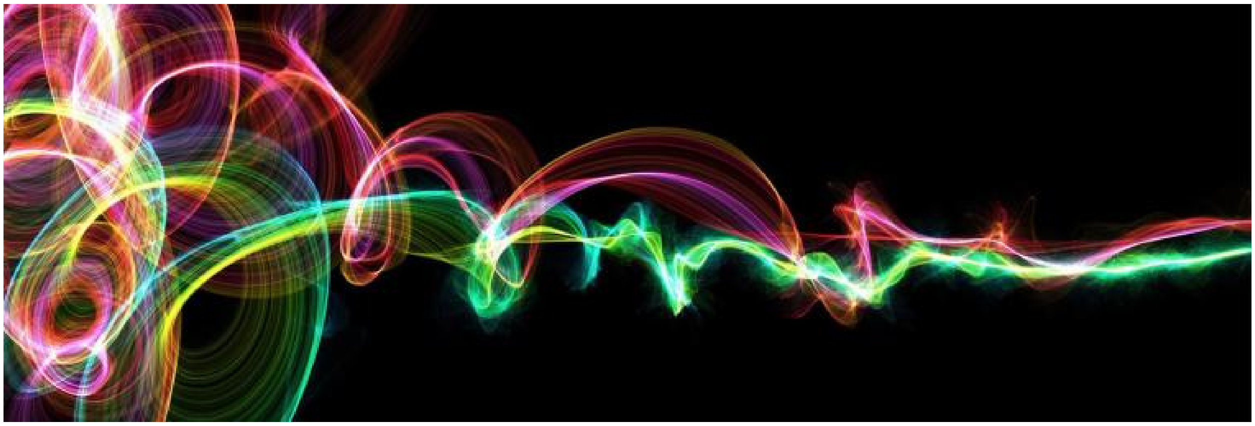
Translating Emotions from Sight to Sound



by

Chase Miller

Ethan Rudmetkin

Translating Emotions from Sight to Sound

An Interactive Qualifying Project

submitted to the Faculty of

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

degree of Bachelor of Science

by

Chase Miller

Ethan Rudmetkin

Date:

3 May 2023

Report Submitted to:

Professor Scott Barton

Worcester Polytechnic Institute

ABSTRACT

Art, whether musical or visual, is fundamentally about emotion. It is still a question of how and if a statement is influenced by its medium. With this question in mind, another is raised: can a statement's emotional value be conserved when expressed through a different medium than it was originally? This project aims to provide an auditory medium that enables both visually impaired and sighted individuals alike to experience the emotion that a work of two-dimensional visual art produces. In an effort to uncover a unifying theme, research will be done to ascertain what elements have an impact on the emotions that music and two-dimensional visual art each individually create. The process of converting visual art into music will be automated through a created algorithm, and the program's effectiveness will be assessed.

EXECUTIVE SUMMARY

There are many different interpretations and ways to understand art, which is a complex and individualized aspect of human culture. Some people place more importance on deciphering the meaning of a work of art than others, and some people concentrate on the physical aspects of the work of art and how they affect the viewing experience. These physical characteristics include scale, arrangement, composition, lighting, texture, color, and form. People can better understand the structure of the artwork and enjoy it more fully by taking these factors into account. Cognitive processes like object recognition, perception of spatial relationships, understanding of form, interpretation of color, and material identification are also involved in parsing a visual scene. These abilities enable people to understand the visual world and the message being presented by a work of art. The rational and emotional aspects of the viewer, such as their past experiences, their familiarity with art history, and their current mood, are also important in how they interpret art.

In terms of what is portrayed in the artwork and how it is viewed, art can also have temporal elements. Inferring past, present, or future events or the lack of temporal progression, time implications can be seen in artwork. A work of art's appreciation and emotional resonance are also influenced by how long it is viewed. People often interact with artworks they like for longer periods of time, which allows for a deeper connection and comprehension of the work.

Art has the ability to evoke a wide range of emotions in viewers, and emotions play a central role in the experience of artworks. Subjective factors, such as personal preferences and artistic appreciation, as well as the objective qualities of the artwork, affect how we feel when we encounter it. In addition to influencing aesthetic judgments, emotional arousal in response to art can cause the viewer to physically change, such as smiling or racing heart.

Capturing the underlying feelings and associated memories is necessary when converting art from one modality to another, such as from visual to auditory. The defining characteristics that unite all works of art are emotion and prior knowledge. Literature can, for instance, vividly describe visual imagery through the technique of ekphrasis and elicit feelings from the reader. The translation of art from one medium to another can be informed by knowledge of how people process visual scenes and the connections between sound and visual elements.

Several problems emerge when previous methods for resolving comparable issues are examined. The emotional and literary qualities of the original work are frequently lost in successful projects, and the translation between various mediums is frequently overly literal and direct. Many projects also exhibit subjective bias and fail to take into account the similarities between media. Furthermore, the level at which the original media is parsed can be either too high or too low, producing either stringent or generalized

results. Future projects should take into account a range of strategies and options that can be tailored to various people and situations in order to address these issues. Overall, being aware of how art is interpreted and translated can help one appreciate and interpret works of art more fully. We can improve our comprehension and produce meaningful translations between various modalities by taking into account the physical properties, temporal aspects, associations, and emotional representations of art.

In our project, we explored the possibility of converting visual art into sound as well as the physical form and dimensions of art. Due to its lower complexity and resource requirements, 2D parsing ultimately won out over 3D parsing for our purposes. We also took into account various types of artwork, such as moving and still images. The translation of moving images into sound in performance arts, such as movies and theatrical productions, presented difficulties. For our project, visual arts like paintings and sculptures were more appropriate. Paintings were our main subject because they allowed for more artistic expression and interpretation.

We looked at impressionism, realism, pop art, minimalism, and expressionism among the various artistic movements. Quick brushstrokes and vibrant colors of Impressionism paintings made them ideal for producing original musical sounds. Paintings that focused on realism, on the other hand, had a harder time converting visual details into sound while still accurately depicting the world. Pop art provided inspiration for meaningful music because it made use of recognizable imagery and vivid colors. With its simplicity and emphasis on material and textural elements, minimalism might also be a good option for music production. However, due to its level of abstraction and complexity, expressionism was ultimately selected as our focus. The abstract nature of music complemented the deeper level of artistic expression offered by expressionist paintings. Expressionist paintings' emotional and distorted qualities provided a wide variety of visual styles that could be translated into sound.

We found that color and shape were crucial elements for artistic retention. In works of visual art, color is used to convey point of view, depth, light, and mood. Contrarily, texture can be used to guide the viewer through a painting and arouse particular feelings. We chose a small set of musical elements for composition that were foundational to music theory. Timbre class, mode, tempo, dynamic variation, and rhythmic regularity were some of these criteria. With the help of these constraints, we were able to capture the key elements of music composition and produce profound musical works.

Our project's overall goal was to investigate the possibilities of converting visual art into sound, with an emphasis on expressionist paintings because of their emotional and abstract qualities. We tried to make a meaningful and expressive sight-to-sound experience by taking into account color, shape, and a few musical parameters. We examined the data from our experiment to learn how participants processed audio stimuli and how they incorporated visual cues like color and shape into their responses. To analyze the data, we used a variety of data analysis techniques, including correlation analysis, cluster analysis, and visualization tools.

We discovered some intriguing connections between hue-timbre, brightness-tempo, and mode-shape in the initial phase of the experiment. For instance, we found a positive correlation between brightness and tempo, suggesting that as the tempo increased, participants used brighter colors. However, we only discovered a small amount of significant information in the second part of the experiment, where we converted images to music. Numerous participants weren't sure of their answers, and the results were varied.

We looked at how the musical elements of rhythmic regularity, dynamic variation, and tempo related to the visual elements of hue, saturation, and brightness. Visual parameters and rhythmic regularity or dynamic variation were not significantly correlated. We did, however, notice a positive correlation between brightness and tempo, which may indicate that participants used brighter colors as the tempo quickened. In order to investigate the connection between visual elements and musical timbre and mode, we also performed cluster analysis. Although we noticed that various timbres might be related to various colors, we did not notice any obvious trends among the various modes. To calculate the separation between points in various sections and comprehend the color distribution, we used scatter plots and heatmaps. We discovered that the spacing between the sections' points remained constant at all tempos, suggesting a consistent color distribution. Low to moderate saturation and brightness levels showed the largest distances between points, indicating more variation in color distribution in these regions. We found that almost all of the figures were centered, despite the fact that each participant approached the assignment in a different way. Additionally, we discovered a strong correlation between the figure's contour and the musical piece's pitch contour. Although not all participants recognized this connection, it appeared that the regularity of the musical pieces' rhythm and the smoothness of the figures' curves were related.

Finally, our analysis shed light on how participants interpreted audio stimuli and incorporated visual cues into their responses. We discovered connections between mode-shape, brightness-tempo, and hue-timbre. The experiment's second phase, however, only produced a small amount of useful data. We observed common characteristics in the figures, such as centeredness and a relationship between contour and pitch contour, despite the fact that each participant approached the assignment in a different way. Our team urges others to continue advancing the project's objectives because the suggested artistic genres, media, and interfaces, along with the relationships discovered, could result in the development of a system that enables people with sight impairments to understand the compositional and emotional elements of visual art. In the end, our research project only touched the surface of this complicated subject; nonetheless, we hope that our conclusions and suggestions will serve as a springboard and source of inspiration for further research.

AUTHORSHIP

The writing of this paper was done by both Chase Miller and Ethan Rudmetkin. Their sections were written separately but in collaboration and conjunction. Each section was reviewed by professor and advisor Scott Barton after each iteration. The distribution of work is listed below.

<table>
<tr><th>Chase Miller</th><th>Ethan Rudmetkin</th></tr>
<tr>
<td>

Background and Literature Review

How Are Visual Scenes Parsed?
Associations in Art and Emotional Representation
How is Art Translated From One Modality to Another
Prior Art

Rule-Based Approaches
Sonification Approaches
Machine Learning and Deep Learning using Neural Networks

The Gap

Methodology

Determining the Method of Translation That Will Result in the Greatest Retention of Identity
Carrying Out a Study on the Relationship Between Audiovisual Parameters

Outlining the Purpose of the Study
Conducting the Study

Results and Analysis

Music to Image Analysis

Figure Analysis

Issues With the Study's Image to Music Section
Conclusion of Results

Conclusion and Future Directions

</td>
<td>

Introduction
Background and Literature Review

How Do Visually-Impaired People Navigate the Visual World?
Musical Representations: Mapping Visual Parameters to Auditory Variables
Prior Art

Artistic Approaches: Sound to Sight

The Gap

Methodology

Determining the Best Candidates of Art That Translate From Sight to Sound
Carrying Out a Study on the Relationship Between Audiovisual Parameters

Determining the Procedure of the Study Execution
Recruiting Participants

Analyzing the Data to Determine the Proper Mappings From Sight to Sound

Results and Analysis

Analysis of Parameters in Response to Auditory Stimuli
Music to Image Analysis

Color Analysis

Lack of Correlations and Further Analysis
Conclusion of Results

</td>
</tr>
</table>

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Visual art has been an integral part of human expression for centuries, but it is often taken for granted by those who are able to see it. However, the experience of visual art is not accessible to everyone, such as individuals who are blind or visually impaired. This raises an important question: can the qualities of art be translated into another medium, such as music, to create a more inclusive artistic experience? This question prompted us to delve into the realm of cross-modal artistic works and explore the possibility of developing a computer program that can translate visual art into musical compositions.

Our original goal was to develop a program that could transform visual art into a musical composition that conveys the same emotional quality as the original piece. However, due to various challenges, we were not able to successfully create the program. Instead, we conducted an experiment to explore the challenges of cross-modal translation and the potential for enhancing the accessibility and inclusivity of the arts for visually impaired individuals. By examining how visual parameters can be mapped to musical parameters, we aimed to gain a deeper understanding of the nature of cross-modal artistic expression. In this paper, we will describe the methods and results of our experiment, including the statistical analysis and visualization techniques we used to identify relationships between visual and musical parameters. Through this work, we hope to contribute to the growing body of research on cross-modal translation and inspire further exploration of cross-modal artistic expression.

CHAPTER 2: Background and Literature Review

Art is a fundamental aspect of human culture that has captivated individuals for centuries. However, the ways in which art is perceived and understood can vary greatly among individuals. In this section we will review the literature on how visual scenes are parsed, the physical and spatial properties of art, the temporal aspects of art, associations in art and emotional representation, and the different approaches that have been used to translate art from one modality to another. We will also explore the prior art in this field, including rule-based sonification, machine learning, and artistic approaches. Finally, we will identify the gap in the current literature and propose our own solution for exploring the translation of visual art to auditory representations.

How Are Visual Scenes Parsed?

While some viewers may prioritize analyzing the meaning of a visual work of art, others are attentive to the physical properties of the artwork itself and how they affect the viewing experience. They consider factors such as the texture, color, form, and composition of the piece, as well as the lighting, scale, and spatial arrangement in which it is presented. By taking these elements into account, they gain a deeper understanding of how the artwork is constructed and how it can be best appreciated.

Physical and Spatial Properties of Art

From a cognitive standpoint, parsing a visual scene is heavily dependent upon recognitive, perceptive, and navigational functions. From these functions stem nine important properties that affect how we as humans decipher a visual scene: edges and junctions, spatial frequency and layout, form, color, material, lighting and shadows, objects, categorical identity, and location (Epstein & Backer, 2019). These properties range in conceptual level from concrete to abstract. The *edges and junctions* property refers to the ability to distinguish the beginning and end of objects by noting where the boundaries exist (Intraub, n.d.). *Spatial frequency and layout* refers to the concept of spatial relationships and perspective, that is, the idea that farther away objects are represented smaller than closer objects, the idea that objects farther away appear closer relative to each other than objects closer to the viewer, et cetera (Torralba., 2019). *Form* refers to the geometries within an image and how their identities are preserved regardless of orientation, scaling, skewing, and other transformational properties (Shapiro & Stockman, 2001). *Color* refers to the concept of how hues are viewed and how they are impacted by changes in lightness and saturation (Valich, 2018; Mather, 2013, Chapter 5). *Material* is similar to color, but also takes into account texturing, which contributes to the depth perceived from a flat image (Valich, 2018; Shapiro & Stockman, 2001). *Lighting and shadows* refers to the ability to distinguish where the brightest and darkest parts of an image are and how the objects in the image contribute to those three-dimensional perceptions (Shapiro & Stockman, 2001). The final three concepts, *objects*, *categorical identity* and *location*, are derived from prior experiences and knowledge of the visual world. People have general understandings of what things are supposed to look like, so big concepts in visual art, such as a house or a car, are able to be easily identified regardless of how realistic they appear in an image. Without these nine aspects, one would not have the capability to truly understand what is being conveyed through the piece of art.

The *pictorial experience* is a kind of aesthetic experience that also impacts how a visual scene is interpreted (Zausner, 2006). Both the pictorial work of art and the observer participate in the experience of the picture. It is performed through the interaction between the characteristics of the work of art and the personality of the observer. In that process, the exchanges of emotional and rational components like age, previous experience in communicating with visual arts, knowledge of art history, art theory, art practice, art language, some personality traits, the current mood of the recipient as well as belonging to a particular social, cultural and historical context affect the experience (Hagtvedt et al., 2008).

Temporal Aspects of Art

Art generally has two temporal aspects. Time is typically referenced in works of art through implications (All Answers Ltd, 2022), which take the form of what has happened, what is going to happen, or the lack of temporal progression (Various Ways the Passage of Time Is Portrayed, n.d.). A great example of these implications is in Gerrit van Honthorst's *Old Woman Examining a Coin by a Lantern (Sight or Avarice)* (1623) (See Figure 1).

Figure 1. *Old Woman Examining Coin by a Lantern (Sight or Avarice)* by Gerrit van Honthorst, 1623

It is implied that the old woman has found this coin and lit a candle in a lantern so that she may examine the coin under better lighting. It is also implied that either this candle has been lit before or it has been burning for a while, since it is short and melted. This painting also implies the lack of a passage of time. As the woman stares at the coin, it is all she is focused on; time stands still (figuratively) around her. Finally, it is implied that eventually she will stop examining the coin, whether it be on her own accord or the candle completely burns away. This concept of temporal progression within what appears to be an invariable image allows for the viewer to experience the piece in a linear fashion common to everyday life.

The other temporal aspect is more concerned with how the art is viewed than it is with what is portrayed in the work. Shared understanding, potential appreciation, and emotional resonance is heavily affected by the amount of time invested in viewing a work of art. It is important to note that the more time you spend looking at a piece of art, the higher the chance that you will gain an appreciation for it. Conversely, the more that you appreciate a piece of art, the more likely you will spend more time with it. (Intraub, 2010). This explains the concepts of sustained and repeated engagement. People frequently immerse themselves in a piece of art because it either has a deeper meaning for them or elicits a strong emotion, usually a positive one.

The final temporal aspect of art to be investigated is how long a viewer typically spends with a work of art. Claus-Christian Carbon conducted an observational study to identify trends in how people engage with paintings in a museum setting. 225 museum visitors with an average age of 43.3 were observed for the study. According to the study, people typically look at a painting for between 25.7 and 41.0 seconds on average. More than half of the visitors returned to paintings they had previously viewed, and the average time spent viewing a painting a second time was 25.8 seconds, or roughly 31.9% less time than the initial viewing, according to the study. People become more enthralled and appreciated when they spend a good amount of time viewing the artwork.

Associations in Art and Emotional Representation

Artworks are used to arouse emotions. A work of art has the ability to evoke some type of emotion, i.e. sadness, happiness, pity, etc. It is interesting that art can actually motivate feeling a range of emotions, just by observing the piece. Musical works of art specifically can evoke a wide range of emotions, based on the tempo instruments, modes, and rhythms used. Emotion is a central part of dealings with artworks (Carroll, 2003).

Hichem Naar, an author for The Internet Encyclopedia of Philosophy, has written on our motives for seeking and valuing an emotional connection with artwork.. Narr states that emotions are subject to norms of correctness. For instance, if the object is not threatening or dangerous, a feeling of fear is inappropriate. However, viewpoints on a piece of art can vary. Additionally, there are various emotional intensities depending on the observer. When someone views a work of art, they might experience melancholy while someone else might experience emptiness. It's important to keep in mind that these feelings fall under the same emotional category. This suggests that, regardless of the medium, there is a genuine connection and relationship we experience between our emotions and the art being observed.

Emotional arousal in those engaged in art is a key characteristic of aesthetic experiences. Subjective and physical responses to the emotion the art is meant to evoke have an impact on aesthetic judgments (Silvia, 2005). Art can make people feel emotions in a way that is not just in their head. It can make them feel things on a deeper level, both inside and outside. This is different from just looking at art and thinking about what emotions it represents. (Pelowski & Leder, n.d.). Individualistic factors, such as empathy and a general appreciation of the arts, have an impact on this interplay. When someone views a work of art, their aesthetic observations and emotional states are closely linked. Even if the observation is subjective, it can still cause physical changes in the person's emotions and expressions, such as smiling or an increase in heart rate. The EVAlab at the University of Vienna has studied the relationships between art and human emotion. They said when an experience is viewed from an aesthetic perspective, more complex emotional states appear to be possible, including a combination of positive and negative responses. For instance, when someone likes a work of art, their emotion changes and they most likely smile. Also, if a perceiver observes a piece of art with a negative tone, they will frown or feel sadness while looking at it.

How is Art Translated From One Modality to Another?

As previously stated, emotion and prior experience have a large impact on how a piece of art is viewed and interpreted. However, one might consider whether there is a defining characteristic that unites all works of art. It turns out that prior knowledge and emotion are the two characteristics that tie art together. Despite the different qualities that the art form takes on, the emotion is still the underlying theme. If that emotion can be encapsulated, the translated piece of art can have a meaningful representation and the same feeling will then be expressed and transferred. A relevant example of this is ekphrasis, which is the concept of describing something so vividly that it can easily be imagined and visualized (Ekphrasis - Writing About Art, n.d.). This is essentially the translation between literature and "images". Description is an important part of this, but emotion is a driving force to the success of this. One can take the words and use their emotional context to derive the visual imagery. The other aspect would be one's prior experiences. Your brain automatically forms connections in general, just because of its pattern seeking nature (Understanding Multimodalities in Arts and Social Sciences, 2019). An efficient translation is not something that would happen deliberately for this aspect, but rather it is a shared common knowledge that would be subconsciously embedded within the translation.

To understand how artists can translate sound into visual forms we must think about the visual items people think about when they hear music. PerMagnus Lindborg and Anders K. Friberg conducted an experiment in 2015 where participants manipulated the color and size of an on-screen circle to match the theme. The experiment found that happy music was associated with yellow, music expressing anger with large red color patches, and sad music with smaller patches towards dark blue. This provided further evidence of a significant emotion mediation mechanism where people tended to match color association with the perceived emotion in the music (Lindborg & Friberg, 2015). This association between music and color can be defined under chromatism where color is received through a response to stimuli that contain no elements of color. Chromatism is more broadly defined as synesthesia which is one sensory or cognitive pathway that leads to involuntary experiences in another sensory or cognitive pathway. Four to six percent of the population claim to feel synesthesia and synesthesia is found to be the most common among artists. To translate sight to sound the artist experiences synesthetic perceptions and creates artworks to invoke synesthetic experience in the viewer (Watson, 2018).

Having an understanding of how someone parses through a visual scene through its physical and spatial properties and knowing how artists are able to translate their art from one modality to another we can examine musical pieces done by artists and how they translate sound to visual items.

How Do Visually-Impaired People Navigate the Visual World?

While mapping visual parameters to auditory variables can create new possibilities for experiencing and understanding visual art, it also has important implications for visually-impaired individuals who rely on auditory cues to navigate the world. Understanding the relationship between the visual and auditory domains can provide valuable insights into how people with visual impairments navigate the visual world.

Recent studies have explored how sensory deprivation affects tactile modality, particularly among individuals who are blind. The study looked at 14 blind, 20 deaf, and 22 controlled participants performing a tactile working memory task. A working memory task examines an individual's ability to manipulate information for a short period of time. The participants sat in front of the examiner with a table between them. On that table was a computer keyboard where the testers put four fingers of each hand on the upper row keys and the examiner touched their fingers in a certain order for 1 second. They were then asked to press the keys in the order that the examiner touched them. Results showed that the blind group performed better than both groups, indicating that a lack of sensory ability may improve working memory (Heled, 2022).

Brain imaging studies found that a loss of vision does not lead to an inactive visual cortex. Functional magnetic resonance imaging was used on blind and sighted subjects. The subjects were required to say and read verbs and nouns. In the second part of the study, the blind and sighted subjects listened to verbs and nouns. Both studies found that blind people had an active visual cortex that was similar to the activity of the sighted subjects. It was found for the Braille task reading for blind individuals the brain imaging showed that the subjects were right-hemisphere dominant. An active visual cortex in blind subjects is important to note because the cortex is involved in cognitive processes such as visual perception and working memory (Ulmer, 2015).

Research has shown that visually-impaired individuals are able to perceive sound and music through other senses such as touch and vibration. For example, some visually-impaired people use vibrotactile aids which convert sound signals into vibrations that can be felt on the skin to navigate and interpret their surroundings (Girard & Collignon, 2017). Studies have also shown that blind individuals have an enhanced ability to process and differentiate sounds suggesting that they may rely more heavily on auditory perception to understand the world around them (Collignon, 2013). These findings suggest that there is potential for developing new technologies and methods for translating visual art into sound that take into account the unique sensory experiences of visually-impaired individuals.

Musical Representations: Mapping Visual Parameters to Auditory Variables

In exploring the relationship between music and visual perception, we can examine how auditory elements can create mental images or visualizations in the listener's mind. By analyzing specific musical pieces such as Claude Debussy's *La Mer*, John Williams' *The Imperial March*, and Gustav Holst's *Venus, the Bringer of Peace*, we can see how composers use different auditory techniques to evoke specific visualizations or mental images. Through the following examples, we can gain a deeper understanding of how music can create a sensory experience that extends beyond sound.

As stated in Judd's article titled "La Mer": Debussy's Sonic Portrait of the Sea Claude Debussy's *La Mer* ("The Sea") is an example of how a visual item such as the sea is translated into music. La Mer is separated into 3 symphonic sections that stray away from the traditional symphonic form. Instead, Debussy uses techniques such as dissonant chords or non-harmonic tones and poly-tones to manipulate the keys in his music to create ambiguity in the perception. In the introduction, Debussy looks to create new sounds and textures by blending sonorities of timpani and double basses, and the echo effect of the two harps playing a beat apart, and the overlapping lines of cellos and violas help to visualize the experience of darkness before dawn at sea. The introduction then moves into the early body of the piece which resembles the sunrise at sea. Instead of presenting two contrasting themes from the first movement to the second movement, Debussy reveals continuously developing sounds that wash over the orchestra similarly to waves. This effect was achieved by accentuating the sound of a melody before letting it fade away. Debussy creates this musical idea by not switching or finishing another idea before it was fully developed creating a fading effect in the sound (Judd, 2018).

According to Brown, The Imperial March by John Williams is a musical piece that is often associated with the evil empire in the Star Wars universe due to its ominous and menacing tone. This tone is achieved through various auditory elements such as timbre, harmony, melody, rhythm, and dynamics. The use of low brass instruments like trombones and tubas creates a deep and powerful sound that evokes the image of large, imposing objects or structures. The harmony, based on minor keys and simple chord progressions, creates a sense of darkness or foreboding, which could be associated with ominous visual imagery. The melody of the march is created by a repeating bass line that gives the piece a sense of power and unity. The strong marching beat can create the mental image of a large army or group of soldiers marching forward. (Brown, 2019).

In the musical piece *Venus, the Bringer of Peace*, the auditory elements such as harmony, melody, rhythm, and dynamics can evoke certain visualizations or mental images in the listener's mind. In Duffie's analysis the use of a diverse orchestral ensemble, including strings, woodwinds, and harp, can create the mental image of a serene and beautiful landscape (Duffie, 2015). In his book "Gustav Holst: The Planets," Kelly notes that the use of curvy and melodic lines between various sections of the orchestra in Venus, the Bringer of Peace, can create a sense of fluid and graceful movements. Additionally, the leisurely and fluid pulse of the rhythm can evoke the image of gentle waves or a peaceful flowing river. These auditory elements work in harmony to produce a serene and calming musical experience that can stimulate certain visualizations or mental images in the listener's mind (Kelly 2016).

Prior Art

In exploring the challenge of navigating the visual world, there have been a range of projects and innovations aimed at developing effective solutions. By examining prior approaches and building upon existing knowledge, we can identify strategies that have potential to improve navigation for people with visual impairments, as well as benefit individuals with other visual capabilities.

Rule-Based Approaches

A rule-based system represents knowledge via rules. A rule-based system is a system that applies rules to store, sort, and manipulate data. It does this by having a source of data and a set of rules for manipulating the data. These rules can be broken down into “if statements” where if X happens then Y happens. A rule-based system follows the human thought process so the implementation of the rules can be considered easy for developers. By knowing what rules you plan to implement you can optimize your system and provide outputs at a fast rate. Using a rule-based system you can ensure a high level of accuracy and fewer errors. Scenarios covered by a rule-based system will provide high accuracy because of the set of predefined rules. A rule-based system cannot accurately simulate intelligence. AI and machine learning have the ability to learn and adapt but a rule-based approach can only do what it has been instructed to by the coder. Also when a rule is introduced that contradicts another, or when the system confronts an issue for which no rule has been defined, problems occur and the program will break down (Rule-Based System, n.d.).

Rule-based systems are usually modeled in the form of deterministic finite-state machines and Markov chains. The latter approach has been utilized in the creation of music. Alvin Lin, a software engineer from New York City fed his program samples of musical works in an effort to find the probabilistic transitions from one note to another. He then created a Markov chain from this data. Using the Markov chain, he started with a random note and continued adding notes based on the highest probability of transition success. The outcomes, while not very dulcet, are interesting and a consideration for an approach to our project (Lin, 2021).

Sonification Approaches

Sonification is the process of communicating information or perceiving facts through audio. Typically, this audio is of a musical nature and is almost always devoid of speech (Rais, 2021). The Mind Music Machine Lab at Virginia Polytechnic Institute and State University created a guide to link visual parameters to auditory variables. They suggest linking hue to pitch, brightness to tempo and loudness, saliency to pitch and mode, size to pitch, and art style to timbre and musical composition (Chihab Nadri et al., 2019). Dimitrios Margounakis and Dionysios Politis at Aristotle University of Thessaloniki have taken a similar approach, however they suggest mapping brightness to duration, color to pitch and timbre, size to duration and tempo, and complexity to pitch (Margounakis & Politis, 2006). Margounakis and Politis went on to apply these suggestions to one example image of a pixelated smiley face. Their sonification algorithm groups pixels in a horizontal line with similar colors into 'bricks' of pixels. The algorithm uses these bricks as it reads the image left-to-right top-to-bottom to inform what pitch and duration should logically be appended to the melody.

Woon Seung Yeo and Jonathan Berger at the Center for Computer Research in Music and Acoustics at Stanford University approached the problem differently when developing a system for converting time-independent images into music. They consider two common methods of sonification mappings: scanning and probing. Scanning is when data is sonified in a fixed, non-modifiable order. Probing is when distinct points are defined and data is gathered from these points. Yeo and Berger acknowledge that neither of these methods alone are sufficient to translate images into the time domain, but they have found that in conjunction, these methods are successful (Yeo & Berger, 2005). SonART is a multi-purpose multimedia environment that Yeo and Berger note has the potential to apply their mode of combined methods to actual artistic works. SonART can create an arbitrary number of layered canvases, which is essentially any number of copies of the artistic work. Each layer has independent control over the visual parameters and can transmit or receive data to and from other layers. Yeo and Berger say that each layer can be thought of as an audio channel. This means one layer might focus on probing RGB values while another scans opacity in the same area. As these two layers share their data between each other, different audio artifacts emerge and combine the characteristics of each channel.

Machine Learning and Deep Learning using Neural Networks

Machine learning is the application and creation of automation capable of learning and reacting without specific guidance, by analyzing and drawing inferences from patterns in data using algorithms and statistical models. There are four main types of machine learning algorithms: supervised, semi-supervised, unsupervised, and reinforcement (Wakefield, n.d.). Each of these approaches have their pros and cons, and thus they all should be considered in the creation of a machine learning application. Supervised learning is when the computer must achieve a given output when provided with its corresponding input. Semi-supervised learning is similar to supervised learning, except some data is unlabeled, meaning the computer must use the labeled data in an attempt to find a way to give the unlabeled data meaning. Unsupervised learning is when the computer is given only unlabeled data and analyzes that data for patterns. Reinforcement learning is when the computer is given a set of operations, attributes, and outputs. The computer must then employ trial and error to produce the desired results. The most popular reinforcement learning method is the genetic algorithm, which is a subclass of the evolutionary algorithm, another kind of reinforcement learning algorithm.

Deep learning is a subset of machine learning which uses artificial neural networks. Node layers in neural networks include an input layer, one or more hidden layers, and an output layer. Each node has a unique threshold value, also known as a static-bias value. The nodes are connected to one another with an associated weight value. If the output of any particular node exceeds its given threshold level, that node is activated and transfers data to the network's next tier. After generating the output, the mean-squared error is determined, and backpropagation is conducted, with the weights modified to minimize the loss. The overall procedure is based on determining ideal weight values for each connection.

While there are nine main types of neural networks, two are most prevalent to this project: the convolutional neural network and the autoencoder (Warr, 2019, Chapter 4). A convolutional neural network (CNN) is a class of artificial neural network that takes in an input, assigns importance to various aspects of the input, and differentiates one from the other using these aspects. Their application is mostly in image recognition, classification, and segmentation. Autoencoders are artificial neural networks that are used to develop optimal encodings of unlabeled data. The encoding is verified and improved In an effort to recreate the input. By training the network to disregard inconsequential input or "noise," the autoencoder produces a model for a data set.

Maximilian Muller-Eberstein and Nanne van Noord from the University of Amsterdam have tackled this idea by using synesthetic variational autoencoders, which they call SynVAE. The goal of their project was to have an autoencoder learn a consistent mapping between visual and auditory sensory modalities in the absence of paired datasets. They specifically used an image autoencoder (VisVAE) and a music autoencoder (MusicVAE) to form a reasonable connection and translate between the original visual artwork and the reconstructed visual artwork. They were able to eliminate the necessity for subjective image-music associations by employing these two single-modality models. An image is encoded into data by the VisVAE encoder first. Then the data is decoded by the MusicVAE. In order to guarantee losslessness, the new data from the MusicVAE is passed into the VisVAE and the image is reconstructed. If the reconstruction matches the original image by a certain margin, the conversion process was successful. The accuracy of the reconstructions using the Behance Artistic Media dataset is displayed in Figure 2. Muller-Eberstein and van Noord conducted a classification study in which participants were presented with three images and a short musical excerpt. The respondents were to identify which image was used to generate the excerpt as well as associate the piece with one of the following emotions: "scary", "peaceful", "happy". The green bars indicate what percentage of participants identified the same image as the SynVAE system. The average accuracy among trials was 71%. These results suggest that the SynVAE was able to preserve consistent cross-modal latent space while translating with enough information content (Muller-Eberstein & van Noord, 2019). In other words, the system was able to translate the images to music with enough precision that humans could perceive the link between the two works.

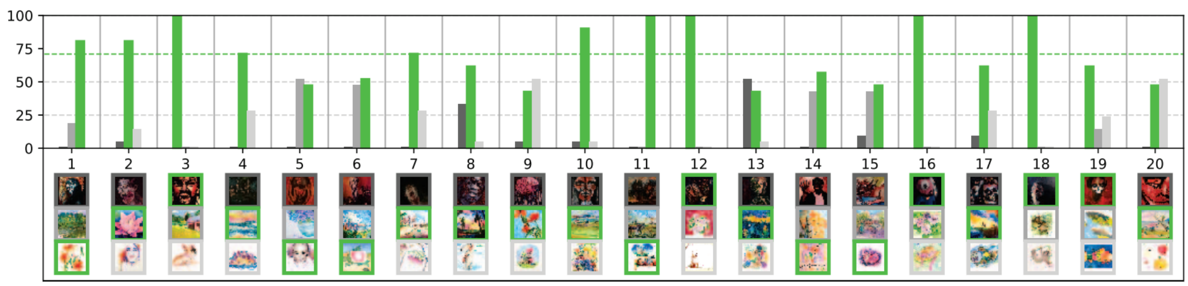


Figure 2. Muller-Eberstein & van Noord, 2019

Román Saldaña and Gerardo Mendizabal-Ruiz from the University of Guadalajara offer an approach based on convolutional neural networks for generating synthetic sounds from shapes. They defined each geometric shape they were investigating to a set of synthesizer parameters that they felt best represented the shape. They then generated a series of images containing one shape of random orientation and size and allowed the neural network to output a short frequency pattern based on the defined parameters of the shape. The final experiment involved adding additional geometrical figures that weren't part of the original dataset. The spectrograms that this trial produced are shown in Figure 3. The data demonstrates the potential of this modeling approach, which excels at pattern recognition (Saldaña & Mendizabal-Ruiz, 2020).

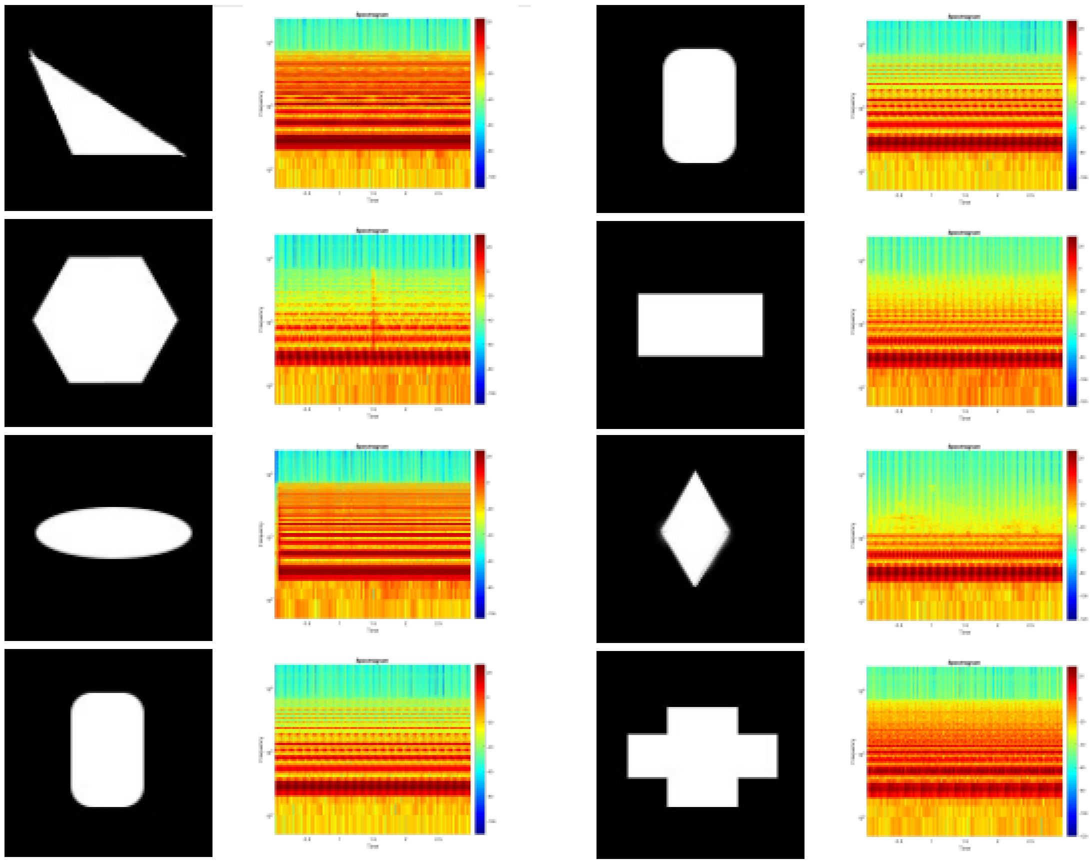


Figure 3. Saldaña & Mendizabal-Ruiz, 2020

Imaginary Soundscape is an online sound installation that emphasizes noises tuned out by selective listening. This approach is built on recent advances in deep learning-based cross-modal information retrieval. The system was trained using two models: given video inputs, one convolutional neural network reads the audio as spectrogram images, evolving so that the distribution of its output is as near as possible to that of the first one, while the other model processes the frames. Once trained, the two networks search through the enormous ambient sound collection and find the sound file that most closely matches a situation. The AI-generated soundscapes tend to be of high interest, but they also occasionally disregard the cultural and geographical context (Imaginary Soundscape, n.d.).

Artistic Approaches: Sound to Sight

Having knowledge of ekphrasis and synesthesia can enhance our understanding of the process of transforming art from one form to another. These concepts help us to better examine artworks that effectively translate sound into visual elements. By examining these pieces, we gain insight into the ways that artists are able to convey the sensory experience of one medium through another.

In the article, "The Reflektor Distortion: An Artistic Interpretation of Distorted Reality," the author describes the components and theme behind the artwork. As seen in Figure 4 The Reflektor Distortion consists of a rotating water-filled basin that is inspired by the shape of a parabolic mirror. The work consists of three main components: mirror, image, and reflection. The theme behind the artwork is distorted perception of reality. The water basin has speed and integrated resistors that continuously change the mirrored reflection of the water. The water's surface is further distorted by resonating low sound frequencies that are played through a speaker below the basin. The sound created from the speakers can be viewed as a visual translation. Sound is a vibration that is translated through a medium however when carried through air it can be difficult to visualize it as we cannot visually see it. But when translated through the water basin the energy of the sound is seen through the vibrations in the water. The low sound frequencies also enhance the theme of the piece as the vibrations from the sound add to the distortion making the water rippled manipulating the reflection (Carsten Nicolai: Reflektor Distortion, 2016).

Carsten Nicolai's Reflektor Distortion provides a unique approach to the translation of visual art into sound. The artwork combines visual and auditory stimuli, creating a multisensory experience for the viewer/listener. The artwork also explores the concept of distorted perception which could inspire our team's project to explore similar themes and ideas. However, the translation of visual art into sound is limited to the use of a water-filled basin and low-frequency sound waves which could limit the scope of your project. The Reflektor Distortion artwork is complex and requires specific equipment and technical expertise to create, which could pose a challenge for us as we are working with limited resources. In addition, the interpretation of the sound translation may vary from person to person, which could impact the effectiveness of your project in conveying a specific message.

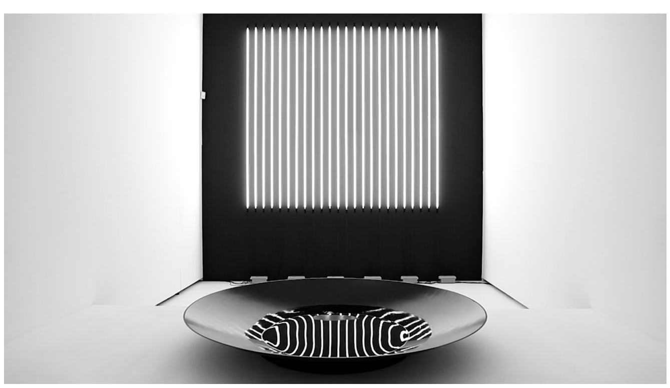


Figure 4. *Reflektor Distortion* by Carsten Nicolai (2016).

According to a source for Neil Harbisson website, Harbisson is an artist who, due to a rare condition called achromatopsia, was born completely colorblind. In Figure 5, the *eyeborg* device developed by Neil Harbisson and Adam Montandon is shown. The device has a camera that detects colors in front of him and a microchip that converts them into musical notes, which are transmitted to Harbisson's inner ear through bone conduction. To create musical compositions based on the colors he encounters, Harbisson has developed a system of musical notation called the *chromatic scale for colorblind people*. This system assigns a specific note to each color in the visible spectrum based on the frequency of light waves associated with that color. Harbisson uses this system to create music inspired by various colors and believes it allows him to perceive and understand color in a new way (Harbission, 2012).

The use of a device like the eyeborg, which has a camera that detects colors in front of the artist and a microchip that converts them into musical notes, could be something to consider in your project of converting visual art to sound. This device provides a unique way of translating colors into sound, which could expand the range of visual art that can be effectively translated. Additionally, the use of a device like the eyeborg provides a concrete and technical approach to the translation of visual art into sound, which could be useful in ensuring consistent and accurate translations.

Figure 5. *Eyeborg* by Neil Harbission (2004)

As noted by Cunningham DJ Christian Marclay is an artist who has pioneered a unique form of turntablism that has inspired his creation of Recycled Records. His work as a photographer, sculptor, installation artist, and performer has informed the development of these artworks, which consist of sliced and broken records that are reassembled to form unique pieces of visual art. Marclay's turntablism practice, which involves manipulating sounds and mixing music together, could be seen as the inspiration behind these recycled record pieces. By using his skills in turntablism to create visual art, Marclay is able to merge two different forms of artistic expression into one (Cunningham, 2022). Through the manipulation and mixing of the vinyl records, Marclay creates something entirely new and visually stunning. The use of recycled records as a medium also adds a layer of sustainability to his work, repurposing discarded materials and giving them new life as pieces of art.

Christian Marclay's Recycled Records offer a unique approach to merging turntablism and visual art. However, implementing some of the concepts used in his pieces into a more systematic and programmable approach for converting visual art to sound may prove to be challenging. This is due to the physical manipulation and mixing of vinyl records, which is not directly transferable to a digital or programmable system. Additionally, the uniqueness of each physical record used in Marclay's artwork makes it difficult to create a standardized approach for translating visual art into sound.

Figure 7. *Recycled Records* by Christian Marclay (1980-1986)

Christine Sun Kim's art centers around systemic barriers attached to deafness. Kim discusses that closed captions with dialogue are acceptable however closed captioning of sound does not help people who are deaf. For example, if violin music begins to play in a movie the closed captions do not help in explaining what the sound is made of, how it moves, or its personality. The closed captions read music or violin music which is not very helpful to someone who is deaf. In her art pieces, she describes the sound of a particular scene through her perception in closed captions (Smithsonian American Art Museum, n.d.).

While Christine Sun Kim's art centers around the systemic barriers attached to deafness, her approach to closed captions can be seen as an inspiration for developing a program that considers how blind people think. Kim highlights the limitations of closed captions for sound and how they do not provide the necessary information to fully understand the characteristics of the sound. Similarly the program we create that converts visual art to sound should consider how blind people will interpret it.

Figure 6. *Close Readings* by Christine Sun Kim (2015)

The Gap

After considering the core ideas behind how art is interpreted and viewed, in conjunction with projects that have already dabbled in the domain of translating art from one medium to another, several issues become apparent. One glaring issue is that the projects that can be considered successful do not relate to what was fed into it. In other words, the resultant piece does not conserve the emotional and literary values of the original piece.

A majority if not all projects that convert a visual medium to sound do not incorporate a tactile system to interpret the sound created. As discussed previously a blind person's ability to feel and sense through touch is important in their ability to store that information and perform cognitive tasks. When parsing a visual scene a blind person needs these cognitive functions in order to better understand the visual piece they plan to interpret through sound or music.

A rule-based approach can possibly make it difficult to translate sight to sound as the way someone parses through a visual scene is complex. As mentioned previously the way someone parses through a visual scene consists of nine different properties. The most notable properties that can possibly be difficult to translate are objects, categorical identity, and location. These properties might be difficult to parse through using a rule-based approach as they are derived from prior experiences and knowledge of the visual world. People have a general understanding of what objects like a house and car are supposed to look like so they are easily identifiable regardless if they are realistic or not. Under a rule-based approach, would it be difficult to identify these objects in the picture as a rule-based approach generates its results based on rules rather than knowledge and experience?

Another problem is that the mapping from one form to another is too direct and literal. Most of the example projects, especially those taking a sonification approach, directly correlate one feature of the medium to another of the translated medium. This is a problem because it derives too much emphasis on what is different between the mediums rather than considering what each medium has in common figuratively. This issue can also be viewed from the lens that most mappings are one-to-one, rather than one-to-many or many-to-many. This is a problem because it does not allow for the possibility of multiple interpretations of the same piece of art. This leads into the second issue observed.

Many of these projects are derived from subjective bias by its creators. While everyone has their own interpretations of art and the meaning of its characteristics, there is some ubiquitous understanding of what aspect of a piece of art relates to. It is not feasible for a project of this nature to solely rely on the "objective" characteristics of art, but it is important that subjective interpretations do not control the outcomes, as that hinders the purpose of the project.

Another issue is that the original media are either parse at too high of a level or too low of a level. This creates an either too generalized or illiberal (stringent) outcome. For example, some projects parse the image down to the geometric or even pixel level, while others parse the image to the object level.

While each of the projects mentioned may be effective in addressing their specific goals, they may also be limited in their scope. While this does not initially sound like a bad idea, it fosters an environment of rigidity and inflexibility. The main conceptual approaches all have promising attributes as well as drawbacks. These drawbacks are often either overlooked in the rush to implement the system or not identified at all. The result is a system that is not as effective as it could be. Therefore, in defining our project objectives, we will aim to consider a variety of approaches and solutions that can be adapted to different individuals and contexts.

There is one additional factor for this project that is most likely taken into account by the other projects but is not stated explicitly. In his 1997 book *Notes on the Cinematographer*, French director Robert Bresson — who is generally regarded as one of the greatest filmmakers of all time— discussed this idea. "The truth of cinematography cannot be the truth of theater, nor the truth of the novel, nor the truth of painting. (What the cinematographer captures with his or her own resources cannot be what the theater, the novel, painting capture with theirs.)" (Bresson, 1997, Page 20). Bresson asserts that each form of art has inherent, distinctive qualities that are unique from a network of qualities shared by other forms of art. It is noted that, while these qualities that are not realistically able to be translated in an accurate sense, it is possible to map one property to another that is similar in terms of the way it is perceived by the viewer. For example, the way a theatrical performance uses the space provided on a stage is not remotely comparable to the way a painting uses the space provided on a canvas, but the way a painting uses color is comparable to the way a theatrical performance uses lighting.

Keeping these problems and considerations in mind, our project will define its objectives by focusing on creating a solution that is inclusive, accessible, and user-friendly for visually-impaired individuals. The solution will aim to bridge the gap between the visual and auditory world, providing a means for those with visual impairments to interact with and navigate their surroundings. To achieve this objective, the project will involve researching various technologies and techniques that can effectively translate visual information into auditory signals.

CHAPTER 3: Methodology

Objectives

1. Determine the best candidates of art that translate from sight to sound
    a. Determine the physical form and dimensions of the art
    b. Consider the different art genres
    c. Determine the genre of the art and consider the abstractness and complexity
2. Determine the method of translation that will result in the greatest retention of identity
    a. Determine which visual qualities should be retained through translation
    b. Determine which auditory qualities could facilitate translation
    c. Determine the algorithmic design and sub-approaches
    d. Determine the best interface
    e. Determine which audiovisual parameters the user can manipulate
3. Carry out a study on the relationship between audiovisual parameters
    a. Outline the purpose of the study
    b. Determine the procedure of the study execution
    c. Recruit participants
    d. Conduct the study
4. Analyze the data to determine the proper mappings from sight to sound
    a. Determine the visual and musical parameters that will be used for mappings
    b. Determine statistical analysis methods to identify relationships between parameters
    c. Determine visualization methods to visualize the relationships between parameters
    d. Develop the best approach using a combination of statistical analysis and visualizations

Determining the Best Candidates of Art That Translate From Sight to Sound

Determining the Physical Form and Dimensions of the Art

To determine the physical form and dimensions of the art we looked at the possibility of dealing with art in 2D or 3D forms. 3D parsing using deep learning approaches has been considered however these methods with fully supervised models require manually annotated point-wise supervision which is extremely user-unfriendly and time-consuming to obtain. A study was done on developing a system that required limited supervision. Although the system was more user-friendly, developing it was still time-consuming (Sun, 2022).

2D parsing is easier to develop due to lower level complexity and analyzing an image in 2D rather than an object in 3D takes less resources and code development. Our team decided to go with 2D parsing as it takes less time and resources to develop compared to 3D parsing.

Considering the Different Types of Art

Our team considered different types of art including performance-based and static visual works. Performing arts would focus on films and theatrical performances. This form of art can include music, sound, and visual dance/acting scenes. Translating performing arts into sound could be highly difficult as the performance consists of moving images and parsing through those scenes to develop music for those images could prove to be difficult. Visual arts consist of pictures, paintings, photographs, and sculptures. Photographs and pictures could be difficult as the images might be too high of quality. Both photographs and paintings can be complex and contain a lot of visual information however photographs tend to be more detailed and high-resolution compared to paintings. This is because photographs capture an image in a literal and precise way, while paintings can be more interpretive. As a result, photographs tend to have larger file sizes and contain more visual data that needs to be processed in order to analyze their essence. This can make it more challenging and time-consuming to translate photographs into sound, compared to paintings or other forms of visual art.. Sculptures are a type of 3D visual art and as we discussed previously 3D parsing requires more time and resources to develop. Our team decided to focus on paintings as the resources and time put into parsing through these works we thought would be possible in the given timeframe. They also are more open to interpretation and tend to have some kind of artistic expression and emotion behind the work (7 Different Forms of Art, n.d.).

Determining the Genre of Art and Considering Its Abstractness and Complexity

As we explore the possibilities of translating visual art into sound, it is important to consider the different genres of art and their potential for sonic representation. We will be looking at various art genres, from the abstract and minimalistic to the complex and detailed, in order to determine which ones may lend themselves best to aural interpretation. By examining the inherent qualities of each genre, such as form, color, and texture, we can assess their potential for translation into sound. Additionally, we will consider the challenges and limitations that may arise with each genre, such as the level of detail or complexity, in order to identify the most suitable candidates for our project.

Impressionism was one of the art genres our team looked at. The art movement focused on painting everyday life using the effects of light and emphasizing the vibrancy of colors to reflect from the artist's perspective the world in which they live. Characteristics of these paintings included quick loose brush strokes, bright paintings, and relative color. Impressionistic paintings can be considered to be mid to low level of abstraction. The paintings present a fairly accurate depiction of visual reality as the paintings are based on the real world and the color of the paintings is similar to how it is presented in reality. There is some deviation from visual reality though as the art pieces consist of loose brush strokes making pieces look soft and blended contrary to how they are presented in the real world. Figure 7 can be seen as an example of a painting with a high level of complexity as it contains a large number of objects arranged in a complex composition. The painting appears to depict a large gathering of people participating in various activities such as eating, dancing, and conversation. On the other hand, Figure 8 can be seen as an example of a painting with a low level of complexity as it contains only a few simple objects arranged in a relatively straightforward composition. The painting appears to depict a few lilypads on the water. The use of vivid bright colors in impressionistic paintings could provide a better oppurtunity for creating better music as different vivid colors can provide more variation in creating unique musical sounds. Possible limitations within the works are that with the use of quick loose brush strokes there are difficulties in identifying objects within the painting (The Editors of Encyclopaedia Britannica, 2022).

Figure 7. *Bal du moulin de la Galette* by Pierre-Auguste Renoir (1876)

Figure 8. *Water Lilies, Evening Effect* by Claude Monet (1897-1899)

Realism is an art genre that focuses on accurately representing the world through painting using techniques such as capturing light, shadow, and texture with flat colors and shading (Tate, 2016). White realism paintings can be considered a low level of abstraction as they look to accurately represent the world through painting, the level of complexity can vary depending on the piece of art, as seen in Figures 9 and 10. Figure 9 depicts a highly complex piece of realism artwork with many objects and activities taking place in the painting, while Figure 10 is a simpler piece with fewer objects and a more straightforward composition. While the techniques used in realism paintings could potentially make it easier to identify objects within the painting, our team felt that creating a sight-to-sound experience with realism paintings could pose difficulties. The accuracy of the representation and the lack of abstraction could limit the creativity in the translation process, as there may not be much room for interpretation. Additionally, the realism style tends to focus on visual details such as light, shadow, and texture, which may not translate as effectively to sound. These visual elements are not easily translated into sound as sound does not have the same physical properties as light. While it is possible to create soundscapes that evoke certain moods or emotions associated with the visual details in a realism painting, it is much more difficult to create a direct translation of these visual elements into sound. Therefore, it may be challenging to create a meaningful sight-to-sound experience with realism paintings.(The Editors of Encyclopaedia Britannica, 1999).

Figure 9. *The Daughters of Edward Darley Boit* by John Singer Sargent (1882)

Figure 10. *The Red Bell Peppers* by Marlene Llanes (2015)

Pop art was another genre we looked at as a possible option. Pop art includes imagery from mass cultures such as advertising, comic books, and mundane mass-produced products. Characteristics of pop art include recognizable imagery, bright colors, irony, and satire (Pop Art Movement Overview, 2012). Pop art can be considered to have a medium level of abstraction. Pop art's use of recognizable imagery is consistent with visual reality but the use of vivid colors associated with these images and icons is not accurate to how it is presented in the real world. Pop art can also be considered to have a high level of complexity as usually a multitude of different icons and images throughout mass culture are used to create the art piece. Pop art's use of a wide range of bright colors could provide us with a better opportunity of creating meaningful music pieces as there are more colors to work with. One of the issues with pop art is identifying what sounds should be created out of icons and symbols from popular media.

Figure 11. *Cambell' Soup Cans* by Andy Warhol (1968)

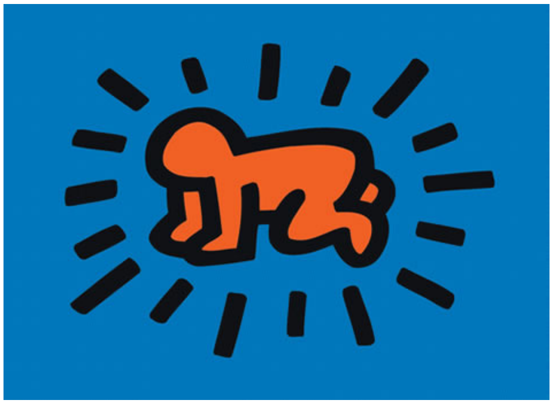


Figure 12. *Radiant Baby* by Keith Haring (1990)

Minimalism is typically composed of simple geometric shapes with a limited monochromatic color palette consisting of limited features and graphic elements. Minimalism can be considered highly abstract as it removes forms of self-expression and individuality and focuses more on the material and textural elements. It also can be considered to have a low-level complexity as there tends to be a small number of shapes used to compose the painting and these shapes tend to be very simplistic (Minimalism Movement Overview, 2022). Minimalism could be a viable option for creating music. The use of limited features and low-level complexity in minimalism can make it easier to create music, as it removes the need for complex melodies and harmonies, and instead focuses on the arrangement and manipulation of sound elements. In a similar way to minimalistic art, minimalistic music may be more focused on the material and textural elements, and less on emotional expression. This could result in more abstract and experimental sound pieces that prioritize the arrangement of sound elements over a coherent melody or harmony.

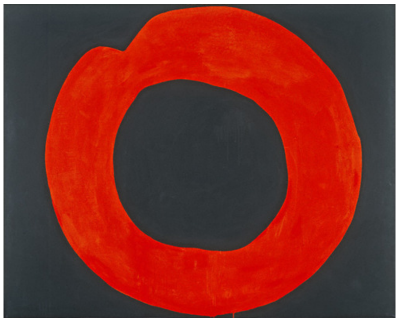


Figure 13. *Red Circle on Black* by Jiro Yoshihara (1965)

Figure 14. *Nine-Sided Figure* by Sol LeWitt

The last art genre we looked at was expressionism. Expressionism is paintings where reality is distorted through the artist's perspective. Characteristics of expressionism paintings are jagged/distorted lines, rapid brushwork, and jarring colors. Expressionist painting can be considered to have mid to high-level abstraction as the paintings tend to have reality distorted as it is told through the artist's perspective (Meleca, 2022). Expressionists try to convey emotion and meaning rather than reality. Complexity can vary depending on the artistic piece. Works can consist of a large or small number of objects that make up the artistic piece. These objects can also be complex depending on the color, textures, and shapes used to create them. With a wide range of colors and textures mapping these characteristics to music could make for better music as there are more elements to work with. With paintings having a medium level of abstraction it could also be easier to create music. If the painting was too concrete there is an exact meaning behind it which makes it harder because we have to capture that meaning. With some level of abstractness it allows for variability in interpretation.

Figure 15. *Castle and Sun* by Paul Klee (1928)

Figure 16. *Blue Horse I* by Franz Marc (1911)

From the options of art genres our team could possibly explore we decided to go with expressionism as we thought it had the best opportunity from the qualities that were represented in the painting. Although all artwork can be viewed as unique, certain genres or styles of art may have more similar elements or techniques that are used to create the artwork. For example, a collection of realistic paintings may all use similar techniques to depict the subject matter, such as the use of light and shadow or realistic color palettes. Parsing through paintings that could be similar in these qualities could make it difficult to uniquely identify each one as results would have to be more concrete and specific. Expressionism focuses on the artist's emotional and psychological expression, rather than on realistic representation of the world. This can result in a wider range of visual styles and characteristics, as different artists explore different ways of conveying their emotions.

We also chose expressionism because of its general level of abstractness and complexity. Expressionist paintings can have a low to high level of abstractness and complexity. With this, we can look for paintings with a level of abstraction that allows for a more open interpretation of the piece while still having enough information to identify what is happening in the art. Paintings that are low in abstraction such as realism makes it difficult to create music based on the piece because there is a more exact meaning behind the work and capturing that could prove difficult. Music can be highly abstract and emotional, and relies on the composer or musician's unique perspective to convey meaning and feeling. This could be an advantage when creating music in response to expressionist paintings. The abstract and emotional nature of music allows for a deeper level of artistic expression that can pair well with the abstract qualities of expressionist paintings. Just as an expressionist painting may use bold colors, exaggerated forms, and heavy texture to evoke a particular emotional response, a musical composition may use unconventional harmonies, dissonant sounds, and experimental techniques to create a similar effect. The abstract nature of music can complement and enhance the abstract qualities of expressionist paintings. This could allow for a better experience that combines visual and auditory elements in a meaningful and expressive way.

Determining the Method of Translation That Will Result in the Greatest Retention of Identity

Determining Which Visual Qualities Should Be Retained Through Translation

The properties that were considered for artistic retention were color and shape. Color is one of the major qualities that should be retained through translation as artists use it to portray mood, light, depth, and point of view in a work or art. The human brain associates warm colors such as red, orange, and yellow with a range of feelings, including passion, comfort, anger, and power. Cool colors such as blue, green, and purple have the opposite effect, creating a calming atmosphere that counteracts feelings of anxiety. Artists can manipulate and use colors within the painting to help convey a particular emotion (Mohr 2020). Color can also be used in visual art to simulate texture. Texture is a tool used by artists to lead the viewer through their creations. Smooth or rough textures can be used by artists to change the mood, draw attention to specific areas of the painting, or deflect our gaze from other elements. Although there are many different color models that can be used to define color (such as RGB, CMYK, and CIELAB), the HSV color model is the best representation of how the human brain encodes and decodes colors because it makes tints, tones, and shades easy to visualize and calculate.

The other essential component in the creation of visual art is shape, also referred to as form. It can have an impact on composition, help maintain the harmony of a piece, give the impression of movement, add texture and depth to an image, imply a mood or emotion, or draw attention to a particular subject (Zeng 2015). The composition of a piece of art is used to assess its completeness and how well each component contributes to the painting's overall impact. The degree of smoothness or sharpness of a component curve's curves, as well as the shape's orientation, dimensionality, and relative location within a piece, can all be used to define form.

Determining Which Auditory Qualities Could Facilitate Translation

In music composition, there are countless musical parameters that could be utilized to create a piece of music. In designing the study, it was crucial to select a limited number of musical parameters. The reason for this was that the study needed to be conducted within a reasonable time frame. However, it was also important that the chosen parameters were comprehensive enough to capture the essential aspects of music theory. As a result, the selected musical parameters were fundamental to music composition and served as the main building blocks of music theory. The selected parameters were chosen because they are all integral aspects to music composition and are considered to be some of the primary building blocks of music theory. The chosen musical parameters include timbre class, mode, tempo, dynamic variation, and rhythmic regularity.

Timbre class refers to the quality of sound produced by a musical instrument or a voice. It is determined by the instrument's construction, the way it is played, and its harmonic content. Timbre class is essential in differentiating between instruments and voices, and it can significantly impact the emotional and expressive qualities of a piece of music. Mode and scale are two closely related parameters that are also crucial to music composition. The mode of a piece of music refers to the set of pitches used in a musical composition, while the scale refers to the ordering of these pitches. Different modes and scales have distinct tonal qualities and can significantly affect the mood and emotional impact of a piece of music. Tempo is another essential parameter that determines the speed or pace of a piece of music. It is measured in beats per minute and can be used to create different moods and feelings in a musical composition. Dynamics, on the other hand, refer to the variation in loudness or intensity of a musical performance. This parameter can be used to create contrast and tension in a musical piece. Finally, rhythmic regularity refers to the degree of consistency and predictability in the rhythmic patterns of a musical composition. It is a fundamental aspect of music and can significantly impact the overall structure and coherence of a piece. Together, these parameters form the foundation of music composition and play a critical role in shaping the listener's experience of a musical work.

By exploring these musical parameters, the study aimed to investigate how they are related to visual parameters such as shape, color, and form. The findings of the study could have significant implications for music composition and theory, as well as for the fields of visual arts and psychology. Overall, the selection of these fundamental musical parameters was crucial for the study's success and the understanding of the relationship between music and visual art.

Determining the Algorithmic Design and Sub-Approaches

As mentioned previously, for us to create a system to convert sight to sound we would expect the need for a combination of rule-based, sonification, and machine-learning approaches. We first determined that we wanted a finite-state automata system for the rule-based section. For our input method, we had the option of going with a deterministic or non-deterministic approach. In a deterministic approach, every state has exactly one transition for each possible input while in a non-deterministic approach an input can lead to one, more than one, or no transition for a given state. Our team decided to go with a non-deterministic approach as the randomness allows for multiple interpretations of the art piece. For our output method, we had the option of choosing acceptors, transducers, or sequencers. Acceptors produce binary output, indicating whether or not the received input is accepted. Transducers produce output based on a given input and/or a state using actions and can translate between inputs and outputs. Sequencers produce only one sequence which can be seen as an output sequence of acceptor or transducer outputs. Our team decided to go with transducers that are used to map between parameters.

There are three sonification approaches we considered. Audification is a process that turns a one-dimensional signal (or a two-dimensional signal-like data set) directly into sound. Without any modification or transformation, the data itself serves as the sound source in audification. For instance, the data could be used to represent amplitude over time and played back on a loudspeaker for listening purposes. Data that already has a distinct auditory signature, like sound recordings or time-series data, is best suited for audification. Parameter-mapping sonification is a technique for mapping specific data parameters to sound parameters, such as pitch, volume, or timbre. Parameter-mapping sonification often involves the use of scales or musical metaphors to create meaningful relationships between the data and the resulting sound. Parameter-mapping sonification is best suited for data that has clear numerical values that can be mapped onto sound parameters. Model-based sonification involves creating a mathematical model of the data that represents the underlying relationships and patterns in the data. The model is then used to generate sounds that convey information about the data. This method necessitates proficiency in mathematical modeling and signal processing, as well as a thorough understanding of the data and its underlying structure. Model-based sonification is an effective tool for data exploration and analysis because it frequently involves the creation of intricate sound structures that convey multiple layers of information. Complex data or models that are tricky to interpret using other techniques are best suited for model-based sonification. Our group ended up choosing parameter-mapping sonification as it is a better option when working with multivariate data and when parsing through a visual scene the system is analyzing qualities like color and shape within the piece.

Machine learning has a few approaches to consider. Supervised learning is where the computer must achieve a given output when provided with its corresponding input. Semi-supervised learning which is similar to supervised learning, except some data is unlabeled, meaning the computer must use the labeled data in an attempt to find a way to give the unlabeled data meaning. Unsupervised learning is where the computer is given only unlabeled data and analyzes that data for patterns. Reinforcement learning using an evolutionary algorithm the computer is given a set of operations, attributes, and outputs; the computer must then employ trial and error to produce the desired results. Because the data being analyzed would largely be unclassified in terms of what relationships exist, our team decided to use unsupervised learning because we wanted the computer to use pattern recognition to identify these relationships.

Determining the Interface For the System

Our team looked at multiple options for what interface we wanted for our system. The most simple option our group came up with was a button the user would press that would play the music. It is an option that requires minimal resources and is easy to develop. However, it lacks tactile function. Blind people rely on the ability to touch in feel in order to better understand the visual world. Touch allows blind people to activate cognitive functions which are important when analyzing visual scenes through sound. It also is a more passive approach and lacks the interactive experience we wanted in order to build more interest into exploring the art piece through sound. Another option for an interface was a controller with buttons and knobs. The controller is also one of the more simple options and would be easy to integrate within our planned system. There are a variety of configurations to work with so it allows customization on the parameters we want to be changed using the button and knobs. Issues with using the controller is that the users will have to be educated on which button and knobs perform certain actions.

We also considered a touchscreen interface. The touchscreen is a fairly simple interface. The user can easily navigate a touchscreen monitor and takes less concentration than using a keyboard, mouse, or multiple buttons or knobs. The touchscreen is engaging and interactive as users have the ability to touch and drag fingers across the touchscreen to identify different sections throughout the artwork creating an engaging and interactive experience. Limitations with touchscreen is the accuracy and precision depends on the size of the icons or objects. If objects are too small it may be difficult to sense with someone with larger fingers. Another limitation is that there is no tactile feedback; the touchscreen is just a flat surface.

A motion or proximity sensor is another interface our team considered. To implement this system the sensor would need to be mounted on some sort of movable platform or device that could traverse the surface of the artwork. The sensor itself would likely be a proximity sensor that uses electromagnetic radiation or other techniques to detect the presence of physical objects or surfaces. As the sensor moves across the surface of the artwork, it would collect data on the proximity and position of the surface. Then processing the sensor data and triggering an action or response based on the position and proximity of the sensor. For example, play a musical track or sound effect when the sensor reaches a particular location on the artwork.

By incorporating a movable sensor into a visual art installation, artists and designers can create a more dynamic and interactive experience for viewers. The sensor allows viewers to engage with the artwork in a more tactile and sensory way, creating a more engaging experience. A movable sensor can also improve accessibility for individuals with disabilities, such as visually impaired or blind individuals. By providing tactile feedback or audio cues based on the position of the sensor, the artwork can be made more accessible and engaging for a wider range of people. This can help to create a more inclusive and diverse art experience.

Some limitations with the sensor is that highly conductive materials can interfere with the sensor's ability to detect the presence of a physical object. This is because conductive materials can absorb or reflect electromagnetic radiation, making it more difficult for the sensor to detect changes in proximity or position. Additionally, highly reflective or transparent materials can also pose challenges for the sensor, as they can cause interference or reflections that make it difficult to accurately interpret the sensor data (Agarwal, 2021). Proximity sensors are a promising option for enhancing the experience for blind people by providing an engaging and interactive experience without the need for physical contact

The last interface our team looked at was an eye tracker. Eye-tracking technology helps observe and measure eye movements, pupil dilation, point of gaze, and blinking to see where subjects of a study focus their visual attention, what they engage with, and what they ignore (SR Research Ltd., 2022). This information can be used to trigger specific audio or musical cues based on the viewer's point of gaze. For example, if the viewer is focusing on a particular element within the artwork the eye-tracking technology could trigger a musical response that is tailored to that element. Similarly, if the viewer is ignoring a particular element, the technology could adjust the audio response to better capture their attention. One possible limitation with eye-tracking is that it is an engaging interactive experience but could possibly be limited to sighted individuals. Eye-tracking technology can be expensive and difficult to obtain for a research project. The cost of the equipment can vary depending on the type and quality of the eye tracker, and can range from a few hundred to several thousand dollars. Additionally, the technology can require specialized training and expertise to set up and operate, which can add to the overall cost and complexity of the project.

From the options we considered our team decided to go with the button and knobs interface. The use of buttons and knobs provides a tactile system that allows us to change musical qualities to enhance the listener's experience. Limitations with a singular button, motion/proximity sensor, and eye tracker is that we would have to build a separate interface to change these musical qualities which would require more work and resources. The touchscreen was a possible option as creating an interface for users to manipulate music qualities would be built into the touchscreen. However, we thought the use of buttons and knobs provided better tactile feedback than touchscreen use.

Determining Which Audiovisual Parameters The User Can Manipulate

Using a button-knob interface, our team has to determine musical parameters the user can manipulate to enhance their listening experience. Structural features in music like tempo, tonality, pitch range, timbre, and rhythmic structure should be left unchanged as those parameters are determined by the system to translate the visual image to music that will invoke a particular emotion. We would not want the user to change these parameters because adjusting them could change the emotion and affect the music that is created from the image. Parameters the user can manipulate would be the section or location the user is able to parse through in the image. Being able to move through different sections throughout the image can enhance the user's listening experience as different sections could be composed of different visual qualities that result in separate musical sounds. The plan for our interface is that the user can turn the knobs on the controller to change what section they are in the paintings by changing x and y coordinates.

Another parameter the user can manipulate is the volume of the music. The user should be able to change the overall volume of the music while retaining the internal dynamics. The dynamics is important as the loud and soft music influences the listeners experience. Loudness is expected to contribute positively to both duration and pace perceptions, because loud (versus soft) music confronts sensory receptors with more salient stimulus information. Louder, more salient music should evoke higher levels of attention, processing, and recall of the stimulus event. Being able to change the overall volume while maintaining the velocity ensures that the variance between loudness and softness of music is still maintained.

The last parameter the user can manipulate is the time spent listening to the music. Although mapping from how long someone looks at a painting to how long they should listen to music is not as simple as a one to one, from this information we can determine that the time spent to analyze the painting for people is different for everyone. With this information we would want to create a similar experience for our listener to determine how long they would like to listen to the music to feel that they could get a good grasp of it.

Carrying Out a Study on the Relationship Between Audiovisual Parameters

Outlining the Purpose of the Study

The purpose of this study was to investigate the relationship between visual and auditory parameters in order to gain a deeper understanding of how they interacted with one another. Specifically, the study aimed to determine which of the visual parameters — hue, saturation, brightness, curvature, and size — were related to which of the auditory parameters — timbre class, musical mode, overall volume, tempo, dynamic variation, and rhythmic regularity — and to what extent. By examining these relationships, the study hoped to shed light on how individuals perceived and processed sensory information, and how this may have impacted their overall experience of music and visual media. Ultimately, the study sought to contribute to the development of new approaches for creating audio-visual experiences and exploring the links between sound and image in artistic expression. The findings of this study may also have implications for the design of multimedia experiences, as well as our understanding of the role that sensory integration played in our perception of the world around us.

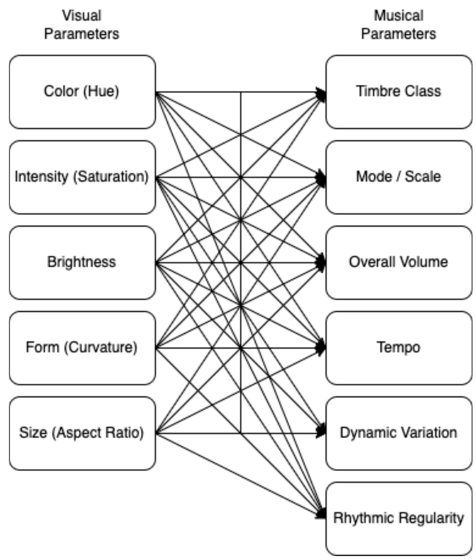


Figure 17. Parameter mappings for study

Determining the Procedure of the Study Execution

Arrival of Participants

After setting a time to meet with participants, participants arrived at the principal investigators' lab (Alden Hall B21). Participants were seated in a location within the office and given a desk and chair that had been properly sanitized. Participants were also instructed to bring their own headphones. If they did not have access to headphones, they were provided and were properly sanitized. After being settled, they were instructed to fill out some pre-experiment questions using Qualtrics. The pre-experiment survey will be comprised of demographic questions (ie. age, gender, ethnicity, etc.), reflection of visual art and musical experience, and acknowledgement of current emotional state. The specific questions can be found in Appendix B.

Execution of the Study

The research was divided into two sections, each comprising fifteen trials of equal complexity. During the first part of the study, known as Music to Image (M2I), we aimed to investigate how people visually represent music. Participants were presented with a series of short musical excerpts, each between 3 and 7 seconds in length. After listening to each excerpt, participants were instructed to draw a shape that they felt represented the piece using several cubic Bézier curves. To aid in this task, a white rectangle was provided as a base shape that could be manipulated in terms of its shape, color, and size. Through analyzing the resulting shapes created by the participants, we were able to gain insights into the ways in which people visually represent different aspects of music, such as its rhythm, tempo, and mood.

During the second part of the study, known as Image to Music (I2M), participants were presented with a series of figures differing in size, shape, and color. The participants were then instructed to adjust the knobs on a custom-built interface in order to create a short musical excerpt that they felt corresponded to the visual stimulus. Specifically, the knobs controlled the instrument, subset of pitches, speed, dynamics, and the rhythm of the composition. The objective of this part of the experiment was to explore the relationship between visual and musical parameters and to investigate how small changes in form and color could impact the participant's interpretation and creation of a musical piece. By analyzing the resulting musical compositions, we were able to identify patterns in the way that participants translated visual information into sound.

Conclusion of the Study

After finishing both parts of the experiment, the program ended. Student investigators collected any items given to the participant, and they were properly cleaned and handled. Discussions were held with the participant to ask how they thought the experiment went and what could have been improved.

Recruiting Participants

Our experiment aimed to gather data from a diverse group of individuals, including adolescents, college students, and selected adults. Adolescents were recruited from others we knew, and we obtained parental consent for any juveniles who participated. College students and professors were from WPI. Flyers were also posted around campus for those who simply wished to participate.

Conducting the Study

During the study, the program used to conduct the experiment was created by one of the investigators using JavaFX. The investigator utilized the javax.sound library and JFugue API to play and compose the musical excerpts used in the study. The source code and components for the application can be found in Appendix A. The program was run on the investigator's laptop throughout the study.

Music to Image (M2I)

After properly sanitizing our computer, we ran the program and provided it to the participants. The program started by prompting the user to click the start button, followed by providing instructions on how to use the interface's tools. The participants were next directed to the experimental interface where they were prompted to listen to an audio file by pressing the play button located at the bottom of the screen. The users had the ability to manipulate the form and color of a white rectangle presented on the screen by moving black dots, anchor points, and adjusting the curvature and size of the lines. Participants could add more anchor points by clicking on a specific line in the shape. The curvature of the lines was determined by the grey dots, referred to as control points, associated with each anchor point that could also be manipulated by the user. The color of the figure they created could be adjusted by clicking on the palette icon on the bottom right of the interface, which would open the color picker panel and allow participants to choose the hue, saturation, and brightness of the color. Once participants were satisfied with their image, they submitted the spline equation and color to the database file, assessed their confidence in their response, and moved on to the next audio clip. This process was repeated for all of the audio clips presented. The specific data that was collected is listed in Appendix C, along with the initial parameters of each trial.

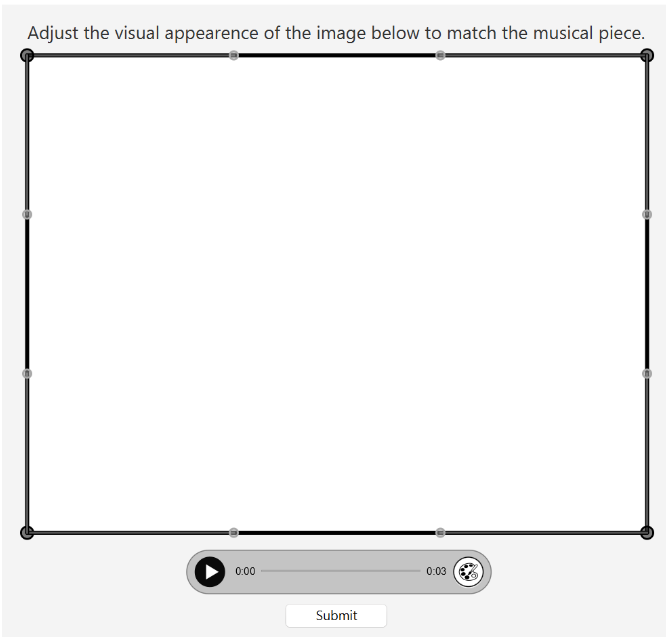


Figure 18. Interface of the first section of the experiment, Music to Image

Image to Music (I2M)

After completing the M2I task, participants then started the I2M portion of the experiment. This part of the experiment aimed to explore the relationship between visual and musical parameters. Participants were presented with fifteen shapes (five shapes with three variations) and were asked to create their own musical piece based on what they saw. The variations of the shapes were included in the experiment to see how small changes in form could affect the participant's interpretation and creation of a musical piece. The experiment was conducted using the same computer program housed on one of the student investigators' computers. The program was still running from the first part of the experiment. Participants had the ability to manipulate various musical parameters such as dynamic variation, rhythmic regularity, tempo, timbre, and mode through adjusting knobs on the interface. The knobs were set to random values at the start of each trial to eliminate any potential biases. However, the knob configurations were the same for each participant to preserve the integrity of the study. Participants were asked to listen to the excerpt associated with the figure and make adjustments to the musical parameters until the excerpt matched the figure properly. Once participants were satisfied with their musical excerpt, they submitted the knob values to the database file, assessed their confidence in their response, and moved on to the next figure. This process was repeated for all of the figures presented. The specific data that was collected is listed in Appendix C, along with the starting state of each trial.

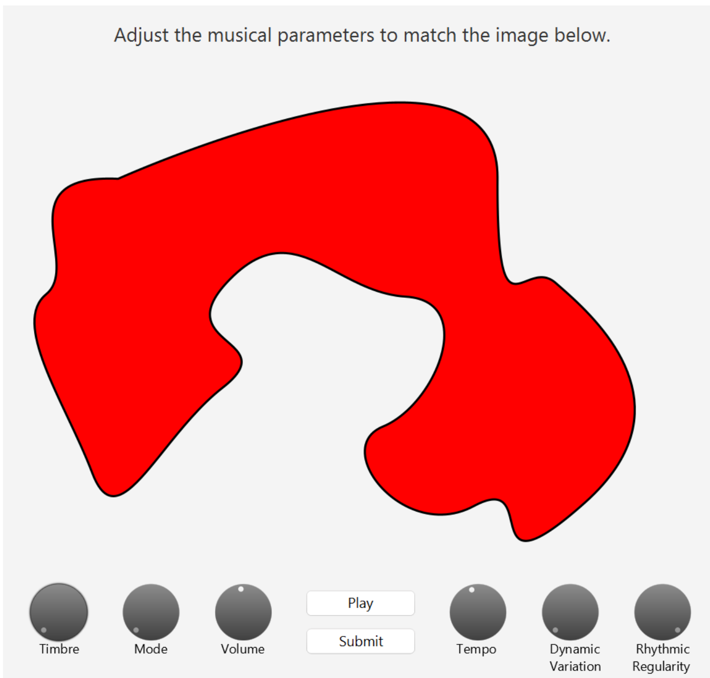


Figure 19. Interface of the second section of the experiment, Image to Music

Analyzing the Data to Determine the Proper Mappings From Sight to Sound

Mapping sight to sound involves a complex process of identifying the relationships between visual and musical parameters and creating mappings that accurately convey these relationships. In order to develop effective mappings, it is necessary to analyze the data to determine the appropriate parameters for the mapping process. This involves identifying the visual and musical parameters that will be used in the mappings, as well as selecting statistical analysis methods and visualization techniques to identify and visualize the relationships between these parameters. By using a combination of statistical analysis and visualization techniques, it is possible to develop an approach that accurately represents the relationships between visual and musical elements.

Determine the Visual and Musical Parameters That Will be Used For Mappings

In this study, we have selected visual and musical parameters based on the data we collected. Specifically, we have defined and collected data on visual parameters such as hue, saturation, brightness, size, and form, as well as musical parameters such as timbre, mode, volume, dynamic variation, and rhythmic regularity. These parameters were selected based on their relevance to the visual stimuli and musical compositions used in this study. There is also potential to create meaningful mappings between visual and musical elements. By using these parameters we can ensure that our mappings are grounded in the data and accurately represent the relationships between visual and musical elements.

Determining Statistical Analysis Methods to Identify Relationships Between Parameters

Correlation analysis is a statistical method that can be used to calculate the correlation between different variables in a data set. By quantifying the relationship between visual and musical parameters, we can identify patterns and trends that may exist within our data set. For example, we can determine which visual parameters are most closely related to certain musical musical parameters such as between musical tempo and visual brightness. To perform correlation analysis, we first quantify categorical data, such as assigning a numerical value for each mode or tempo. We can then create a correlation matrix that displays the correlation coefficients between all pairs of variables in the data set. This matrix can help us to identify which variables are strongly correlated, which variables are weakly correlated, and which variables have no correlation at all. By analyzing the correlation matrix, we can form predictions and analyze patterns within the data. For example, if we observe a strong positive correlation between a particular musical parameter and a visual parameter, we can predict that changes in that musical parameter will be associated with corresponding changes in the visual parameter. We can also use the correlation matrix to identify any outliers or unusual patterns in the data set that may require further investigation.

In addition to correlation analysis, we can also use cluster analysis to identify patterns and relationships between visual and musical parameters. Cluster analysis is a statistical method that groups similar data points together into clusters or categories based on their characteristics or attributes. By applying cluster analysis to our data set, we can identify clusters of visual and musical parameters that share similar characteristics or that are closely related. This can help us to identify patterns and trends in the data that may not be immediately apparent from correlation analysis alone. For example, we may discover that certain visual and musical parameters tend to occur together more frequently than others, or that certain combinations of parameters are more strongly associated. By identifying these patterns and relationships, we can develop more accurate and effective mappings between visual and musical elements. This can result in a more immersive and engaging experience for the audience. Cluster analysis can be sensitive to the number and choice of variables used in the analysis. If the wrong variables are selected or if the number of variables is too large or too small, the resulting clusters may not accurately represent the underlying patterns in the data. It can also be difficult to interpret especially if the resulting clusters are complex or difficult to visualize. This can make it challenging to communicate the results of the analysis to others, and may require additional effort to develop clear and effective visualizations to communicate the findings.

Another possible solution for analyzing the data would be to develop a machine learning algorithm to identify patterns and relationships in the data. For example, you could use a decision tree algorithm to identify the most important factors that predict the visual responses to musical stimuli. Machine learning algorithms can identify complex relationships between variables that may not be apparent through other statistical methods. Algorithms can make accurate predictions based on the patterns identified in the data. Possible issues could be overfitting and to develop machine learning algorithms could require lots of data. Limited knowledge within machine learning algorithms could also pose a challenge given the time constraint.

Determining Visualization Methods to Visualize the Relationships Between Parameters

3D Scatter Plot

One potential visualization technique that we could use to investigate the relationship between musical and visual parameters in our study is a 3D scatter plot. By visualizing the distribution of hue, saturation, and brightness values in a 3D scatter plot, we can observe the correlation between certain musical and visual parameters in more detail and identify any patterns or trends that may exist. For example, we may observe that as the musical tempo increases, the distribution of colors in the dataset shifts towards brighter, more saturated hues. This could indicate a relationship between faster tempos and more energetic, vibrant visual elements. It is important to keep in mind that 3D scatter plots can be complex and difficult to interpret, especially if there are many data points or if the distribution is highly variable. Despite this a 3D scatter plot can be a powerful tool for identifying patterns and relationships in the data and can help to guide the development of more accurate and effective mappings between visual and musical elements.

2D Scatter Plot

A 2D scatter plot can be a useful visualization technique for investigating the relationship between musical visual parameters in our study. By plotting two visual parameters, such as saturation and brightness, against each other in a 2D scatter plot, we can visually examine the relationship between these parameters and musical tempo. For example, we could divide the 2D scatter plot into low, moderate, and high sections for each variable, and then count the number of data points in each section for all three tempos. This would allow us to compare the distribution of visual parameters across different tempos and identify any patterns or trends that may exist. Additionally, we could calculate the correlation coefficient between the two visual parameters and musical tempo to quantify the strength of the relationship between these variables. This information can help us to better understand the mapping between visual and musical parameters and guide the development of more accurate and effective mappings in our study.

Heatmap

Another visualization technique that we could use to investigate the relationship between musical and visual parameters is a heatmap. Heatmaps are useful for visualizing the distribution of values across multiple variables, and can be especially helpful for identifying patterns or trends in datasets. Heatmaps can also be used to visualize the relationship between multiple variables, such as hue, saturation, brightness, and form. By using color coding or other visual cues, we can represent multiple variables in a single heatmap and identify patterns or trends in the data. Heatmaps can be sensitive to the choice of color scale or other parameters, and may require careful selection and interpretation to avoid misinterpretation or bias. However it can still be a good tool for identifying patterns and trends in the data.

Pie Chart

Pie charts are a useful visualization technique for representing the distribution of categorical data in our study, such as different colors or shapes. By categorizing the data points based on their hue values, shape, or other relevant parameters, we can count the number of data points that fall within each category and represent the data in the form of a pie chart. The size of each slice of the pie chart would correspond to the percentage of data points in each category, making it easy to visualize the relative distribution of each category within the dataset. Pie charts can be helpful for identifying patterns and trends in the data that may not be apparent from other types of visualizations. Pie charts can be prone to misinterpretation or oversimplification, especially if there are many categories or if the distribution is highly variable. Still a very useful tool in simplifying complex data in a meaningful way.

Defining the Final Approach for Mapping Sight to Sound Based on the Data Analysis

For our study, we chose to use a combination of correlation analysis and cluster analysis to identify the most significant visual parameters that are associated with specific musical parameters. Additionally, we used a variety of visualization techniques such as 3D scatter plots, 2D scatter plots, heatmaps, and pie charts to help us better understand the relationships between variables and to communicate our findings more effectively. By using a combination of statistical analysis and visualization techniques, we were able to gain a more comprehensive understanding of the mapping between visual and musical parameters and to develop more accurate and effective mappings in our study.

CHAPTER 4: Results and Analysis

Analysis of Parameters in Response to Auditory Stimuli

We conducted a thorough analysis of the data collected from our experiment, specifically focusing on the colors and shapes/forms created by participants in their responses to the audio stimuli. We used various data analysis and visualization techniques such as correlation analysis, cluster analysis, 3D scatter plots, 2D scatter plots, heatmaps, and pie charts to analyze the data and present our findings. The results of our analysis provided insights into how participants perceived and interpreted the audio stimuli, as well as how they used visual parameters such as color and shape/form to create their own responses. We found several interesting relationships between hue-timbre, brightness-tempo, and mode-shape. However, in the second part of our experiment, which involved converting images to music, we obtained limited meaningful information. Even though some of the relationships seen in the first section were still loosely present, the results were highly diverse, and many participants were unsure about their responses.

Music to Image Analysis

Figure 20. The figures constructed by the participants during the first portion of the study. Row letters represent one participant and column numbers indicate the trial within that section of the study.

Color Analysis

In our study, we examined the relationship between visual and musical parameters, specifically the parameters of hue, saturation, and brightness in relation to rhythmic regularity, dynamic variation, and tempo. We used correlation analysis to identify any patterns or trends that may exist between these parameters. Our analysis revealed that there was no significant correlation between hue, saturation, and brightness and rhythmic regularity or dynamic variation. However, we did observe a positive correlation between tempo and brightness, indicating that as the tempo increased, participants tended to use brighter colors in their responses. We also conducted cluster analysis to explore the relationship between visual parameters and musical timbre and mode. We took the hue, saturation, and brightness values for each data point and grouped them by the different timbres and modes using a 3D scatter plot. We did not observe any clear trends or findings among different modes however we did observe possible colors being associated with different timbres.

Relationship Between Musical Tempo and Visual Brightness

To determine the relationship between musical tempo and visual brightness, our team created three separate 3D scatter plots of hue, saturation, and brightness values of the entire music to image dataset, one for each tempo category: slow, medium, and fast. The colors of each point are the actual colors created by the user as we used the hue saturation and brightness values to give us more information about the colors created by participants to see if there were any correlations. It was determined that in the 3D scatter plots that can be seen in Figure 21 points where the tempo was fast there were a greater number of points at higher brightness values compared to that of slow and moderate tempos.

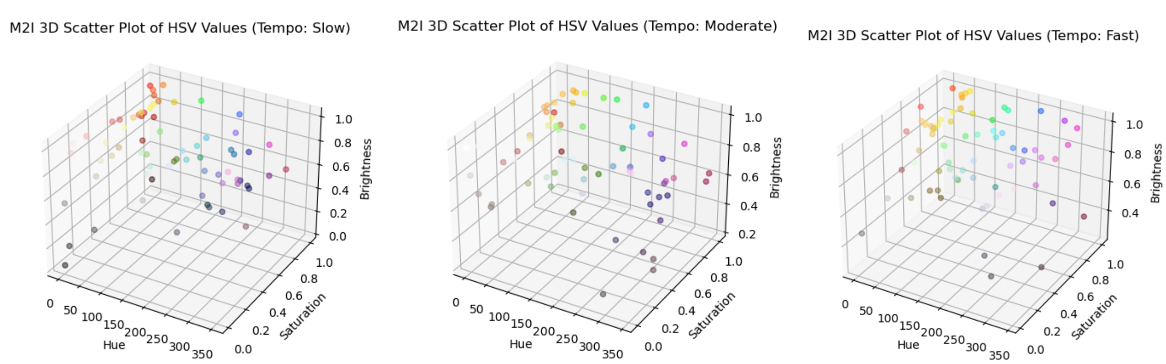


Figure 21. 3D Scatter plots of hue, saturation, and brightness values for music-to-image (M2I) dataset categorized by tempo: slow, medium, and fast. The colors of each point represent the actual colors created by participants.

While the 3D scatter plot provided a view of the relationship between hue, saturation, and brightness for the entire dataset, the 2D scatter plots in Figure 22 allowed us to focus specifically on the relationship between saturation and brightness for each tempo. The scatter plot analysis revealed that data points for fast tempo music had higher values compared to slow and moderate tempo music, indicating an average brightness value of 0.93. In contrast, the average brightness value for slow and moderate tempo music was found to be in the range of 0.8 to 0.86.

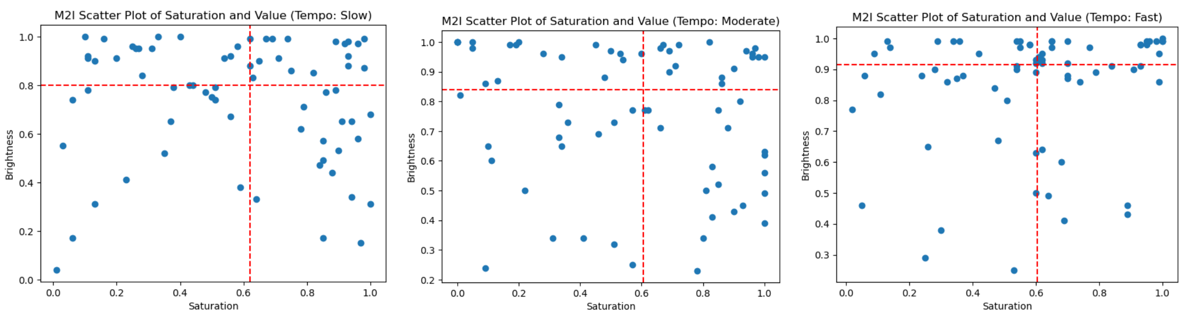


Figure 22. 2D Scatter plots of brightness and saturation for each of the three tempo categories: slow, medium, and fast. The plots highlight the trend of increasing brightness values as the tempo of the music increases.

Based on the 2D scatter plot we created, our team separated the plot into three different sections based on saturation and brightness values low, moderate, and high for each variable. We counted all the points that fell within each section for all three tempos, and created a heatmap to visualize the results. These graphs allow us to understand the distribution of saturation and brightness values in our dataset and identify any patterns or trends that may exist among the different tempos. The heatmap showed that for the fast tempo, there was a large grouping of clusterings in the moderate/high saturation and brightness sections with count of points 26 and 21. The groupings were more spread out among the slow and moderate tempos with large groupings in high brightness and low saturation sections. Slow had the highest count of 14 while fast tempo had the lowest count of 10.

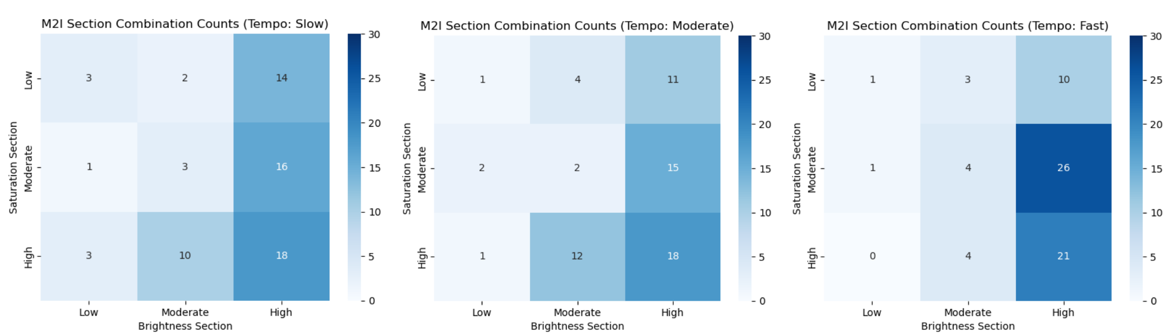


Figure 23A. Heatmap showing the distribution of colors in the music to image dataset based on saturation and brightness values for three different tempo categories: slow, medium, and fast.

Using the sections from the 2D scatter plot, our team also created another heatmap to measure the distance between points in these different sections. Measuring the distance between points in each section can provide a deeper understanding of the color distribution within that section. It allows us to identify any outliers or unusual patterns that may not be immediately apparent from the scatter plot From the graphs created, we noticed that the distance between points in the sections discussed previously remained fairly constant across all tempos. This suggests that the distribution of colors in each section was similar for all three tempos. We also observed that the largest distances between points were observed in the low to moderate saturation and brightness levels, which may indicate a greater variability in color distribution in these sections.

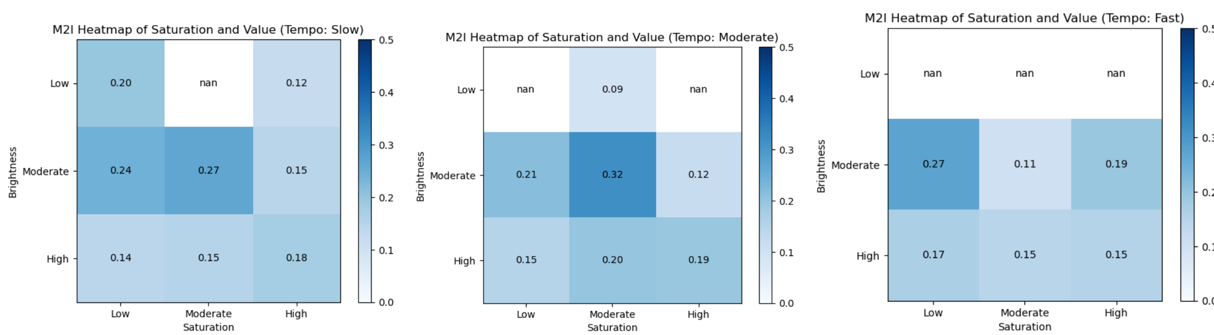


Figure 23B. Heatmap showing the Euclidean distance between points in different sections of the music to image dataset based on saturation and brightness values for three different tempo categories: slow, medium, and fast.

Relationship Between Musical Tempo and Visual Brightness Image to Music

Our team created a 2D scatter plot of saturation and brightness for the Image-to-Music (I2M) dataset and divided it into separate sections based on the saturation and brightness values. We then counted the number of points that fell within each section and created a heatmap to visualize the results. Similar to our findings with the M2I dataset, we observed that the distribution of points in the I2M dataset was also highest in the high saturation and high brightness sections. Our team also found that there was a gradual increase in the number of points in the high saturation and high brightness section as the tempo increased. This finding is similar to what we observed in the M2I dataset, and suggests that there may be a clear correlation between the tempo of music and the use of high saturation and high brightness colors. Specifically, it appears that faster tempos may be associated with the use of more vibrant and bright colors.

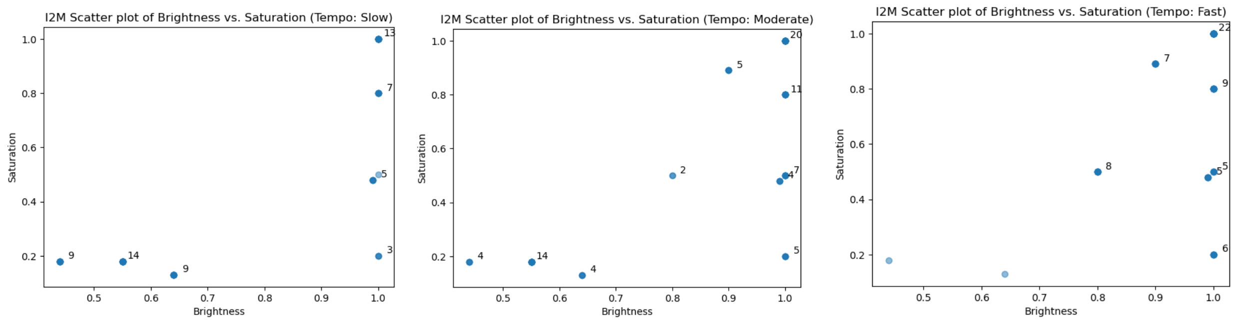


Figure 24A. 2D Scatter plot of saturation and brightness for the I2M dataset with count of number of data points that fall under a certain saturation and brightness value

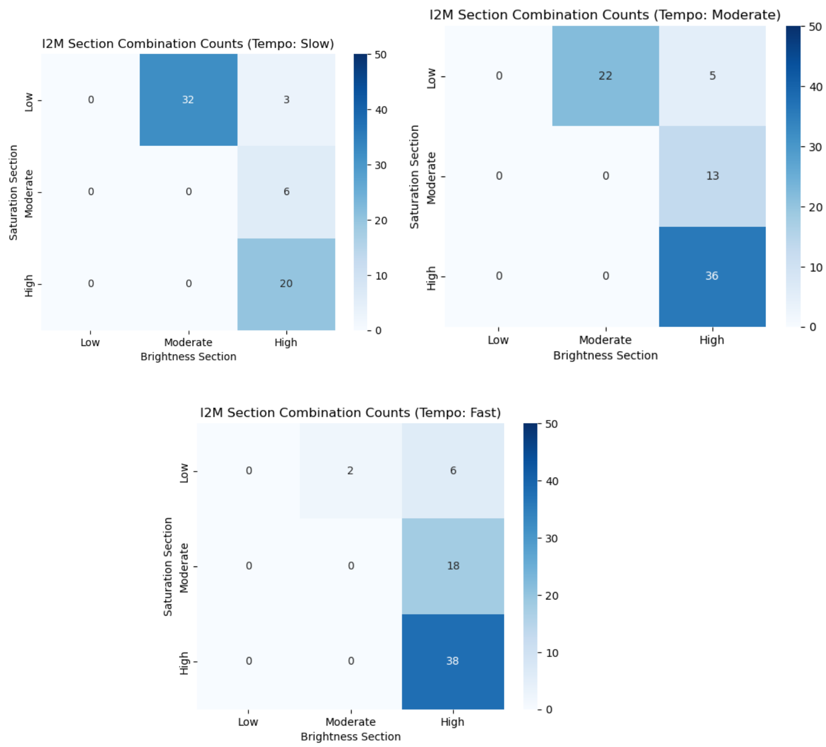


Figure 24B. Heatmap showing the distribution of colors in the I2M dataset based on saturation and brightness values.

Relationship Between Musical Timbre and Visual Hue

After conducting an initial cluster analysis of the Hue, Saturation, and Brightness values for all timbres our team decided to focus on analyzing the Brass and Percussion timbres further. This decision was made after examining the 3D scatter plots of the HSV values for each timbre, which revealed that these two timbres had the most convincing color distributions.

In Figure 25A our team created a 3D scatter plot filtered by the timbre class brass. In this scatter plot we observed strong clustering of warmer colors such as red, orange, and yellow. A pie chart using the data points from the 3D scatter plot is shown in Figure 25B. We categorized the points into different colors based on their hue values. If the point fell within a certain hue value range we would keep a count of that instance. Our team observed similar findings in the pie chart as the color red made up 54.8% of the data points. This finding supports the idea that the use of warm and bold colors may be more common in the Brass timbre.

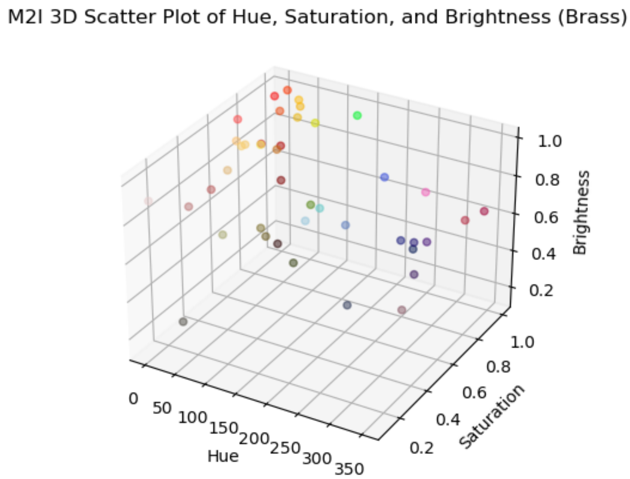


Figure 25A. 3D Scatter plots of hue, saturation, and brightness values for music-to-image (M2I) dataset categorized by Timbre: Brass. The colors of each point represent the actual colors created by participants.

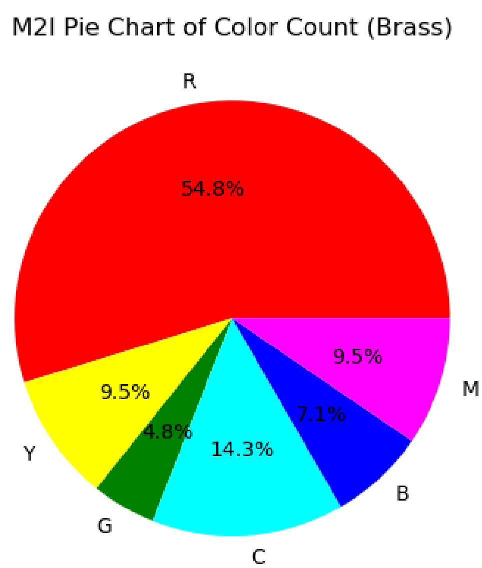


Figure 25B. Pie chart generated from the data points in the corresponding 3D scatter plot. Each color in the pie chart corresponds to a range of hue values, and the size of each slice represents the proportion of data points falling within that range.

In Figure 25C, our team plotted a 2D scatter plot of the saturation and brightness values for the most common hue value range, which was red. We split up the scatter plot into three distinct sections, similar to what was done previously in the tempo analysis, and calculated the Euclidean distance between points in these sections. In Figure 25D, we created a heatmap of the Euclidean distance values for each section. The findings show that at high to moderate saturation and brightness levels, there are closer groupings. At high saturation and moderate brightness, the distance value was 0.01, indicating a possible strong clustering in that region. Similarly, at high saturation and high brightness, the value was 0.05. Higher distance values of 0.08 and 0.09 were observed at lower to moderate saturation and brightness values. These results suggest that there may be significant clustering within certain sections of the red hue value range, particularly at high saturation and brightness levels.

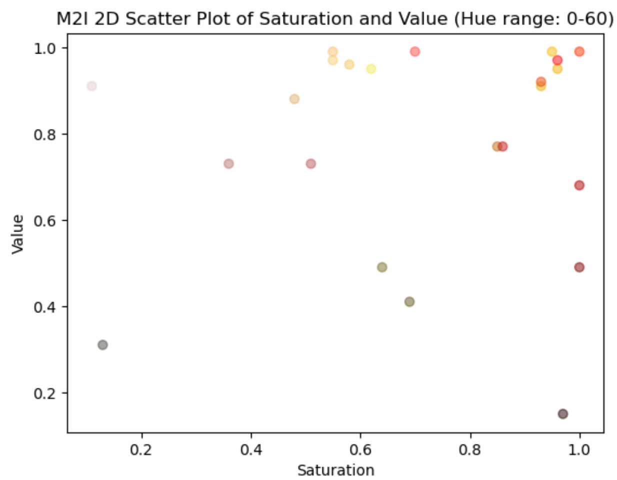


Figure 25C. 2D scatter plot of the saturation and brightness values for the most common hue value range, red, in the Brass timbre of the M2I dataset.

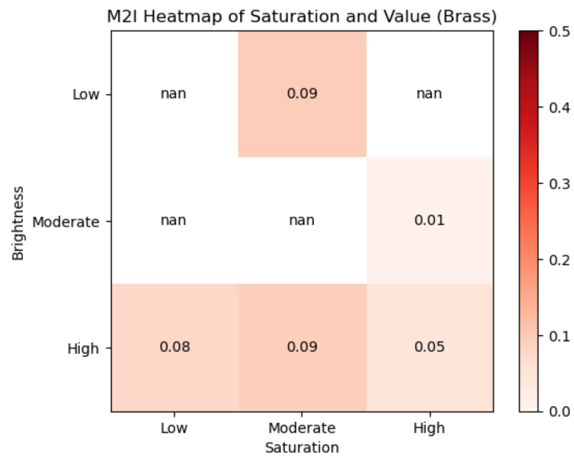


Figure 25D. Heatmap of mean euclidean distance between points in the 2D scatter plot. The heatmap is split into 3 sections based on saturation and brightness levels

Our team performed the same analysis on the Percussion timbre of the M2I dataset. In Figure 26A a 3D scatter plot revealed strong clustering and groupings with cooler colors such as purple and blue. Magenta and blue made up a large portion of the distribution of colors with magenta making up 31% and blue making up 26.2% (Figure 26B). There was also a significant portion of red making up 23.2%. This finding suggests that the use of cool and muted colors may be more common in the Percussion timbre. In Figure 26C, our team plotted a 2D scatter plot of the saturation and brightness values for the most common hue value range, which was blue and magenta. In Figure 26D, we created a heatmap similarly to the above section. The findings show that at high saturation and brightness levels, there are closer groupings. At high saturation and high brightness, the distance value was 0.05 indicating a possible strong clustering in that region. Higher distance values of 0.08 to 0.1 were observed at lower to moderate saturation and brightness values.

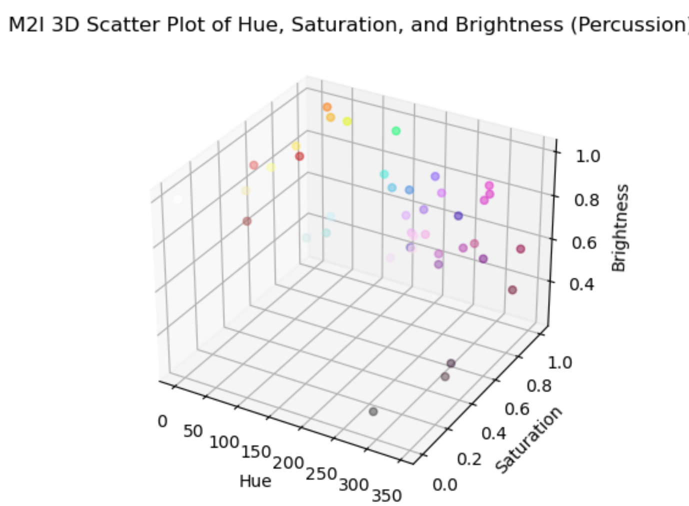


Figure 26A. 3D Scatter plots of hue, saturation, and brightness values for music-to-image (M2I) dataset categorized by Timbre: Percussion. The colors of each point represent the actual colors created by participants.

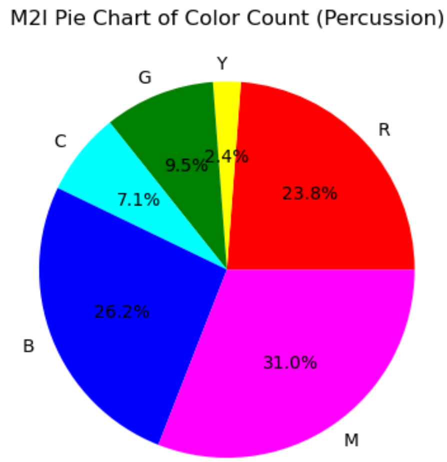


Figure 26B. Pie chart generated from the data points in the corresponding 3D scatter plot. Each color in the pie chart corresponds to a range of hue values, and the size of each slice represents the proportion of data points falling within that range.

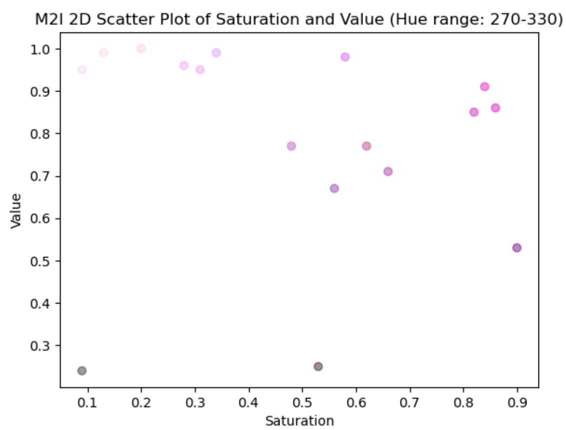


Figure 26C. 2D scatter plot of the saturation and brightness values for the most common hue value range blue and magenta in the Percussion timbre of the M2I dataset.

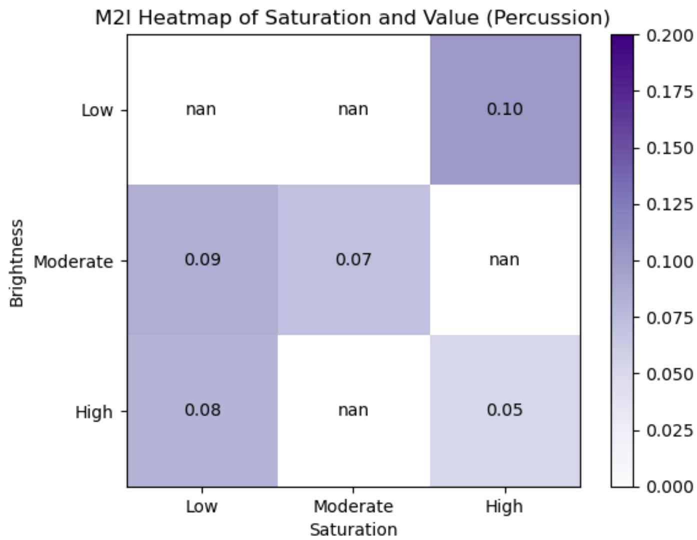


Figure 26D. Heatmap of mean euclidean distance between points in the 2D scatter plot. The heatmap is split into 3 sections based on saturation and brightness levels.

The findings of our study have revealed important insights into the relationship between visual and musical parameters. Specifically, we observed a correlation between musical tempo and visual brightness, with faster tempos associated with higher levels of brightness. Our team also identified specific color associations with different instrument categories. From the visualizations created we found that the timbres of brass and percussion had the most convincing color distributions, with brass being highly associated with red and percussion being associated with blue and magenta.

Figure Analysis

Participant Technique and Approach

Despite receiving the same task, it is important to recognize that each participant approached the assignment differently. This led to a wide range of artistic representation and stylistic choices within each trial. Despite this, there are still some traits that the participants shared in common. These are what will be made use of when we analyze the data. The first observation about the figures is that almost all of them are centered. This might be due to the design of the study: ie. the starting state, examples in tutorial, etc. Because of this, we cannot make any assumptions on the relationship between orientation and location of the figures and other musical parameters.

Melodic Contour

A clear association emerges when looking at the contour of the figures drawn by the participants. The most obvious examples are those produced by Participant B. When comparing their figures to the piano roll, a visual representation of the pitch and timing of notes in a musical composition, of the associated musical piece, it is clear that the shape of the figure matches that of the pitch contour of the piece.Some select examples can be seen in Figure 27.

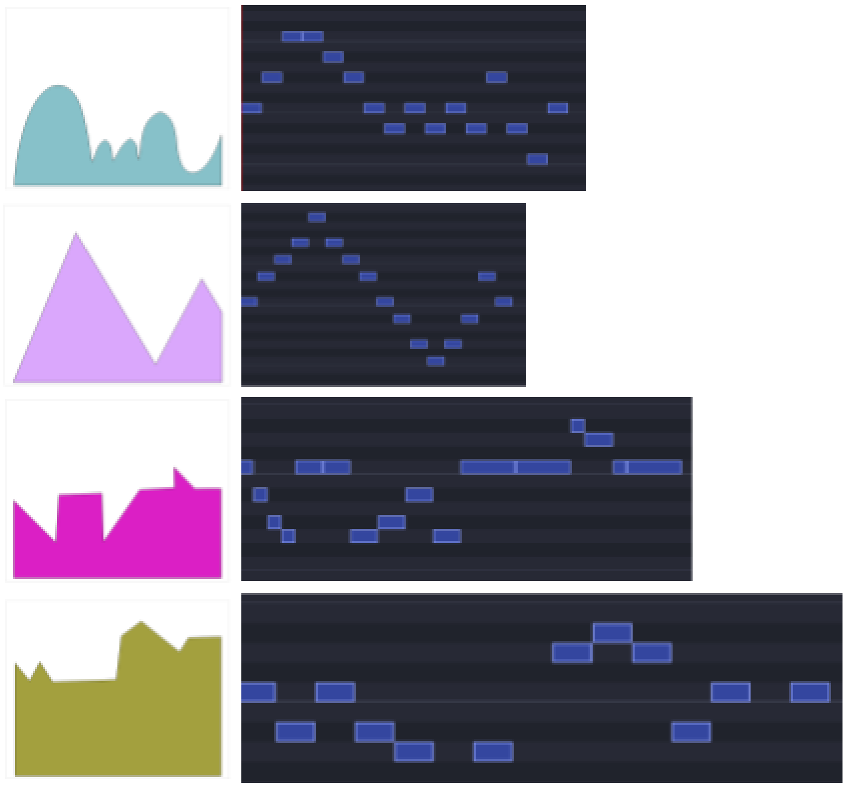


Figure 27. A selection of Participant B's figures compared against the melodic contours of the associated pieces.

The remaining participants also constructed the figures in a similar manner, albeit slightly differently. Some of these participants reflected the musical pieces' pitch contours more loosely, capturing general shapes rather than matching them to scale and exact linear timing. There are a few standout examples on the left side of Figure 28. The melodic contours were radially reflected by other participants. Once more, the fundamental shapes are captured, but it is obvious that the contour is similar. Some notable examples of this can be seen on the right side of Figure 28.

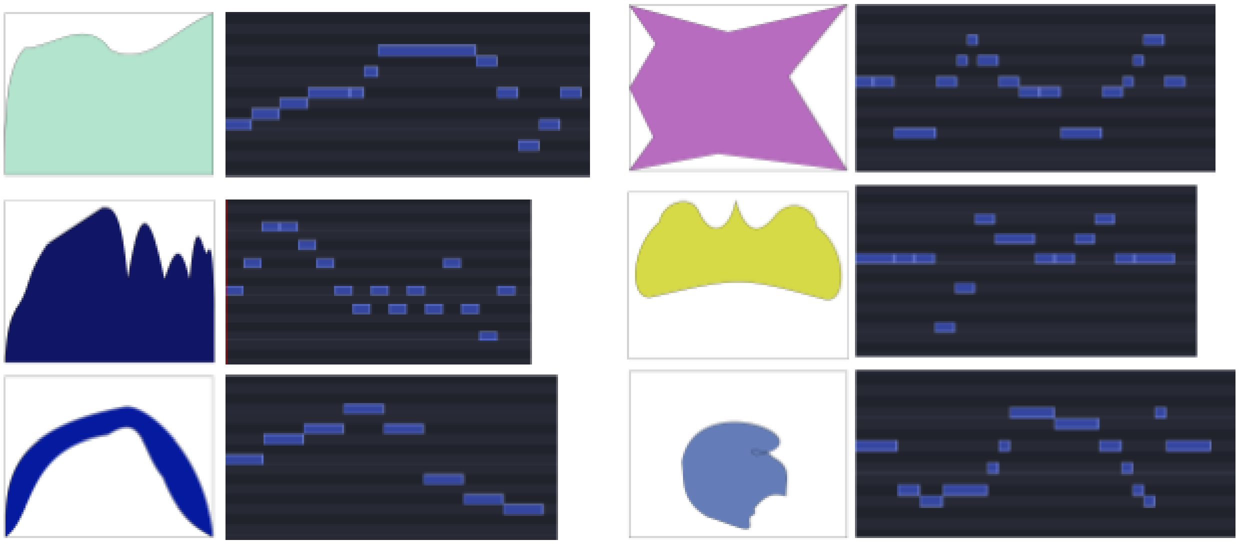


Figure 28. A selection of figures by other participants compared against the melodic contours of the associated pieces.

Rhythmic Regularity and Curvature

The smoothness of the curves in the figures and the rhythmic regularity of the musical pieces also appear to be related. Even though not all of the participants followed this relationship, its occurrence makes it noteworthy. The relationship is more obvious when the figures are divided into groups according to the rhythmic regularity of the corresponding musical piece. It seems that a lot of the figures connected to a regular rhythm only have sharp or linear curves, most of the figures linked to an irregular rhythm have a combination of sharp and smooth curves, and the majority of the figures connected to a syncopated rhythm only have smooth, rounded curves. A curve is essentially a product of lines. This finding makes sense because the curvature of a line is somewhat a function of its distortedness, or how much it deviates from the direct path between endpoints. The lines become more curved as the rhythm becomes further from the grid.

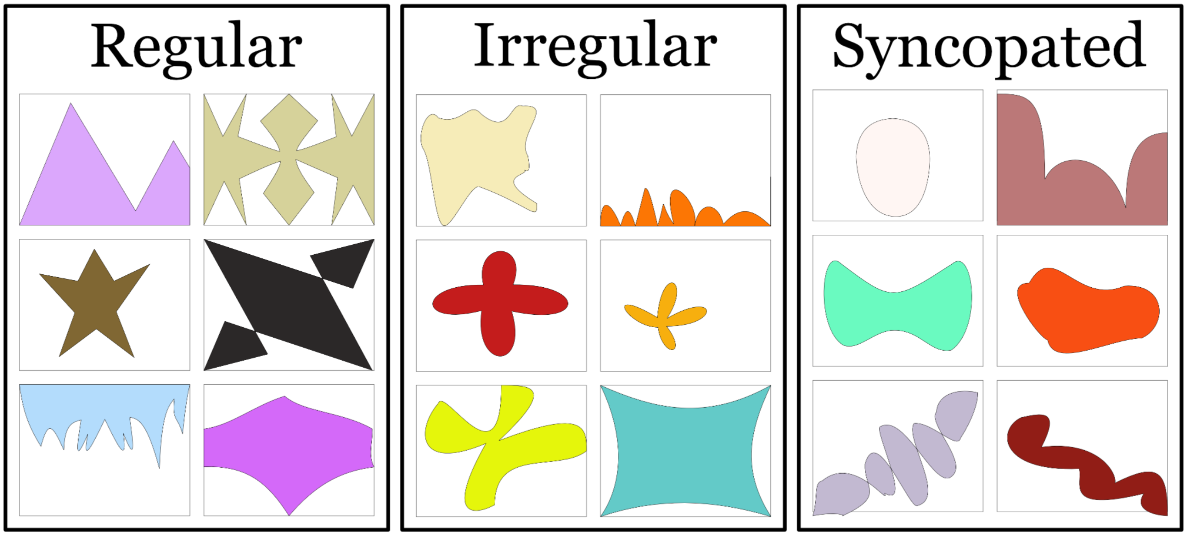


Figure 29. A selection of figures categorized by their associated piece's rhythmic regularity.

Issues With the Study's Image to Music Section

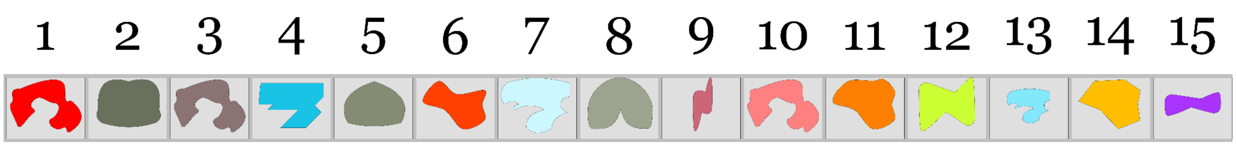


Figure 30. The figures provided to the participants during the second portion of the study. Column numbers indicate the trial within that section of the study.

The study's second section, which involved converting images to music, produced limited insights. Although some relationships observed in the first section, such as brightness-tempo and curvature-rhythmic regularity, remained, the outcomes varied significantly, and few commonalities emerged among responses. A very small minority of responses exhibited the previously observed relationships, and therefore cannot be used to reinforce those relationships further. These revelations raise the possibility that data analysis might be biased if based on the conclusions from the M2I section. Moreover, about 79.04% of participants were unsure about their response, likely due to the lack of control over the music creation tool. Numerous participants mentioned that they felt their options were extremely constrained and that the majority of the combinations offered did not correspond to how they interpreted the figures. The main issue, we believe, was the fact that the basis for the constructed musical excerpts were randomly generated for each trial, with the pitches determined by the contour of the figure presented. In light of this, we believe that the experimental design of this section of the study was biased, or at least not constructed in the best possible manner to produce meaningful data.

If this portion of the study were to be run again, we would revise how it was set up. An example mock-up is presented in Figure 31.

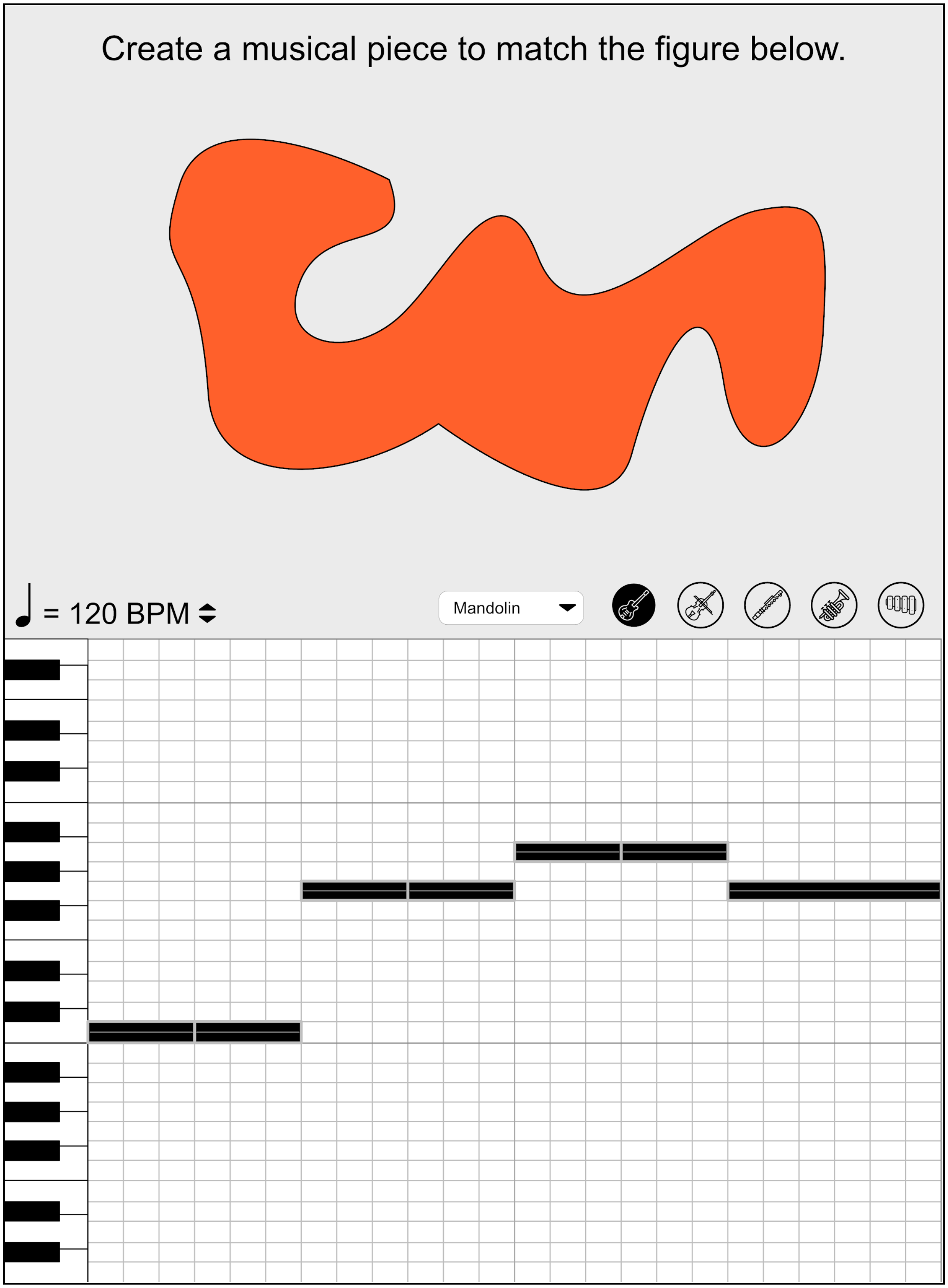


Figure 31. A mock-up of a revised I2M interface based on participant feedback.

In this model, the user would have direct control over selecting the instrument, the tempo, the durations of each note, and the pitches associated with the figure. This is thought to produce a better, more statistically significant set of data while maintaining the desired relationships. The musical mode could be identified from the pitches, the rhythmic regularity from the durations, and the dynamic variance and overall volume from the velocities of the individual notes. Additionally, this model makes it simpler to set the tempo by presenting the user with a numerical value, as well as simpler to select an instrument by using buttons that represent the available options. The use of adding the pitches manually to the piano roll can also be utilized to see if the pitch contour is still relevant. There are also multiple timbre options within each timbre class, which is another aspect of this redesign that might not be evident from Figure 31. The lack of a broad enough selection of timbres was a major complaint we heard from participants. For instance, many participants said they desired a woodwind instrument but did not specifically want a flute. This adaptation lets the participants feel more creative and assured with their creation, which may result in improved data. It also enables us as researchers to establish the relationship between visual parameters and timbre class. This design does have one drawback, however, in that participants will essentially be drawing a shape again, only this time there will be a reference shape, which could lead to bias. A direct copy of the figure onto the piano roll is not possible, though, as almost all of the presented figures are non-functional relations. Ignoring this problematic factor, the overall redesign of this study's section should produce data that can be better analyzed.

Lack of Correlations and Further Analysis

While we were able to identify some clear relationship between certain musical and visual parameters, we also encountered analytical challenges due to the complexity and subjectivity of these creative processes. In this section, we will discuss our analysis of the correlations between music and visual parameters, highlighting the insights we gained and the challenges we faced in our research.

Our team conducted an analysis of the correlation between spectral centroid and brightness in the music-to-image dataset. Spectral centroid is a measure of the frequency center of a sound, while brightness is a measure of the overall lightness of an image. We created a scatter plot of the two variables to visualize any potential correlation between them. Our initial hypothesis was that as spectral centroid increased, the brightness of the images created would also increase. The scatter plot did indicate a positive correlation between spectral centroid and brightness, as we observed a general trend of increasing brightness as spectral centroid increased. However, we also noted that there was a lack of data points towards the higher end of the spectral centroid range, which may have impacted the accuracy of the analysis. This lack of data points may have led to the linear regression line indicating a stronger correlation than actually exists. Given these findings, we believe that further analysis and testing may be needed to determine the true nature and strength of the correlation between spectral centroid and brightness in the music-to-image dataset. Nonetheless, our analysis provides a valuable starting point for future research into the relationship between musical parameters and visual aesthetics.

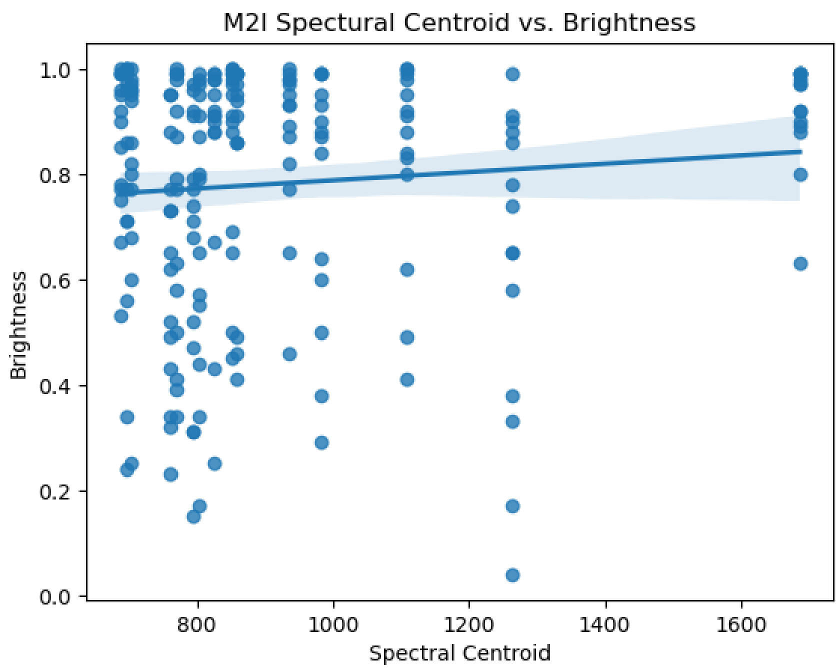


Figure 32. Scatter plot of Spectral Centroid vs Brightness in the M2I Dataset.

Our team also conducted an analysis of the different modes in the M2I dataset, creating a 3D scatter plot to visualize any potential relationship between the visual parameters and different modes of music. While we observed some trends in the data, we did not find any conclusive evidence of a strong correlation between the modes of music and the visual parameters. To further analyze the relationships between color and the different modes of music, we also created a bar chart showing the count of points in different color sections. Our analysis did not reveal any clear trends or patterns, as similar counts of colors were observed in all sections. While this lack of clear correlation may indicate that the relationship between music and color is complex, we believe that further analysis of the saturation and brightness among the different modes could be done to uncover any possible correlations.

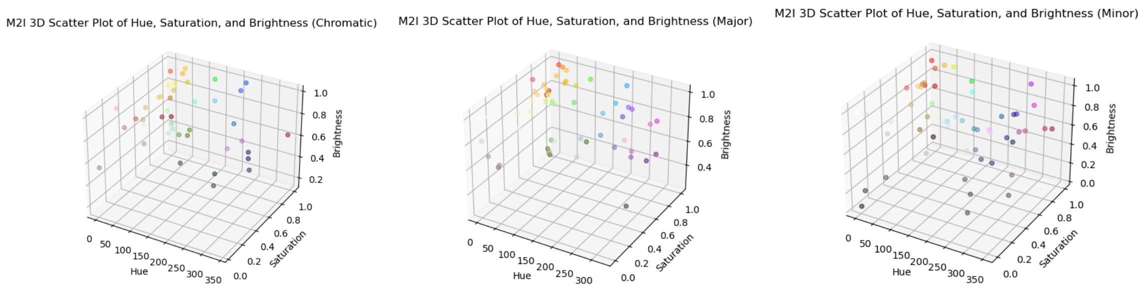


Figure 33. 3D Scatter plots of hue, saturation, and brightness values for music-to-image (M2I) dataset categorized by Mode: Chromatic, Major, Minor. The colors of each point represent the actual colors created by participants.

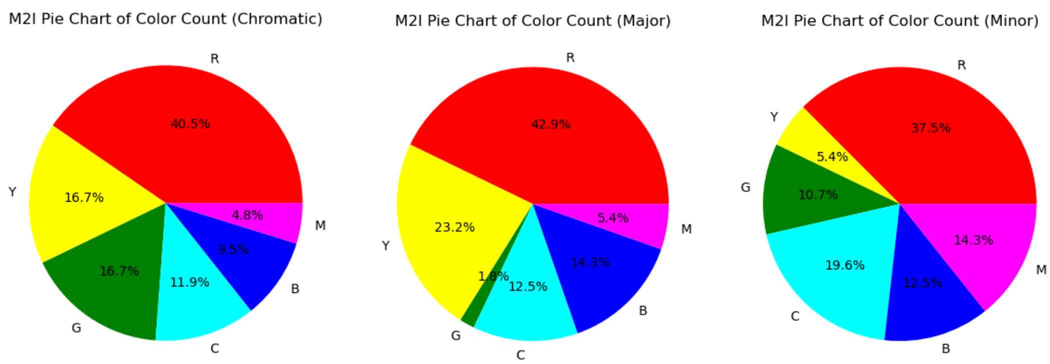


Figure 34. Pie chart generated from the data points in the corresponding 3D scatter plot. Each color in the Pie chart corresponds to a range of hue values, and the size of each slice represents the proportion of data points falling within that range.

To investigate the potential correlation between dynamic variation and rhythmic regularity in the M2I and I2M datasets, our team created correlation matrices. After generating the correlation matrices, we observed that the values for the correlation coefficients were close to 0 for both datasets, indicating no statistically significant correlation between dynamic variation and rhythmic regularity on HSV. This finding suggests that, in the context of the M2I and I2M datasets, there is no strong correlation between dynamic variation and rhythmic regularity in determining the visual parameters of hue, saturation, and value.

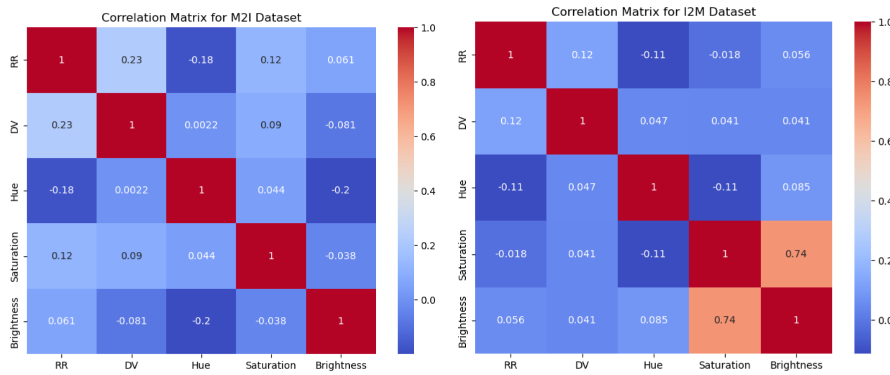


Figure 35. Shows the correlation matrices for dynamic variation (DV) and rhythmic regularity (RR) in the M2I and I2M datasets.

Conclusion of Results

Based on the results of the study, several findings were identified regarding the relationship between musical and visual parameters.

1. The first finding suggests a correlation between musical tempo and visual brightness. The study found that as the tempo of the music increased, the values of brightness in the visual images created by the participants also increased. This finding indicates a possible relationship between the tempo of music and the intensity of visual perception, with higher tempo music being associated with more vivid and brighter visual images.
2. The second finding from the study suggests a correlation between the use of specific colors and different instrument categories. The study found that certain instrument categories, such as Percussion and Brass, were associated with particular colors, and red and orange for Brass.
3. The third finding suggests that there may be a relationship between the pitch contour in the generated music and the overall shape of the visual figure being translated. For example, a figure with a lot of curves may correspond to a more varied pitch contour in the generated music, whereas a figure with more straight lines may correspond to a flatter pitch contour. This is because the pitch contour of a musical piece is essentially the shape of the music.
4. The fourth finding suggests a relationship between the rhythmic regularity in the generated music and the curvature of the figure's components. This means that a figure with more curved components may correspond to a more irregular or complex rhythm in the generated music, while a figure with more straight components may correspond to a more regular or predictable rhythm.

These findings suggest that there is a possibility of a strong relationship between musical and visual parameters that we identified.

CHAPTER 5: Conclusion and Future Directions

This research project began with a fundamental question: is it possible to translate one art form into another without losing its emotional and compositional qualities? Throughout our investigation, we have discovered that the answer to this question is nuanced. To some degree, emotional elements of one piece of art can be conveyed through another art form, at a very fundamental level, such as the category of emotion, tone, mood, and symbolism. However, there are limitations to how much of the original artwork can be preserved during translation. It is essential to recognize that any translation will inevitably involve an implicit bias, therefore making it have a sense of mimicry.

The possibilities for future research in the field of artistic translation are numerous and exciting. The current study focused on translating visual art into sound, but there are many other forms of art that could be explored, such as literature, poetry, dance, and theater. Investigating these pairings could shed more light on the broader topic of artistic translation as a whole and provide insights into the similarities and differences between different forms of art. Furthermore, the relationships between other audiovisual parameters could also prove fruitful avenues for future research. As we stated previously, there are a great amount of parameters in the realm of visual art and music that we neglected for the sake of simplicity and time. Exploring these parameters could uncover previously unexplored links between different art forms, leading to a deeper understanding of the role that emotion and composition play in art. It would also be beneficial to examine the degree and valence of emotion retained when translating art from one medium to another. While the current study found that emotion can be retained to some degree, through anecdotal evidence, the extent to which this is possible remains unexplored. Finally, it would be interesting to find out which mediums translate more successfully. This direction of research could help to identify which mediums are more adaptable and which ones require more effort and creativity to translate successfully. The study of successful examples of translation could be a crucial step in developing effective methods for translating art. To advance these topics, utilizing different techniques and methodologies could be valuable. For example, future studies could employ machine learning algorithms to analyze and translate works of art, or they could use brain imaging techniques to study the emotional and cognitive processes involved in translating art between different mediums.

Our team encourages others to continue building upon the project's goals, as the recommended genres of art, artistic medium, and interfaces in conjunction with the uncovered relationships could lead to creating a system that allows sight-challenged individuals to experience the emotional and compositional aspects of visual art. Ultimately, our research project has only scratched the surface of this complex topic, and we hope that our findings and recommendations will serve as a starting point and inspiration for future studies.

REFERENCES

7 different forms of art: An easy guide. (n.d.). Eden Gallery. https://www.eden-gallery.com/news/7-different-forms-of-art

All Answers Ltd. (2022, July 29). *Concept of Time Portrayed in Art.* https://www.ukessays.com/essays/arts/the-concept-of-time-in-different-works-art-essay.php

Agarwal, T. (2021, December 20). *Difference Between Position Sensor, Motion Sensor and Proximity Sensor*. ElProCus - Electronic Projects for Engineering Students. https://www.elprocus.com/difference-motion-sensor-position-sensor-proximity-sensor/

Bal du moulin de la galette - by Pierre-Auguste Renoir. (n.d.). https://www.renoir.net/bal-du-moulin-de-la-galette.jsp

Blue Horse I by Franz Marc. (n.d.). https://www.franzmarc.org/Blue-Horse.jsp

Bresson, R. (1997). Notes on the Cinematographer. Green Integer. Retrieved May 3, 2023, from http://artsites.ucsc.edu/faculty/gustafson/FILM170B.W12/bresson1.pdf

Brown, M. A. (2019). John Williams and the music of Star Wars. Oxford University Press.

Campbell's Soup Cans-by Andy Warhol. https://www.moma.org/collection/works/79809

Carbon, C.-C. (2017). *Art Perception in the Museum: How We Spend Time and Space in Art Exhibitions*. I-Perception, 8(1). https://doi.org/10.1177/2041669517694184

Carroll, N. (2003). Art and Mood: Preliminary Notes and Conjectures. *The Monist*, *86*(4), 521–555. http://www.jstor.org/stable/27903843

Carsten Nicolai: Reflektor Distortion | My Art Guides. (n.d.). My Art Guides | Your Compass in the Art World. Retrieved May 3, 2023, from https://myartguides.com/exhibitions/berlin/carsten-nicolai-reflektor-distortion/

Castle and Sun by Paul Klee. (n.d.). https://www.paul-klee.org/castle-and-sun/

Chihab Nadri, Chairunisa Anaya, Shan Yuan, & Myounghoon Jeon. (2019, July 25). *Preliminary guidelines on the sonification of visual artworks: Linking music, sonification & visual arts.* https://smartech.gatech.edu/handle/1853/61536

Christine Mohr. *Similar Patterns of Emotion Associations with Color Patches.* National Library of Medicine https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7027086/

Collignon, O., Charbonneau, G., Lassonde, M., Lepore, F. (2013). Early visual deprivation alters multisensory processing in perceiving audiovisual emotions, 14-20.

Cooper, N. (2018). The sound of the Dark Side: analyzing the music of the Star Wars prequels. In The Music of Star Wars: 30th Anniversary Collector's Edition.

Cunningham, J. M. (2022, January 7). *Christian Marclay | Biography, Art, The Clock, & Facts.* Encyclopedia Britannica. https://www.britannica.com/biography/Christian-Marclay

Duffie, B. (2015). Gustav Holst's "The Planets." Retrieved from https://www.npr.org/sections/deceptivecadence/2015/05/23/408743957/gustav-holsts-the-planets.

eigen+art - Carsten Nicolai. (n.d.). https://eigen-art.com/en/exhibitions/archive/carsten-nicolai/

Ekphrasis - Writing About Art. (n.d.). Retrieved May 3, 2023, from https://writingaboutart.org/pages/ekphrasis.html

Epstein, R. A., & Backer, C. I. (2019, June 21). *Scene perception in the human brain.* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6989029/

Girard, S. & Collignon, O. (2017). The sound of touch: vision changes how sound is processed. Current Opinion in Behavioral Sciences, 16, 75-80.

Great Big Story. (2017, June 19). *Longplayer Is a Symphony That Will Play for a Thousand Years.* YouTube. https://www.youtube.com/watch?v=71bp0x8Ly-Y

Great Learning Team. (2022, October 25). *Types of Neural Networks and Definition of Neural*

Network. Great Learning Blog: Free Resources What Matters to Shape Your Career! https://www.mygreatlearning.com/blog/types-of-neural-networks/

Harbisson, N. (2012). I listen to color. TEDGlobal. https://www.ted.com/talks/neil_harbisson_i_listen_to_color

Hagtvedt, H., Patrick, V. M., & Hagtvedt, R. (2008). The Perception and Evaluation of Visual Art. *Empirical Studies of the Arts*, *26*(2), 197–218. https://doi.org/10.2190/em.26.2.d

Heled, E. (2022, June). *Working Memory in Intact Modalites Among Individuals with Sensory Deprivation* ScienceDirect https://www.sciencedirect.com/science/article/pii/S2405844022008465

Imaginary Soundscape. (n.d.). Retrieved May 3, 2023, from https://imaginarysoundscape.net/

Intraub, H. (n.d.). *Visual Scene Perception*. Retrieved May 3, 2023, from https://cpb-us-w2.wpmucdn.com/sites.udel.edu/dist/3/6297/files/2017/10/2002-Intraub-Encyclopedia-of-Cognitive-Sci-1uop5d8.pdf

Intraub, H. (2010). Rethinking Scene Perception. *The Psychology of Learning and Motivation*, 231–264. https://doi.org/10.1016/s0079-7421(10)52006-1

Judd, T. (2018, May 16). *"La Mer": Debussy's Sonic Portrait of the Sea*. The Listeners' Club. https://thelistenersclub.com/2018/05/16/la-mer-debussys-sonic-portrait-of-the-sea/

Kelly, T. (2016). Gustav Holst: The Planets. Cambridge University Press.

Laurentis, C. de. (2022, May 6). *Synesthesia – The Surprising and Powerful Fusion of Senses in Art.* DailyArt Magazine. Retrieved May 3, 2023, from https://www.dailyartmagazine.com/synesthesia-in-art/

Library Stack: Close Readings. (n.d.). LIBRARY STACK. https://www.librarystack.org/close-readings/

Lin, A. (2021, February 5). *Generating Music Using Markov Chains - HackerNoon.com*. Medium.

https://medium.com/hackernoon/generating-music-using-markov-chains-40c3f3f46405

Lindborg, P., & Friberg, A. K. (2015). Colour Association with Music Is Mediated by Emotion: Evidence from an Experiment Using a CIE Lab Interface and Interviews. *PLOS ONE*, *10*(12), e0144013. https://doi.org/10.1371/journal.pone.0144013

Margounakis, D., & Politis, D. (2006). CONVERTING IMAGES TO MUSIC USING THEIR COLOUR PROPERTIES. *International Conference on Auditory Display*. http://cogsci.eecs.qmul.ac.uk/oldpages/icad2006/proceedings/papers/f66.pdf

Mather, G. (2013). *The Psychology of Visual Art: Eye, Brain and Art*. Cambridge University Press.

Meleca, A., MA. (2022.). *Expressionism Art Movement – History, Artists and Artwork*. Artlex. https://www.artlex.com/art-movements/expressionism/

Millet, J. (n.d.). *The Angelus - Jean-François Millet*. Google Arts & Culture. https://artsandculture.google.com/asset/the-angelus/CgHjAgexUzNOOw?hl=en

Minimalism Movement Overview. (n.d.). The Art Story.

https://www.theartstory.org/movement/minimalism/

Muller-Eberstein, M., & van Noord, N. (2019). Translating Visual Art Into Music. *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*.

https://doi.org/10.1109/iccvw.2019.00378

Naar, H. (n.d.). *Art and Emotion*. The Internet Encyclopedia of Philosophy. Retrieved May 3, 2023, from https://iep.utm.edu/art-and-emotion/

Nine-Sided Figure - Paula Cooper Gallery. (n.d.). https://www.paulacoopergallery.com/

Pelowski, M., & Leder, H. (n.d.). *Art & Emotion*. EVAlab. Retrieved May 3, 2023, from

https://aesthetics.univie.ac.at/research/art-emotion/

Pop Art Movement Overview. (n.d.). The Art Story.

https://www.theartstory.org/movement/pop-art/

Pluta, J. (2016). A Guide to Gustav Holst's The Planets. Oxford University Press.

Radiant Baby from Icons series | Keith Haring. (n.d.). https://www.haring.com/!/art-work/773

Rule-Based System. (n.d.). Engati. Retrieved May 3, 2023, from https://www.engati.com/glossary/rule-based-system

Saldaña, R., & Mendizabal-Ruiz, G. (2020, September). *Deep Learning for Image to Sound Synthesis*. https://cce.cinvestav.mx/images/archivos/papers-20/CCE_2020_paper_80.pdf

Shapiro, L. G., & Stockman, G. C. (2001). *Computer Vision* (1st ed.). Pearson.

Silvia, P. J. (2005). Emotional Responses to Art: From Collation and Arousal to Cognition and Emotion. *Review of General Psychology*, *9*(4), 342–357. https://doi.org/10.1037/1089-2680.9.4.342

Smithsonian American Art Museum. (n.d.). *Close Readings (with Jeffrey Mansfield, Ariel Baker-Gibbs, Alison O'Daniel, Lauren Ridloff)*. Retrieved May 3, 2023, from https://americanart.si.edu/artwork/close-readings-jeffrey-mansfield-ariel-baker-gibbs-alison-odaniel-lauren-ridloff-117969

Sun, W. (2022, June 18). 3D Face Parsing via Surface Parameterization and 2D Semantic. . . arXiv.org. https://arxiv.org/abs/2206.09221

Synesthete, O. C. A. (2022, November 3). *Vivaldi.* http://creationsofasynesthete.blogspot.com/2016/11/vivaldi.html

Tate. (n.d.). *Realism*. https://www.tate.org.uk/art/art-terms/r/realism

The Daughters of Edward Darley Boit – Works – Museum of Fine Arts, Boston. (n.d.). https://collections.mfa.org/objects/31782

The Editors of Encyclopaedia Britannica. (2022, October 25). *Impressionism | Definition, History, Art, & Facts*. Encyclopedia Britannica. https://www.britannica.com/art/Impressionism-art

The Editors of Encyclopaedia Britannica. (1999, July 28). *Realism | History, Definition, & Characteristics*. Encyclopedia Britannica. https://www.britannica.com/art/realism-art

The Record: Contemporary Art & Vinyl. (2019, August 15). Nasher Museum of Art at Duke University. https://nasher.duke.edu/exhibitions/record-contemporary-art-vinyl/

The Water Lilies by Claude Monet | Musée de l'Orangerie. (n.d.). https://www.musee-orangerie.fr/en/node/197502

Tobi Zausner. (2006). Process and meaning: nonlinear dynamics and psychology in visual art. *Nonlinear Dynamics, Psychology, and Life Sciences*, *11*(1), 149–165. https://europepmc.org/article/MED/17173732

Torralba., A. (2019, September 16). *Visual and auditory scene parsing*. https://dspace.mit.edu/handle/1721.1/122101?show=full

Ulmer, J. (2015, May). *Imaging of Functional Visual System* National Library of Medicine https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4551406/

Understanding Multimodalities in Arts and Social Sciences. (2019, October 29). Arts and International Affairs. https://theartsjournal.net/2019/03/28/editorial-2/

Valich, L. (2018, December 20). *The science of seeing art and color*. News Center. https://www.rochester.edu/newscenter/the-science-of-seeing-art-color-354182/

Various Ways the Passage of Time is Portrayed. (n.d.). Google Arts & Culture. Retrieved May 3, 2023, from https://artsandculture.google.com/usergallery/2AJyQspY_1P4Iw

Wakefield, K. (n.d.). *A guide to machine learning algorithms and their applications*. SAS. Retrieved May 3, 2023, from https://www.sas.com/en_us/insights/articles/analytics/machine-learning-algorithms-guide.html

Warr, K. (2019). *Strengthening Deep Neural Networks: Making AI Less Susceptible to Adversarial Trickery*. Van Duuren Media.

Watson, K. (2018, October 25). *What Is Synesthesia?* Healthline.
https://www.healthline.com/health/synesthesia

Xianyi Zeng. *Aesthetic Perceptions of Visual Texture*. The National Library of Medicine
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4631837/

Yeo, W. S., & Berger, J. (2005). *APPLICATION OF IMAGE SONIFICATION METHODS TO MUSIC.*

APPENDIX A: Study Application Source Code and Components

In lieu of pasting the source code and component images here, all source code and components of the study application can be found on GitHub at https://rb.gy/6exvf. This repository contains both the study application (in the `study` branch) and the backbone of the translation program this IQP originally set out to create (in the `main` and `develop` branches).

APPENDIX B: Pre-Experiment Survey Questions

Participant demographic information

- What is your age?
- What is your gender?
  - Male
  - Female
  - Other (Please Specify)
  - Prefer not to say
- What is your ethnicity or cultural connection?
  - Native American or Alaskan Native
  - Asian or Pacific Islander
  - Middle Eastern or North African
  - Black or African American
  - Hispanic or Latino
  - Caucasian
- What is your geographical connection?
  - Northeast US
  - Southeast US
  - Midwest US
  - Southwest US
  - West Coast
  - Northern Territories
  - Prairie Provinces
  - Central Canada
  - Atlantic Region
- What is your highest degree of education?
  - Did not attend school
  - Elementary-Middle School

- High School
- Associate Degree
- Bachelor's Degree
- Master's Degree
- Doctoral Degree

Participants' prior experience with music

- What musical instruments do you play or have experience with?
- How well do you think you can distinguish between different musical instruments or voices in a piece of music?
- On a scale of 1 to 5, how familiar are you with different types of scales, such as major and minor modes?
- How frequently do you actively listen to music, and upon which aspects of the music do you focus?

Participants prior experience with visual art

- How frequently do you visit art museums or galleries, and do you typically focus on particular styles or artists?
- Have you ever had experience creating your own form of visual art?
- If you have experimented with creating your own art, and how did you go about doing it?
- On a scale of 1 to 5, how familiar are you with the following forms of visual art: drawings, paintings, digital images, photography, sculptures

Participants' self-reported mood or emotion before the experiment

- On a scale of 1-10, how would you rate your emotional valence, where 1 is extremely negative and 10 is extremely positive?
- Please describe your current mood and emotional state.

APPENDIX C: DATA TABLES

The raw data from the M2I and I2M portions of the study are not present here due to their size. The data tables can be found at https://rb.gy/wldrf and https://rb.gy/egnm3, respectively.